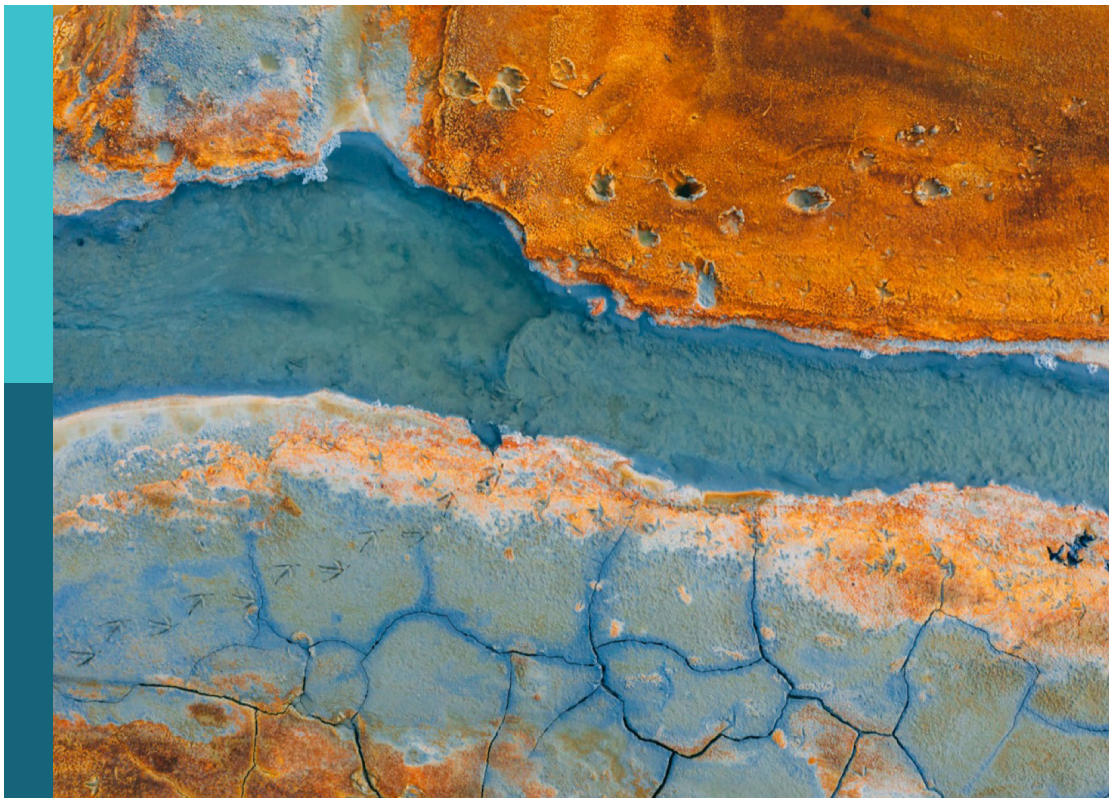
frontiers | Research Topics

Advances in integrated surface—subsurface hydrological modeling

Edited by
Matteo Camporese, Oliver S. Schilling and Hoori Ajami

Published in
Frontiers in Water



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6842-2
DOI 10.3389/978-2-8325-6842-2

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in integrated surface—subsurface hydrological modeling

Topic editors

Matteo Camporese — University of Padua, Italy

Oliver S. Schilling — University of Basel, Switzerland

Hoori Ajami — University of California, Riverside, United States

Citation

Camporese, M., Schilling, O. S., Ajami, H., eds. (2025). *Advances in integrated surface—subsurface hydrological modeling*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-6842-2

Table of contents

frontiers | Frontiers in Water

TYPE Editorial
PUBLISHED 29 August 2025
DOI 10.3389/frwa.2025.1680183

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Harrie-Jan Hendricks Franssen,
Helmholtz Association of German Research Centres (HZ), Germany

*CORRESPONDENCE
Matteo Camporese
matteo.camporese@unipd.it

RECEIVED 05 August 2025
ACCEPTED 15 August 2025
PUBLISHED 29 August 2025

CITATION
Camporese M, Schilling OS and Ajami H (2025) Editorial: Advances in integrated surface—subsurface hydrological modeling. *Front. Water* 7:1680183.
doi: 10.3389/frwa.2025.1680183

COPYRIGHT
© 2025 Camporese, Schilling and Ajami. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Advances in integrated surface—subsurface hydrological modeling

Matteo Camporese[1]*, Oliver S. Schilling[2,3] and Hoori Ajami[4]

[1]Department of Civil, Environmental and Architectural Engineering, University of Padova, Padua, Italy, [2]Hydrogeology, Department of Environmental Sciences, University of Basel, Basel, Switzerland, [3]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, [4]Department of Environmental Sciences, University of California, Riverside, Riverside, CA, United States

KEYWORDS

integrated surface-subsurface hydrological modeling, model coupling, subsurface dynamics, climate sensitivity, operational hydrological forecasting

Editorial on the Research Topic
Advances in integrated surface—subsurface hydrological modeling

Aims and scope of the Research Topic

Integrated surface–subsurface hydrological models (ISSHMs) are at the heart of contemporary water science, combining surface flow dynamics with complex groundwater interactions at multiple scales, from hillslopes to entire continents. As global water systems are increasingly impacted by climate change, anthropogenic pressures, and infrastructure development, the demand for holistic, process-based modeling frameworks has intensified. This Research Topic in *Frontiers in Water* assembles cutting-edge contributions that advance the theoretical, methodological, and practical applications of ISSHMs. Collectively, these works address challenges in coupling, scaling, data assimilation, model accuracy, and computational efficiency, with implications spanning flood forecasting, groundwater management, sediment transport, and climate resilience (Figure 1).

Innovations in model coupling and subgrid representation

One of the persistent challenges in integrated modeling lies in resolving hydrological processes across spatial and temporal scales. Peeples and Maxwell tackle this directly through the development of a subgrid channel formulation within the ParFlow model, offering a refined approach to representing channel hydraulics in coarse-resolution models. Their method significantly reduces flow bias compared to conventional coarse grids, thereby enhancing the accuracy of flood and streamflow simulations in large domains.

Similarly, Paniconi et al. explore model response and numerical challenges across hillslope and catchment scales using CATHY, revealing the hidden instabilities and limitations in ISSHMs under highly heterogeneous and nonlinear conditions. Their empirical findings underscore the necessity of robust coupling strategies, particularly in long-term simulations with dynamic boundary conditions and complex feedbacks.

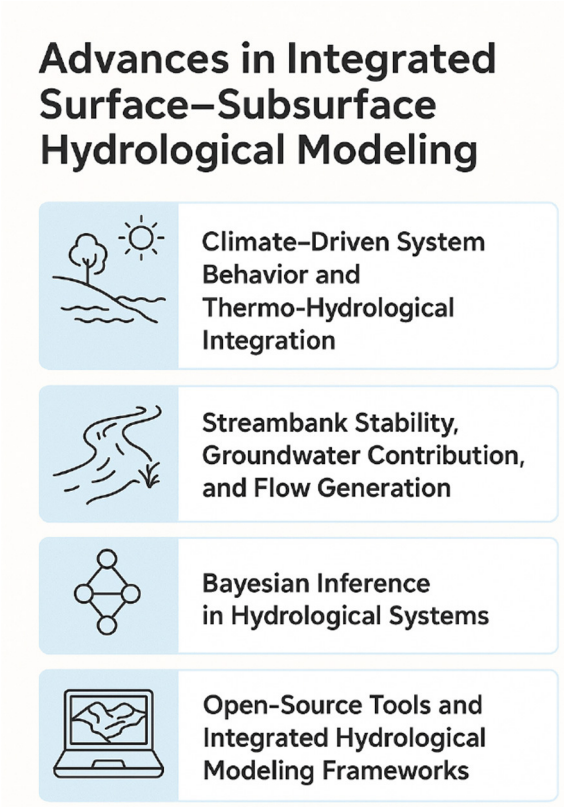


FIGURE 1
Summary of main contributions to the Research Topic.

Subsurface dynamics, climate sensitivity, and hydrological function

Understanding how subsurface systems respond to climatic, structural, and anthropogenic forcings remains a critical frontier in hydrological science. Tsypin et al. address this through a 62-year thermo-hydrological simulation of the North German Basin, combining climate-driven recharge variability from a hydrologic model with geologic heterogeneity in a groundwater flow model. Their modeling results highlight how permeability contrasts, structural dips, and recharge dynamics jointly shape groundwater flow and heat propagation, providing insight into subsurface climate memory and geothermal potential.

In the southern hemisphere, Moore et al. employ a Bayesian hydrochemical approach (BACH) to distinguish shallow and deep groundwater flow contributions to streamflow in New Zealand. Their findings reveal that deep, slow groundwater contributes significantly to streamflow—even during high-flow conditions—demonstrating the year-round relevance of deep aquifers.

Similarly, von Trapp et al. quantify bedrock groundwater contributions in a mountainous catchment using Radon data with a combined mass-balance and 1D steady-state advective solute transport model, accounting for groundwater inflow into the stream. Their findings reveal that bedrock groundwater discharge can supply up to 44% of streamflow during dry seasons, highlighting the hydrologic importance of deep and fractured bedrock in flow generation, especially in headwater systems.

Wei et al. introduce a streambank stability module into HydroGeoSphere, coupled with the surface water operations model, OASIS. Their study shows how reservoir releases and groundwater pumping interact with bank geometry to drive slope instability, emphasizing the need to include infrastructure and anthropogenic stressors in river basin models.

Advances in forecasting and operational hydrological modeling

ISSHMs also play a pivotal role in near real-time forecasting. Patakchi Yousefi et al. investigate the utility of deep learning-based precipitation correction in enhancing integrated hydrologic model accuracy. Although their U-Net CNN model effectively reduces forecast bias, the study finds that uncorrected ECMWF precipitation often leads to better soil moisture simulations, highlighting a paradox where better input statistics do not guarantee better hydrological performance. This result underscores the complex, nonlinear relationships between meteorological input and hydrological response, and the need for careful validation of machine learning corrections in operational workflows.

Goergen et al. demonstrate the predictive capabilities of ParFlow in reconstructing the catastrophic July 2021 floods in the Eifel-Ardennes region. Their ensemble hindcast approach captures both timing and magnitude of the floods without tuning, illustrating the potential of physics-based ISSHMs for disaster reconstruction and analysis. They also provide new perspectives on soil buffering capacity and the role of antecedent saturation in modulating flood severity.

Outlook and future directions

The collective contributions in this Research Topic reflect a vibrant and interdisciplinary community pushing the frontiers of integrated surface-subsurface hydrological modeling. Future work will need to address persistent challenges in:

- coupling hydrological, ecological, and socio-economic processes;
- incorporating machine learning within physically-based frameworks;
- improving computational efficiency for high-resolution, long-term simulations;
- quantifying uncertainties in scenarios of climate and land-use change.

As data availability expands and computing power increases, these integrated frameworks will become central to the design of resilient, data-informed water management systems.

Author contributions

MC: Writing – original draft. OS: Writing – review & editing. HA: Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 21 February 2024
DOI 10.3389/frwa.2024.1353394

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Heather Sheldon,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Jannis Epting,
University of Basel, Switzerland

*CORRESPONDENCE
Mikhail Tsypin
✉ tsypin@gfz-potsdam.de

RECEIVED 10 December 2023
ACCEPTED 31 January 2024
PUBLISHED 21 February 2024

CITATION
Tsypin M, Cacace M, Guse B, Güntner A and Scheck-Wenderoth M (2024) Modeling the influence of climate on groundwater flow and heat regime in Brandenburg (Germany). *Front. Water* 6:1353394. doi: 10.3389/frwa.2024.1353394

COPYRIGHT
© 2024 Tsypin, Cacace, Guse, Güntner and Scheck-Wenderoth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Modeling the influence of climate on groundwater flow and heat regime in Brandenburg (Germany)

Mikhail Tsypin[1,2]*, Mauro Cacace[1], Björn Guse[1,3], Andreas Güntner[1,4] and Magdalena Scheck-Wenderoth[1,5]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute of Applied Geosciences, Technische Universität Berlin, Berlin, Germany, [3]Department of Hydrology and Water Resources Management, Christian-Albrechts-University of Kiel, Kiel, Germany, [4]Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany, [5]Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen, Germany

This study investigates the decades-long evolution of groundwater dynamics and thermal field in the North German Basin beneath Brandenburg (NE Germany) by coupling a distributed hydrologic model with a 3D groundwater model. We found that hydraulic gradients, acting as the main driver of the groundwater flow in the studied basin, are not exclusively influenced by present-day topographic gradients. Instead, structural dip and stratification of rock units and the presence of permeability contrasts and anisotropy are important co-players affecting the flow in deep seated saline aquifers at depths >500 m. In contrast, recharge variability and anthropogenic activities contribute to groundwater dynamics in the shallow (<500 m) freshwater Quaternary aquifers. Recharge fluxes, as derived from the hydrologic model and assigned to the parametrized regional groundwater model, reproduce magnitudes of recorded seasonal groundwater level changes. Nonetheless, observed instances of inter-annual fluctuations and a gradual decline of groundwater levels highlight the need to consider damping of the recharge signal and additional sinks, like pumping, in the model, in order to reconcile long-term groundwater level trends. Seasonal changes in near-surface groundwater temperature and the continuous warming due to conductive heat exchange with the atmosphere are locally enhanced by forced advection, especially in areas of high hydraulic gradients. The main factors controlling the depth of temperature disturbance include the magnitude of surface temperature variations, the subsurface permeability field, and the rate of recharge. Our results demonstrate the maximum depth extent and the response times of the groundwater system subjected to non-linear interactions between local geological variability and climate conditions.

KEYWORDS

groundwater modeling, climate, groundwater level, geothermal potential, groundwater recharge, mesoscale Hydrological Model (mHM), North German Basin, Brandenburg

1 Introduction

Aquifers are sometimes perceived as more resilient and protected from the impacts of climate change and extreme weather events compared to surface waters (Rodella et al., 2023). This is despite the fact that climate variability directly affects the net balance between precipitation and evapotranspiration, which, in turn, defines groundwater recharge into aquifers. In addition, heavy precipitation events can lead to an increase in the runoff component of the water cycle, thereby leaving less water for infiltration (Taylor et al., 2013).

Other feedback effects between climate and groundwater are more complex to quantify. Climate determines human demand for water: during droughts, the demand for irrigation and public water supply results in increased pumping, which might lead to groundwater depletion and land subsidence. Conversely, return flows from surface-water-fed irrigation can cause groundwater accumulation (Taylor et al., 2013). Removal of water from terrestrial water storage contributes to sea-level rise, in addition to thermal expansion and ice melting (Konikow, 2011). Global sea-level rise induces saltwater intrusion into coastal freshwater aquifers (Werner and Simmons, 2009). Changes in the thermal state of an aquifer, as triggered by ongoing global warming, may affect groundwater quality by altering chemical equilibrium and microbiological activity (Riedel, 2019).

Groundwater flow dynamics is the result of non-linear interactions between three primary forcings: climate, geology, and topography (Toth, 1963; Condon and Maxwell, 2015) with a growing influence from human activities (e.g., irrigation, pumping, etc.). The effects of climate change on groundwater dynamics can be assessed by monitoring variations in groundwater levels and near-surface temperatures (Chen et al., 2004; Benz et al., 2017) or, on a more global scale, by analyzing changes in the Earth's gravity field (Thomas and Famiglietti, 2019; Güntner et al., 2023). Projected feedback effects on groundwater resources from climatic forcing are usually quantified by coupling groundwater and surface flow (hydrologic) models (Goderniaux et al., 2011; Dams et al., 2012) or, more recently, by relying on machine learning-assisted regression analysis trained on historical weather and groundwater records (Wunsch et al., 2022). These methods typically estimate changes in groundwater level and recharge/discharge in response to precipitation and surface temperature scenarios. Temperature data is mainly used in hydrologic modeling for computing evapotranspiration. Including thermo-hydraulic coupling into fully-saturated subsurface models opens up a set of additional applications such as estimation of historic recharge rates and tracing groundwater flow (Anderson, 2005; Mather et al., 2022).

In this study, we aim at quantifying how climate-driven forcing (in terms of time- and space-varying recharge and temperature), basin-scale geology (e.g., aquitard discontinuities, deformed strata, and contrasts in rock properties between geological units), and topographic gradients interact with each other to modify the regional groundwater flow and the thermal field. To achieve this, we look at the North German Basin (NGB) beneath Brandenburg (NE Germany) as an example of a porous aquifer system with a shallow-lying water table under a humid continental climate.

Brandenburg is one of the driest federal states in Germany, characterized by a negative and declining climatic water balance, i.e., with potential evapotranspiration (PET) exceeding precipitation (LfU Brandenburg, 2022). The annual mean surface temperature in the region has increased by $\sim$1°C over the past 70 years, with additional recent evidence of rising temperatures at the water table (SenStadt, 2020). Since the onset of continuous observations, precipitation exhibited no statistically confident annual trend, rather showing a high annual variability with a 20% increase of winter precipitation and a minimal increase in the number of days with $>$10 mm of precipitation (Deutscher Wetterdienst, 2019; GERICS-Climate-Service Center Germany, 2019). The estimated groundwater recharge varies between 80 and 150 mm/a, compared to an average of 50–300 mm/a for the entire Germany (Bundesanstalt für Geowissenschaften und Rohstoffeand, 2019). In Brandenburg the majority of groundwater observation wells and groundwater-fed lakes showcase a decrease in water level with a long-term mean rate of 1–3 cm/a since 1970's. Though further changes in atmospheric temperature, precipitation volumes and intensity are expected to put additional stress on the availability of groundwater resources in the region, such effects so far remain to be quantified.

Previous thermo-hydraulic modeling studies in NGB addressed the role of aquitard discontinuities (Noack et al., 2013), of faults (Cherubini et al., 2014), of salt diapirs (Magri et al., 2005; Kaiser et al., 2013), and of interactions with surface waters (Frick et al., 2019) on groundwater and heat transport. These studies aimed to investigate the overprint of the regional gravity-driven groundwater flow and free convection in response to fluid density variations on the conductive heat transport regime in the subsurface. The underlying models commonly assumed steady-state pressure and temperature boundary conditions and approximated the hydraulic head by the surface elevation, therefore forcing the water table to coincide with the topographic relief.

However, to properly evaluate the impact of climate change on the groundwater flow dynamics and thermal regime requires to consider variable (in space and time) boundary conditions across the top of the groundwater model. The absolute position of the water table is a function of topographic elevation, water level in surface water bodies, recharge, and hydraulic conductivity, and is subject to diurnal, seasonal, and longer-term fluctuations. Under specific conditions (e.g., arid climate, high permeability, and steep relief) the water table may get disconnected from topography and become influenced predominately by recharge variability (Haitjema and Mitchell-Bruker, 2005). At the same time deep groundwater dynamics may carry memory effects from past climates, as evidenced in parts of Northern and Central Europe that are still in the process of pressure and temperature re-equilibration following the last glacial maximum, as suggested by postglacial seismicity (Sirocko et al., 2008), presence of relict permafrost in deep wells in Poland (Szewczyk and Nawrocki, 2011) and sub-glacial groundwater modeling (Frick et al., 2022). In order to understand the transient effects of such a complex non-linear system, we coupled a subsurface 3D groundwater model of Brandenburg with a near-surface distributed hydrologic model and simulated the 62-year-long behavior (1953–2014) of the coupled dynamics of groundwater flow and thermal field. The novelty of our approach stems from the integration of basin-scale flow modeling principles (thermo-hydraulic coupling, capturing heterogeneity of an entire sedimentary fill) with modeling tools typically applied in catchment hydrology (transient hydrometeorological forcing, utilizing the true depth of the water table as the boundary condition, rather than assuming the water table is a subdued replica of topography). To our knowledge, this is the first model of this kind, at least in the selected geographic area.

2 Materials and methods

2.1 Geological model

Geologically, Brandenburg belongs to NGB, the largest sub-basin within the intracontinental Central European Basins System (CEBS). It hosts up to 8 km-thick Permian to Cenozoic fill, structurally shaped during its multiphase tectonic evolution by rifting, thermal subsidence, salt tectonics, and more recently, by several stages of glaciation (Littke et al., 2008).

In our study, we make use of an available 3D structural model of the subsurface beneath Brandenburg (Noack et al., 2013). The model was built by integrating a set of depth structure maps and exploration wells and was further constrained against potential field data (Scheck and Bayer, 1999; Maystrenko et al., 2010; Stackebrandt and Manhenke, 2010; Figure 1). The original model spans over parts of five German states and the westernmost part of Poland. In this study we limited its spatial extent within administrative boundaries of the state of Brandenburg in order to ensure consistency of groundwater monitoring data, being collected and stewarded by the countries' administrative bodies independently. The model differentiates 12 geological units, which depth extent has been limited to a constant level of −6,000 m a.s.l. in order to encompass all permeable strata above.

The lowermost model unit includes all formations below the Zechstein salt: the Paleozoic basin fill (Rotliegend sandstone of Permian age, Permo-Carboniferous volcanics, and Carboniferous molasses) is lumped with the Variscan basement, given the high consolidation state and low permeability of these units (see Section 2.4). These rocks are overlain by the Upper Permian Zechstein formation, which dominantly consists of halite with some anhydrite and carbonate. Salt tectonics resulted in numerous salt diapirs, pillows, and walls from post-depositional mobilization, with vertical thickness locally reaching up to 4,000 m. Rim synclines around these salt structures provided accommodation space, which have been filled by Mesozoic-Cenozoic sediments (Scheck-Wenderoth and Maystrenko, 2013).

Stratigraphically above the Zechstein, the lower Triassic Buntsandstein unit is the deepest regional aquifer considered in the model. This succession, dominated by non-marine sandstones and siltstones, is an important target for deep geothermal energy in the NGB (Huenges and Ledru, 2011; Franz et al., 2018). Above it, the middle Triassic Muschelkalk unit, composed of limestones and interbedded marls, acts as a regional aquitard. The overlying units, Upper Triassic Keuper, Jurassic, Lower Cretaceous, Upper Cretaceous, and pre-Rupelian Tertiary, have a heterogeneous composition with permeable sediments represented by sand, silt, and limestone. These deposits form a single brackish to saline aquifer megacomplex, with salinity increasing with depth from ∼1 g/l to >200 g/l (Gaupp et al., 2008). These saline aquifers are a potential target for a number of subsurface utilization projects (geothermal energy, aquifer thermal energy storage, H_2 storage, CO_2 sequestration; Bruhn et al., 2023). Due to salt movements, basin inversion, glacial erosion, and postglacial isostatic rebound, the thicknesses of each individual unit vary significantly with local manifestation of non-deposition and complete erosion of stratigraphic units.

The Keuper-to-pre-Rupelian Tertiary complex is overlain by the Rupelian Clay unit, which acts as the main aquitard separating the freshwater aquifers above from the saline aquifers below. The Rupelian aquitard is eroded locally, providing domains for cross-formational flow between the Tertiary—Quaternary section above the Rupelian Clay and the units below the base Rupelian unconformity. These hydrogeological windows are the widest in the southern part of the study area, where a greater stratigraphic succession is missing, such that Quaternary deposits lay directly on Paleozoic rocks (Figure 2A).

The post-Rupelian Tertiary unit is mainly composed by fluvial and lacustrine clastic deposits. Tertiary lignite has been mined in the region since the end of the eighteenth century, with large open-case mines altering groundwater quality and levels from the second half of the twentieth century (Benthaus and Totsche, 2015; Tissen et al., 2019). Quaternary is the uppermost modeled unit, with its top defined by the surface topography. Lithologically, it consists of glacial, fluvio-glacial, and alluvial sands, silts, and muds with a total thickness of up to 540 m (Figure 2B). Maximum thickness is reached within NE-SW oriented glacial channels, likely formed by subglacial meltwater erosion (Sirocko et al., 2008). Quaternary channels incised deep into underlying Tertiary rocks, sometimes eroding the Rupelian Clay.

The present-day relief in Brandenburg has been shaped by the last three Pleistocene glacial and intraglacial cycles (Stackebrandt and Manhenke, 2010). Consequently, elevated belts of moraines align preferentially along WNW-ESE oriented axes, separated by wide ice-marginal valleys. Absolute elevations range from 0 to 200 m a.s.l., with elevations above 50 m being typical for plateau belts, and those below characteristic for valleys and floodplains. Large differences in height occur over short distances (Stackebrandt and Manhenke, 2010).

2.2 Model parametrization

We have assigned constant physical properties to each stratigraphic unit, chosen consistently with its predominant lithology (Table 1). Despite recent advances in facies-dependent petrophysical characterizations for specific sites and formations within the NGB (BGR-SGD, 2019; Norden et al., 2023), we chose to rely on a homogeneous parameterization due to the regional span of the model, which together with an uneven data coverage hinders a more differentiated litho- or sequence-stratigraphical mapping at a basin scale. The selected rock properties are consistent with published parametrized models of the same geographic area and with the results of sensitivity analyzes conducted herein (Noack et al., 2013; Frick et al., 2019). Permeability is typically much higher parallel to bedding than orthogonal to it. We approximated this anisotropy, by orienting principal directions of permeability vertically and horizontally. Assigned vertical component (Kz) is set one order of magnitude smaller than lateral components (Kxy). This is a common assumption for stratified sedimentary materials, while Kz/Kx(y) ratios measured in cores may vary substantially depending on mineralogy and lithology (Domenico and Schwartz, 1997). We assign an isotropic permeability of 1e-19 m^2 to both, the pre-Salt and Zechstein units, which act as

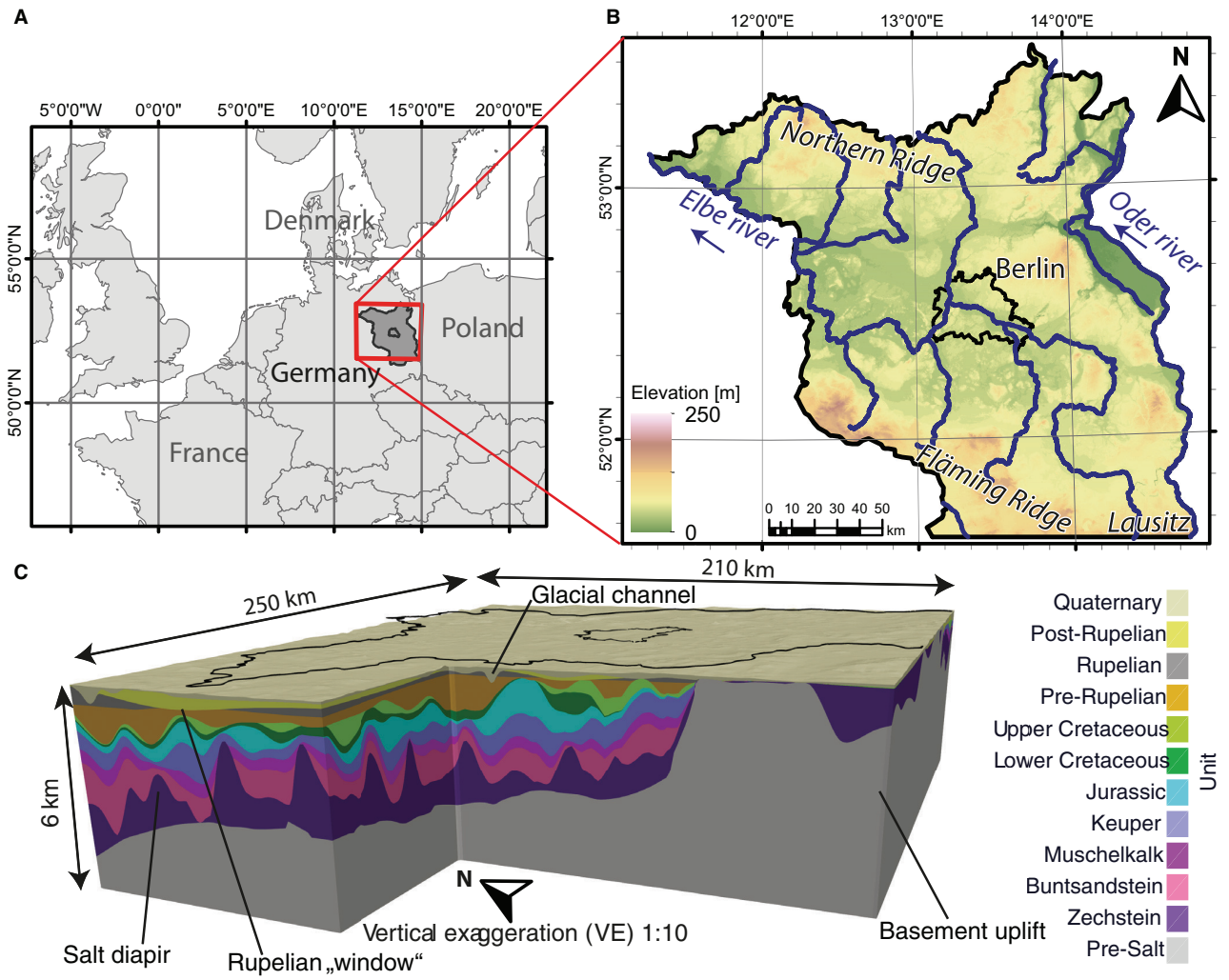


FIGURE 1
(A) Location of the study area; **(B)** topographic map; **(C)** 3D view of the structural model. Black outline follows Brandenburg administrative boundary, except the southern latitudinal edge, which corresponds to the model 3D model limit.

a natural impermeable base to fluid flow, while still allowing for heat conduction. Consequently, the model does not account for groundwater movement in Rotliegend sandstones, marginal carbonate-dominated facies of Zechstein formation, or weathered basement uplifts. Due to its composition, the Zechstein unit has a higher thermal conductivity than the rest of the basin fill, with salt structures acting as chimneys for conductive heat transfer on a basin scale (Noack et al., 2013).

Fluid properties are considered constant for all units (Table 1), with viscosity 0.001 Pa·s and fluid modulus 2.18e+9 Pa, except for the Quaternary unit, for which we set a higher fluid modulus equal to 2.18e+5 Pa to approximate the higher storativity in the aquifer with unconfined conditions, consistent with an equivalent specific yield of 0.1. In our study, we neglect variations in fluid density and viscosity with depth due to salinity and temperature stratification. Our choice stems from previous studies, which demonstrated how density driven free convection is unlikely to be a significant heat transport mechanism in the NGB (Kaiser et al., 2011). Given the modeled temperature range of interest (~200°C), an ~1 order of magnitude variations in water viscosity is expected (Kestin et al., 1978). Therefore, we may underestimate the hydraulic conductivity at greater depths. However, the impact of such variations in fluid viscosity on the regional Darcy flux can be considered of secondary relevance than that resulting from the heterogeneous permeability field, which also decreases with depth.

2.3 Numerical model set-up

We have conducted all simulations presented below with GOLEM, a Finite Element Method (FEM) modeling platform for thermal-hydraulic-mechanical and non-reactive chemical processes in fully-saturated porous media (Cacace and Jacquey, 2017). GOLEM relies on a

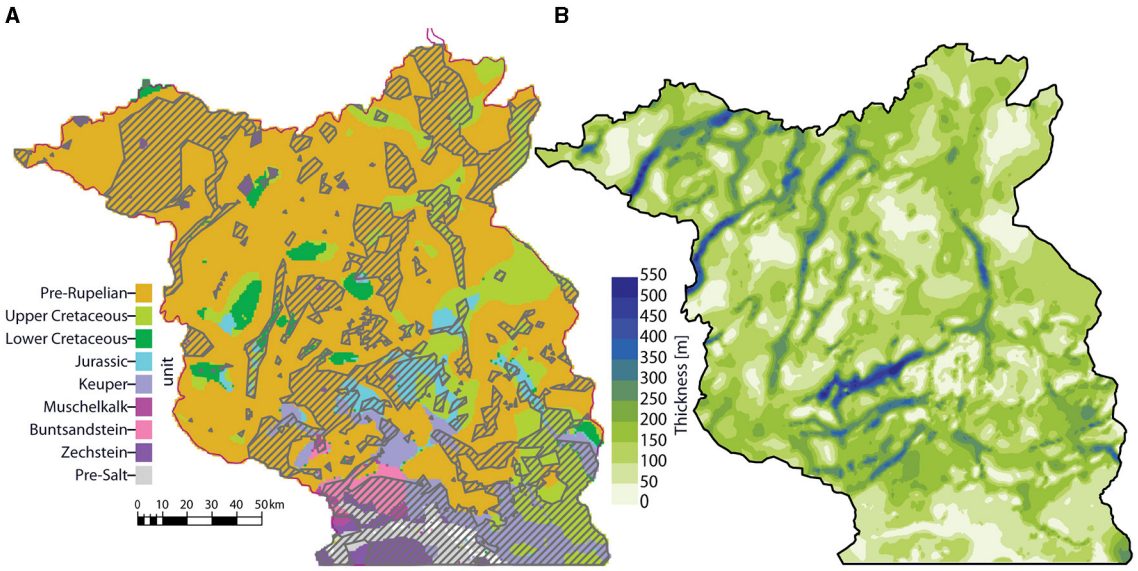


FIGURE 2
Geologic controls on groundwater flow, as extracted from the 3D geological model, relevant surface-groundwater interactions: **(A)** Rupelian subcrop map with dashed areas corresponding to absence of Rupelian aquitard; **(B)** Quaternary thickness map, where thicknesses of 300–550 m correspond to SW-NE-oriented glacial channels.

TABLE 1 **Petrophysical properties of the model units assigned in the numerical simulations.**

Unit	Lithology	Thermal conductivity λ [W/(m*K)]	Porosity ϕ [%]	Radiogenic heat production Q_r [W/m^3]	Specific heat capacity c [J/kg·K]	Bulk density ρ [kg/m^3]	Permeability k_{xy} [m^2]
Quaternary	Sand, silt, clay	1.5	23	7.0E-07	1,620	2,100	1.0E-13
Tertiary, post-Rupelian	Sand, silt, clay	1.5	23	7.0E-07	1,640	2,100	5.0E-14
Rupelian	Clay	1.0	20	4.5E-07	1,810	2,400	1.0E-16
Tertiary, pre-Rupelian	Sand, silt, clay	1.9	10	3.0E-07	1,700	2,100	1.0E-14
Upper Cretaceous	Limestone with marl	1.9	10	3.0E-07	2,290	2,215	1.0E-13
Lower Cretaceous	Marl with claystone	2.0	13	1.4E-06	2,290	2,215	1.0E-13
Jurassic	Claystone with silt- and sandstone	2.0	13	1.4E-06	2,250	2,450	1.0E-13
Keuper	Claystone with marl and gypsum	2.3	6.0	1.4E-06	2,320	2,530	1.0E-14
Muschelkalk	Limestone with marl	1.85	0.1	3.0E-07	2,250	2,530	1.0E-18
Buntsandstein	Sandstone, siltstone	2.0	4.0	1.0E-06	2,390	2,530	1.0E-14
Zechstein	Salt, gypsum, evaporite	3.5	~0	9.0E-08	1,940	2,150	1.0E-19
Pre-Salt	Compacted clastics, volcanics	2.5	~0	1.5E-06	2,600	2,600	1.0E-19
Water		0.65			4,180	1,000	

Thermal conductivity, porosity, permeability, and radiogenic heat production values are taken from Noack et al. (2013), specific heat capacity estimates taken from Frick et al. (2019), bulk density values are taken from Frick et al. (2022), and Noack et al. (2013).

pressure-based formulation to describe the groundwater mass balance as:

$$\frac{\phi}{K_f}\frac{\partial p_f}{\partial t} + \nabla \bullet \left(\frac{-k}{\mu_f} \cdot \left(\nabla p_f - \rho_f g\right)\right) = 0 \quad (1)$$

where ϕ–porosity, K_f–fluid modulus, ρ_f —fluid density, k—permeability, μ_f–fluid viscosity, g—gravitational acceleration. The first term of Equation 1 describes the change in water storage under the assumption of incompressible rock matrix), and the second term describes the gradient of Darcy flux.

Equation 1 is implicitly coupled with an equation describing the conservation of the internal energy of the porous system, considering change in heat storage, as well as advective and conductive transport components as:

$$(\rho c)_b \frac{\partial T}{\partial t} + \nabla \bullet \left((\rho c)_f \frac{-k}{\mu_f} \cdot \left(\nabla p_f - \rho_f g\right) T\right) - \nabla \cdot (\lambda_b \nabla T) - Q_r = 0 \quad (2)$$

where c—specific heat capacity, b and f subscripts denote bulk or fluid properties, T—temperature, λ–thermal conductivity, Qr—rate of radiogenic heat production.

An earlier coupled simulation study in NGB suggested that free convection (density-driven flow due to temperature and/or salinity gradients) can be considered as a secondary source of groundwater flow acting only locally within the most permeable and thick Mesozoic section and in the absence of any significant hydraulic gradients (Kaiser et al., 2013). Since our study focuses on climatic drivers over a time window of six decades, which will affect primarily the shallow freshwater aquifers, we decided to neglect double-diffusive convection in our simulation set-up.

When converting the input geological model into a 3D FEM mesh, each modeled stratigraphic unit has been vertically subdivided into four computational layers, while a higher resolution has been applied to the Quaternary unit, which was refined by a factor of 10 to better approximate the dynamics near the surface-subsurface interface. The lateral resolution of the FEM mesh has been kept to 1 km x 1 km as in the original geological model. The final model, divided into 54 computational layers, consists of 1.59 million nodes, giving a total of 3.18 million degrees of freedom.

Equations 1, 2 form an initial boundary value problem, the closed form solution of which requires to specify a set of initial and boundary conditions. We have first derived steady-state conditions by solving separately for the hydraulic (Equation 1) and the thermal (Equation 2 with only the conductive term included). These uncoupled steady-state simulations have been used as initial conditions to run a coupled pseudo-transient simulation, the results of which have been later imposed to initialize the pore pressure and the temperature in the final transient simulation.

For the steady state runs, we assume hydrostatic conditions with respect to a spatially variable hydraulic loading at the surface as given by a fixed hydraulic head. The initial distribution of hydraulic heads for the Quaternary unit across the study area is based on ~100 groundwater level measurements taken in 1953 together with a trend from the water table contour map of 1999 (Landesamt für Umwelt Brandenburg, 2020). The resulting hydraulic heads grid has been converted into a pressure boundary condition (Figure 3A) as:

$$P = -\left(Z - h\right) * \rho g + P_0 \quad (3)$$

where P is the assigned pressure at the top boundary, P_0 is 1 bar (pressure at the water table), Z is ground elevation, h is the hydraulic head.

To better approximate the shallow groundwater dynamics, we additionally integrate the 15 longest rivers in Brandenburg as a node set to which we assign a Type I (Dirichlet) pressure boundary condition with a constant value of 1 bar, thereby forcing groundwater level to match the model top at these locations.

We close the side edges of the 3D model, therefore assuming that no significant lateral flow or heat transport occurs into or out of the domain. This assumption particularly applies for the shallowest aquifer, for which the Oder river in the east, the Elbe river in the west, and watershed divides in the north and south act as natural hydraulic no-flow boundaries (Figure 1B). There is a higher potential of outflow from the deeper aquifers, that, according to regional geology, dip northward, extending beyond the model domain, while onlapping onto the basement high to the south. The impact of our choice of side boundary conditions is discussed further in Section 4.4. A Type I temperature boundary condition is assigned along the top surface as derived from grids of monthly averaged daily air temperature 2 m above the ground from the German Weather Service (Deutscher Wetterdienst—Climate Data Center, 2021) and averaged over a period of 1951–1953.

Along the base of the model, we assign no flow and a Type I temperature boundary conditions (Figure 3B). Given the extremely low permeability at a depth of −6,000 m, flow through this boundary can be indeed considered as negligible. The basal temperatures have been extracted from a lithosphere-scale thermal conductive model of Brandenburg, which has been additionally validated against temperature measurements from deep exploration wells (Noack et al., 2012). Basal temperature ranges from 230 to 160°C, reflecting difference in the thickness of the crystalline crust and the sedimentary cover. Lower temperatures in the southwest corner are caused by an almost complete absence of any sedimentary cover on top of the uplifted basement (Figure 1C). Without insulation from the low-conductive sediments, highly conductive basement rocks efficiently dissipate heat toward the surface. Higher temperatures in the center correlate with a thicker felsic upper crystalline crust, generating more heat from radiogenic decay, covered by 3–5 km of insulating clastic sediments (Scheck-Wenderoth and Maystrenko, 2013).

In a second step, we use the results from the uncoupled thermal and hydraulic runs as initial conditions to a pseudo transient coupled thermo-hydraulic simulation, that is, a transient simulation which has been run till reaching equilibrium conditions. For this run we impose the same set of boundary conditions as described above. We set a constant timestep of 1,000 years and considered steady state conditions if the maximum change in temperature and pressure per simulation step drops to values lower than 0.1°C or 10 kPa, respectively. Steady-state conditions were achieved after ~120,000 years.

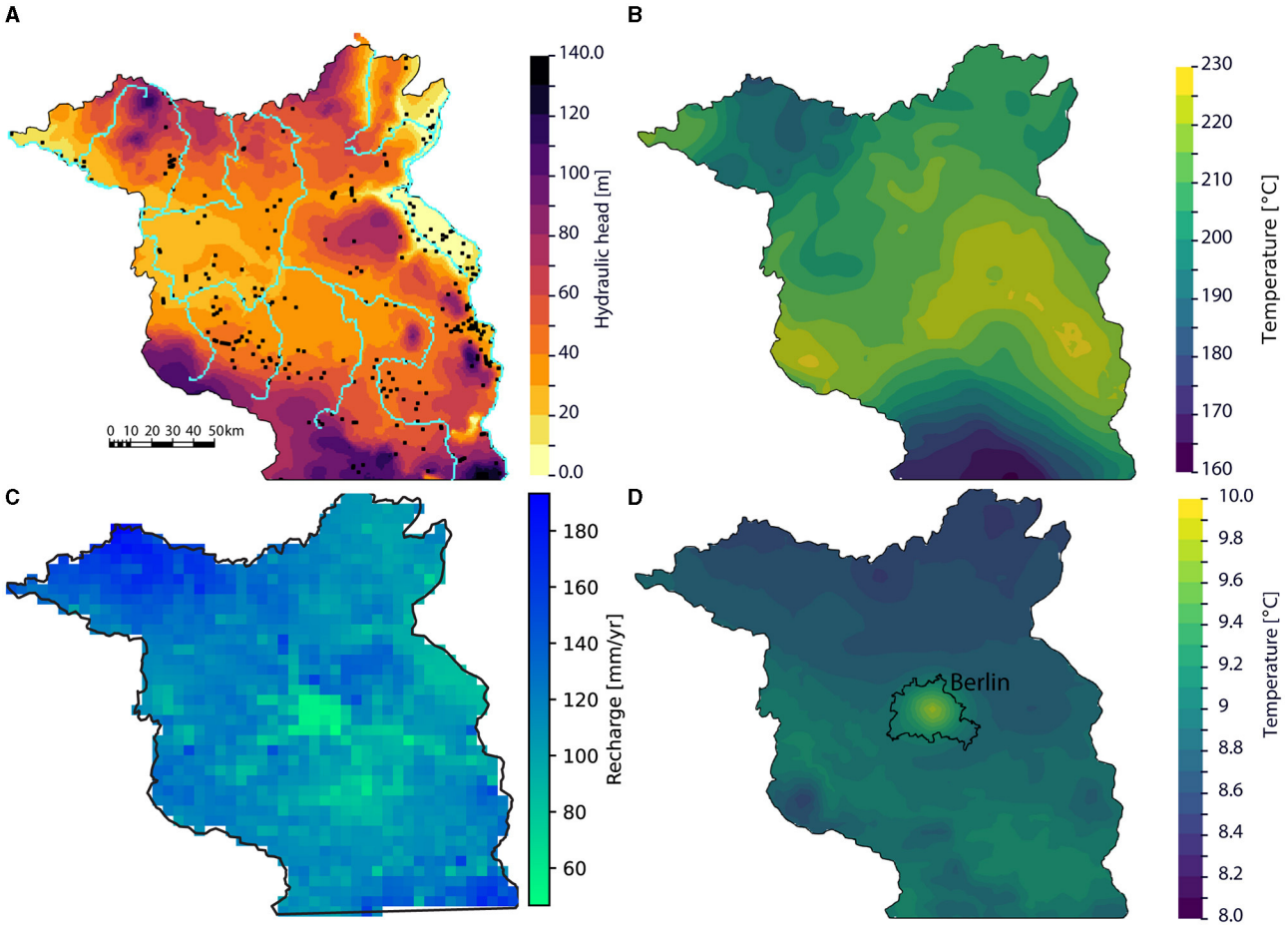


FIGURE 3
Source grids used for boundary conditions assignment: **(A)** pressure at the top boundary with groundwater level measurement points (Landesamt für Umwelt Brandenburg, 2023) and river boundary conditions; **(B)** basal temperature at −6,000 m a.s.l.; **(C)** mean annual groundwater recharge 1953–2014; **(D)** mean surface temperature 1953–2014 (Deutscher Wetterdienst—Climate Data Center, 2021).

In the final simulation step, we take the output of the pseudo transient coupled simulation as initial conditions for the transient thermo-hydraulic simulation to quantify the influence of time-varying climate-driven surface forcing conditions over the period from 1953 to 2014, with a timestep of 1 month. This period was chosen due to availability of all input data required to compute recharge in a hydrological model. As a caveat, the onset of the transient simulation coincides with increasing groundwater abstraction due to population growth and mining activities in Brandenburg.

The temperature boundary conditions imposed along the top surface have been derived from a spatial interpolation of the monitored monthly air temperature time series (Deutscher Wetterdienst—Climate Data Center, 2021). Mean surface temperatures across the study area vary between +8.5 and +10°C, with a decreasing trend from south to north and a positive thermal anomaly in the center of the study area associated with the urban heat island of the capital city of Berlin (Figure 3D). Time series of the surface temperatures indicates an annual difference between the coldest and the warmest months varying from 15 to 21°C, and an increase of mean annual temperature of ∼1°C during the simulated period (Figures 4A, B).

Many hydrologic models that employ water balance components, climatic data, and physics of the unsaturated zone, are able to simulate groundwater recharge from watershed to global scales. For the current study, we make use of the mesoscale Hydrological Model (mHM; Samaniego et al., 2010). mHM is a spatially distributed hydrologic model that uses gridded observed surface temperature and precipitation as inputs. We make use of the results from a Germany-wide realization of mHM to derive time and space varying water fluxes, which we translate into boundary conditions at the top of our groundwater model. The output of the mHM model consists of monthly 5 km x 5 km grids as used in Rakovec et al. (2016) and Samaniego et al. (2019) covering the period 1953–2014, with the following response fluxes relevant to the current study, as defined in Samaniego et al. (2010):

$$C(t) = \beta_{22} x_5 (t-1) \quad (4)$$

$$q_4(t) = \beta_{23} x_6 (t-1) \quad (5)$$

where C is percolation from the soil layers to the groundwater reservoir (i.e., groundwater recharge), [mm/d]; q_4 is baseflow,

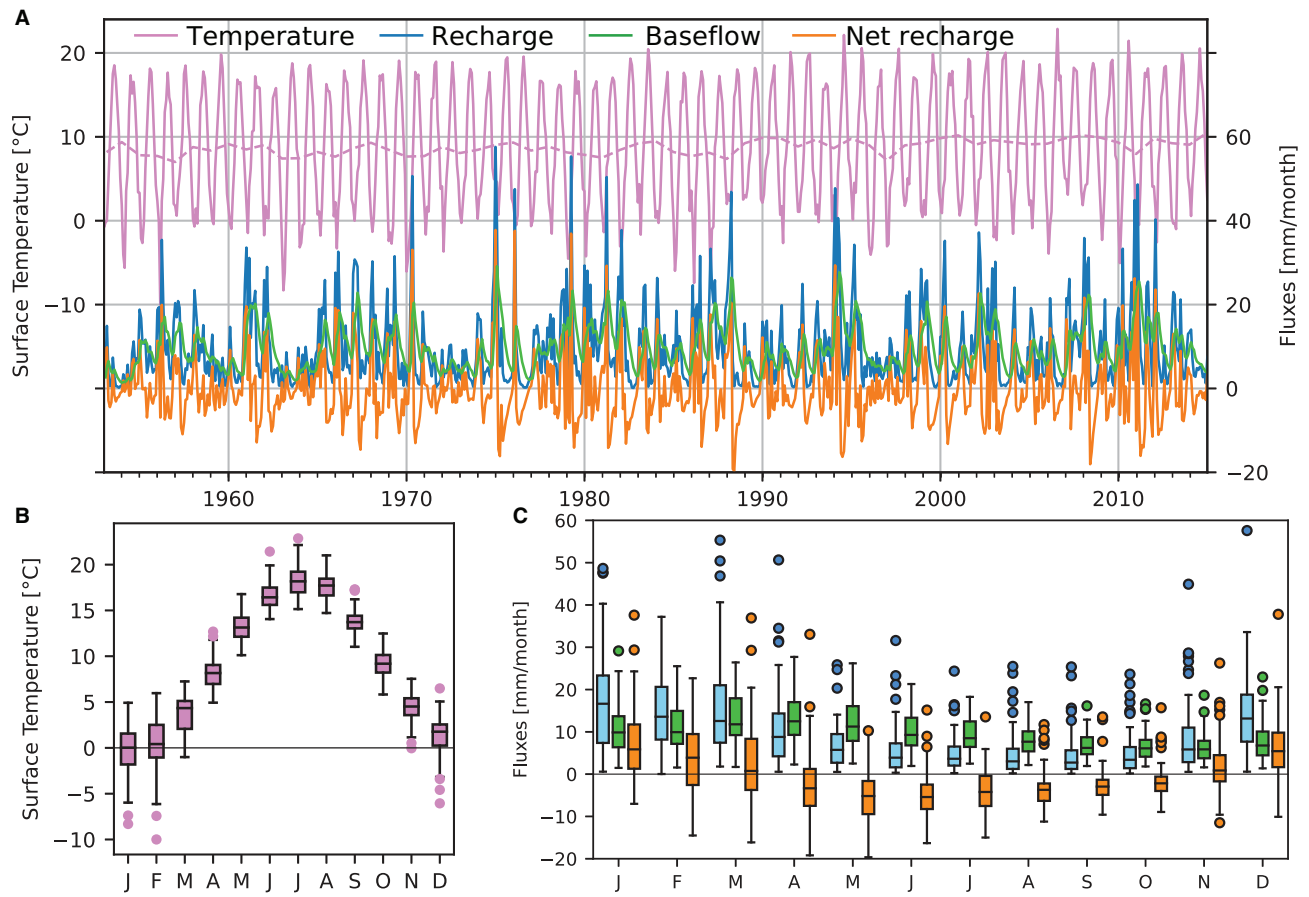


FIGURE 4
Hydrometeorological monthly time series, mean values across the study area. **(A)** Surface temperature and mHM output fluxes; dashed line corresponds to annual temperature mean; **(B)** boxplots of monthly temperature **(B)** and mHM output fluxes **(C)** across the study area for the entire simulation period.

[mm/d]; β_{22} is an effective percolation rate [1/T], which depends on the hydraulic conductivity of the soil; β_{23} is a baseflow recession rate [1/T], which defines how the release of water from the groundwater reservoir to surface runoff in the stream network must be distributed over time; x_5 defines the amount of water storage in the lower soil reservoir, expressed in equivalent water height, [mm]; and x_6 the amount of the water storage in the groundwater reservoir, [mm]; t is time index for each time interval.

mHM is a process-based bucket-type hydrological model with a spatial regularization on regular grids. All hydrological processes are calculated first separately for each individual grid. The flow from different runoff components (e.g., fast and slow interflow and baseflow) is then routed from cell to cell, according to the largest topographic gradient as derived from the digital elevation model. The spatial parameterization approach (multiscale parameter regionalization, MPR) leads to spatially consistent outputs of hydrological states and fluxes (Samaniego et al., 2010; Kumar et al., 2013). There is no subsurface lateral flow redistribution between grid cells in mHM, though this may be relevant especially at the regional scale, where groundwater recharge and discharge areas are usually situated away from each other (Jing et al., 2018).

We apply the net recharge flux, being the difference between recharge and baseflow fluxes in an mHM cell at any given time, as a Type II (Neumann) pressure boundary condition, where positive net flux indicates that recharge outweighs baseflow, and vice versa (Figure 4A). Given that for each grid cell of mHM the sole source of baseflow is recharge in the corresponding cell, over a sufficiently long period the net flux will be close to zero. It follows that cells with a high recharge flux also have a high baseflow component and vice versa. Lateral groundwater flow along a topographic gradient is not represented in the hydrological model (no lateral/horizontal fluxes between adjacent model cells). To alleviate for this limitation of the hydrologic model and to allow for lateral groundwater flow we have integrated the major rivers in the study area into the groundwater model with a constant pressure (1 bar), ensuring the presence of a regional hydraulic gradient at all times.

For the modeling and data analysis, we limited the mHM output to the spatial extent of the 3D geological model, and interpolated the data onto the mesh. Long-term recharge increases in the NW direction up to a maximum of >160 mm/a and spatially correlates with a seaward increase in precipitation (Figure 3C). Southern Brandenburg, characterized by a hilly terrain, also has relatively higher recharge rates. The lowest recharge (<60 mm/a) is modeled for the Berlin metropolitan area due to its urban land cover.

Figure 4C illustrates the 12-month long time series derived from averaging the mHM fluxes across the whole study area and the entire simulation period. Groundwater recharge peaks between January-March, and reaches its minimum in August-September, being therefore temporally linked to the seasonality of evapotranspiration rather than precipitation. The peak in baseflow typically lags 1–3 months behind the peak in recharge followed by a 7–11-month long recession. According to the sign of net recharge flux, groundwater typically gains storage from November to March and loses storage from April to October.

2.4 Data for model validation

The final output of thermo-hydraulic simulation contains modeled pressure and temperature at every mesh node for monthly timesteps. We use the open-source ParaView software (Ayachit, 2015) for post-processing and analysis, including back-calculation of hydraulic heads from the pressure field.

We validated the model results against published groundwater level and temperature data. The groundwater monitoring network of Brandenburg comprises ∼1,900 observation wells and piezometers with daily to quarterly measurements (Landesamt für Umwelt Brandenburg, 2023). While the earliest groundwater level records date back to the 1920's, a more uniform spatial coverage (>500 localities) is only available from 1970 onward, making possible to have a higher-confidence validation dataset for the modeled time series (Figure 3A). Observation wells tap exclusively Quaternary and Tertiary unconfined, confined, and perched aquifers. Differences in well construction, spatial clustering of sometimes conflicting data, and uncertainty in the well penetration levels make it challenging to reconstruct consistent potentiometric maps over the entire Berlin-Brandenburg region for every timestep. We therefore utilize a limited set of gridded equipotential line maps that incorporate quality-controlled groundwater and surface water data (Landesamt für Umwelt Brandenburg, 2020), while relying on individual wells for time series analysis.

Groundwater temperature is monitored in ∼200 observation wells in Berlin and ∼200 wells in Brandenburg and is limited to 80 m below ground (SenStadt, 2020). Static measurements in the deeper pre-Cenozoic section come from corrected bottom-hole temperatures and repeated temperature logs in ∼50 petroleum exploration wells (Förster, 2001; Norden et al., 2008). They have been previously used for calibration of steady-state models of Brandenburg (Noack et al., 2013), from which we derived the lower thermal boundary condition for calculating the deep thermal field in the present study (Supplementary Figure 1).

3 Results

3.1 Groundwater flow: the initial state

The aim of this model realization is to capture equilibrium groundwater conditions under a fixed, though spatially variable reference head and surface temperature prior to the onset of the transient simulation (1953). As such, it provides a first-order description of the groundwater dynamics and resulting thermal field through the entire stratigraphic section without considering the influence of climatic cyclicity (i.e., neglecting seasonal changes in recharge and surface temperature) or continuous climate warming.

Analytical models of gravity-driven flow in a homogeneous isotropic basin suggest that the superposition of the regional slope and local water table relief is responsible for forming flow cells that differ in distance between recharge and discharge areas and their depth of penetration (Toth, 1963). Consequently, groundwater flow in sedimentary basins is often subdivided into regional, intermediate, and local flow systems. In what follows we examine the results based on computed groundwater flowlines and Darcy flux, focusing on three stratigraphic intervals of interest: Buntsandstein, Keuper—pre-Rupelian, and post-Rupelian—Quaternary (Figure 5).

As an example of a regional groundwater system, we look at the Buntsandstein unit (Figure 5A), the deepest modeled aquifer, bounded below and above by impermeable Zechstein and Muschelkalk units. The spatial extent of individual flowlines (computed as the distance between the entry and exit points of the groundwater flow) spans over 100 km, indicating a regional, long-distance flow with long residence times. The dominant direction of the flow is from the south to the center and then to the northwest and northeast. Such an orientation is defined by the reservoir dip, which is, in turn, controlled by the basin configuration. In the southern marginal part of the basin, the Buntsandstein is uplifted to <-1,000 m a.s.l., while in the northwestern part it is buried below −3,500 m a.s.l. The Buntsandstein aquifer is likely recharged in the south of Brandenburg, where it lies at depths of <200 m below ground directly underneath a thin Cenozoic cover in the absence of Muschelkalk and Rupelian aquitards. Following the steeply dipping Buntsandstein, Darcy flux ranges between 1e-4 and 1e-3 m/s, and water can freely flow downward. As groundwater flow becomes dominated by a more lateral component, Darcy flux decreases to 1e-5 – 1e-4 m/s. Groundwater flow is diverted locally, in the northeastern quadrant of the model, where large Zechstein salt diapirs and salt walls are present. Salt structures act as a local gradient anomaly being associated to elevated Darcy flux. Groundwater tends to sink into the salt rim synclines and to rise over salt structures, following the folded topology of the Buntsandstein reservoir. Due to the no-flow boundary along the model edge, groundwater leaves the aquifer via cross-flow to the adjacent units, eventually discharging at the surface. A larger basin model of the whole Northern Europe suggested that this deep groundwater flow actually continues north-west toward the most buried offshore parts of the NGB (Frick et al., 2022).

The intermediate flow system includes units from Keuper to pre-Rupelian Tertiary and lies between two regional aquitards, the Muschelkalk below and the Rupelian above. The thickness of this aquifer complex increases from <1,000 m south of Berlin to more than 2,000 m in the north-western part of Brandenburg, where fluid circulation is also more active (Figure 5B). Erosional windows in the Rupelian aquitard provide pathways for groundwater cross-flow to and from the post-Rupelian—Quaternary aquifer complex. The intermediate flow system is recharged, where hydrological windows lie below topographic highs, while discharge occurs where the windows are found below topographic lows. Therefore, the flow travel distances structurally depend on the distance between

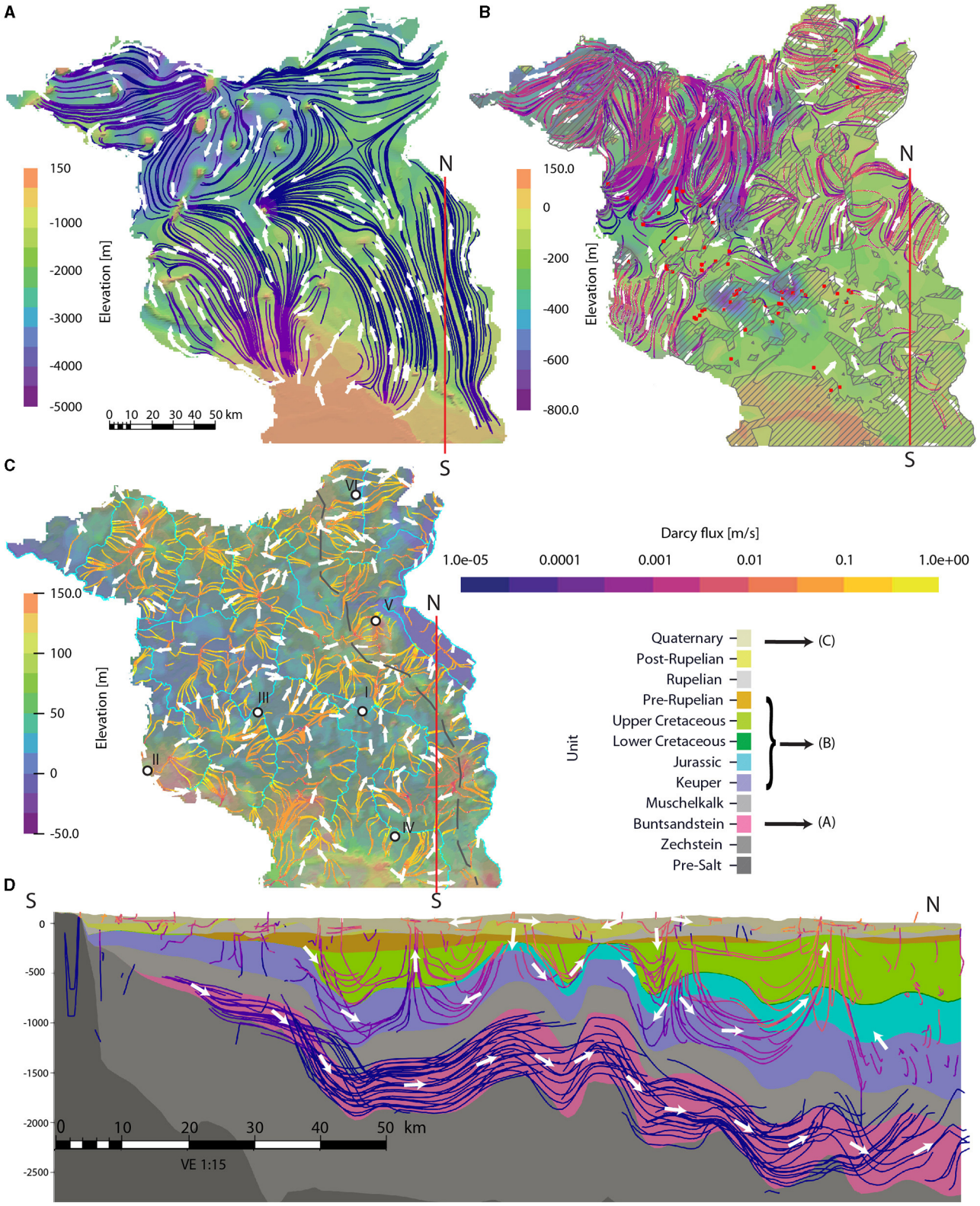


FIGURE 5
Groundwater flowlines under steady-state conditions: in Buntsandstein **(A)**, in Keuper to pre-Rupelian Tertiary untis **(B)**, in Quaternary **(C)**, in all units along a S-N slice **(D)**. Coloring of the maps show top aquifer elevations. Roman number annotations refer to locations shown on Figures 7, 10. Dashed areas refer to Rupelian windows, arrow glyphs indicate flow direction, streamline color indicate Darcy flux, dashed gray line—Oder-Elbe watershed divide, red squares—evidence of saltwater discharge at the surface as given by Rößling et al. (2010).

adjacent erosional windows. The presence of saltwater springs and halophilic plants on the surface as mapped by Rößling et al. (2010) is consistent with the upward flow through Rupelian windows in the central and northeastern parts of the study area. Darcy flux ranges from 1e-5 to 1e-2 m/s with highest flux in the permeable Jurassic and Cretaceous units. Individual salt structures cause local diversion of the flow from the shortest path between inflow and outflow points. In rim synclines the vertical flow dominates over the lateral component of the flow.

Groundwater dynamics of the local flow system, that includes post-Rupelian Tertiary and Quaternary units, is fundamentally different from what has been described above (Figure 5C). Higher Darcy flux (1e-2 – 1e-1 m/s) and shorter travel distances lead to shorter groundwater residence times. There is no predominant flow direction or continuous streamlines at the scale of the whole Brandenburg. Instead, the present-day watersheds configuration controls the orientation of groundwater flow, which follows the topographic gradient, from elevated areas (glacial plateaus, acting as watershed divides) toward the nearest adjacent valley or stream. Due to the "hummocky" post-glacial relief of Brandenburg, groundwater forms numerous domains of recharge and discharge. A N-S drainage divide (gray line on Figure 5C) separates groundwater catchments contributing to the Elbe and Oder river basins.

Groundwater levels vary from −5 to 140 m a.s.l. across the study area (Figure 6A). Depth to the water table, calculated as the difference between surface elevation and groundwater level, lies predominantly between 0 and 30 m, with 5% of model cells having a deeper water table (Figure 6B). In places where recharge derived from mHM is abundant, the water table is shallower and the groundwater level shows a preferential correlation with the topographic elevation, i.e., the potentiometric surface replicates the topographic relief. As recharge decreases, groundwater level correlates less with the topography. At surface elevations >100 m, the groundwater level becomes more variable and does not follow the surface relief even in those model cells characterized by a high estimated recharge (Figure 6A). Model cells with a surface elevation to groundwater level ratio >2 can be explained by the additional artificial lowering of the water table from pumping and drainage.

3.2 Groundwater flow: transient state

In order to visualize how variations of the imposed boundary condition over time affect the local groundwater system, we plot in Figure 7 time series of computed Darcy flux for two selected locations, on a glacial plateau (point V in Figure 5C) and in a river valley (point VI in Figure 5C). We found that variations in the vertical component of the Darcy flux show a high degree of correlation with variations in the net recharge. In late winter and spring, when recharge is higher than baseflow (i.e., net inflow), the vertical downward component of the Darcy flux is enhanced, while showing an opposite trend in summer and early fall. The lag time between net recharge and Darcy flux peaks at a depth of 25 m below ground varies from 1 to 6 months, being shorter in the valley location. Such correlation is even more evident if looking at specific years characterized by exceptionally high recharge (as in 1994 and in 2002), where we observed a reversal in the vertical flow component in the valley location, otherwise dominated by an upward flow (marked by green boxes in Figure 7). Seasonal changes in groundwater flux magnitudes and directions are limited to the post-Rupelian section: the annual cycle in the surface flux signal does not propagate through erosional windows, as can be seen from steady velocities at −200 m, corresponding approximately to the top of pre-Rupelian unit for both locations. We therefore observe no modification in the regional and intermediate groundwater flow during the whole 62-year time period investigated.

Figure 8 depicts computed differences between modeled and observed water levels at the end of the transient simulation. The Pearson correlation coefficient of 0.98, with a RMSE of 1.92, indicates that the model predicts the observed groundwater level well with +/- 2 m of misfit over the study area. An anomalously high misfit of ±10 m in two localities in the southeast of the study area is associated with lignite mining activity (Lausitz region), which was on the rise until 1985, followed by a gradual closure and recultivation of open-cast mines. To access the coals seams in Tertiary and Quaternary units, the water level had to be lowered by 30–50 m. The area of misfit spatially correlates with depression cones, which extend well over the borders of the opencast mines (Benthaus and Totsche, 2015).

In valleys and floodplains, where groundwater level is near the surface and therefore less variable, the misfit between simulated and observed level lies within a confidence interval of ± 2 m. A misfit up to +8 m is found around topographic heights, where the groundwater level is also deeper. According to hydrogeological mapping, near-surface porosity and permeability is lower on the highs than in the river valleys (BGR-SGD, 2019). Therefore, we can anticipate higher magnitude fluctuations in these regions.

The amplitude of hydraulic head variations in the model is a function of the net recharge flux at the upper boundary and of the aquifer storativity. As commonly done, we can also make use of the relationship that estimates recharge following a Water Table Fluctuation (WTF) method (Healy and Cook, 2002). For a specific yield of 0.1, as the one assumed for the Quaternary unit, and given recharge rates as typical for Brandenburg (60–180 mm/yr) the expected seasonal range of water table fluctuation, i.e., the difference between the highest and the lowest position of the water table in a given year, is between 0.3 and 0.9 m, assuming an equal duration of rise and fall of 6 months. If we compute the seasonal fluctuations based on all measured wells, we arrive at values of 0.17 m [10th percentile (P10)], 0.68 m (P90), and 0.41 (P50). Modeled average seasonal fluctuations in the groundwater level range between 0.25 m (P10) and 0.48 m (P90) with P50 of 0.36 m, thereby in agreement with measured values. We note a positive correlation between magnitudes of seasonal fluctuations and recharge, with the maximum fluctuations in NW and SE, and the lowest fluctuations in the center (Berlin) and in E-NE (Figure 9). We also note an overprint of the local geology: in the Quaternary glacial channels the average seasonal fluctuations are lower than outside of the channels because of the greater porosity and higher storage.

To further validate the model outcomes, we cross-plot the simulated groundwater level time series against measurements

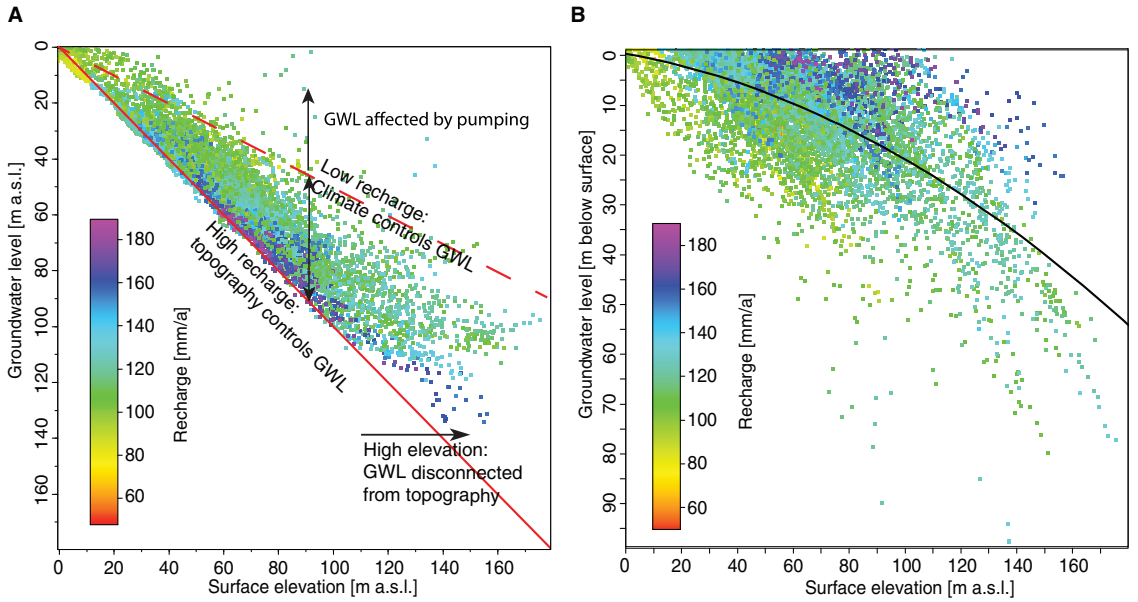


FIGURE 6
Controls on steady-state groundwater level. **(A)** Topography and groundwater level elevation, with 1:1 and 2:1 ratios shown as solid and dashed red lines for reference in solid red; **(B)** topography and depth to water table; a line of best fit shown in black. Each point represents a single grid cell of the groundwater model.

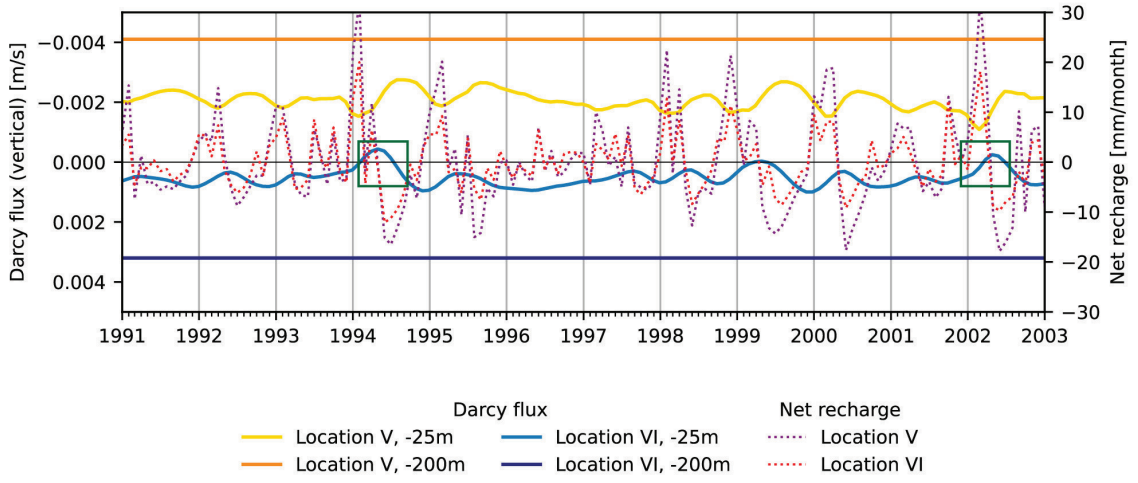


FIGURE 7
Time series of net recharge flux and simulated vertical Darcy flux at two locations, shown on Figure 5C. Location V is on top of a glacial plateau, while location VI is in a river valley. Positive net recharge flux corresponds to the gain of groundwater storage, negative Darcy flux corresponds to downward flow.

for several representative observation wells, within different geomorphological and hydro-geological settings (Figure 10).

Well I is representative of a river valley or floodplain setting, characterized by a very shallow water table (0–5 m below ground). The wellhead location is ~0.5 km away and 4 m above a river. We observe a good agreement between modeled and measured groundwater level behavior with short-wavelength fluctuations correlating to the assigned net recharge flux. The seasonal magnitude in groundwater level change ranges between 0.1 and 1 m with highest levels in spring and lowest in late summer—early fall. Our model also captures years with exceptionally high levels (e.g., 1975, 2011) and the extended periods of low level running throughout 1989–1994. Local disagreements between model and observations of up to 0.5 m can be seen in the period between 1972 and 1974. We also note that the modeling results can capture the longest observed period of sustained decline in the levels from 2011 to 2015.

Well II is representative of a hilly terrain setting, characterized with a deeper water table (>20 m below surface). The well is located on the Fläming ridge, SW Brandenburg. We note a poor correlation between observed and modeled groundwater level dynamics, with the measurements lacking any clear indication of seasonal variability but rather following a 7–8-year cycles with a smooth and likely delayed response to large recharge pulses (arrows on Figure 10). This smoothed response is likely due to the travel time of infiltrating meteoric water through a thicker unsaturated

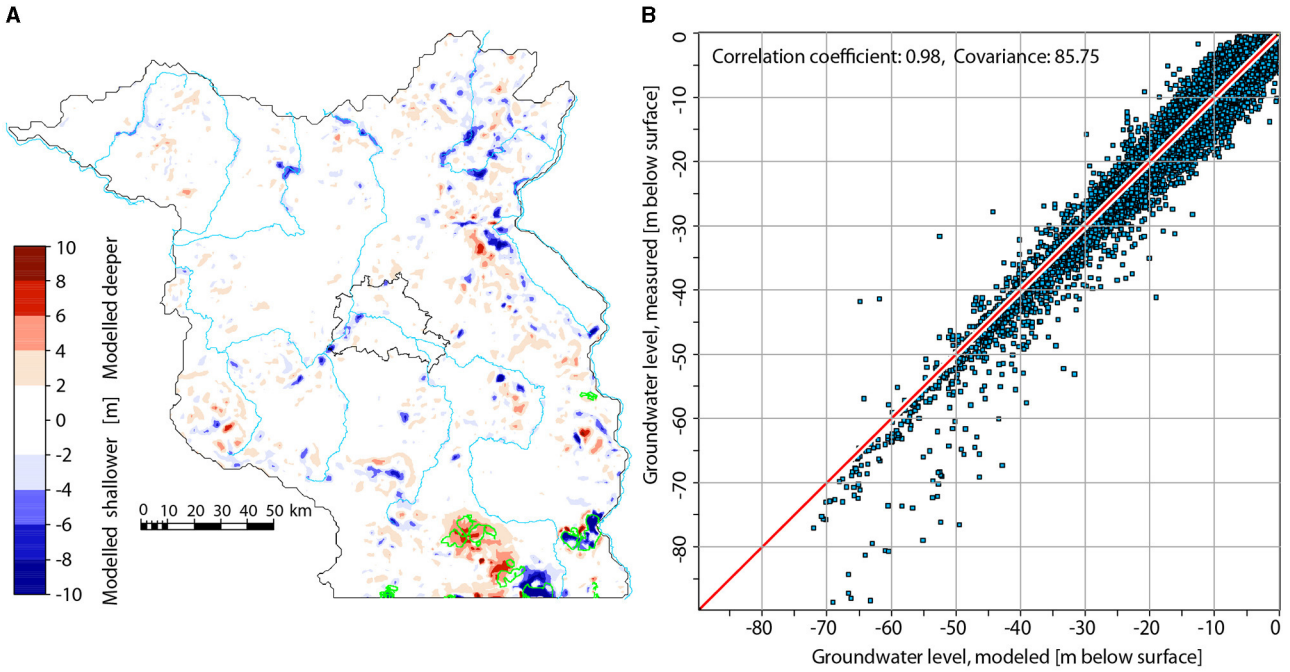


FIGURE 8
Difference between measured hydraulic heads and those predicted by the transient simulation at the end of the transient simulation. **(A)** Map **(B)** cross-plot with each point corresponding to a grid cell of the groundwater model, red line is 1:1 line. Green polygons outline open-cast mines.

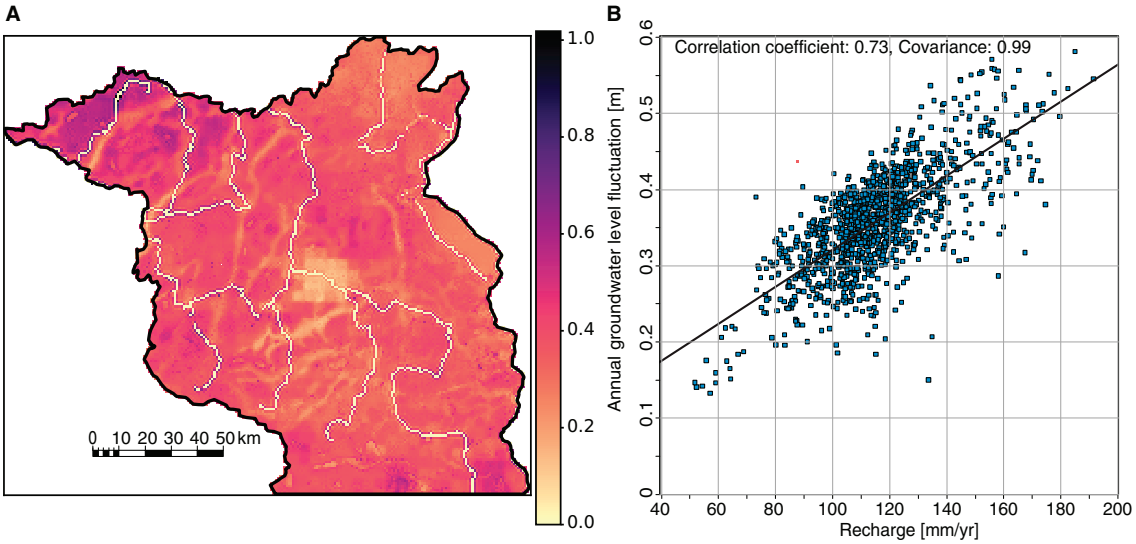


FIGURE 9
Average groundwater level difference between an annual maximum and minimum for the transient simulation: **(A)** spatial distribution of the seasonal groundwater level fluctuations. **(B)** Cross-plot between annual groundwater level fluctuation and recharge, with each point corresponding to a grid cell of the groundwater model; black line is a line of best fit.

zone, a process not explicitly considered by mHM, and therefore not reflected in the net recharge or modeled level time series. Indeed, after applying a moving average filter with a 48-month window to the modeled heads, we have been able to reproduce the periodicity in the modeled time series (dashed line). We also note a ~2 m groundwater level decline as monitored over the entire time window, while the modeled level exhibited a weaker long-term trend. A stronger decline of the measured levels could be explained by the fact that groundwater has been used for crop irrigation in the surrounding area on top of the Fläming ridge since the 1990's, representing an additional sink which is not captured by our model (NABU-Brandenburg, 2021).

Well III represents an example of long-term water table behavior in an urban area (SE Potsdam). The water table is shallow,

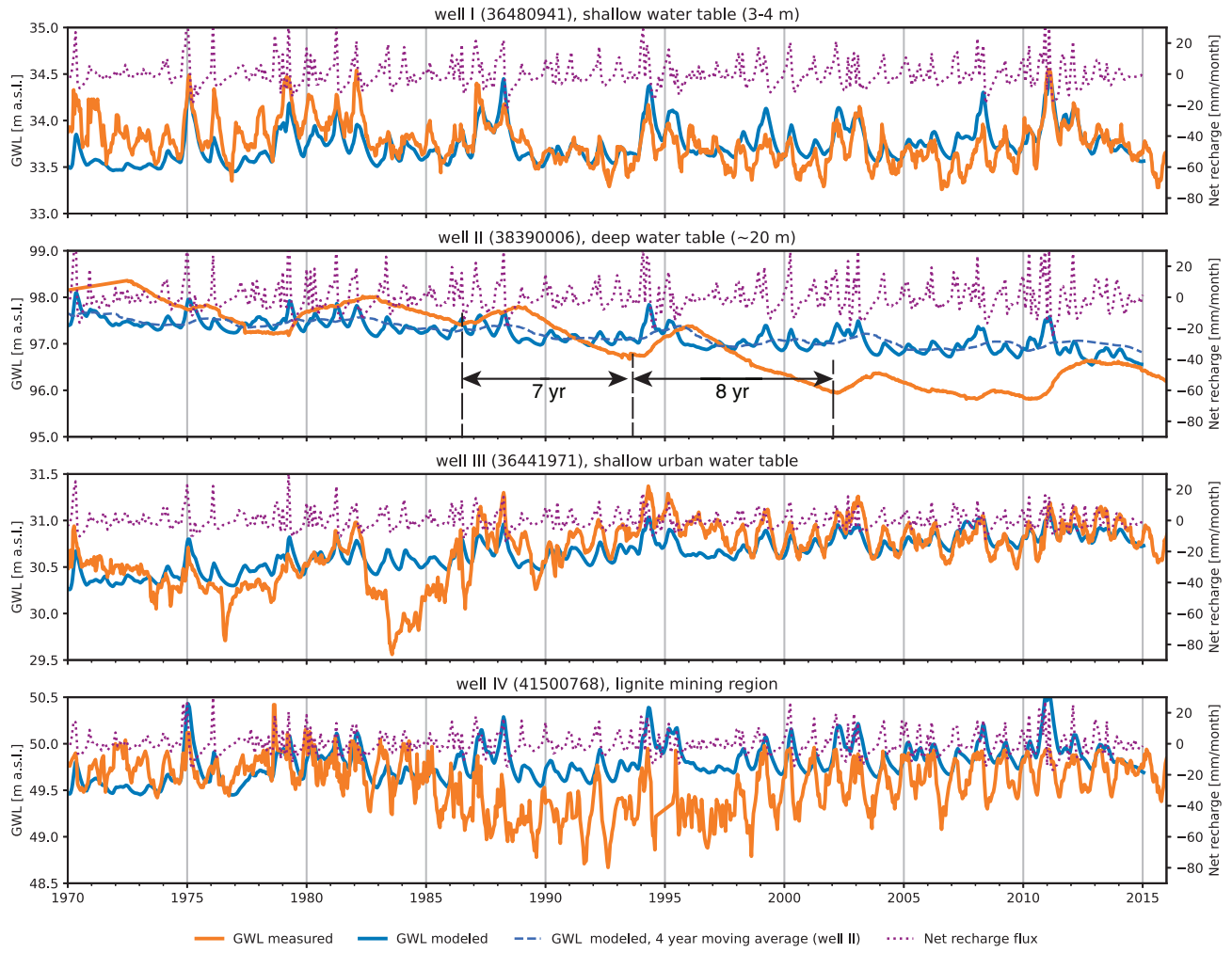


FIGURE 10
Groundwater levels (GWL): measured and modeled, and net recharge flux time series for selected observation wells shown on Figure 5C. All presented wells are screened in the Quaternary unit. Numbers in the brackets refers to well ID in Brandenburg Water Information Platform (Landesamt für Umwelt Brandenburg, 2023), from where the monitoring data is available for download.

3–4 m below ground. During two periods, in 1976 and 1983, modeled levels deviate from the historical record by up to +0.7 and +1 m, respectively. The modeled trend is solely driven by climate forcing, as the groundwater model does not include other sinks, such as pumping. Conversely, the climatic variability of the measured levels shows a clear interference typically caused by a drawdown from groundwater abstraction. The affected time period is indeed associated with the maximum population growth and new housing development in post-war Potsdam. The average annual city water consumption rose during these times from 36 k m^3/d in 1967 to its peak of 65 k m^3/d in 1988 before dropping sharply in the 1990's (Zühlke, 2018). In the remaining period (from 1988 to 2014) modeled and observed groundwater level behavior match, an indication that by that time groundwater extraction has stabilized, and that the annual hydrologic cycle has become the primary control to the observed fluctuations.

Well IV denotes the influence of open-cast mining on groundwater levels in Brandenburg's Lausitz region. The observation well with a shallow water table (1–2 m deep) is located 3 km away from an ex-coal mine, which was in operation between 1983 and 1996 (Seese-Ost), and then was partially filled with water to form artificial lakes (lakes Bischdorf and Kahnsdorf). The measured levels began declining in 1985, followed by a recovery since 1995. In the same period, modeled levels exhibited only seasonal fluctuations. The maximum deviation between measured and modeled levels of 1.1 m was reached in 1992. Despite the strong anthropogenic overprint, recharge influence on groundwater levels was not suppressed as seen from peak levels in 1994 and relatively stable low levels during a dry period in 1996–1997.

3.3 Groundwater temperature

The steady-state thermal field beneath Brandenburg has been extensively studied in previous works (Noack et al., 2012, 2013), which compared the effect of basement heterogeneity on the conductive heat transport and modifications to the resulting basin-wide thermal regime by additional advective forces in the

sedimentary section. Here we focus on discussing the transient behavior of shallow groundwater temperatures due to the spatially and temporally varying recharge and surface temperatures.

In doing so, we based our analysis on temperature maps extracted at different depths relative to the surface (500, 100, and 25 m), rather than depth maps relative to sea level or top aquifer maps since this enables more consistently quantify the overprint of surface processes on the thermal distribution. We therefore compared the temperature at the start (1953) and the end (2014) of the transient simulation as well as the difference between summer and winter seasons. We further note that at 500 and 100 m below the surface, the temperature distribution does not exhibit any significant temporal fluctuation, so for those intervals we discuss only the initial-state maps (Figure 11).

At depths of 100 m and greater, temperature patterns are primarily the result of contrast in thermal rock properties between salt, basement and the sedimentary overburden (Figures 11A, B). Lower temperatures in the south are caused by an uplift of the highly conductive crystalline basement in the presence of a thin sedimentary cover. In the rest of the study area, a thick sedimentary cover provides enough thermal insulation to accumulate heat and maintain relatively higher temperature (Figure 12A). The presence of small-wavelength positive thermal anomalies can be associated with occurrences of Zechstein salt structures acting as conductive heat transport chimneys, leading to a dipole shaped heat anomaly: on top of the salt structures, the temperatures are increased due to heat refraction within the highly conductive salt rock and then "trapped" by the less conductive post-Zechstein sandstones, mudstones and limestones.

The additional role played by heat advection from imposed pressure (hydraulic head) gradients can also be noticed at all depth levels. Low-wavelength negative temperature anomalies could be associated to domains characterized by higher hydraulic heads, characteristic for topographically elevated areas (NW, NE, and the south), where hydraulic gradients drive downward flow (recharge areas). Most of these anomalies have a circular shape, and they don't align spatially with hydrogeological windows in the Rupelian aquitard. An opposite thermal signature occurs, for example, below a N-S river valley with lower heads in the NE corner of the model, where upward groundwater flow causes a local warming.

At depths shallower than 100 m, the temperature distribution carries both geological and climatic signatures. For example, at 25 m below ground, the mean temperature for January is +9.9°C, and for July it is +10.8°C (Figures 11C, D), while mean air temperature is −0.2°C and +18.2°C, respectively. The basement high in the SW quadrant and elevated areas with higher head in the east and northwest cause a relative cooling down by 1–1.5°C below the background. Above shallow salt diapirs and in the deeply incised valley in the northeast, temperatures are above the average, reaching up to +15–20°C. Additionally, we note a northward cooling trend of 0.5–1°C, which resembles the pattern of the average surface temperature imposed as the upper boundary condition (Figure 3D).

According to the analytical solution proposed by Stallman (1965), the magnitude of seasonal surface temperature change and total thermal conductivity provide first-order control to the depth to the neutral zone, defined here as a depth where seasonal temperature fluctuations can be considered as negligible (<0.1°C), while the additional advective component due to downward or upward flow is capable of deepening or raising its depth level. Based on the Quaternary rock properties and the annual average delta in surface temperature for Brandenburg (18°C), the top of the neutral zone is expected to be at 20 m below surface, with the additional effect from downward groundwater flow of 0.01 m/d lowering it to 30 m below surface. Monthly temperature-depth profiles measured in a selection of wells in Berlin find the top of the neutral zone, between 15 and 20 m below ground (SenStadt, 2020). Accordingly, the comparison of July and January temperature distribution at 25 m below ground reveals that most of the study area exhibit no or minimum seasonal change (Figures 11C, D). Patchy areas where the seasonal difference is well-pronounced could be linked to higher heads (Figure 3A) and to glacial channels, where the porous Quaternary unit is the thickest. However, we must note that the effect of surface temperatures within the channels is likely to be overestimated due to the lower vertical mesh resolution. Thirty-five percent of the cells in the uppermost model layer have a thickness >15 m, and 15% of the cells >20 m. Thus, the vertical resolution does not allow to entirely map the top of the neutral zone or temperature at the water table. This said, we can nevertheless determine areas where the seasonal fluctuations propagate substantially deeper than the first computational layer (Figure 12B). Seasonal fluctuations, reaching a maximum depth of 100–150 m, are associated with higher hydraulic gradients (their magnitudes being proportional to the imposed flux), and occur almost entirely within the Quaternary unit, with the Rupelian aquitard effectively limiting the downward propagation of the advective front.

In order to analyze the climate-induced warming effects over the 62-year period, we calculated the difference in the average groundwater temperatures between the last and first 3 years of the simulation and looked at this difference at the depth of 25 m below surface (as a proxy for a top of the neutral zone) (Figure 13A). In most of Brandenburg, the temperature increase is <0.17°C, i.e., the temperature below the neutral level was not significantly affected during the simulation timeframe. A scattered groundwater heating of up to 2°C at this depth level is observed only locally across Brandenburg and it is associated with elevated hydraulic heads, being additionally facilitated in case of reduced thickness of the Quaternary unit thus forcing flow downward around less permeable underlying rocks (post-Rupelian Tertiary and Rupelian; Figure 13B). Such mechanism can propagate the six decades-long groundwater heating signal of at least 0.2°C to the depths of 150 m below the surface. In instances where these advective forces overlap with hydrogeological windows, the warming signal may propagate deeper into the saline aquifers, where the second regional aquitard, the Muschelkalk, provides an additional barrier to the advective front.

In Berlin, the most populous city within the model domain, the urban heat island effect due to land use and waste heat adds onto the climatic warming signal, as reflected in the upper boundary condition grids (Figure 3D). The annual mean surface temperature in Berlin has increased during the modeling period by 2°C. After the year 2000, at any month, the temperature in the center of the city is 1–2°C above the Brandenburg average. The

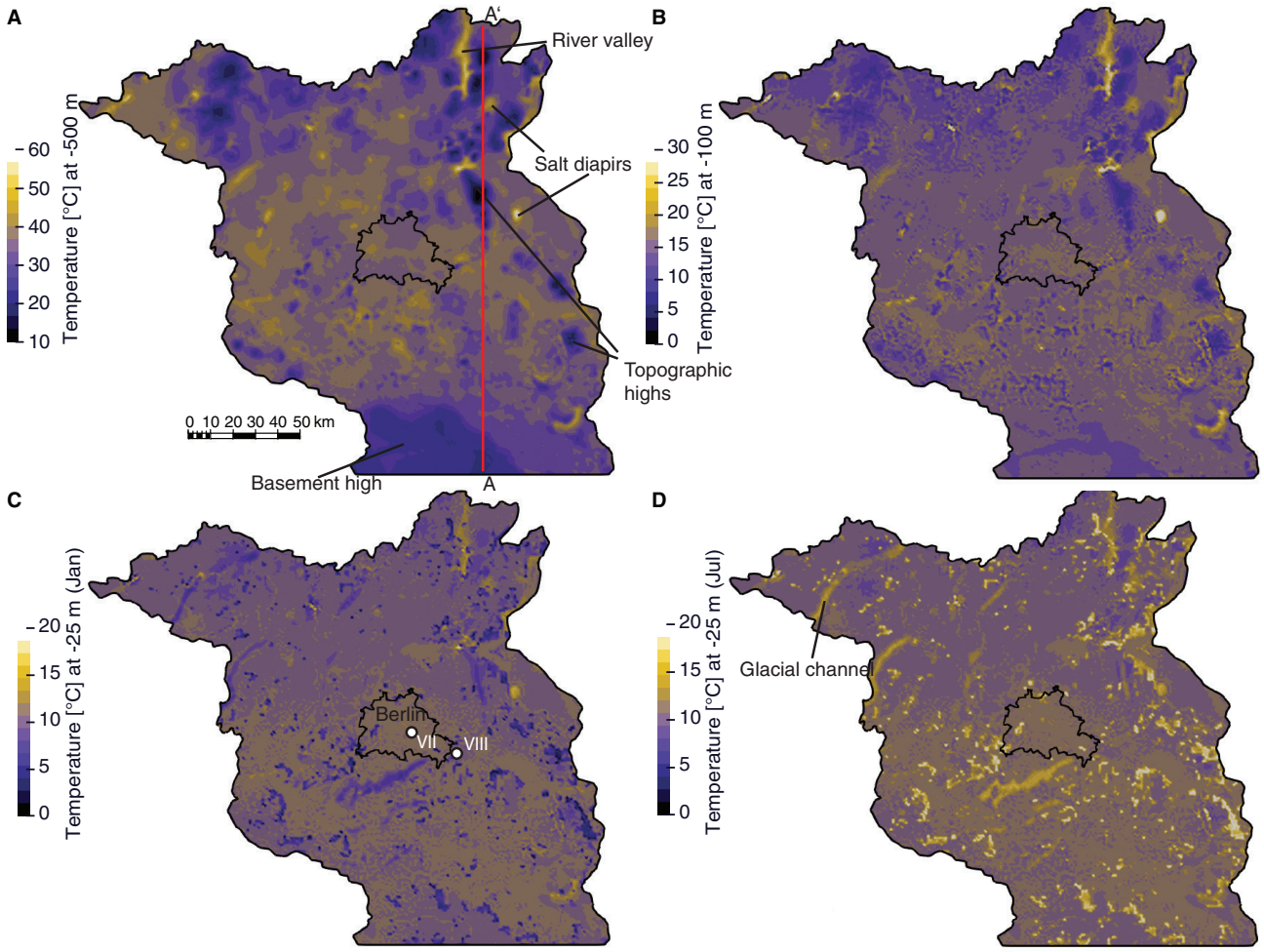


FIGURE 11
Temperature at 500 m **(A)** and 100 m **(B)** below ground at the start of the simulation (no seasonal difference) and average temperature in 1980–2014 in January **(C)** and in July **(D)** at 25 m below ground. A-A' refer to the cross-sections on Figure 12; VII, VIII—well locations, for which temperature time series are presented on the Supplementary Figure 2.

warming trend is also reflected in groundwater temperatures: well measurements are indicative of an increase of ~1.5°C in the past three decades (SenStadt, 2020; Supplementary Figure 2). Unlike groundwater level data, groundwater temperature monitoring in Berlin and Brandenburg is sparse, and the depth of the temperature probe relative to the water level in the well is not always documented. This hinders a quantitative comparison of modeled results against the available measurements. This said, though on a more qualitative level, we can note that modeled temperature time series exhibit depth-depended damping of temperature fluctuations, seasonal phase shifts, and a long-term warming trend (Supplementary Figure 2).

4 Discussion and conclusions

Based on a 3D thermo-hydraulic groundwater simulator one-way coupled to a distributed hydrologic model, we have been able to reconstruct the time-dependent pressure and temperature evolution within the sedimentary units in the region of Brandenburg. In this section, we explore model results further with a focus on the relative control that the imposed recharge, the topographic relief, and the heterogenous hydrostratigraphy have on the shallow thermal and pressure fields. Unlike in synthetic box models or 1D analytical solutions, extracting individual effects from our data-driven 3D model is challenging. As shown in the previous section, all driving factors at play, that is the thickness of the Quaternary unit, discontinuities in the Rupelian aquitard, ground steepness and elevation, and net recharge, are quite variable in space (and time). In the following discussion, we aim at discussing these driving forces more systematically and decoupling what can be considered primary and secondary factors controlling flow configuration (4.1), groundwater temperature (4.2), and groundwater level fluctuations (4.3).

4.1 Groundwater flow configuration

We found that hydraulic gradients, acting as the main driver to the groundwater flow in the studied basin, are not

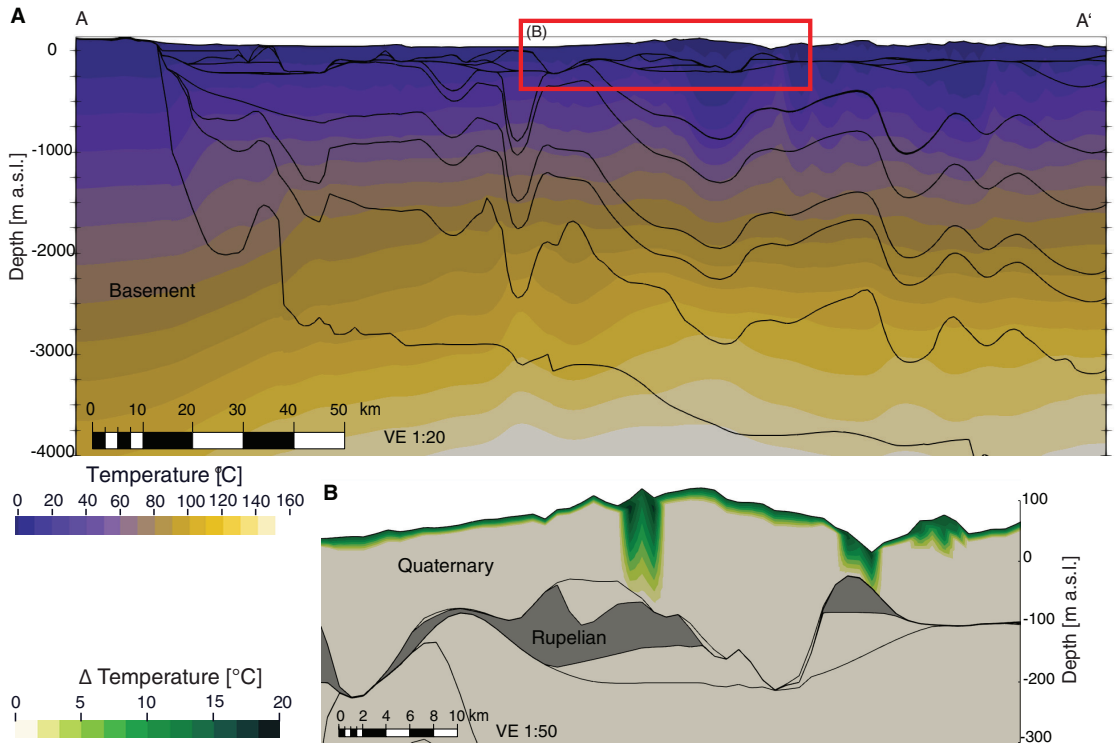


FIGURE 12
(A) Steady-state temperature distribution in the basin. **(B)** Average temperature difference between January and July (1980–2014). Location of cross-sections shown on Figure 11.

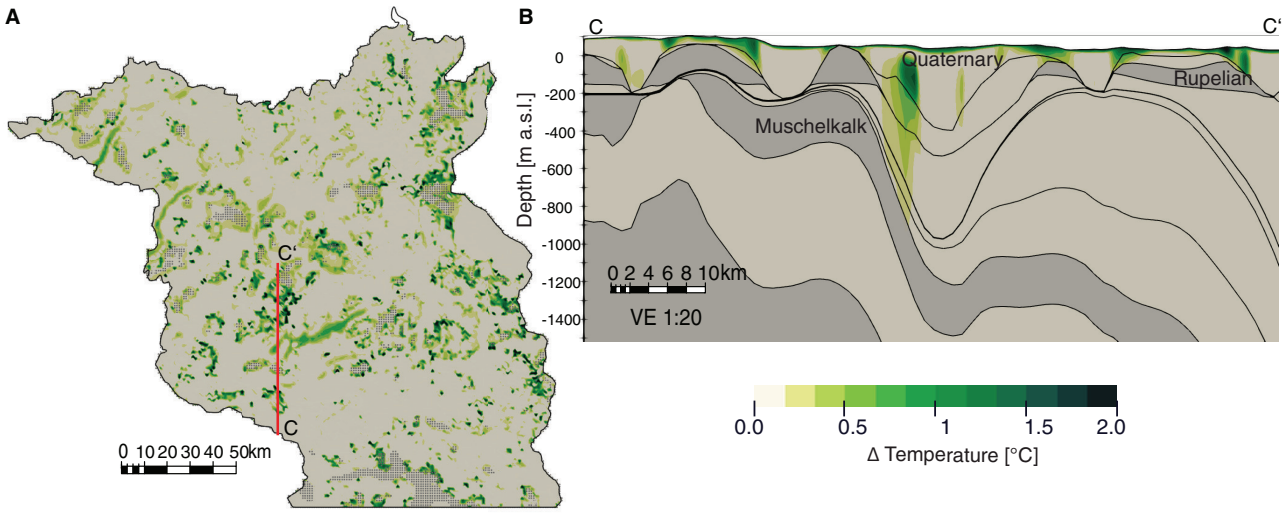


FIGURE 13
Temperature difference between the first and last 3 years of the transient simulation (1953–1956 and 2011–2014). The plotted data is averaged annually to remove the effect of seasonal fluctuations. **(A)** Map at 25 m below the surface. Gray dots indicate areas of reduced Quaternary thickness (<25 m). **(B)** Cross-section; main aquitards are shown in gray.

exclusively influenced by present-day topographic gradients. Instead, structural dip and stratigraphic layering of geological units and the presence of permeability contrasts and anisotropy all concur to shape the flow in pre-Rupelian aquifers, while recharge variability and anthropogenic groundwater use can be pointed as the sources of hydraulic head and Darcy flux variations in the Quaternary unit.

Several stages of structural deformation and the paleogeographic evolution of the NGB led to a high level of hydrogeological stratification and variable thickness of the

stratigraphic formations. This geological heterogeneity is reflected in a relatively complex pressure field and a groundwater system not bounded to individual reservoir units, but characterized by extensive cross-flows and sharp changes in the main flow direction. Regional aquitards (Rupelian, Muschelkalk, and Zechstein) act as hydraulic barriers, responsible for flow reorganization. Vertical flow is more vigorous within steeply dipping strata, adjacent to salt structures or along the southern rim of the basin, and where hydrological windows provide hydraulic communication between otherwise stratigraphically disconnected compartments. The resulting vertical flux has the greatest potential to locally overprint the conductive thermal signature (see also Section 4.2).

The regional flow system in the Buntsandstein and the intermediate flow system, spanning from Keuper to pre-Rupelian Tertiary are not in direct contact with surface waters, with a possible exception of the southern basement uplift, and are dominated by groundwater storage. In these aquifers, groundwater ages are older than the simulated period of six decades and therefore unlikely to be affected by the recent climate variability.

In the Buntsandstein, the deepest modeled aquifer, groundwater flow is controlled by the geometry of the geological units. Indeed, groundwater flows predominantly from south to north/northwest following the regional dip of 1–2°. Consequently, groundwater age also increases from south to north. Inflow of groundwater into the aquifer occurs in the south of Brandenburg, where the Buntsandstein subcrops directly below a thin Quaternary unit due to erosion and non-deposition of the strata in between. This is the only area where we would expect a potential overprinting of the regional flow modulated by long-term climatic variations.

Inflow into the Keuper—pre-Rupelian Tertiary aquifer complex occurs preferentially via leakage from the overlying Quaternary and post-Rupelian Tertiary aquifers. As such, it is genetically connected to two conditions: presence of hydrogeological windows in the Rupelian aquitard and the existence of a hydraulic gradient forcing a downward groundwater flux (Figure 5D). The latter is controlled to a large degree by the present-day topographic relief. Areas where the hydraulic gradient forces upward groundwater fluxes can provide flow paths for outflow of deep Cl^--rich groundwater via hydrogeological windows thereby posing a potential risk of salinization to the drinking-water aquifers. Initial evidence from brackish springs and lakes above the shallow-buried salt domes has been verified by groundwater sampling (Rößling et al., 2010) and modeling (Kaiser et al., 2013; Chabab et al., 2022). At the typical depths of this aquifer complex (>500 m), we did not observe any influence of recharge variability on the groundwater dynamics during the simulation period. However, sustained lower recharge rates over a longer period of time can potentially lead to a pressure decline in freshwater aquifers, and force further brine uprising from saltwater aquifers.

The situation is different if considering the local flow system. In Quaternary and post-Rupelian Tertiary aquifers, groundwater flows inside catchments, from watershed divides toward the drainage network. Those aquifers receive recharge across the upper boundary, either in terms of percolation from the unsaturated zone or due to leakage from incorporated rivers. As a result of the one-way coupling to mHM, recharge occurs in our model where and when the net flux of percolation and baseflow to/from the groundwater storage in mHM is positive. As such, the amount of recharge estimated by mHM only partly follows the actual topography, being principally a function of precipitation, evapotranspiration, soil properties, and land use. As an example, consider the Fläming ridge in SW Brandenburg (Figure 1B). This is the most elevated area in Brandenburg, and yet it does not receive as much recharge as the hills in the NW of Brandenburg, which experience a greater impact of oceanic air masses and receive more precipitation than average (Figure 3C). As a result, the depth to the water table in the south is greater than that in the north, despite a comparable topographic elevation. This leads to different shallow groundwater dynamics than when considering a constant-head boundary condition as a subdued replica of the topography, as usually done in basin-scale groundwater modeling studies. In the latter, recharge is limited to high-elevation areas (highest head) and discharge to low-elevation areas (lowest head), without any consideration of the amount of water effectively reaching the water table.

4.2 Overprint of recharge on shallow groundwater temperatures

A gradual and continuous groundwater warming trend as well as the overprint of seasonal fluctuations are strongest in areas of elevated hydraulic head and hydraulic gradients. The factors controlling the penetration depth of the temperature disturbance are imposed groundwater recharge, Darcy flux, and the magnitude of surface temperature variations.

Advective heat transport by groundwater flow can significantly overprint the conductive subsurface temperature field. This is particularly the case beneath recharge and discharge areas, where advective processes depress or increase heat flow, respectively (Ingebritsen et al., 2006). In practice, such deviations in groundwater temperatures profiles from a purely conductive geotherm can be used to estimate vertical water fluxes, and thus present and past recharge rates (Anderson, 2005; Kurylyk et al., 2017). Our initial simulation, that considered constant annual surface temperatures as Type I boundary condition, demonstrates how topographic highs of only a few tens of meters, if sufficient time for equilibration is given, can result in localized negative temperature anomalies throughout the entire sedimentary succession (Figure 12A), with high Peclet numbers pointing to their advective nature (Supplementary Figure 3). However, it is clear that climatic cyclicity and oscillations hinder temperatures in any groundwater system from reaching a true steady state.

In the transient simulation, we have demonstrated that the seasonal groundwater temperature cycling due to conductive heat exchange with the atmosphere is occurring in the upper 20 m and that an additional advective transport component can deepen this feedback. Given that the Quaternary unit is modeled with a homogeneous vertical permeability and a typical thickness >100 m, most of the change is indeed occurring in this unit, with the key factors controlling the depth of the neutral level being the hydraulic gradient, defined either by hydraulic heads or recharge flux, and the magnitude of the annual air temperature

variation. Hydrogeological windows in the Rupelian Clay that were thought to play a key role in the vertical advective flow (Noack et al., 2013) cannot trigger this process, but can deepen the thermal front propagation only if the local hydraulic conditions are favorable. In some exceptional cases of very high Darcy flux in the absence of Rupelian aquitard, seasonal temperature fluctuations could reach greater depths, as deep as 100–150 m in our model. The same mechanism is also responsible for the mean increase of ∼1.5°C in modeled groundwater temperature from 1953 to present-day in response to the global warming and the urban heat island effect below the city of Berlin, with a modeled anomaly lying within ±0.5°C of the monitored values (Figures 11C, D; Supplementary Figure 2). Additional point sources to the near-surface warming within the study area were attributed to tunnels, underground car parks, and mining operations (Henning and Limberg, 2012; Tissen et al., 2019; Noethen et al., 2023), which have not been incorporated in our model.

Advection-driven undulations of the neutral zone and continuous groundwater warming may impact the state-wide utilization of shallow geothermal heat in Brandenburg, which, as of 2022, consisted of more than 20,000 ground source heat pumps (Ministerium für Wirtschaft, 2023), that, by design, target depths with constant mean annual temperatures.

4.3 Groundwater level trends

Despite the lack of modeled heterogeneity in the Quaternary reservoir architecture, and neglecting the details of the hydrodynamics within the unsaturated zone, we have been able to reproduce the seasonal amplitudes of observed groundwater level fluctuations over several decades by properly selecting effective rock properties and considering the assigned net fluxes. However, approximating the multi-annual periodicity and long-term trends with the model has proven to be more challenging.

Modeled groundwater levels predominantly follow the seasonal recharge cycle, showing no delay, with highs during February-March and lows in August-October. Nevertheless, wells characterized by a deeper water table show a smoother and delayed response to recharge pulses (e.g., Well II on Figure 10). We interpret this behavior as an indication of a damped signal from precipitation due to the hydrodynamic dispersion of the advancing wetting front as water infiltrates through a thick unsaturated zone. Indeed, we found such effects to positively correlate with the thickness of the unsaturated zone, becoming increasingly visible for a thickness higher than 10 m. Lischeid et al. (2021), on the basis of principal component analysis of wells and groundwater-fed lakes in Brandenburg, correlated such attenuation to low-frequency patterns in the groundwater time series. Thus, it can be argued that the "deeper" wells are likely to be more suitable to analyze long-term water level trends, since the seasonal signal has been, at least partially, filtered out (Dong et al., 2019; Lischeid and Steidl, 2023).

Inter-annual groundwater level variations in Brandenburg have been attributed to progressive accumulation of soil moisture deficits in the deeper unsaturated zone (Lischeid and Steidl, 2023) or teleconnections between precipitation and the North Atlantic Oscillation (Landesamt für Umwelt Gesundheit und Verbraucherschutz Brandenburg, 2014). mHM recharge fluxes for Brandenburg follow a 5–7-yr periodicity, but in the absence of any damping effect, the seasonal signal is stronger, therefore masking the low-frequency signal in modeled groundwater time series.

In Figure 14 we compare long-term groundwater level trends estimated from a regression analysis, following the procedure described in Länder-Arbeitsgemeinschaft Wasser (2011). The trends based on actual measurements between 1976 and 2013 are presented in accordance with the results from Landesamt für Umwelt Gesundheit und Verbraucherschutz Brandenburg (2014). One caveat of ∼37 years of head time series with a single linear regression is evidenced by a reversal in the groundwater trends as seen in some of the wells. For example, the Lausitz mining district underwent a period of mine drainage, followed by mine flooding (Benthaus and Totsche, 2015; Well IV on Figure 10), and the Berlin metropolitan area saw high abstraction rates in the post-war era to supply its growing population and recovering industry, followed by a recovery phase starting from 1990's onward as more sustainable water management strategies were introduced (Frommen and Moss, 2021).

Despite these assumptions, the majority of wells depicts a decrease in the groundwater level at a rate of 1–3 cm/yr (Figure 14A). Wells with similar behavior are often clustered within a radius of several kilometers. Falling levels are found to correlate spatially with areas where the surface elevation is greater than ∼50 m. Highest rate of decline (>10 cm/yr) is found in areas of active or past coal mining in the south of Brandenburg, while lowest values (close to negligible) are seen in the river valleys, occasionally neighbored by wells with rising levels. Falling groundwater levels cannot be directly associated with decline in annual precipitation, which does not exhibit any statistically significant trend (Figure 14C) in the time interval considered. Instead it can be explained by (a) changes in water balance; (b) anthropogenic factors; (c) memory effects in the unsaturated zone.

Regarding the climatic water balance, the observed rise in surface temperatures due to global warming caused an increase of PET in the study area (Figure 14C), which likely led to an increase in the actual evapotranspiration, given no obvious trend in annual precipitation. Coupled groundwater-surface water models from other regions have been able to demonstrate how a higher evapotranspiration can deplete groundwater due to reduced infiltration through the soil and water loss from a shallow-lying water table (Condon et al., 2020).

Local anthropogenic factors depleting groundwater storage in Brandenburg include pumping, coal mine drainage, wetland drainage and other land use changes (Germer et al., 2011). As an evidence, it is typical for neighboring monitoring wells in Brandenburg to show opposing water level trends (Lischeid et al., 2021). Recent high magnitudes of water level decline in local groundwater-fed lakes could only be simulated by superposition of several anthropogenic and natural factors altering the water balance (Somogyvári et al., 2023).

Finally, an apparent monotonic decades-long decline of groundwater levels may arise from memory effects, for example, due to a progressive soil moisture deficit accumulation with depth

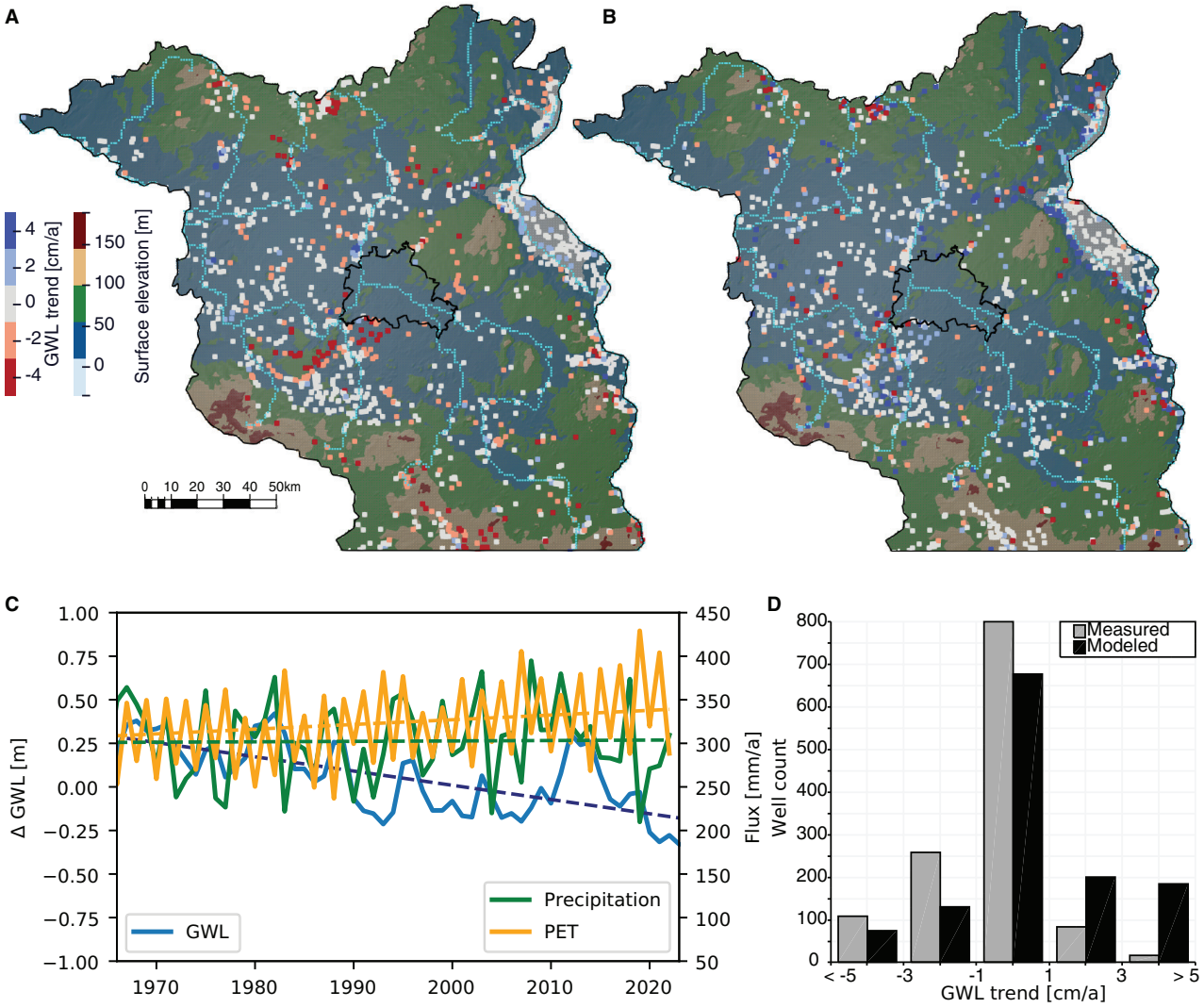


FIGURE 14
Groundwater level trend (1976–2013). **(A)** Calculated from well measurements (Landesamt für Umwelt Brandenburg, 2023); **(B)** Extracted from the model at well locations; **(C)** annual average hydraulic heads in observation wells in Brandenburg, mHM input fluxes: precipitation and PET; **(D)** Distribution of modeled and measured groundwater trends.

initiated by a meteorological drought, an effect sometimes referred to as the "drought cascade" (Changnon, 1987; Lischeid and Steidl, 2023).

While inspecting the long-term head behavior based on the model results, we note a higher spatial variability with opposing trends in neighboring grid cells and less clustering than in the observation data (Figure 14B). Clustering can only be seen for wells showing a weak trend (−1 to 1 cm/yr) mostly within lowlands, where the imposed fixed head in the river cells limits the actual rise or fall of groundwater levels. When compared to the observations, the model underestimates the long-term groundwater level decline (Figure 14D). The main explanation is that the net recharge flux derived from mHM does not exhibit a statistically significant trend over time, neither for the entire Brandenburg (Figure 4A), nor in specific locations (Figure 10). Instead, the inflow and outflow components of the groundwater storage in mHM balance each other out on the long-term. Indeed, if we compute the total volume of simulated recharge and discharge over the whole study area for the 1953–2014 period we arrive at an estimate of 2.560e+5 km^3 and 2.558e+5 km^3, respectively. This results in a cumulative sum of the net recharge fluxes at the scale of individual cells at the end of the simulation to be only −12 to +31 mm. The associated change in groundwater levels would be therefore minimal, that is in the order of few centimeters. Consequently, modeled long-term groundwater level trends are generally weaker than measured ones. This comes back to some inherent limitations of mHM (and of any bucket hydrologic models), that is, in the absence of a deep percolation component, each grid cell of a groundwater reservoir maintains a mass balance between recharge and baseflow

over time. Therefore, any single-year increase or decrease of recharge will be compensated by a corresponding baseflow change such that the long-term annual average head will not be significantly affected.

To address these challenges, decoupling cell-based recharge and discharge by lateral redistribution of subsurface water fluxes between grid cells may improve modeling of the long-term groundwater level evolution. On a watershed scale, baseflow routing can be applied to assign Type II boundary condition along river nodes (Jing et al., 2018), but for the Germany-wide mHM model, the routing scheme has not been implemented so far. An alternative could be to assign only positive recharge fluxes from the hydrologic model as a boundary condition while computing the corresponding discharge rates as a function of the hydraulic conductivity, hydraulic gradient, and river network configuration directly in the groundwater model. In such case, a careful calibration of the aquifer parameters becomes crucial to close the water balance.

In addition to input fluxes (precipitation and evapotranspiration), land cover is another transient parameter affecting recharge flux in mHM. Three land cover classes (natural, agricultural, and artificial) are defined for four periods: 1950–1999, 2000–2005, 2006–2011, 2012–2015. Changes in land cover have an immediate impact, resulting in a step-like change of output water fluxes. However, land cover class changes occurred in only 6% of the cells, mostly in the vicinity of Berlin (agricultural to artificial) and in the mining district (natural to artificial; Supplementary Figure 4). Thus, the impact of land cover on the model results is limited.

4.4 Model limitations and uncertainties

The uppermost model unit (Quaternary) encompasses three aquifers, all of which are utilized for water supply in Brandenburg, along with corresponding aquitards separating them (Stackebrandt and Manhenke, 2010). In our study, we have approximated this complex architecture of fluvio-glacial deposits characterized by mixed lithologies and abundant discontinuities via unit-averaged effective rock properties, typical for poorly consolidated sand. Recent efforts have been focused on providing a more accurate parameterization of these aquifers, either through higher-resolution structural models (Frick et al., 2019; Chabab et al., 2022) or via stochastic parameter distribution (Truong et al., 2023).The incremental level of heterogeneity in the shallow geology is expected to locally decelerate or enhance surface-to-groundwater interactions.

The sensitivity of the simulation outputs to the mesh resolution was conducted during the FEM mesh set-up. Initially, we carried out test simulations without any vertical refinement, that is the FEM resolution equals the vertical stratification of the input geological model. We found that these runs provided with erroneous output P/T values. In a second stage, we gradually increased the vertical resolution of the numerical mesh until numerically stable and physically consistent (against observations) results were attained. Further mesh refinement is recommended for greater resolution and thus, the accuracy of temperature and Darcy flux profiles in the first tens of meters below the surface, which are sensitive to the high-frequency variations in the upper boundary conditions.

The structural framework of the model does not incorporate faults, which are attributed to basement uplifts, salt diapirism and Pleistocene isostatic adjustment in NGB (Sirocko et al., 2008; Scheck-Wenderoth and Maystrenko, 2013). Some of these faults exhibit significant displacement and a surface expression, however data on faults permeability is not available. Introducing faults would add an extra degree of freedom to the model and can significantly overprint the thermo-hydraulic state at least locally. Previous modeling studies indicated that warm fluids rising along presumably permeable faults can be responsible for a positive anomaly of up to 15°C at a depth of 450 m and a similar magnitude of advective cooling within recharge domains (Cherubini et al., 2014).

Our model does not consider groundwater extraction as an additional sink in the groundwater balance. For a single watershed in Brandenburg (Nuthe), groundwater extraction is estimated to be 8% of groundwater recharge with a total rate of 6 mm/a Landesumweltamt Brandenburg (2009). In Lausitz, the drainage of opencast mines produced a widespread depression in the water levels. Future efforts should characterize such additional fluxes as groundwater extraction, irrigation, riverbank filtration, and lateral inflow/outflow across the model boundaries. Satellite gravimetry may provide an independent, though lower resolution, evidence of groundwater storage reduction in Brandenburg, while multi-scale modeling seems appropriate for evaluating a contribution of the basin-scale flow to the model's water balance A screening analysis suggested that the effect of open vs. closed boundaries on the local groundwater system is minimal since the flow is bounded by smaller-scale catchments and hydraulically connected to the resolved rivers (Figure 5C). The situation differs in regional groundwater system, which showcases an evidence of lateral outflow beyond the model domain (Supplementary Figure 5). However, these boundary effects can be considered of secondary relevance for current study, given that the internal flow geometries within the limits of Brandenburg are not significantly affected by our choice of open vs. closed boundary conditions, and, in principle, in agreement with a flow pattern obtained from the larger-scale model of CEBS (Frick et al., 2022; Supplementary Figure 5).

Our model simulates groundwater flow in a fully saturated domain. Fluxes applied at the top of the model have been derived by one-way coupling the groundwater model to a hydrologic model simulation (mHM) with its own uncertainties and limitations. The mHM setup for Germany used in the present study includes a large number of discharge gauges from upland and headwater regions for model calibration. In contrast, the contribution of pure lowland catchments is rather low and thus limits the performance of the model in lowlands. However, due to the MPR, the effect of parameter changes on the model output is lower in mHM compared to other hydrological models. Simulated fluxes in mHM, like in many other hydrologic models, are calibrated against river discharges but in the majority of the cases not against groundwater levels or other measurements from the subsurface, and are thus subject to uncertainty in soil and rock properties. The travel-time of infiltrating water through the thick

unsaturated zone and the delay of recharge pulses are likely to be underestimated.

Given that groundwater accounts for 87.6% of Brandenburg's public water supply and 55.9% of its non-public water supply (Statistisches Bundesamt, 2022), forecasting the system's evolution with respect to climate scenarios becomes crucial for water management. On one hand, climate models suggest an increase in winter precipitation for Brandenburg (GERICS-Climate-Service Center Germany, 2019), which would lead to an overall increased groundwater recharge. On the other hand, rising temperatures will lead to higher evapotranspiration, having an opposite effect on groundwater recharge. Additionally, an increase in the intensity of individual precipitation events may cause a higher surface runoff, further reducing flux through the soil to the water table. Consequently, projecting the trend of groundwater levels due to climate change is challenging, even without considering pumping rates and irrigation. The same applies to groundwater temperatures, which, despite an overall warming trend, may be further altered by advection from evolving recharge rates and their seasonal distribution. Increasing shallow aquifer temperatures might affect groundwater composition, ecosystem, and energy potential, prompting the need for innovative adaptation strategies and heat utilization concepts (Riedel, 2019; Tissen et al., 2019). Present-day subsurface conditions reconstructed by this study can serve as a starting point and input for the next generation of predictive models exploring these effects over longer simulation periods (i.e., 100's—1,000's of years).

In addition to freshwater availability in Quaternary aquifers, climatic impacts on the groundwater dynamics is of relevance for site selection and proper planning and design of geothermal energy utilization concepts as well as in relation to the integrity of subsurface storage projects in saline and brackish formations. In this regard, future work is required to improve current modeling solutions in order to better approximate hydraulic and thermal heterogeneity, and anisotropic conditions in target reservoirs and neighboring aquitards. A use of layer-parallel orientation of Kxy, rather than horizontal, has proven to produce more accurate fluid flow vectors (Poulet et al., 2023). Sensitivity analysis of model parameters, such as permeability, thermal conductivity, and storativity, would become instructive in establishing confidence in model forecasts.

Data availability statement

Publicly available datasets were analyzed in this study. Input data for model set-up and output files can be found at: https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=cf5927b9-b491-11ee-967a-4ffbfe06208e.

Author contributions

MT: Conceptualization, Writing – original draft, Data curation, Investigation, Visualization. MC: Software, Writing – review & editing, Conceptualization, Investigation, Supervision. BG: Writing – review & editing, Investigation. AG: Writing – review & editing, Investigation. MS-W: Supervision, Writing – review & editing, Project administration.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding was provided by Helmholtz Center Potsdam—GFZ German Research Center for Geosciences.

Acknowledgments

The authors would like to acknowledge Charlotte Krawczyk and Guido Blöcher for their valuable consultations. Additionally, the authors thank Landesamt für Umwelt Brandenburg for kindly providing groundwater temperature data. Constructive suggestions made by two reviewers are greatly appreciated.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2024.1353394/full#supplementary-material

References

Anderson, M. P. (2005). Heat as a ground water tracer. *Groundwater* 43, 951–968. doi: 10.1111/j.1745-6584.2005.00052.x

Ayachit, U. (2015). *The Paraview Guide: a Parallel Visualization Application*. Clifton Park, NY: Kitware, Inc.

Benthaus, F. C., and Totsche, O. (2015). *The Groundwater Raise in Lignite Mining Induced Areas in Lusetia–Actions Taken to Reduce the Follows Ups*. Mining Report (Essen: Bergbau-Verwaltungsgesellschaft mbH), 522530.

Benz, S. A., Bayer, P., and Blum, P. (2017). Global patterns of shallow groundwater temperatures. *Environ. Res. Lett.* 12:034005. doi: 10.1088/1748-9326/aa5fb0

Bruhn, D. K., Silke, K., Regenspurg, S., and Schumann, F. (2023). *Wärmeversorgung in Berlin und Brandenburg durch Geoenergie, Position Paper*. Available online at: https://www.gfz-potsdam.de/fileadmin/gfz/sec48/pdf/de/Netzwerke/GEB/Posititionspapier_GEB__2.pdf (accessed December 1, 2023).

Bundesanstalt für Geowissenschaften und Rohstoffeand (2019). *Hydrogeologische Karten für den Hydrologischen Atlas von Deutschland*.

Bundesanstalt für Geowissenschaften und Rohstoffeand and Staatliche Geologische Dienste (2019). *Hydrogeologische Übersichtskarte von Deutschland 1:250.000 (HÜK250). Digitaler Datenbestand, Version 1.0.3*. Hannover.

Cacace, M., and Jacquey, A. B. (2017). Flexible parallel implicit modelling of coupled thermal–hydraulic–mechanical processes in fractured rocks. *Solid Earth* 8, 921–941. doi: 10.5194/se-8-921-2017

Chabab, E., Kühn, M., and Kempka, T. (2022). Upwelling mechanisms of deep saline waters via Quaternary erosion windows considering varying hydrogeological boundary conditions. *Adv. Geosci.* 58, 47–54. doi: 10.5194/adgeo-58-47-2022

Changnon, S. A. (1987). "Detecting drought conditions in Illinois," in *Circular no. 169* (Champaign, IL: Illinois State Water Survey).

Chen, Z., Grasby, S. E., and Osadetz, K. G. (2004). Relation between climate variability and groundwater levels in the upper carbonate aquifer, southern Manitoba, Canada. *J. Hydrol.* 290, 43–62. doi: 10.1016/j.jhydrol.2003.11.029

Cherubini, Y., Cacace, M., Scheck-Wenderoth, M., and Noack, V. (2014). Influence of major fault zones on 3-D coupled fluid and heat transport for the Brandenburg region (NE German Basin). *Geoth. Energ. Sci.* 2, 1–20. doi: 10.5194/gtes-2-1-2014

Condon, L. E., Atchley, A. L., and Maxwell, R. M. (2020). Evapotranspiration depletes groundwater under warming over the contiguous United States. *Nat. Commun.* 11:873. doi: 10.1038/s41467-020-14688-0

Condon, L. E., and Maxwell, R. M. (2015). Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model. *Water Resour. Res.* 51, 6602–6621. doi: 10.1002/2014WR016774

Dams, J., Salvadore, E., Van Daele, T., Ntegeka, V., Willems, P., and Batelaan, O. (2012). Spatio-temporal impact of climate change on the groundwater system. *Hydrol. Earth Syst. Sci.* 16, 1517–1531. doi: 10.5194/hess-16-1517-2012

Deutscher Wetterdienst (2019). *Klimareport Brandenburg, 1 Edn*. ed. D. Wetterdienst (Deutschland: Offenbach am Main).

Deutscher Wetterdienst—Climate Data Center (2021). *Grids of Monthly Averaged Daily Air Temperature (2m) Over Germany, Version v1.0*. (Offenbach am Main).

Domenico, P. A., and Schwartz, F. W. (1997). *Physical and Chemical Hydrogeology*. Hoboken, NJ: John Wiley & Sons, Inc.

Dong, Y., Jiang, C., Suri, M. R., Pee, D., Meng, L., and Rosenberg Goldstein, R. E. (2019). Groundwater level changes with a focus on agricultural areas in the Mid-Atlantic region of the United States, 2002–2016. *Environ. Res.* 171, 193–203. doi: 10.1016/j.envres.2019.01.004

Förster, A. (2001). Analysis of borehole temperature data in the Northeast German Basin: continuous logs vs. bottom-hole temperatures. *Petrol. Geosci.* 7, 241–254. doi: 10.1144/petgeo.7.3.241

Franz, M., Barth, G., Zimmermann, J., Budach, I., Nowak, K., and Wolfgramm, M. (2018). Geothermal resources of the North German Basin: exploration strategy, development examples and remaining opportunities in Mesozoic hydrothermal reservoirs. *Geolog. Soc.* 469, 193–222. doi: 10.1144/SP469.11

Frick, M., Cacace, M., Klemann, V., Tarasov, L., and Scheck-Wenderoth, M. (2022). Hydrogeologic and thermal effects of glaciations on the intracontinental basins in central and Northern Europe. *Front. Water* 4:818469. doi: 10.3389/frwa.2022.818469

Frick, M., Scheck-Wenderoth, M., Schneider, M., and Cacace, M. (2019). Surface to groundwater interactions beneath the city of Berlin: results from 3D models. *Geofluids* 2019:4129016. doi: 10.1155/2019/4129016

Frommen, T., and Moss, T. (2021). Pasts and presents of urban socio-hydrogeology: groundwater levels in Berlin, 1870–2020. *Water* 13:2261. doi: 10.3390/w13162261

Gaupp, R., Moeller, P., Lüders, V., di Primio, R., and Littke, R. (2008). "Fluids in sedimentary basins: an overview," in *Dynamics of Complex Intracontinental Basins: The Central European Basin System*, eds. R. Littke, U. Bayer, D. Gajewski, and S. Nelskamp (Berlin: Springer Science & Business Media), 345–366.

GERICS-Climate-Service Center Germany (2019). *Klima-Ausblick Brandenburg*. Hamburg.

Germer, S., Kaiser, K., Bens, O., and Hüttl, R. F. (2011). Water balance changes and responses of ecosystems and society in the Berlin-Brandenburg Region—a review. *J. Geogr. Soc. Berlin* 142, 65–95.

Goderniaux, P., Brouyère, S., Blenkinsop, S., Burton, A., Fowler, H. J., Orban, P., et al. (2011). Modeling climate change impacts on groundwater resources using transient stochastic climatic scenarios. *Water Resour. Res.* 47:10082. doi: 10.1029/2010WR010082

Güntner, A., Sharifi, E., Behzadpour, S., Boergens, E., Dahle, C., Darbeheshti, N., et al. (2023). Global Gravity-based Groundwater Product (G3P). V. 1.11. *GFZ Data Serv*. doi: 10.5880/G3P.2023.001

Haitjema, H. M., and Mitchell-Bruker, S. (2005). Are water tables a subdued replica of the topography? *Groundwater* 43, 781–786. doi: 10.1111/j.1745-6584.2005.00090.x

Healy, R. W., and Cook, P. G. (2002). Using groundwater levels to estimate recharge. *Hydrogeol. J.* 10, 91–109. doi: 10.1007/s10040-001-0178-0

Henning, A., and Limberg, A. (2012). Veränderung des oberflächennahen Temperaturfeldes von Berlin durch Klimawandel und Urbanisierung. *Brandenburgische Geowiss. Beitr* 19, 81–92.

Huenges, E., and Ledru, P. (2011). *Geothermal Energy Systems: Exploration, Development, and Utilization*. Hoboken, NJ: John Wiley & Sons.

Ingebritsen, S. E., Sanford, W. E., and Neuzil, C. E. (2006). *Groundwater in Geologic Processes*. Cambridge: Cambridge University Press.

Jing, M., Heße, F., Kumar, R., Wang, W., Fischer, T., Walther, M., et al. (2018). Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5.7) to the groundwater model OpenGeoSys (OGS). *Geosci. Model Dev.* 11, 1989–2007. doi: 10.5194/gmd-11-1989-2018

Kaiser, B. O., Cacace, M., and Scheck-Wenderoth, M. (2013). Quaternary channels within the Northeast German Basin and their relevance on double diffusive convective transport processes: constraints from 3-D thermohaline numerical simulations. *Geochem. Geophys. Geosyst.* 14, 3156–3175. doi: 10.1002/ggge.20192

Kaiser, B. O., Cacace, M., Scheck-Wenderoth, M., and Lewerenz, B. (2011). Characterization of main heat transport processes in the Northeast German Basin: constraints from 3-D numerical models. *Geochem. Geophys. Geosyst.* 12:3535. doi: 10.1029/2011GC003535

Kestin, J., Sokolov, M., and Wakeham, W. A. (1978). Viscosity of liquid water in the range −8 °C to 150 °C. *J Phys. Chem. Ref. Data* 7, 941–948. doi: 10.1063/1.555581

Konikow, L. F. (2011). Contribution of global groundwater depletion since 1900 to sea-level rise. *Geophys. Res. Lett.* 38:48604. doi: 10.1029/2011GL048604

Kumar, R., Samaniego, L., and Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resour. Res.* 49, 360–379. doi: 10.1029/2012WR012195

Kurylyk, B. L., Irvine, D. J., Carey, S. K., Briggs, M. A., Werkema, D. D., and Bonham, M. (2017). Heat as a groundwater tracer in shallow and deep heterogeneous media: analytical solution, spreadsheet tool, and field applications. *Hydrol. Process.* 31, 2648–2661. doi: 10.1002/hyp.11216

Länder-Arbeitsgemeinschaft Wasser (2011). "Fachliche Umsetzung der Richtlinie zum Schutz des Grundwassers vor Verschmutzung und Verschlechterung (2006/118/EG)", in *Teil 5, Bundesweit einheitliche Methode zur Beurteilung des mengenmäßigen Zustands* (Sachstandsbericht vom).

Landesamt für Umwelt Brandenburg (2020). *Hydroisohypsen und Messwerte des oberen genutzten Grundwasserleiters im Land Brandenburg - Datenserie*. Available online at: https://geoportal.brandenburg.de/detailansichtdienst/render?url=https%3A%2F%2Fgeoportal.brandenburg.de%2Fgs-json%2Fxml%3Ffileid%3D2435B954-BA43-4598-B6E1-06AA36F2BAB7 (accessed December 12, 2023).

Landesamt für Umwelt Brandenburg (2022). "Klimawandelmonitoring im Land Brandenburg - Aktualisierungsbericht," in *Fachbeiträge des LfU* (Potsdam).

Landesamt für Umwelt Brandenburg (2023). *Auskunftsplattform Wasser (APW)*. Available online at: https://apw.brandenburg.de/ (accessed December 12, 2023).

Landesamt für Umwelt Gesundheit und Verbraucherschutz Brandenburg (2014). "Die Wasserbilanzen der Grundwasserkörper im Land Brandenburg," in *Fachbeiträge des Landesamtes für Umwelt, Gesundheit und Verbraucherschutz* ed S. Wieneke (Potsdam), 154.

Landesumweltamt Brandenburg (2009). *Umweltdaten Brandenburg 2008/09*. Potsdam.

Lischeid, G., Dannowski, R., Kaiser, K., Nützmann, G., Steidl, J., and Stüve, P. (2021). Inconsistent hydrological trends do not necessarily imply spatially heterogeneous drivers. *J. Hydrol.* 596:126096. doi: 10.1016/j.jhydrol.2021.126096

Lischeid, G., and Steidl, J. (2023). *Why Do Our Models Underestimate Regional Groundwater Trends*? Vienna: EGU General Assembly.

Littke, R., Bayer, U., Gajewski, D., and Nelskamp, S. (2008). *Dynamics of Complex Intracontinental Basins: the Central European Basin System*. Berlin: Springer Science & Business Media.

Magri, F., Bayer, U., Jahnke, C., Clausnitzer, V., Diersch, H., Fuhrman, J., et al. (2005). Fluid-dynamics driving saline water in the North East German Basin. *Int. J. Earth Sci.* 94, 1056–1069. doi: 10.1007/s00531-005-0497-9

Mather, B., Müller, R. D., O'Neill, C., Beall, A., Vervoort, R. W., and Moresi, L. (2022). Constraining the response of continental-scale groundwater flow to climate change. *Sci. Rep.* 12:4539. doi: 10.1038/s41598-022-08384-w

Maystrenko, Y., Bayer, U., and Scheck-Wenderoth, M. (2010). *Structure and Evolution of the Central European Basin System According to 3D Modeling*. DGMK research report (Hamburg) 577-2/2-1, 90.

Ministerium für Wirtschaft, Arbeit und Energie Brandenburg. (2023). *Geothermie*. Available online at: https://mwae.brandenburg.de/de/geothermie/bb1.c.478390.de (accessed December 6, 2023).

NABU-Brandenburg (2021). *Wir haben ein Wasserproblem!* Available online at: https://nabu-bb.de/wasser-in-brandenburg/ (accessed November 13, 2023).

Noack, V., Scheck-Wenderoth, M., and Cacace, M. (2012). Sensitivity of 3D thermal models to the choice of boundary conditions and thermal properties: a case study for the area of Brandenburg (NE German Basin). *Environ. Earth Sci.* 67, 1695–1711. doi: 10.1007/s12665-012-1614-2

Noack, V., Scheck-Wenderoth, M., Cacace, M., and Schneider, M. (2013). Influence of fluid flow on the regional thermal field: results from 3D numerical modelling for the area of Brandenburg (North German Basin). *Environ. Earth Sci.* 70, 3523–3544. doi: 10.1007/s12665-013-2438-4

Noethen, M., Hemmerle, H., Menberg, K., Epting, J., Benz, S. A., Blum, P., et al. (2023). Thermal impact of underground car parks on urban groundwater. *Sci. Tot. Environ.* 2023:166572. doi: 10.1016/j.scitotenv.2023.166572

Norden, B., Bauer, K., and Krawczyk, C. M. (2023). From pilot knowledge via integrated reservoir characterization to utilization perspectives of deep geothermal reservoirs: the 3D model of Groß Schönebeck (North German Basin). *Geothermal Energy* 11:1. doi: 10.1186/s40517-022-00242-2

Norden, B., Förster, A., and Balling, N. (2008). Heat flow and lithospheric thermal regime in the Northeast German Basin. *Tectonophysics* 460, 215–229. doi: 10.1016/j.tecto.2008.08.022

Poulet, T., Sheldon, H. A., Kelka, U., and Behnoudfar, P. (2023). Impact of permeability anisotropy misalignment on flow rates predicted by hydrogeological models. *Hydrogeol. J.* 31, 2129–2137. doi: 10.1007/s10040-023-02708-4

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., et al. (2016). Multiscale and multivariate evaluation of water fluxes and states over European River Basins. *J. Hydrometeorol.* 17, 287–307. doi: 10.1175/JHM-D-15-0054.1

Riedel, T. (2019). Temperature-associated changes in groundwater quality. *J. Hydrol.* 572, 206–212. doi: 10.1016/j.jhydrol.2019.02.059

Rodella, A.-S., Zaveri, E. D., and Bertone, F. M. D. C. (2023). *The Hidden Wealth of Nations: The Economics of Groundwater in Times of Climate Change-Executive Summary*. Washington, DC: World Bank.

Rößling, H. B., Albrecht, B., Haarring, C., Hermsdorf, A., Herrmann, A., List, U., et al. (2010). Regionale Überblicksdarstellungen und Gebietssteckbriefe. *Naturschutz Landschaftpflege Brandenburg.* 19, 52–118.

Samaniego, L., Kumar, R., and Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* 46:7327. doi: 10.1029/2008WR007327

Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., et al. (2019). Hydrological forecasts and projections for improved decision-making in the water sector in Europe. *Bullet. Am. Meteorol. Soc.* 100, 2451–2472. doi: 10.1175/BAMS-D-17-0274.1

Scheck, M., and Bayer, U. (1999). Evolution of the Northeast German Basin — inferences from a 3D structural model and subsidence analysis. *Tectonophysics* 313, 145–169. doi: 10.1016/S0040-1951(99)00194-8

Scheck-Wenderoth, M., and Maystrenko, Y. P. (2013). Deep control on shallow heat in sedimentary basins. *Energy Proc.* 40, 266–275. doi: 10.1016/j.egypro.2013.08.031

SenStadt (2020). *Berlin Environmental Atlas*. Berlin: SenStadt, Senate Department for Urban Development, Building and Housing.

Sirocko, F., Reicherter, K., Lehné, R., Hübscher, C., Winsemann, J., and Stackebrandt, W. (2008). "Glaciation, salt and the present landscape," in *Dynamics of Complex Intracontinental Basins. The Central European Basin System*, eds R. Littke, U. Bayer, D. Gajewski and S. Nelskamp (Berlin; Heidelberg: Springer), 233–245.

Somogyvári, M., Scherer, D., Bart, F., Fehrenbach, U., Okujeni, A., and Krueger, T. (2023). A hybrid data-driven approach to analyze the drivers of lake level dynamics. *EGUsphere* 2023, 1–23. doi: 10.5194/egusphere-2023-2111

Stackebrandt, W., and Manhenke, V. (2010). *Atlas zur Geologie von Brandenburg im Maßstab 1:1,000,000*. Kleinmachnow: Landesamt für Geowissenschaften und Rohstoffe.

Stallman, R. W. (1965). Steady one-dimensional fluid flow in a semi-infinite porous medium with sinusoidal surface temperature. *J. Geophys. Res.* 70, 2821–2827. doi: 10.1029/JZ070i012p02821

Statistisches Bundesamt (2022). *Wasserwirtschaft*. Available online at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/Wasserwirtschaft/_inhalt.html#236444 (accessed December 10, 2023).

Szewczyk, J., and Nawrocki, J. (2011). Deep-seated relict permafrost in northeastern Poland. *Boreas* 40, 385–388. doi: 10.1111/j.1502-3885.2011.00218.x

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., et al. (2013). Ground water and climate change. *Nat. Clim. Change* 3, 322–329. doi: 10.1038/nclimate1744

Thomas, B. F., and Famiglietti, J. S. (2019). Identifying climate-induced groundwater depletion in GRACE observations. *Sci. Rep.* 9:4124. doi: 10.1038/s41598-019-40155-y

Tissen, C., Benz, S. A., Menberg, K., Bayer, P., and Blum, P. (2019). Groundwater temperature anomalies in central Europe. *Environ. Res. Lett.* 14:104012. doi: 10.1088/1748-9326/ab4240

Toth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. *J. Geophys. Res.* 68, 4795–4812. doi: 10.1029/JZ068i016p04795

Truong, T. M., Guadagnini, A., and Engelhardt, I. (2023). *Characterization of Spatial Heterogeneity of Geomaterials in Large Scale Groundwater Bodies Through a Compositional Data Approach*. Vienna: EGU General Assembly.

Werner, A. D., and Simmons, C. T. (2009). Impact of sea-level rise on sea water intrusion in coastal aquifers. *Groundwater* 47, 197–204. doi: 10.1111/j.1745-6584.2008.00535.x

Wunsch, A., Liesch, T., and Broda, S. (2022). Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nat. Commun.* 13:1221. doi: 10.1038/s41467-022-28770-2

Zühlke, K. (2018). "Ressourcenmanagement in der Wasserwirtschaft - Wie ist Potsdam auf das Wachstum und mögliche Folgen vom Klimawandel vorbereitet?," in *REKLIM Regionalkonferenz 25. September 2018*. Potsdam: Energie und Wasser.

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 02 October 2024
DOI 10.3389/frwa.2024.1439906

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Sanjeev Kumar Jha,
Indian Institute of Science Education and Research, Bhopal, India
Fabio Oriani,
Agroscope, Switzerland

*CORRESPONDENCE
Kaveh Patakchi Yousefi
✉ kpatakchi@gmail.com

RECEIVED 28 May 2024
ACCEPTED 09 September 2024
PUBLISHED 02 October 2024

CITATION
Patakchi Yousefi K, Belleflamme A, Goergen K and Kollet S (2024) Impact of deep learning-driven precipitation corrected data using near real-time satellite-based observations and model forecast in an integrated hydrological model. *Front. Water* 6:1439906. doi: 10.3389/frwa.2024.1439906

COPYRIGHT
© 2024 Patakchi Yousefi, Belleflamme, Goergen and Kollet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Impact of deep learning-driven precipitation corrected data using near real-time satellite-based observations and model forecast in an integrated hydrological model

Kaveh Patakchi Yousefi[1,2]*, Alexandre Belleflamme[1,3], Klaus Goergen[1,3] and Stefan Kollet[1,2,3]

[1]Agrosphere (IBG-3), Research Centre Jülich, Jülich, Germany, [2]Meteorological Institute, University of Bonn, Bonn, Germany, [3]Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Jülich, Germany

Integrated hydrological model (IHM) forecasts provide critical insights into hydrological system states, fluxes, and its evolution of water resources and associated risks, essential for many sectors and stakeholders in agriculture, urban planning, forestry, or ecosystem management. However, the accuracy of these forecasts depends on the data quality of the precipitation forcing data. Previous studies have utilized data-driven methods, such as deep learning (DL) during the preprocessing phase to improve precipitation forcing data obtained from numerical weather prediction simulations. Nonetheless, challenges related to the spatiotemporal variability of hourly precipitation data persist, including issues with *ground truth* data availability, data imbalance in training DL models, and method evaluation. This study compares three (near) real-time spatiotemporal precipitation datasets to be used in the aforementioned IHM forecast systems: (1) 24h precipitation forecast data obtained by ECMWF's 10-day HRES deterministic forecast, (2) H-SAF h61 satellite observations as reference, and (3) DL-based corrected HRES precipitation using a U-Net convolutional neural network (CNN). As high-resolution data, H-SAF is used both as a reference for correcting HRES precipitation data and as a stand-alone candidate for forcing data. These datasets are used as forcing data in high-resolution (~0.6km) integrated hydrologic simulations using ParFlow/CLM over central Europe from April 2020 to December 2022. Soil moisture (SM) simulations are used as a diagnostic downstream variable for evaluating the impact of forcing data. The DL-based correction reduces the gap between HRES and H-SAF by 49, 33, and 12% in mean error, root mean square error, and Pearson correlation, respectively. However, comparison of SM simulations obtained from the three datasets with ESA CCI SM data reveals better agreement with the uncorrected HRES 24-h forecast data. In conclusion, H-SAF satellite-based precipitation data falls short in representing precipitation used for SM simulations compared to 24h lead time HRES forecasts. This emphasizes the need for more reliable spatiotemporally continuous high-resolution precipitation observations for using DL correction in improving precipitation forecasts. The study demonstrates the potential of DL methods as a near real-time data pre-processor in quasi-operational water resources forecasting workflows. The quality of the preprocessor is directly proportional to the quality of the applied observation.

KEYWORDS

integrated hydrological model, convolutional neural network, soil moisture simulation, precipitation correction, bias correction, water resources forecasting, near real-time hydrological forecasting, operational hydrological modeling

1 Introduction

Water resources forecasting is important for many sectors. Information on the availability and distribution allows for a sustainable management of water, for example in agriculture, domestic and industrial water supply, or ecosystem functioning. IHMs play a vital role in water resources forecasting, predicting the impacts of climate change on water availability and assessing risks associated with hydrological extremes (Qi et al., 2016; Tuo et al., 2016). The accuracy of simulations obtained by these models heavily relies on the quality of atmospheric forcing data from numerical weather prediction and climate models (Bennett et al., 2022). Amongst the forcing data, precipitation is key, the main driver for the terrestrial surface and subsurface water budgets and also impacting the land surface energy balances (Fekete et al., 2004; Fersch et al., 2020; Jabbari et al., 2019; Pan et al., 2010; Tanhapour et al., 2023). While enhancing precipitation accuracy alone does not ensure improved hydrological predictions, errors in model-based precipitation data, especially for heavy rainfall events, strongly affects the accuracy of hydrological predictions (Li et al., 2023; Qi et al., 2016; Saadi et al., 2023a, 2023b).

Operational, near real-time, high-resolution water resources forecasting systems that integrate coupled atmosphere-land-subsurface processes on a (sub-) continental scale offer comprehensive information to a wide array of stakeholders by representing a wider range of water and energy processes within the terrestrial water cycle (Tijerina-Kreuzer et al., 2021). Two representative examples of impact-scale hydrological forecasting systems are the national water model (NWM) over the US by National Oceanic and Atmospheric Administration (NOAA) based on WRF-Hydro (Cosgrove et al., 2024; Gochis et al., 2020; Gochis and Chen, 2003; Senatore et al., 2015; Towler et al., 2023; Yucel et al., 2015) and a monitoring and forecasting system based on the IHM ParFlow/CLM (Kollet et al., 2010; Kollet and Maxwell, 2006; Kuffour et al., 2020) in a setup for subsurface water resources over central Europe (DE06) as described by Belleflamme et al. (2023). These systems primarily operate with a one-way coupling approach, where atmospheric forcings drive land and subsurface processes without explicit consideration of feedbacks from the land surface and subsurface to the atmosphere.

In flood forecasting and rainfall-runoff studies, the role of precipitation accuracy and its correction using conventional, or machine learning methods have been explored (Huang et al., 2023; Saadi et al., 2023a; Tri et al., 2022; Wijayarathne et al., 2020; Xu et al., 2023). Yet, there is a notable scarcity of research especially on DL-based precipitation correction in water resources forecasting systems. For example, the correction of short-term, medium or long-term precipitation forecasts used as forcing data within the previously mentioned systems is commonly implemented through a statistical bias correction or adjustment, or data assimilation (DA). For reference, within the context of the forecasting systems mentioned above, NWM 2.1 utilizes quantile mapping bias correction (adjustment) for long-range precipitation and other atmospheric forcing data from Climate Forecast System (CFS) forecasts (Cosgrove et al., 2024; Panofsky and Brier, 1968). The ParFlow/CLM DE06 experimental 10-day deterministic forecasts use as atmospheric forcing mainly the high-resolution deterministic medium-range forecasts (HRES) from the European Centre for Medium-Range Weather Forecasts (ECMWF) (refer to Belleflamme et al., 2023) that are based on a 4D variational Data Assimilation (Bannister, 2001).

The effectiveness of precipitation correction methods is measured through statistical evaluation, e.g., comparing probability distributions, error metrics, and spatiotemporal correlations between corrected precipitation and observed data (Li et al., 2021; Liu et al., 2020; Patakchi Yousefi and Kollet, 2023). However, implementing and evaluating these correction methods as well as nowcasting, forecasting, and downscaling applications of precipitation imposes challenges such as the availability and quality of ground truth data, data imbalance (e.g., in data-driven methods), and the selection of meaningful evaluation metrics, which is well known from flood forecasting applications (Hess and Boers, 2022; Lam et al., 2023; Ravuri et al., 2021; Wang et al., 2021). In the following paragraphs, these challenges are explained in greater detail.

High spatiotemporal variability of precipitation necessitates reliable ground truth data from *in-situ* observations, yet conventional sources of measuring precipitation have limitations, such as spatial representativity issues in rain gauge observations and beam-blockage gaps in weather radar data (Kidd et al., 2017; Yaswanth et al., 2023; Yousefi et al., 2023). Although satellite data cover larger areas, uncertainties stemming from cloud cover and retrieval algorithms affect their accuracy or lead to data gaps (Tian et al., 2009). Addressing data gaps involves estimating missing data through statistical or data-driven algorithms (Sattari et al., 2020; Mital et al., 2020). However, many data-driven bias correction or forecasting studies rely on reanalysis data sources such as COSMO-REA6, COSMO-REA2, ERA5, and ERA-interim, offering consistent and continuous coverage but constrained by coarse resolution and model-related errors (Bi et al., 2023; Han et al., 2021; Patakchi Yousefi and Kollet, 2023).

There is an inherent data imbalance in precipitation data, where heavy rainfall (i.e., larger than 10 mm/h) occurrences are relatively rare compared to lighter events. This poses a challenge to traditional machine learning loss functions which are commonly designed to perform well on balanced datasets (Dablain et al., 2023; You et al., 2023). Customized loss functions such as treat score, dice loss, and weighted loss have shown promise in mitigating data imbalance issues and directing model training toward heavy rainfall events (Hess and Boers, 2022; Larraondo et al., 2019; Li et al., 2021; Rojas-Campos et al., 2023; You et al., 2023).

Typical metrics such as root mean squared error are not ideal but useful for training data-driven methods, but evaluation using these metrics may not capture a comprehensive assessment. On the other hand, temporal metrics averaged over time mapped over space overlook the spatial structure in evaluation. Multi-component spatial

metrics such as SPAEF address the problem of accounting for the spatial and temporal features (Dembélé et al., 2020; Demirel et al., 2018; Koch et al., 2018; Yorulmaz et al., 2023). Such metrics can be used for indirect evaluation using hydrological models to simulate various states and fluxes within the land and subsurface compartments, encompassing variables such as soil moisture (SM) content, groundwater levels, surface runoff, streamflow, and other related processes. Model results and diagnostics are then compared against observed data, serving as a benchmark to estimate the effectiveness of precipitation data and/or correction methods (Casanueva et al., 2016; Fang et al., 2015; Lafon et al., 2013; Luo et al., 2018; Pan et al., 2010; Teng et al., 2015).

Within the ML realm, Deep Learning (DL)-based precipitation correction methods using Convolutional Neural Networks (CNNs) have gained popularity due to their independence from statistical assumptions, and ability to learn complex non-linear error relationships (Hess and Boers, 2022; Kim et al., 2021; Ronneberger et al., 2015; Sun et al., 2019; Wang et al., 2023; Zhang et al., 2023). In this context, Patakchi Yousefi and Kollet (2023) utilized U-Net architecture in a merging framework to learn and correct errors between model- and reanalysis-based daily precipitation data; in this study, U-Net, a type of CNN, was shown to outperform the commonly used quantile mapping bias correction (Cannon et al., 2015; Piani et al., 2010). However, the impact of DL-based precipitation correction on hydrological simulations has not been evaluated so far.

This study introduces novelty by integrating U-Net architecture known for its efficiency against quantile mapping and versatility among DL methods as a corrector of atmospheric forcing data in operational IHM forecasting system. We introduce a dynamic mask in training the DL network on available space–time grids to address the data gaps in satellite-based precipitation. The hydrological impact of precipitation data corrected using DL and used as forcing data in operational, near real-time, high-resolution hydrological forecasting systems. For the impact assessment, the precipitation datasets including model-based short-term forecasts, near real-time satellite-based observations, and DL-based corrected precipitation derived from the first two datasets are compared against each other. Our analysis primarily focuses on assessing SM as a downstream variable due to its immediate hydrological response from precipitation, rather than a more complicated variable such as evapotranspiration influenced by a larger number of states and fluxes. SM reference data is available as a spatial representation for comparison with simulated data. The evaluation serves two main objectives: first, to investigate the effectiveness of DL-based correction on improving precipitation data, and second, to compare the influence of different (DL-corrected) precipitation data on the fidelity of hydrological forecasts. Overall, this research contributes to advancing our understanding on the role of precipitation data in hydrological forecasting—operational or experimental—and thereby supports informed decision-making in water resource management.

The manuscript is organized as follows: section 2 presents the methodology, including the study domain and data, the DL-based precipitation correction, and the hydrological simulations. Section 3 presents the results and discussion on hyperparameter tuning, on the evaluation of precipitation correction, on the comparison of SM simulations, and the validation of precipitation products. Section 4 summarizes the conclusions drawn from the study.

2 Methodology

We begin with a general introduction to our methodology in section 2.1 and the study area and data in section 2.2. In section 2.3, we explain how we utilize the DL method to correct precipitation simulations. In section 2.4, we describe the hydrological model setup. Finally, in section 2.5, we describe the evaluation methods used in the study.

2.1 General methodology

The goal of this study is to evaluate the hydrological impact of various precipitation data, and DL-based precipitation correction in an experimental, near real-time, high-resolution hydrological IHM forecasting system. We evaluate three precipitation datasets used as forcing data: (1) short-term atmospheric forecast data, (2) near real-time observations, and (3) DL-based corrected precipitation using the first two datasets. To achieve these goals, we need short-term atmospheric forecast data, near real-time observational precipitation data, an operational hydrological forecasting system, a DL framework for precipitation correction, and an evaluation method to implement and investigate our methodology (Figure 1).

For the atmospheric forecasting and observational data (Figure 1A), we choose the HRES forecasts by ECMWF (European Centre for Medium-Range Weather Forecasts, 2016) and the near real-time satellite-based precipitation product (h61) by Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) (EUMETSAT, 2021a, 2022; Martins Costa Do Amaral et al., 2018). Apart from ENS (ensemble probabilistic forecasts), and SEAS (seasonal forecasts) used for probabilistic and/or longer term forecasts, HRES is commonly used for short-term forecasts up to 10 days. Our focus is the first 24-h is due to the increased complexity and uncertainty associated with longer lead times, requiring sophisticated methods to address lead time errors. This study focuses on short-term forecasts to assess the immediate impact of DL-based correction.

The choice of H-SAF data is motivated by its high resolution among satellite data, spatial coverage, and near real-time data availability, which makes it a candidate for use in near-real time hydrological forecasts. The alternatives, such as lower-resolution data from other satellites or radar-based datasets covering only specific regions, are deemed less suitable for our study.

As a demonstrator for the hydrological forecasting, we choose the ParFlow/CLM IHM with its DE06 setup, tailored to water resources forecasts and implemented in an experimental near-real time workflow over central Europe (Figure 1D; Belleflamme et al., 2023). The ParFlow/CLM model is known for its versatility and applicability indicating that our findings can offer insights with broader usefulness across different geographical contexts.

We use the DL framework proposed by Patakchi Yousefi and Kollet (2023) to learn and correct the mismatches between HRES precipitation forecast and H-SAF observations (Figure 1B). The method was initially applied to gridded daily reanalysis data without space–time gaps. However, here we implement it with satellite-based data at a higher spatial resolution and at hourly time scale. This introduces new challenges, such as unavailable data in time and space in H-SAF satellite-based observations and more

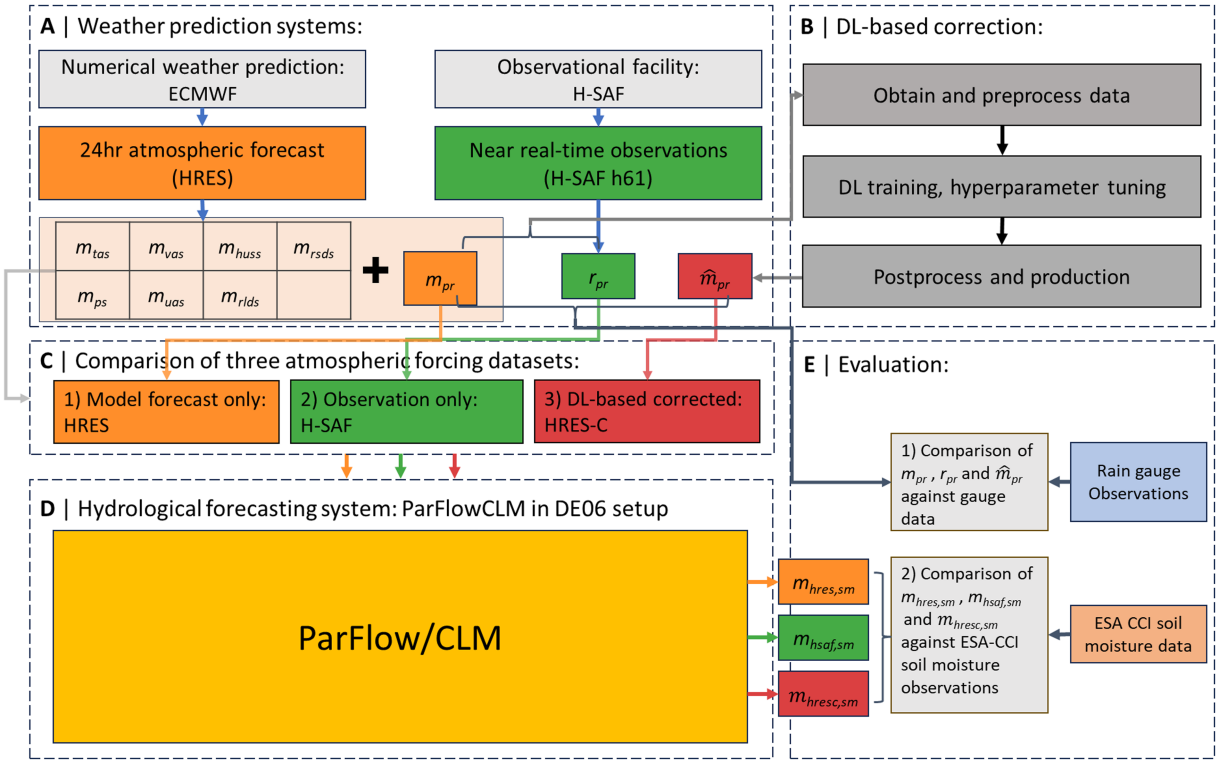


FIGURE 1
Schematic diagram of the overall methodology and data flow path of the study: weather prediction **(A)** and experimental hydrological forecasting systems **(D)**, DL-based correction of precipitation **(B)**, and implementation and evaluation of three types of precipitation data used among the other atmospheric forcing data **(C,E)**. m_{pr}, r_{pr}, and m_{pr} represent precipitation data obtained by HRES model, H-SAF satellite observations, and corrected using DL method, respectively. $m_{hres,\,sm}$, $m_{hsaf,\,sm}$, and $m_{hresc,\,sm}$ represent SM simulations (obtained from ParFlow pressure field outputs) by ParFlow/CLM given the three corresponding atmospheric forcings. m_{tas}, m_{ps}, m_{vas}, m_{uas}, m_{huss}, m_{rsds}, and m_{rlds} represent the short-term atmospheric forecast obtained from HRES for air temperature (*tas*), surface air pressure (*ps*), meridional (*vas*) and zonal (*uas*) wind speed, specific humidity (*huss*), visible (*rsds*) and infrared (*rlds*) downward radiation at the surface, respectively.

data imbalance because of the higher temporal resolution. To manage the gaps and data imbalance problems, we introduce customized loss function.

The evaluation of the hydrological impact of the various precipitation datasets and DL-based precipitation correction involves comparing the three precipitation datasets against rain gauge observations and assessing the SM simulations derived from these datasets against ESA CCI SM data used as a reference (Figure 1E).

2.2 Study domain and data

The study domain mainly consists of Central Europe, with two different grid definitions. The first grid definition is for the implementation of the DL-based framework on precipitation forcing data (Figure 1C), and the second one is for high-resolution 3D ParFlow/CLM simulations (Figure 1D).

The first grid definition (Figure 2A) is a 2D grid with a resolution of $0.1° \times 0.1°$, consisting of 196×125 grid points within the geographical bounds of 1.1°W–18.4°E and 44.1°N–56.5°N. This grid corresponds to a subset of HRES global grid. H-RES, apart from the ECMWF's other ensemble runs, provides a high-resolution and deterministic 10-day weather forecast. Data assimilation in HRES involves a sequential 4D-Var assimilation scheme that updates a previous model forecast with new observations including ground-based, satellite-based and other meteorological measurements for initialisation (Bannister, 2001; Bonavita and Laloyaux, 2020). While sharing core characteristics with other forecast ensemble members, HRES stands out with its more detailed analysis, higher resolution, and improved representation of land-sea processes (Owens and Hewson, 2018). HRES data is downloaded for the given study domain (Figure 2A) and is preprocessed to be used for analysis and DL training. The preprocessing steps involve converting the precipitation units from m to mm, cumulative into instantaneous, and selecting the first 24 h forecast for each day, and merging them into a single NetCDF file using Climate Data Operators (CDO).

Satellite-based data from H-SAF, P-AC-SEVIRI-PMW (referred to as H61B in this section and H-SAF throughout the rest of the manuscript) is an hourly accumulated precipitation dataset. H61B is derived by integrating Passive Microwave (PMW) sensor and SEVIRI (visible and infrared) instrument data. H61B improves prior H-SAF precipitation data by classifying rainfall as convective or stratiform through adjustments in the rain rate-brightness temperature relationship (EUMETSAT, 2021b, 2022). H61B data spans from latitude 60°S to 67.5°N and longitude 80°W to 80°E, and is rectified, projected, and resampled into a defined grid (Figure 2B). The spatial resolution is ~4.8 km at nadir, but it decreases in areas away from the nadir point, reaching ~8 km over Europe. To obtain H-SAF data for the first domain and grids, a reference algorithm by Mueller et al.

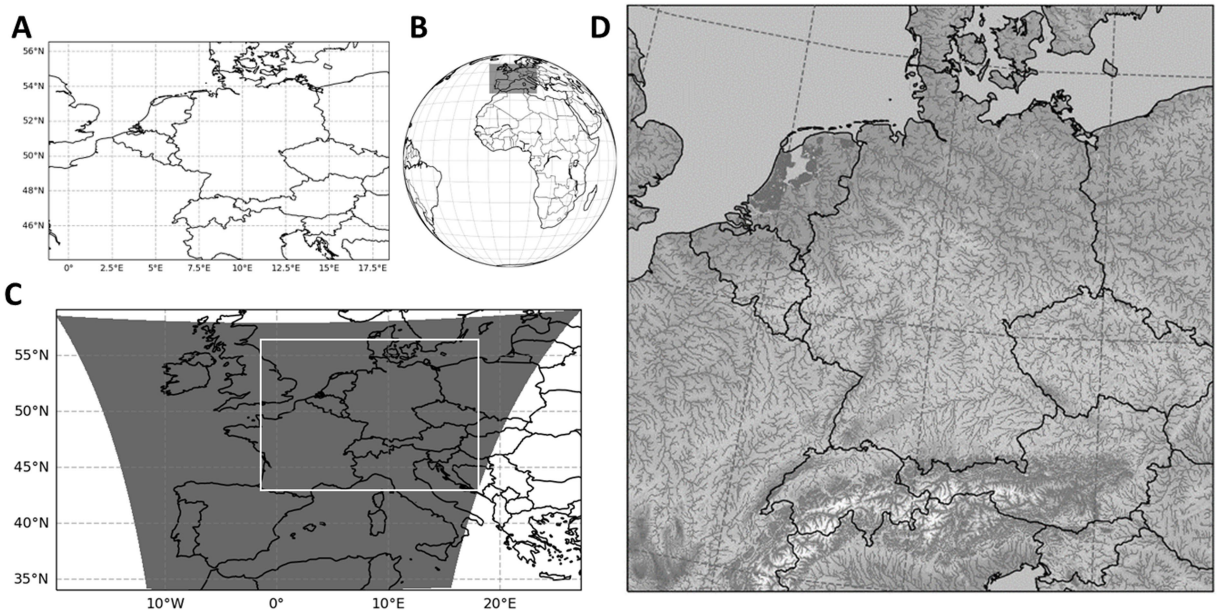


FIGURE 2
Study domain and grids of HRES **(A)**, H-SAF **(B)**, resampled and trimmed H-SAF **(C)** data, and DE06 grid **(D)**. Maps shown on panels **(A,C)** are mapped in Plate Carree projection. Panel **(B)** is shown in geostationary (azimuthal) projection, and **(D)** is shown in EUR-11 rotated-pole projection as used in the DE06 setup.

(2018), is used to establish the relationship between the pixel locations to geographic coordinates. The data from the full disc area is trimmed for size reduction (Figure 2B), resampled to the $0.1° \times 0.1°$ HRES grid using *remapbil* bilinear interpolation function (Figure 2C), and then trimmed to fit the study domain (Figure 2A). Resampling is necessary to match the input and output dimensions of the data for calculating the mismatches in corresponding grids required for the U-Net architecture.

The second grid definition (Figure 2D) has a resolution of $0.0055° \times 0.0055°$ ($\sim 0.611\,\text{km} \times 0.611\,\text{km}$) in the lateral direction, covering 2000×2000 grid cells. This grid extends over 15 terrain-following vertical model layers with increasing thickness from the land surface to the model bottom. The uppermost layer reaches from the surface to 2 cm depth and the lowest layer extends from 42 to 60 m below the surface. This is called the DE06 grid in the ParFlow/CLM setup, a high-resolution hydrological model that simulates the water and energy cycles in the study domain. For more information about the DE06 setup, the reader is referred to Belleflamme et al. (2023). All the precipitation data are remapped using *remapbic* bicubic interpolation function onto the DE06 grid (Figure 2D) to be used in ParFlow/CLM simulations.

2.3 Precipitation correction

2.3.1 DL network setup

We employ a Deep Learning (DL)-based framework introduced by Patakchi Yousefi and Kollet (2023). This approach relies on the non-linear space–time relationships between the extracted features from the input data, and the mismatch between the model-based HRES and H-SAF reference data. The framework consists of two steps (Equations 1, 2). First step is to learn the mismatch data:

$$\mathrm{DL}_{pr}\left(I_{pr,i,t}, w_{pr}\right): I_{pr,i,t} \rightarrow \delta_{pr,i,t}, \text{where } \delta_{pr,i,t} = m_{pr,i,t} - r_{pr,i,t} \quad (1)$$

where $\mathrm{DL}_{pr}\left(I_{pr,i,t}, w_{pr}\right)$ represents the DL network with inputs $I_{pr,i,t}$ to be trained on the mismatch data $\delta_{pr,i,t}$ for precipitation pr as a modeled $m_{pr,i,t}$, and observed $r_{pr,i,t}$ atmospheric variable over time t and location i.

The second step is to remove mismatches from $m_{pr,i,t}$ using the trained weights in the network independent from observations:

$$\hat{m}_{pr,i,t} = m_{pr,i,t} - \hat{\delta}_{pr,i,t} \quad (2)$$

where the corrected HRES precipitation (HRES-C) $\hat{m}_{pr,i,t}$ is obtained by removing the predicted mismatch $\hat{\delta}_{pr,i,t}$ from the original modeled data $m_{pr,i,t}$.

The U-Net architecture employed in this study for precipitation correction follows the work of Patakchi Yousefi and Kollet (2023), featuring a U-Net CNN (Ronneberger et al., 2015) with a distinct feature of squeeze-and-excitation (SE) blocks between the two convolutional functions both in the up-sampling and down-sampling steps. The SE blocks contain global average pooling followed by convolutional operations with relu and sigmoid activations. Global average pooling layer computes the average value of all feature maps for each batch and channel, summarizing the information into a single representative value. This map is reshaped and processed through two 1×1 convolutional layers. The SE blocks perform channel-wise attention, emphasizing informative features and suppressing less useful ones (Hu et al., 2019). The architecture incorporating the layers and blocks discussed above is illustrated in Figure 3.

Previous similar studies emphasize including spatiotemporal information to improve U-Net predictions (Bastos et al., 2021; Teimouri et al., 2019). In our approach, we include a total of seven input channels.

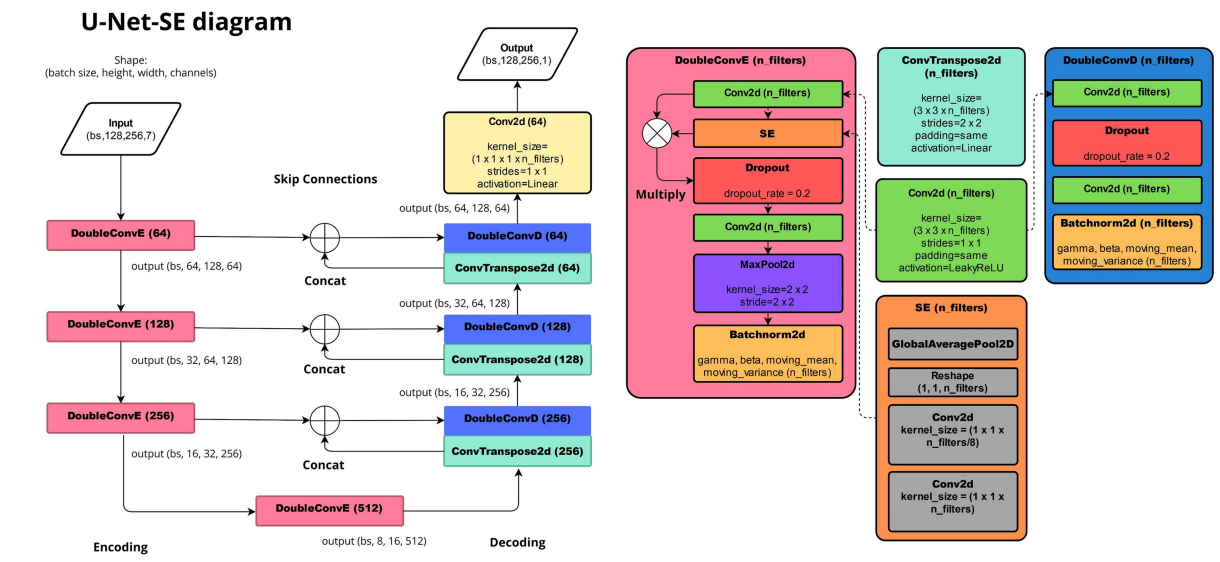


FIGURE 3
Schematic representation of the U-Net convolutional neural network architecture used in this study with Conv2d, SE, Dropout, MaxPool2d, Batchnorm2d, GlobalAveragePool2D, ConvTranspose2d, Concat, and Multiply layers from *Keras* library.

These channels include modeled precipitation data at t and $t-1$, geographical coordinates (latitude and longitude), altitude maps and calendar information specifying year, and day of the year (assigned to all pixels in the corresponding image). Each input channel undergoes batch normalization layer (*Keras* library in *python*), immediately after the input layer to standardize each channel's values by subtracting the batch mean and dividing by the square root of the batch variance. This normalization step ensures that all channels contribute effectively to the learning process by maintaining consistent scaling with trainable parameters. All corresponding input and target data with 196×125 dimension are filled with zero padding to obtain images with the nearest square dimensions of 256×128 for training in U-Net.

2.3.2 U-Net training

We employ and compare two loss functions: Mean Squared Error (MSE) and Weighted MSE (WMSE). We employ and compare two loss functions (Equations 3, 4):

$$MSE_T = \frac{1}{N_T N_S} \sum_{t=1}^{N_T} \sum_{i=1}^{N_S} M_{land,i}\, M_{avail,i,t} \left(\hat{\delta}_{pr,i,t} - \delta_{pr,i,t} \right)^2 \quad (3)$$

where MSE_T represents the loss during the training period; N_T and N_S represent the number of data points in time and space; $M_{land,i}$ and $M_{avail,i,t}$ represent binary masks for land and H-SAF data availability; $\hat{\delta}_{pr,i,t}$ and $\delta_{pr,i,t}$ represent the predicted and actual precipitation mismatches over location i and time t. MSE_V MSE_{TE} are similarly defined for validation data, as well. The MSE loss function is formulated to reduce the mean squared error between predicted and actual precipitation.

The WMSE loss function is defined as

$$WMSE_T = \frac{1}{N_T N_S} \sum_{t=1}^{N_T} \sum_{i=1}^{N_S} W_{pr,i,t}\, M_{land,i}\, M_{avail,i,t} \left(\hat{\delta}_{pr,i,t} - \delta_{pr,i,t} \right)^2 \quad (4)$$

where $WMSE_T$ represents the weighted loss during the training period; $W_{pr,i,t}$ corresponds to the intensity weights calculated for each pixel i and time t over the training period. $WMSE_V$ $WMSE_{TE}$ are similarly defined for validation data, as well.

The WMSE loss function introduces intensity-based weighting to emphasize higher-intensity precipitation events. To calculate $W_{pr,i,t}$, we obtain the probability of H-SAF precipitation categorized into three levels: dry (up to 0.1 mm/h), light (0.1 to 2.5 mm/h), and moderate to heavy (more than 2.5 mm/h). Toprioritize moderate to heavy over dry and light categories, we calculate the $W_{pr,i,t}$ for each category such that the sum of inversely weighted probabilities for each precipitation category equals to one. The calculated weights are presented in Table 1. Furthermore, in both loss functions, we utilize a stationary land and a dynamic data availability mask specifically to train the loss function over land grids where H-SAF data is accessible.

2.3.3 Hyperparameter tuning

A simple hyperparameter tuning (HPT) is employed to optimize the performance of U-Net. The tuned parameters include initial learning rate (LR), batch size (BS), and the number of filters in the first U-Net network layer. The LR values tested are 0.01, 0.001, and 0.0001, the considered batch sizes are 2, 4, 8, and 16, and the number of filters in the first layer of the U-Net architecture is chosen from 8, 16, 32, and 64. The possibility of reaching an optimum result with a wider range

TABLE 1 Probability of occurance (%) for each precipitation intensity over the training dataset for HRES, H-SAF and HRES-C.

Intensity (mm/h)	Probability of HRES	Probability of H-SAF	Assigned weight(W_{pr})
0 <=pr< 0.1	80.167%	79.231%	0.01
0.1 <=pr< 2.5	19.140%	19.980%	0.04
2.5 <=pr	0.685%	0.779%	0.95

of combinations is possible. However, the current selection of 48 different combinations is reasonable, given the computational resources required for more extensive search.

The study period for precipitation correction is between 2020-07-01 13UTC and 2023-04-25 12UTC. The data is randomly split 50/50 into training and validation sets. The validation set consists of multiple clusters of 10 consecutive days (240 h) within the study period with an equal duration of hours as the training set.

In HPT, the goal is to find the best combination of hyperparameters that result in better and robust skill in predicting mismatches, toward the final goal of improving HRES data. Therefore, for each parameter combination, and in each training epoch, model training is performed using the training set and validated with the validation set. If there is no improvement in validation loss after 8 epochs, the training is stopped (early stopping). If the validation loss does not improve after 2 epochs, the learning rate is reduced by a factor of 0.5. Following these runs, we analyze the results to identify the HPT combination that leads to the least validation loss and save it for producing the corrected HRES. The results from HPT are provided in Section 3.1.

2.4 ParFlow/CLM simulations

We employ the ParFlow v3.8.0 integrated hydrological model which utilizes partial differential equations (PDEs) to simulate variably saturated subsurface and groundwater flow, integrated with overland flow, which constitutes the upper boundary condition (Kollet and Maxwell, 2006; Kuffour et al., 2020; Maxwell et al., 2015). The physics-based methodology of ParFlow yields consistent results across time and space scales, ranging from watershed hydrodynamics to large-scale continental simulations (Maxwell et al., 2015; Saadi et al., 2023a). ParFlow can be run efficiently on large HPC systems, including GPUs (Burstedde et al., 2018; Hokkanen et al., 2021; Kollet et al., 2010).

In this study, the model uses the same setup with the same external parameter files for slopes, soil hydraulic properties, land cover, etc. as in the experimental forecasts runs by Belleflamme et al. (2023). The Community Land Model (CLM) in ParFlow/CLM, enhances the representation of energy and mass fluxes by adding land surface exchange processes such as interception and evapotranspiration (Kuffour et al., 2020). We use ParFlow/CLM model outputs from the aforementioned study from 2020-06-30 13UTC up to the next 24 h for our runs as the initial conditions. From 2020-07-01 13UTC and onwards, three different precipitation forcing data are tested with dedicated ParFlow/CLM model runs: H-RES, H-SAF, and corrected HRES (HRES-C). The model is run on the GPU compute nodes of the JUWELS Booster HPC system at the Jülich Supercomputing Centre.

2.5 Evaluation methods

To assess the fidelity of the precipitation data, we employ a variety of evaluation metrics at hourly and monthly time scales such as Mean Error (ME), Root Mean Squared Error (RMSE), and Pearson Correlation (COR). ME quantifies the overall bias against the reference data, indicating whether the predictions tend to overestimate or underestimate actual values. RMSE measures variability and magnitude of deviations from reference data. COR is utilized to indicate the strength and direction of the linear relationship between simulated (or predicted) and reference values. COR and RMSE metrics are also used to evaluate simulated SM. These metrics are calculated according to Patakchi Yousefi and Kollet (2023).

COR and RMSE metrics are used for comparing simulated against observed SM. Additionally, we evaluate the spatial patterns using the SPAEF multi-component metric that accounts for the spatial correlation, variability rate, and histogram match. For further details on the SPAEF metric, refer to Koch et al. (2018).

False Alarm Ratio (FAR) and Probability of Detection (POD) are utilized to evaluate the reliability of specific precipitation thresholds of 0.1, 2.5, and 10 mm/h by different datasets. FAR measures the ratio of falsely identified events above a threshold to the total number of identified events above that threshold. POD represents the proportion of correctly identified precipitation events above a specific threshold out of all events that exceeded that threshold. For more information regarding POD and FAR, the reader is referred to Amjad et al. (2020).

3 Results and discussion

In this section, we present and discuss key results from calculating the intensity weights in section 3.1, network hyperparameter tuning in Section 3.2, precipitation correction in Section 3.3, comparing SM simulations in Section 3.4, and validation of precipitation products in Section 3.5.

3.1 Intensity weights

The inverse intensity weights were calculated by first calculating the probability of occurrence for each precipitation category based on the H-SAF training dataset. The assigned weight is shown for each precipitation category.

3.2 Hyperparameter tuning

While exploring the efficacy of the loss function based on WMSE in HPT experiments, a marginal improvement (less than 1%), between the WMSE and the MSE was observed for both training and validation (not shown). Given the limited impact of the WMSE on overall model performance, we proceed with the MSE results in our subsequent analyses and discussion.

Figure 4A shows the training (solid colored lines) and validation (dotted colored lines) MSE loss over the training epochs for the top five HPT settings according to the validation loss. The remaining gray lines represent the rest of the experimented HPT settings. We found that larger initial filter numbers (IFN) and batch sizes (BS) contribute to a smaller training loss and validation loss (i.e., model robustness). The best HPT setting (LR = 0.001, BS = 16, LRF = 0.5, and IFN = 64) is shown with the blue line.

Figures 4B,C show the mismatch scatter diagram for actual $\delta_{pr,i,t}$ against predicted $\delta_{pr,i,t}$ training and validation datasets. The consistency between the training and validation (in terms of predicting mismatch) sets is reflected by similar MSE values of 0.183 and 0.188 mm/h, respectively. Nevertheless, prediction skill is limited concerning negative mismatches (i.e., instances where HRES indicates

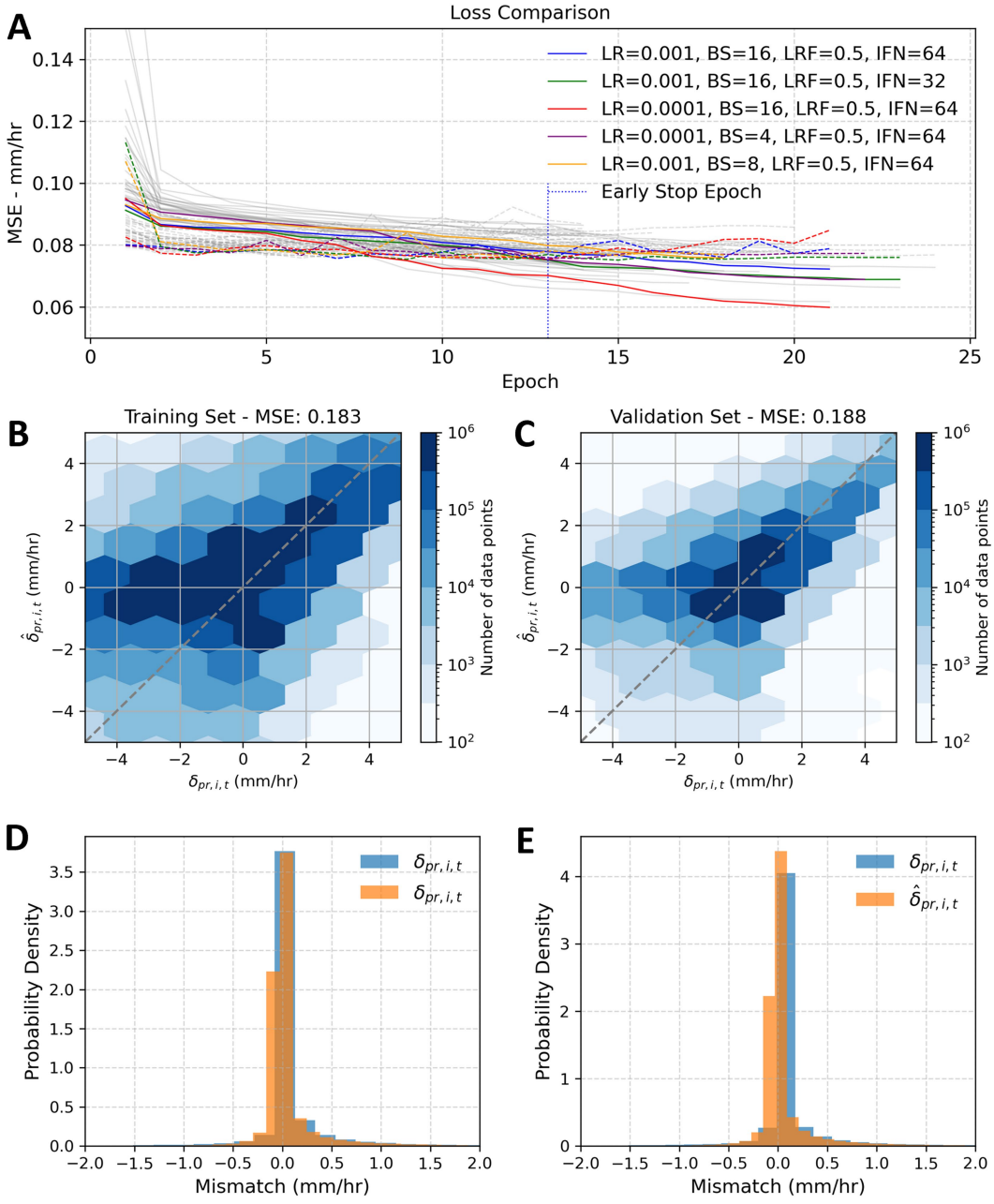


FIGURE 4
Training and validation loss comparisons for hyperparameter tuning (HPT) experiments for MSE loss function. Panel **(A)** shows the training and validation loss over training epochs for the top five HPT settings. Panels **(B–E)**, respectively, represent the scatter and probability density function (PDF) diagrams for predicted $\hat{\delta}_{pr,i,t}$ against actual $\delta_{pr,i,t}$ data.

low or zero precipitation) to be pronounced during seasons characterized by increased spatiotemporal variability, such as summer and fall, with increased likelihood of convective precipitation (Patakchi Yousefi and Kollet, 2023).

Figures 4D,E present the Probability Density Functions (PDFs) for mismatch data in the training and validation datasets, ranging from −2.0 to +2.0 mm/h. The distribution indicates a bias toward positive mismatches, because HRES generally forecasts higher precipitation rates than H-SAF observations. Overall, there is notable agreement between the actual and predicted mismatches, with the majority of the data concentrated between −1 and + 1 mm/h.

3.3 Precipitation correction

In the following, we further discuss our comparison results on the corrected HRES data (HRES-C) and the original HRES data against H-SAF between 2020-07-01 13UTC and 2023-04-25 12UTC. At this point, our assumption is that satellite-based precipitation (H-SAF) represents the reference. So, the better the agreement between HRES-C or HRES against H-SAF, the better the assumed precipitation accuracy.

Figures 5, 6 provide a spatiotemporal assessment of hourly error metrics, mean error (ME), root mean squared error (RMSE), and

Pearson correlation (COR), across four seasons comparing HRES and corrected HRES-C data against H-SAF in the study region over the study period. Figure 5A illustrates the initial overestimation by HRES against H-SAF, depicting ME >0, while Figure 6A shows a post-correction improvement with ME ~0. Seasonal variations of RMSE are shown in Figures 5B, 6B, with increased error rates during the summer season, particularly in the southern regions corresponding to the Alps.

Figure 7 presents maps illustrating the percent improvement achieved through the implementation of correction across four seasons as a quantitative comparison of alignment of HRES-C and HRES with respect to H-SAF, which is assumed as the *ground truth*. The average improvement in ME, RMSE, and COR is generally positive across all seasons when compared to H-SAF. An exception is observed in the summer season, where the average improvement in ME is negative. This discrepancy can likely be attributed to the presence of convective precipitation during summer, which tends to be more localized and intense, introducing uncertainties in both model- and satellite-based data.

3.4 Soil moisture simulations

Volumetric SM is used to study the impact of H-RES, H-SAF, and HRES-C precipitation forcing datasets on ParFlow/CLM simulations. To evaluate the accuracy and consistency of SM simulations, COR and RMSE metrics are used. The comparison is drawn against the ESA CCI SM data from 2020-07-01, to 2022-12-31, using daily and monthly values for comparison. For this, ParFlow/CLM data representing the uppermost soil layer (0–2 cm depth) are aggregated (averaged values of all nearest neighbor grids) to the coarser ESA CCI grid with a resolution of 0.25°.

The results depicted in Figure 8 surprisingly reveal that the SM simulations driven by the HRES 24 h forecast present lower RMSE and higher COR than those generated by satellite-based precipitation (H-SAF) or the corrected forecast (HRES-C). This unexpected outcome is particularly interesting considering the anticipated better agreement in SM simulations with ESA CCI through satellite-driven precipitation or DL-corrected forecasts. The existing 4D-Var assimilation scheme of HRES with a diverse array of observations, including *in-situ* and remotely sensed data, potentially establishes the reliability of the HRES 24 h precipitation forecast in representing SM dynamics as it appears to closely mirror the trends observed in ESA CCI SM data.

Figure 8 reveals consistency in the patterns observed between H-SAF and HRES-C simulation outcomes across daily and monthly correlations, with a slightly better performance observed in the former dataset. This consistency highlights the effectiveness of the DL-based correction method in forcing SM simulations that closely mirror the SM driven by H-SAF dataset.

The HRES dataset shows the best overall SPAEF performance (Figure 9). Both HRES and H-SAF have similar histogram match ratios, but H-SAF shows a higher CV ratio, indicating greater

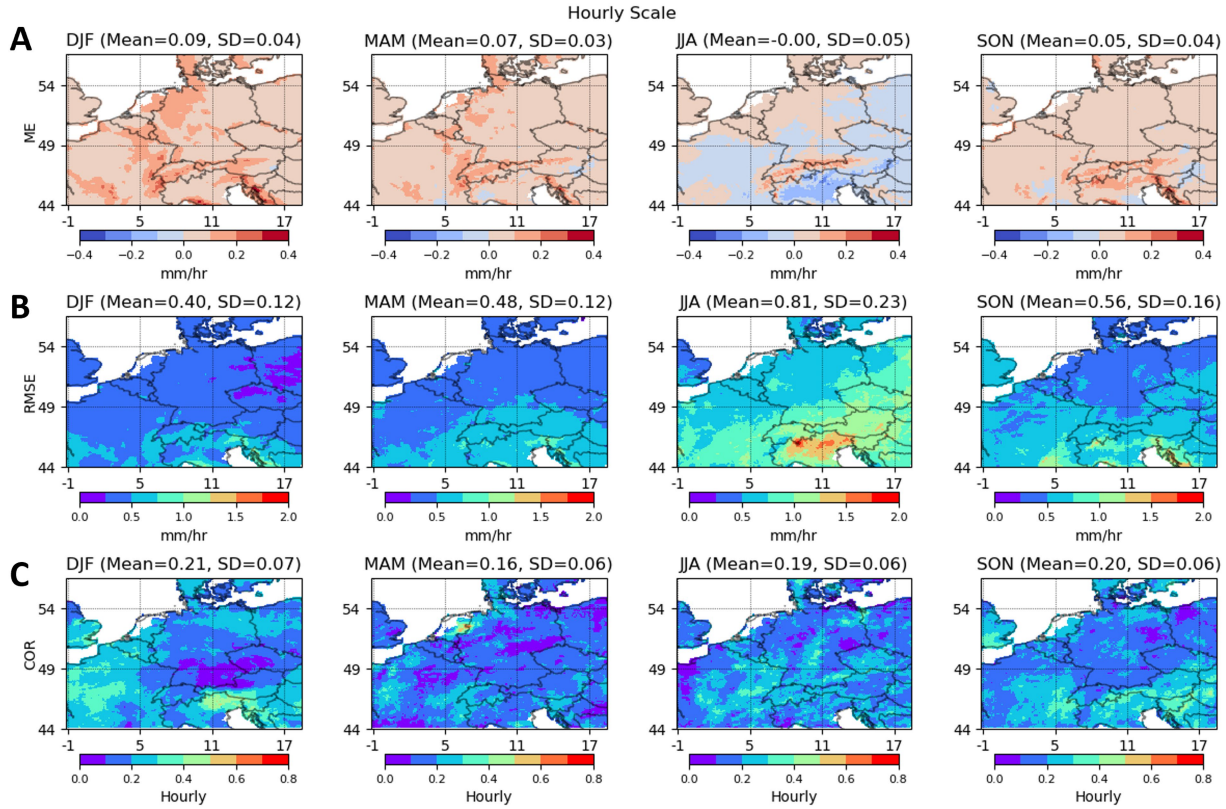


FIGURE 5
Maps of seasonal mean error (ME, **A**), root mean squared error (RMSE, **B**), and Pearson correlation (COR, **C**) metrics shown for HRES against H-SAF over four seasons in hourly scale.

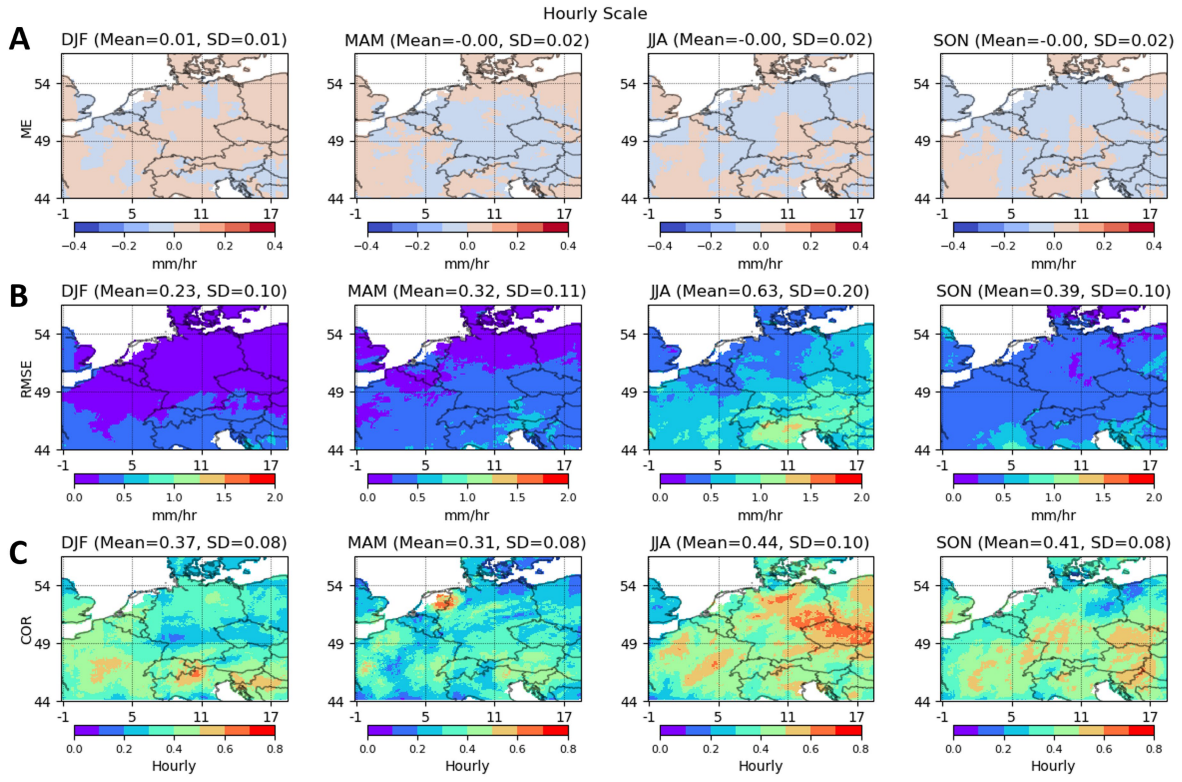


FIGURE 6
Maps of seasonal mean error (ME, **A**), root mean squared error (RMSE, **B**), and Pearson correlation (COR, **C**) metrics shown for HRES-C against H-SAF over four seasons in hourly scale.

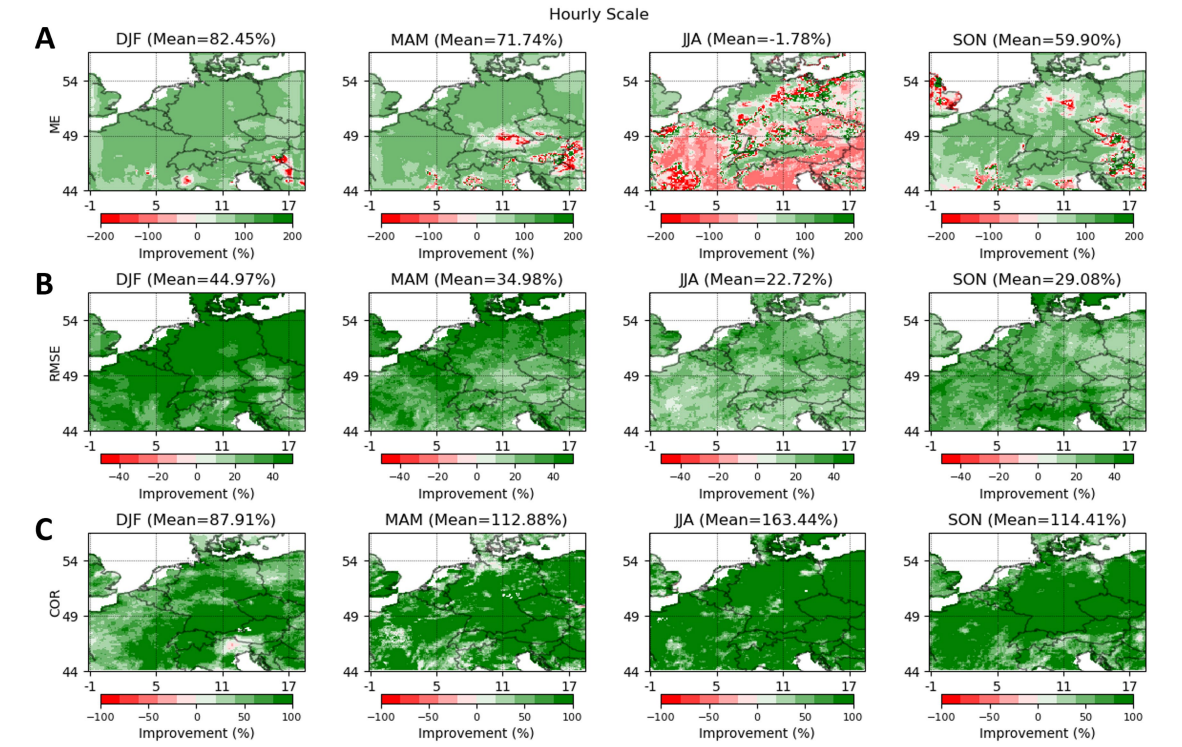


FIGURE 7
Maps of mean improvement [i.e., reduction in mean error (ME, **A**) and root mean squared error (RMSE, **B**) and increase in Pearson correlation (COR, **C**) in hourly scale] of HRES-C over HRES shown in percentage over four seasons.

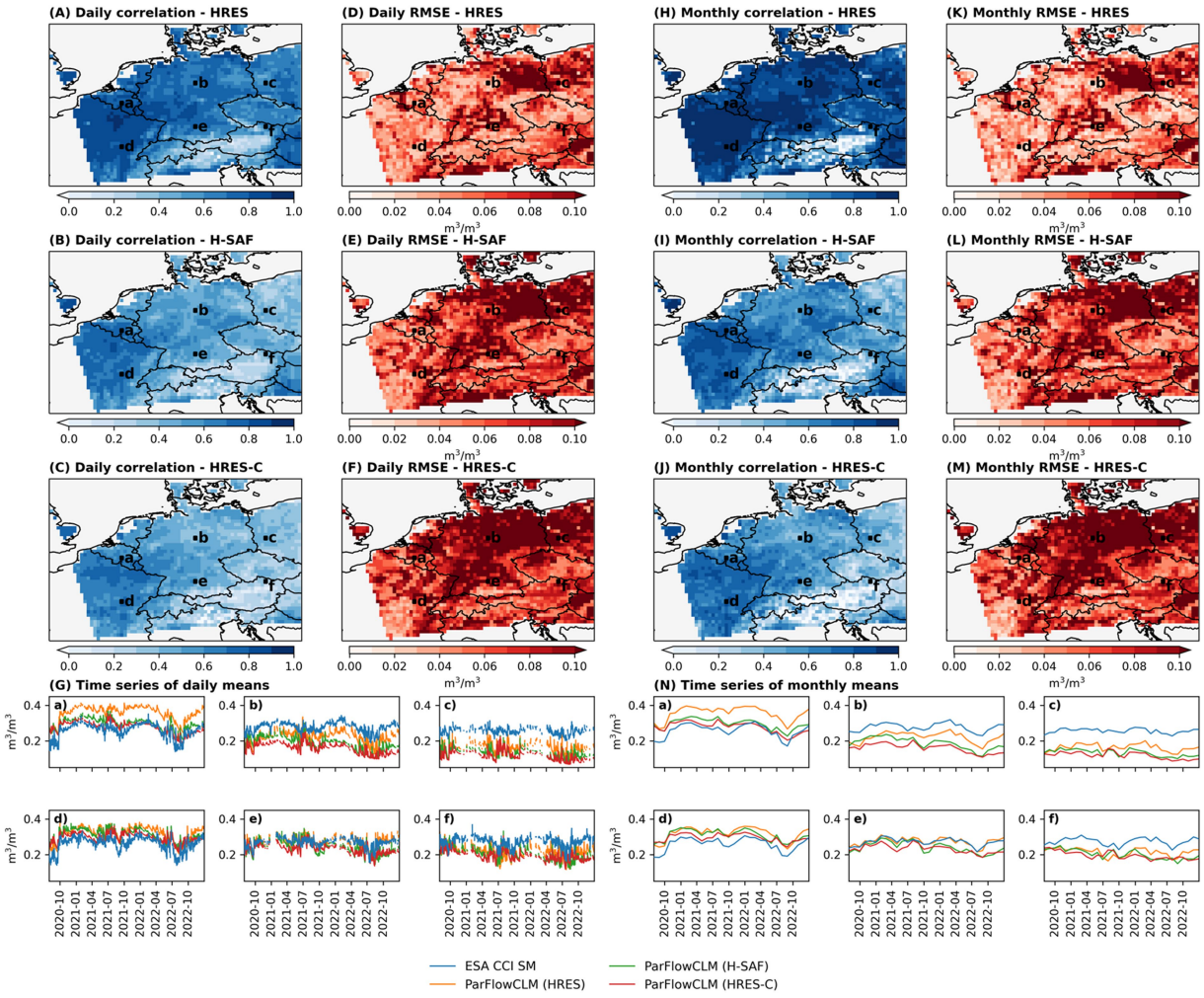


FIGURE 8
Comparison of daily **(A–G)** and monthly **(H–N)** uppermost 2 cm volumetric SM simulations obtained by H-RES, H-SAF and HRES-C precipitation datasets run by ParFlow/CLM against ESA CCI SM. Panels **(A–C)** and **(H–J)** show the distribution of correlation, and panels **(D–F,K–M)** show the distribution of root mean square error of SM values over the study area between 2020-07-01 and 2022-12-31. Panels **(G,N)** show the daily and monthly time-series of SM over six random locations **(a–f)** pinned on the study area.

variability mismatch and lower spatial COR, indicating less consistency in spatial patterns with ESA CCI SM data. While HRES-C dataset achieves the highest spatial COR, indicating accurate spatial pattern representation, it does not lead to an overall improvement in the SPAEF metric.

3.5 Precipitation data validation

In Section 3.2, the assessment reveals that HRES-C agrees well with H-SAF concerning ME, RMSE, and COR. However, as explored in Section 3.3, the alignment of SM simulations with observations (ESA CCI data) is not consistently improved. This discrepancy prompts a re-evaluation of the precipitation data using *in-situ* measurements (i.e., rain gauge station data), focusing on H-RES, H-SAF, and HRES-C, to understand the reasons for the observed gaps.

In the re-evaluation, 16 rain gauges were randomly picked across Germany with available data from July 2020 to the end of 2022 and located the nearest pixel from the precipitation datasets to the station locations. The monthly time-series of the three precipitation datasets and the rain gauge station data as well as H-SAF data quality index (QI) is provided in Figure 10A. Figure 10B shows the locations of these stations over the map of H-SAF mean data availability (mean ratio of available data during the study period) shown in percentage. Figure 10C shows the station locations over the map of mean QI provided by H-SAF metadata and averaged over the study period.

According to Figure 10A, the monthly sum of precipitation obtained by HRES-C generally matches H-SAF, resembling our findings in downstream simulations (Figure 8). However, HRES demonstrates a better agreement with rain gauge data. H-SAF underestimates precipitation compared to the rain gauge data, in line with findings from a validation report noting a slight underestimation in all precipitation classes for the H61B H-SAF product (EUMETSAT, 2022).

The H-SAF Quality Index (QI) steadily increases from March to July 2022, indicating improved data reliability during this period (see the gray lines over time-series in Figure 10A). The mean QI is

10.3389/frwa.2024.1439906

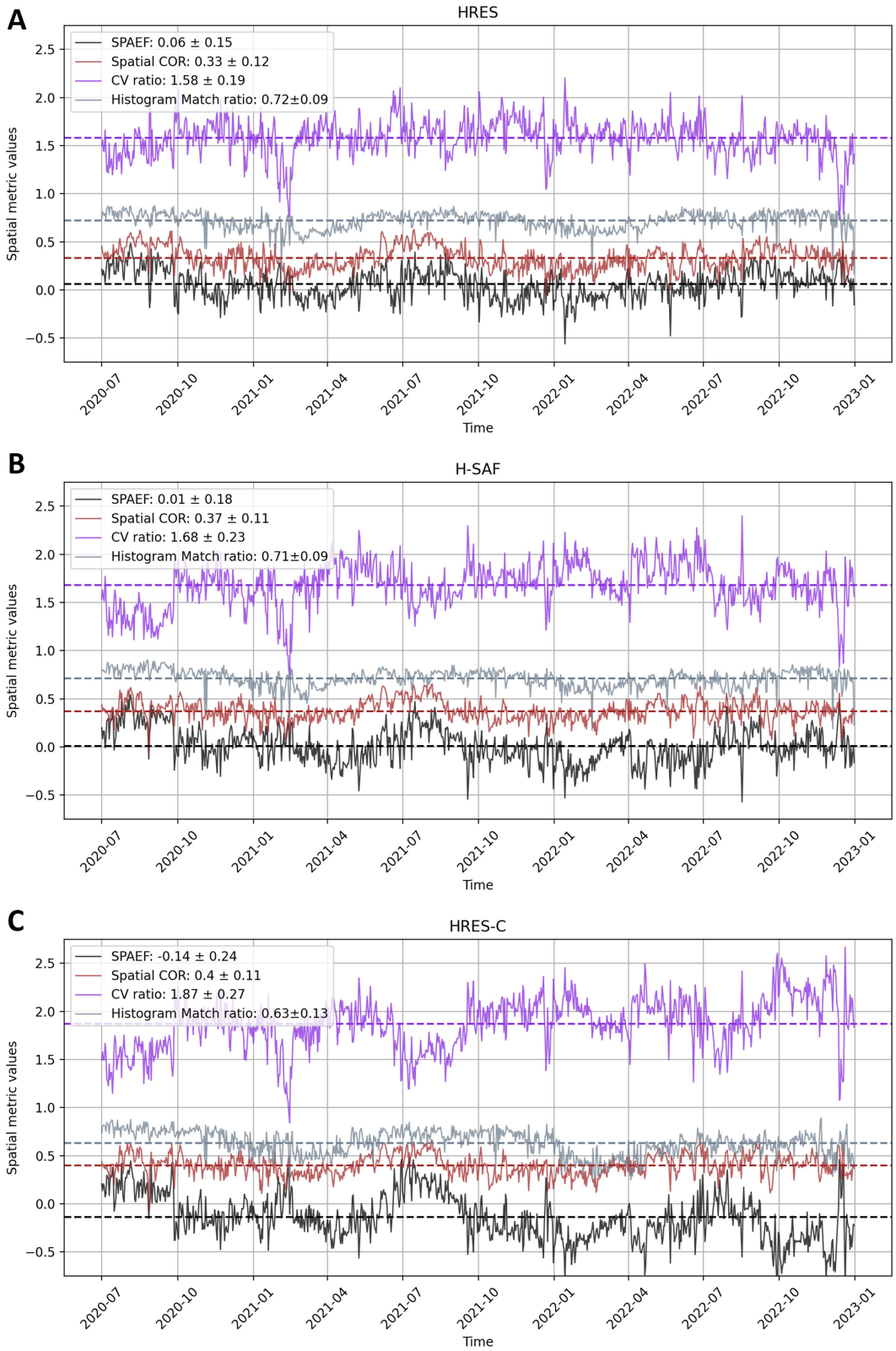


FIGURE 9
Daily time-series the spatial metrics for SM simulations using HRES **(A)**, H-SAF **(B)**, and HRES-C **(C)** precipitation datasets against ESA CCI SM. The metrics include the Spatial Efficiency (SPAEF) and its components: spatial Pearson correlation coefficient (spatial COR), coefficient of variation (CV) ratio, and histogram match ratio. The average values of the metrics are shown in mean ± standard deviation. In figure, we analyze the spatial performance of the SM simulations driven by the HRES, H-SAF, and HRES-C precipitation datasets using the Spatial Efficiency (SPAEF) metric and its components: spatial Pearson correlation coefficient (spatial COR), coefficient of variation (CV) ratio, and histogram match ratio. The SPAEF metric, ranging from $-\infty$ to 1, quantifies the agreement between simulated and observed SM, where higher values indicate better spatial performance.

10.3389/frwa.2024.1439906

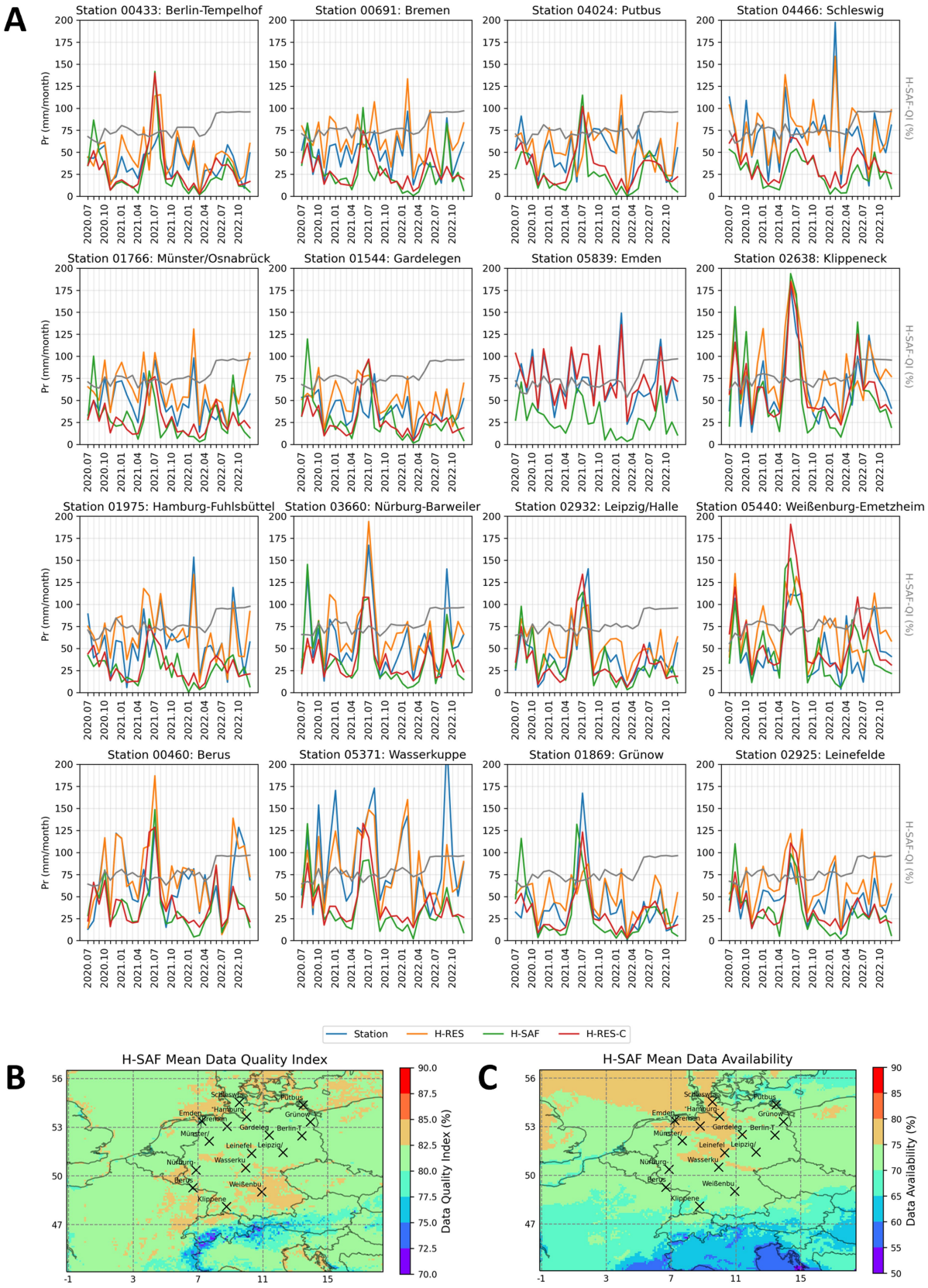


FIGURE 10
Comparison of H-RES, H-SAF, and HRES-C against rain gauge observations. Panel **(A)** shows time-series of monthly precipitation sum (mm/month) as well as average H-SAF quality index (QI) for 16 randomly selected rain gauge stations across the study domain from July 2020 to the end of 2022. Panels **(B,C)**, respectively, show the maps of H-SAF data availability and quality index (H-SAF-QI) across the study domain. Locations of randomly selected rain gauge stations are pinned over the maps.

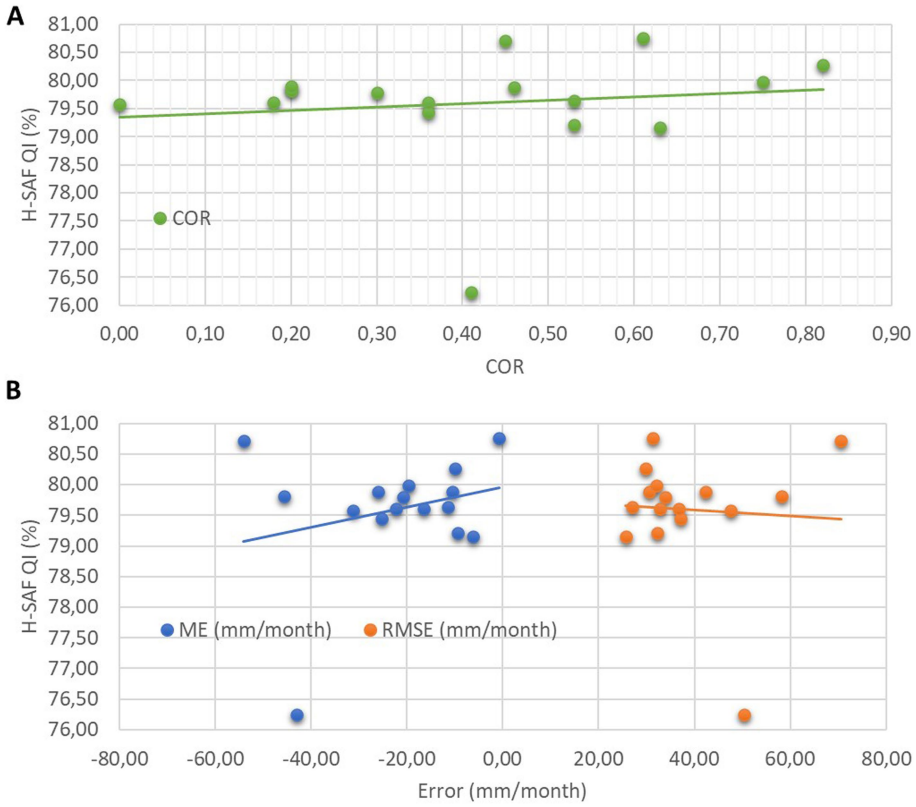


FIGURE 11
Scatterplots representing the relationship between average H-SAF QI (%) and COR **(A)** as well as ME and RMSE **(B)** over 16 rain gauge stations.

calculated over all station locations and compared against the COR in Figure 11A as well as mean RMSE and ME in Figure 11B. There seems to be a positive relationship between the reduction in absolute ME and COR as well as increment in RMSE and the mean QI of each station. Therefore, the quality and accuracy of HRES-C could potentially be attributed to errors and uncertainties in H-SAF both before and after July 2022. While assessing changes in performance post-July 2022 is possible by training the network after this date and evaluating the changes in performance, this is out of the study's scope.

Figure 12 and Table 2 provide an evaluation of hourly error metrics for H-RES, H-SAF-O, H-SAF, and HRES-C. H-SAF-O represents H-SAF data without preprocessing to account for its effects of trimming and remapping on HRES domain and grids. In comparison, HRES demonstrates better performance with lower ME (in absolute value), reduced RMSE, and higher COR against the 16 randomly chosen rain gauge stations.

Preprocessing H-SAF grid-wise to align with HRES exhibits slightly, but not significantly, higher ME and RMSE and a lower COR than H-SAF-O. Moreover, HRES-C shows improvements over both H-SAF and H-SAF-O, showing better metrics in ME, RMSE, COR, and POD of more than 0.1 mm/h.

For light precipitation events (POD >= 0.1 mm/h), HRES exhibits the highest detection rate, followed by HRES-C, while H-SAF and H-SAF-O show significantly lower POD. In terms of moderate precipitation (POD >= 2.5 mm/h), HRES shows higher detection, while the other products have lower detection rates. None of the products effectively detect high-intensity precipitation events (POD >= 10 mm/h). Regarding FAR, HRES and HRES-C have lower false alarm rates for light and moderate precipitation compared to H-SAF and H-SAF-O. For high-intensity events, HRES-C stands out with the lowest FAR, indicating a much lower rate of false alarms compared to H-SAF and H-SAF-O.

4 Conclusion

In this study, we evaluated the immediate and downstream impacts of implementing U-Net CNN DL-based correction on HRES precipitation using the satellite-based H-SAF h61 precipitation observations toward operational hydrological simulations over central Europe. The findings of this study highlight the effectiveness of using the DL-based precipitation correction in improving the accuracy of precipitation forecast. The corrected precipitation data (HRES-C) exhibited greater agreement in mean error, root mean squared error, and correlation with the assumed reference data (H-SAF), compared to the forecast data (H-RES). This shows the potential of DL-driven methods in correcting precipitation data despite the data imbalance of hourly precipitation and spatiotemporal gaps in satellite-based data.

For soil moisture (SM) simulations, the HRES 24-h forecast used as forcing data shows greater spatiotemporal agreement with the referenced ESA CCI SM data compared to both H-SAF and HRES-C. The additional evaluation of these three datasets against 16 rain gauge data supports these findings with higher consistency of

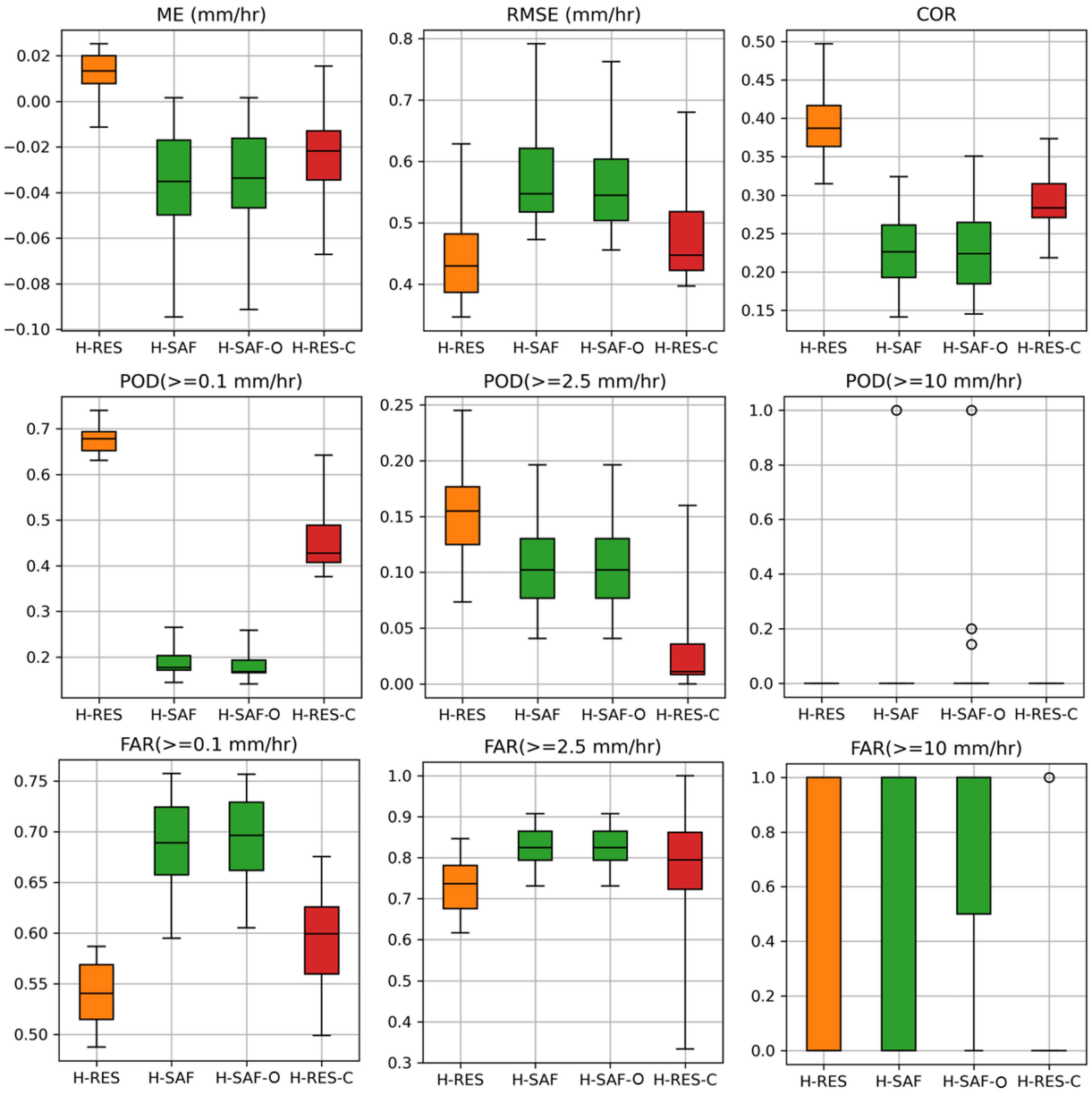


FIGURE 12
Boxplot comparison of H-RES, H-SAF, H-SAF-O, and HRES-C against rain gauge observations. H-SAF-O is the non-preprocessed H-SAF data used to account for the effect of preprocessing the data on evaluation results. POD and FAR, respectively, represent the probability of detection and false alarm ratio for three precipitation rates.

HRES with the rain gauge observations compared to the other datasets. This highlights the importance of reference data quality in DL-based correction approach and challenges the assumptions on the better representativity of H-SAF satellite-based observations as *ground truth* for short-term correction or near real-time application in operational hydrological forecasting.

Future research should focus on the limitations related to deterministic aspects in DL-based correction methods, e.g., to account for uncertainties in the reference data. Furthermore, the accuracy of forecast data is expected to diminish over time, particularly as the forecast lead time extends. We focused on implementing the correction over the initial 24-h forecast from H-RES. Yet, we recognize the potential for improvements in longer-term forecasts.

Code and data availability

The preprocessed data of this article is made available by the authors (Patakchi Yousefi et al., 2024). The codes regarding the data preprocessing, precipitation correction, and post-processing of forcing data can be found at https://gitlab.jsc.fz-juelich.de/kiste/atmoscorrect. ParFlow/CLM documentation and codes are publicly available at https://parflow.org/.

10.3389/frwa.2024.1439906

TABLE 2 Overall hourly error metrics for H-RES, H-SAF, H-SAF-O, and HRES-C precipitation data presented as mean ± standard deviation.

	ME (mm/h)	RMSE (mm/h)	COR	POD > =0.1 mm/h	POD > =2.5 mm/h	POD > =10 mm/h	FAR > =0.1 mm/h	FAR > =2.5 mm/h	FAR > =10 mm/h
HRES	0.01 ± 0.01	0.45 ± 0.08	0.39 ± 0.05	0.74 ± 0.02	0.16 ± 0.05	0.00 ± 0.00	0.63 ± 0.03	0.74 ± 0.06	0.31 ± 0.46
H-SAF	−0.04 ± 0.03	0.59 ± 0.10	0.22 ± 0.05	0.22 ± 0.04	0.11 ± 0.05	0.06 ± 0.24	0.73 ± 0.04	0.83 ± 0.05	0.67 ± 0.46
H-SAF-O	−0.04 ± 0.03	0.57 ± 0.09	0.23 ± 0.05	0.21 ± 0.04	0.11 ± 0.05	0.08 ± 0.24	0.74 ± 0.04	0.83 ± 0.05	0.73 ± 0.43
HRES-C	−0.02 ± 0.02	0.48 ± 0.08	0.29 ± 0.05	0.53 ± 0.09	0.03 ± 0.04	0.00 ± 0.00	0.66 ± 0.04	0.78 ± 0.15	0.06 ± 0.24

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KP: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AB: Conceptualization, Formal analysis, Software, Visualization, Writing – review & editing. KG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing. SK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. We acknowledge receive funding through the BMBF BiooekonomieREVIER funding scheme with its "BioRevierPlus" project (funding ID: 031B1137D/031B1137DX). Furthermore, we gratefully acknowledge the AI Strategy for Earth System Data (KI:STE) and Earth System Modeling (ESM) projects for funding this study by providing computing time the supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Acknowledgments

We thank the ECMWF for providing free access to their weather forecasts, EUMETSAT for satellite observations, Deutscher Wetterdienst (DWD) for rain gauge observations, and ESA CCI for soil moisture data. We also acknowledge the use of OpenAI's ChatGPT versions 3.5 and 4 as a tool in initializing the first draft of this manuscript. Finally, we appreciate the constructive suggestions by the reviewers, which greatly contributed to the improvement of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amjad, M., Yilmaz, M. T., Yucel, I., and Yilmaz, K. K. (2020). Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. *J. Hydrol.* 584:124707. doi: 10.1016/j.jhydrol.2020.124707

Bannister, R. N. (2001). Elementary 4D-VAR, DARC Technical Report No. 2: University of Reading, UK.

Bastos, B. Q., Cyrino Oliveira, F. L., and Milidiú, R. L. (2021). U-convolutional model for spatio-temporal wind speed forecasting. *Int. J. Forecast.* 37, 949–970. doi: 10.1016/j.ijforecast.2020.10.007

Belleflamme, A., Goergen, K., Wagner, N., Kollet, S., Bathiany, S., El Zohbi, J., et al. (2023). Hydrological forecasting at impact scale: the integrated ParFlow hydrological model at 0.6 km for climate resilient water resource management over Germany. *Front. Water* 5:1183642. doi: 10.3389/frwa.2023.1183642

Bennett, A., Stein, A., Cheng, Y., Nijssen, B., and McGuire, M. (2022). A process-conditioned and spatially consistent method for reducing systematic biases in modeled streamflow. *J. Hydrometeorol.* 23, 769–783. doi: 10.1175/JHM-D-21-0174.1

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian,, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619, 533–538. doi: 10.1038/s41586-023-06185-3

Bonavita, M., and Laloyaux, P. (2020). Machine learning for model error inference and correction. *J. Adv. Model. Earth Syst.* 12, 1–22. doi: 10.1029/2020ms002232

Burstedde, C., Fonseca, J. A., and Kollet, S. (2018). Enhancing speed and scalability of the ParFlow simulation code. *Comput. Geosci.* 22, 347–361. doi: 10.1007/s10596-017-9696-2

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? *J. Clim.* 28, 6938–6959. doi: 10.1175/JCLI-D-14-00754.1

Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J. M., Boberg, F., et al. (2016). Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations. *Clim. Dyn.* 47, 719–737. doi: 10.1007/s00382-015-2865-x

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., et al. (2024). NOAA's National Water Model: advancing operational hydrology through continental-scale modeling. *J. Am. Water Resour. Assoc.* 60, 247–272. doi: 10.1111/1752-1688.13184

Dablain, D., Jacobson, K. N., Bellinger, C., Roberts, M., and Chawla, N. V. (2023). Understanding CNN fragility when learning with imbalanced data. *Mach. Learn.* 113, 4785–4810. doi: 10.1007/s10994-023-06326-9

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B. (2020). Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets. *Water Resour. Res.* 56:e2019WR026085. doi: 10.1029/2019WR026085

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., and Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrol. Earth Syst. Sci.* 22, 1299–1315. doi: 10.5194/hess-22-1299-2018

EUMETSAT (2021a). Product user manual (PUM) for product H61B, H90 – P-AC-SEVIRI-PMW/P-AC-SEVIRI_E precipitation rate at ground by GEO/IR supported by LEO/MW. EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Darmstadt, Germany.

EUMETSAT (2021b). Product user manual (PUM) for product H60B, H63/P-IN-SEVIRI-PMW, P-IN-SEVIRI_E, Precipitation rate at ground by GEO/IR supported by LEO/MW. EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Darmstadt, Germany.

EUMETSAT (2022). Product validation report (PVR) for products P-AC-SEVIRI-PMW (H61B) and P-AC-SEVIRI_E (H90) (Product Validation Report). EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management, Darmstadt, Germany.

European Centre for Medium-Range Weather Forecasts. (2016). *ECMWF IFS High-Resolution Operational Forecasts*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.

Fang, G. H., Yang, J., Chen, Y. N., and Zammit, C. (2015). Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrol. Earth Syst. Sci.* 19, 2547–2559. doi: 10.5194/hess-19-2547-2015

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., and Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *J. Clim.* 17, 294–304. doi: 10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2

Fersch, B., Senatore, A., Adler, B., Arnault, J., Mauder, M., Schneider, K., et al. (2020). High-resolution fully coupled atmospheric-hydrological modeling: a cross-compartment regional water and energy cycle evaluation. *Hydrol. Earth Syst. Sci.* 24, 2457–2481. doi: 10.5194/hess-24-2457-2020

Gochis, D., Barlage, M., Cabell, R., Dugger, A., Fanfarillo, A., FitzGerald, K., et al. (2020). WRF-Hydro® modeling system technical description, (Version 5.1.1). NCAR Technical Note. 107 pages.

Gochis, J., and Chen, F. (2003). Hydrological enhancements to the community Noah land surface model. (No. NCAR/TN-454+STR). University Corporation for Atmospheric Research. doi: 10.5065/D60P0X00

Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., et al. (2021). A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Adv. Atmos. Sci.* 38, 1444–1459. doi: 10.1007/s00376-021-0215-y

Hess, P., and Boers, N. (2022). Deep Learning for Improving Numerical Weather Prediction of Heavy Rainfall. *J. Adv. Model. Earth Syst.* 14:e2021MS002765. doi: 10.1029/2021MS002765

Hokkanen, J., Kollet, S., Kraus, J., Herten, A., Hrywniak, M., and Pleiter, D. (2021). Leveraging HPC accelerator architectures with modern techniques — hydrologic modeling on GPUs with ParFlow. *Comput. Geosci.* 25, 1579–1590. doi: 10.1007/s10596-021-10051-4

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2019). Squeeze-and-excitation networks. I*EEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/CVPR.2018.00745

Huang, Z., Wu, H., Gu, G., Li, X., Nanding, N., Adler, R. F., et al. (2023). Paired satellite and NWP precipitation for global flood forecasting. *J. Hydrometeorol.* 24, 2191–2205. doi: 10.1175/JHM-D-23-0044.1

Jabbari, A., So, J.-M., and Bae, D.-H. (2019). Precipitation forecast contribution assessment in the coupled meteo-hydrological models. *Atmos.* 11:34. doi: 10.3390/atmos11010034

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., et al. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bull. Am. Meteorol. Soc.* 98, 69–78. doi: 10.1175/BAMS-D-14-00283.1

Kim, D. K., Suezawa, T., Mega, T., Kikuchi, H., Yoshikawa, E., Baron, P., et al. (2021). Improving precipitation nowcasting using a three-dimensional convolutional neural network model from multi parameter phased array weather radar observations. *Atmos. Res.* 262:105774. doi: 10.1016/j.atmosres.2021.105774

Koch, J., Demirel, M. C., and Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geosci. Model Dev.* 11, 1873–1886. doi: 10.5194/gmd-11-1873-2018

Kollet, S. J., and Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.* 29, 945–958. doi: 10.1016/j.advwatres.2005.08.006

Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., et al. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resour. Res.* 46:2009WR008730. doi: 10.1029/2009WR008730

Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S., and Maxwell, R. M. (2020). Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geosci. Model Dev.* 13, 1373–1397. doi: 10.5194/gmd-13-1373-2020

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C. (2013). Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *Int. J. Climatol.* 33, 1367–1381. doi: 10.1002/joc.3518

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science* 382, 1416–1421. doi: 10.1126/science.adi2336

Larraondo, P. R., Renzullo, L. J., Inza, I., and Lozano, J. A. (2019). A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks. *arXiv: Atmospheric and Oceanic Physics*. doi: 10.48550/arXiv.1903.10274

Li, B., Huang, Y., Du, L., and Wang, D. (2021). Bias correction for precipitation simulated by RegCM4 over the upper reaches of the Yangtze River based on the mixed distribution quantile mapping method. *Atmos* 12:1566. doi: 10.3390/atmos12121566

Li, X., Wu, H., Nanding, N., Chen, S., Hu, Y., and Li, L. (2023). Statistical bias correction of precipitation forecasts based on quantile mapping on the sub-seasonal to seasonal scale. *Remote Sens.* 15:1743. doi: 10.3390/rs15071743

Liu, H., Chen, J., Zhang, X.-C., Xu, C.-Y., and Hui, Y. (2020). A markov chain-based bias correction method for simulating the temporal sequence of daily precipitation. *Atmos.* 11:109. doi: 10.3390/atmos11010109

Luo, M., Liu, T., Meng, F., Duan, Y., Frankl, A., Bao, A., et al. (2018). Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: a case study from the Kaidu River basin in Western China. *Water* 10:1046. doi: 10.3390/w10081046

Martins Costa do Amaral, L., Barbieri, S., Vila, D., Puca, S., Vulpiani, G., Panegrossi, G., et al. (2018). Assessment of ground-reference data and validation of the H-SAF precipitation products in Brazil. *Remote Sens.* 10:1743. doi: 10.3390/rs10111743

Maxwell, R. M., Condon, L. E., and Kollet, S. J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geosci. Model Dev.* 8, 923–937. doi: 10.5194/gmd-8-923-2015

Mital, U., Dwivedi, D., Brown, J. B., Faybishenko, B., Painter, S. L., and Steefel, C. I. (2020). Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests. *Front. Water* 2. doi: 10.3389/frwa.2020.00020

Mueller, J., Fowler, G., Gutierrez Lopez, R., Just, D., Ledez, C., Lekouara, M., et al. (2018). Geostationary projection grids for three generations of Meteosat. Proceedings of the 2018 EUMETSAT Meteorological Satellite Conference, 17–21 September 2018, Tallinn, Estonia.

Owens, R., and Hewson, T. (2018). ECMWF forecast user guide. Reading: ECMWF.

Pan, M., Li, H., and Wood, E. (2010). Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resour. Res.* 46. doi: 10.1029/2009WR008290

Panofsky, H. A., and Brier, G. W. (1968). Some applications of statistics to meteorology: Earth and Mineral Sciences Continuing Education, College of Earth and~….

Patakchi Yousefi, K., Belleflamme, A., Goergen, K., and Kollet, S. (2024). Preprocessed EUMETSAT H-SAF h61 Satellite and ECMWF HRES 24h Forecast Precipitation Datasets for Hydrometeorological Applications over Central Europe [Data set]. Zenodo.

Patakchi Yousefi, K., and Kollet, S. (2023). Deep learning of model- and reanalysis-based precipitation and pressure mismatches over Europe. *Front. Water* 5:1178114. doi: 10.3389/frwa.2023.1178114

Piani, C., Haerter, J. O., and Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* 99, 187–192. doi: 10.1007/s00704-009-0134-9

Qi, W., Zhang, C., Fu, G., Sweetapple, C., and Zhou, H. (2016). Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations. *Hydrol. Earth Syst. Sci.* 20, 903–920. doi: 10.5194/hess-20-903-2016

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597, 672–677. doi: 10.1038/s41586-021-03854-z

Rojas-Campos, A., Langguth, M., Wittenbrink, M., and Pipa, G. (2023). Deep learning models for generation of precipitation maps based on numerical weather prediction. *Geosci. Model Dev.* 16, 1467–1480. doi: 10.5194/gmd-16-1467-2023

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *IEEE Access* 9, 16591–16603. doi: 10.1109/ACCESS.2021.3053408

Saadi, M., Furusho-Percot, C., Belleflamme, A., Chen, J.-Y., Trömel, S., and Kollet, S. (2023a). How uncertain are precipitation and peak flow estimates for the July 2021 flooding event? *Nat. Hazards Earth Syst. Sci.* 23, 159–177. doi: 10.5194/nhess-23-159-2023

Saadi, M., Furusho-Percot, C., Belleflamme, A., Trömel, S., Kollet, S., and Reinoso-Rondinel, R. (2023b). Comparison of three radar-based precipitation nowcasts for the extreme July 2021 flooding event in Germany. *J. Hydrometeorol.* 24, 1241–1261. doi: 10.1175/JHM-D-22-0121.1

Sattari, M. T., Kambiz, F., Ahmet, I., Shahab, S., and Sultan, N. Q. (2020). Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall, *Engineering Applications of Computational Fluid Mechanics*, 14, 1078–1094, doi: 10.1080/19942060.2020.1803971

Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., and Kunstmann, H. (2015). Fully coupled atmosphere-hydrology simulations for the Central Mediterranean: impact of enhanced hydrological parameterization for short and long time scales. *J. Adv. Model Earth Syst.* 7, 1693–1715. doi: 10.1002/2015MS000510

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., et al. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? *Water Resour. Res.* 55, 1179–1195. doi: 10.1029/2018WR023333

Tanhapour, M., Soltani, J., Malekmohammadi, B., Hlavcova, K., Kohnova, S., Petrakova, Z., et al. (2023). Forecasting the ensemble hydrograph of the reservoir inflow based on post-processed TIGGE precipitation forecasts in a coupled atmospheric-hydrological system. *Water* 15:887. doi: 10.3390/w15050887

Teimouri, N., Dyrmann, M., and Jørgensen, R. N. (2019). A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sens.* 11:990. doi: 10.3390/rs11080990

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., et al. (2015). How does bias correction of regional climate model precipitation affect modelled runoff? *Hydrol. Earth Syst. Sci.* 19, 711–728. doi: 10.5194/hess-19-711-2015

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., et al. (2009). Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res. Atmos.* 114:D24101. doi: 10.1029/2009JD011949

Tijerina-Kreuzer, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M. M., Sampson, K., et al. (2021). Continental hydrologic intercomparison project, phase 1: a large-scale hydrologic model comparison over the continental United States. *Water Resour. Res.* 57:e2020WR028931. doi: 10.1029/2020WR028931

Towler, E., Foks, S. S., Dugger, A. L., Dickinson, J. E., Essaid, H. I., Gochis, D., et al. (2023). Benchmarking high-resolution hydrologic model performance of long-term retrospective streamflow simulations in the contiguous United States. *Hydrol. Earth Syst. Sci.* 27, 1809–1825. doi: 10.5194/hess-27-1809-2023

Tri, D. Q., Thai, T. H., and Van Hoa, V. (2022). Bias-correction data of IFS rainfall forecasts for hydrological and hydraulic models to forecast flood events. *Arab. J. Geosci.* 15:1535. doi: 10.1007/s12517-022-10801-3

Tuo, Y., Duan, Z., Disse, M., and Chiogna, G. (2016). Evaluation of precipitation input for SWAT modeling in alpine catchment: a case study in the Adige river basin (Italy). *Sci. Total Environ.* 573, 66–82. doi: 10.1016/j.scitotenv.2016.08.034

Wang, F., Tian, D., and Carroll, M. (2023). Customized deep learning for precipitation bias correction and downscaling. *Geosci. Model Dev.* 16, 535–556. doi: 10.5194/gmd-16-535-2023

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J. (2021). Deep learning for daily precipitation and temperature downscaling. *Water Resour. Res.* 57:e2020WR029308. doi: 10.1029/2020WR029308

Wijayarathne, D., Coulibaly, P., Boodoo, S., and Sills, D. (2020). Evaluation of radar-gauge merging techniques to be used in operational flood forecasting in urban watersheds. *Water* 12:1494. doi: 10.3390/w12051494

Xu, C., Zhong, P., Zhu, F., Yang, L., Wang, S., and Wang, Y. (2023). Real-time error correction for flood forecasting based on machine learning ensemble method and its uncertainty assessment. *Stoch Environ. Res. Risk Assess.* 37, 1557–1577. doi: 10.1007/s00477-022-02336-6

Yaswanth, P., Kannan, B. A. M., Bindhu, V. M., Balaji, C., and Narasimhan, B. (2023). Evaluation of remote sensing rainfall products, bias correction and temporal disaggregation approaches, for improved accuracy in hydrologic simulations. *Water Resour. Manage.* 37, 3069–3092. doi: 10.1007/s11269-023-03486-0

Yorulmaz, E. B., Kartal, E., and Demirel, M. C. (2023). Benchmarking multi-component spatial metrics for hydrologic model calibration using MODIS AET and LAI products. *ESS Open Arch.* doi: 10.22541/essoar.169290537.78371628/v1

You, X., Liang, Z., Wang, Y., and Zhang, H. (2023). A study on loss function against data imbalance in deep learning correction of precipitation forecasts. *Atmos. Res.* 281:106500. doi: 10.1016/j.atmosres.2022.106500

Yousefi, K. P., Yilmaz, M. T., Öztürk, K., Yucel, I., and Yilmaz, K. K. (2023). Time-independent bias correction methods compared with gauge adjustment methods in improving radar-based precipitation estimates. *Hydrol. Sci. J.* 68, 1963–1983. doi: 10.1080/02626667.2023.2248108

Yucel, I., Onen, A., Yilmaz, K. K., and Gochis, D. J. (2015). Calibration and evaluation of a flood forecasting system: utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *J. Hydrol.* 523, 49–66. doi: 10.1016/j.jhydrol.2015.01.042

Zhang, T., Liang, Z., Li, W., Wang, J., Hu, Y., and Li, B. (2023). Statistical post-processing of precipitation forecasts using circulation classifications and spatiotemporal deep neural networks. *Hydrol. Earth Syst. Sci.* 27, 1945–1960. doi: 10.5194/hess-27-1945-2023

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 11 December 2024
DOI 10.3389/frwa.2024.1430374

Check for updates

OPEN ACCESS

EDITED BY
Matteo Camporese,
University of Padua, Italy

REVIEWED BY
Marwan Fahs,
National School for Water and Environmental Engineering, France
Benjamin Mary,
Instituto de Ciencias Agrarias, Spain

*CORRESPONDENCE
Q. Wei
✉ quan.wei@uwaterloo.ca

RECEIVED 16 May 2024
ACCEPTED 18 November 2024
PUBLISHED 11 December 2024

CITATION
Wei Q, Brookfield A and Layzell A (2024) Quantifying the effects of water management decisions on streambank stability.
Front. Water 6:1430374.
doi: 10.3389/frwa.2024.1430374

COPYRIGHT
© 2024 Wei, Brookfield and Layzell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Quantifying the effects of water management decisions on streambank stability

Q. Wei[1]*, A. Brookfield[1] and A. Layzell[2]

[1]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada, [2]Kansas Geological Survey, University of Kansas, Lawrence, KS, United States

Both natural processes and human activities alter streamflow conditions, which can significantly affect streambank erosion and stability, leading to consequences such as sedimentation of reservoirs, contamination of streams, loss of productive land, and damage to infrastructure. Hydrological conditions, which are often controlled by water management decisions and infrastructure (e.g., reservoirs and dams), are a major factor affecting streambank erosion and stability. Extensive research has explored the relationships between hydrology, water management, and streambank stability. However, limited studies directly address the impacts of water management decisions, particularly reservoir operations, on the driving mechanisms of streambank stability such as changes in pore water pressure, pressure differentials between the surface and subsurface, and gravitational forces versus shear stress. This study builds upon these existing concepts by integrating them into a model that accounts for both the effects of water management and inherent hydrologic conditions on streambank stability.

The module estimates streambank stability using a factor of safety approach, with hydrologic conditions derived from an established integrated hydrologic model, HydroGeoSphere, coupled with the surface water operations model, OASIS. This module is validated and then demonstrated using simulations from the Lower Republican River Basin in Kansas, United States. Results indicate that several water management decisions, such as groundwater pumping and timing of reservoir releases, may negatively affect streambank stability by changing pore water pressure, the weight of the bank material, and the pressure differential between the surface and the subsurface. Given that most of the rivers and streams of the world are regulated by reservoir operations, this work demonstrates that water management practices need to be considered in simulations of streambank stability.

KEYWORDS

streambank stability, integrated water resources management, porewater pressure, reservoir operations, water management decisions

1 Introduction

Land use changes, combined with shifting hydrologic conditions, have a significant impact on sediment transport in river systems. Factors such as urbanization, construction of water conservancy projects, groundwater withdrawals, and climate change all contribute to these changing dynamics. For example, the construction of dams and reservoirs was found to decrease downstream sediment loads significantly due to reservoir trapping efficiency (Yang et al., 2007), and groundwater pumping and the subsequent decline in groundwater levels can destabilize streambank materials through pore pressure changes (Casagli et al., 1999). Increased stream discharge due to precipitation events or reservoir releases can also increase streambank erosion through increased shear stress (Springston, 2007). Bank erosion has been

shown to be a major contributor to overall watershed sediment yields and is essentially controlled by two dominant processes: hydraulic forces acting on the channel boundary (e.g., shear stresses) and gravitational forces acting on the channel banks (e.g., hydrostatic stresses) (Simon et al., 2000). When they occur concurrently, streamflow can erode the bank toe, which increases the bank angle and, subsequently, the gravitational forces acting upon the bank, resulting in instability and bank failure when those forces exceed the shear strength of the bank material (e.g., Osman and Thorne, 1988). Therefore, it is critical to understand how water management affects hydrologic conditions, streambank stability, and, ultimately, sediment loading within managed watersheds.

In this study, the focus is on hydrostatic pressure and its role in streambank stability. Hydrostatic pressure results from the weight of water acting directly on the bank, leading to elevated pore water pressure within the soil. This pressure weakens the soil's internal cohesion, making the bank more prone to failure. Unlike velocity-induced shear stress, which results from the dynamic forces of flowing water, hydrostatic pressure is a static force that plays a critical role in destabilizing the streambank by altering pore water conditions (Rinaldi et al., 2004; Fox et al., 2007; Shields et al., 2009). Building on previous work, such as that of Brookfield and Layzell (2019) which focused on shear stresses, this paper emphasizes hydrostatic pressure as the primary factor contributing to streambank failure and explores its implications for future erosion modeling and stability assessments.

Some recent studies, such as those by Bigham et al. (2024) and Kadhim et al. (2024) have investigated sediment dynamics and streambank stability, focusing on factors like flow velocity, slope angle, soil type and flow regulations. However, they do not directly address the impacts of water management decisions, particularly reservoir operations, on the driving mechanisms of streambank stability. This study builds upon these existing concepts by further developing a model that accounts for both the effects of water management and inherent hydrologic conditions, including hydrostatic pressure, on streambank stability.

Streambank failures occur in various forms, with planar failures being particularly common in steep, cohesive banks, and are the focus of this work. These failures involve sections of the bank sliding along a nearly flat plane, forming a wedge shape slide or slab failure, due to a combination of increased pore water pressure and weakened internal cohesion (Osman and Thorne, 1988; Langendoen, 2000; Simon et al., 2000; Chu-Agor et al., 2008; Midgley et al., 2012). In addition to planar failures, other types of failures, such as rotational failure, cantilever failure and piping or sapping failure, can also significantly impact streambank stability (e.g., Langendoen, 2000; Rinaldi and Darby, 2007; Patsinghasanee et al., 2018). A thorough understanding of these failure mechanisms is essential for accurately assessing the stability of streambanks in different hydrological conditions.

Hydrologic models are commonly used to simulate hydrological processes and support water management decisions (Midgley et al., 2012). Many models are available for simulating integrated (surface and subsurface) hydrologic conditions (e.g., Brunner and Simmons, 2012; Maxwell et al., 2015; Taie Semiromi and Koch, 2019) and water management strategies (e.g., Hydrologics, 2009; Valerio et al., 2010; Qiu et al., 2019); however, they do not capture streambank erosion processes. Several models that simulate or quantify streambank stability exist (e.g., Midgley et al., 2012; Chu-Agor et al., 2008; Simon et al., 2000; Langendoen, 2000; Osman and Thorne, 1988); however, these models do not directly link to the influence of water management decisions on the hydrologic conditions driving erosion. Recent studies have explored related topics, such as modeling rainfall-induced landslides using the concept of local factor of safety (Abbasov et al., 2024) and investigating slope stability influenced by reservoir water level fluctuations and precipitation (Kafle et al., 2022), which highlight the effect reservoir operations can have on local slope stability. However, these studies do not provide a modeling framework to simulate both the temporal and spatial changes in groundwater and surface water conditions and their effects on streambank stability.

Some research has examined the links between hydrology, water management, and streambank stability, addressing the importance of factors such as pore water pressure, shear stress, and pressure differentials. The mechanisms driving changes in pore water pressure are often tied to variations in groundwater levels and reservoir releases, which can alter the balance of forces within streambank material, affecting stability (e.g., Casagli et al., 1999). Additionally, the pressure differential between the surface and subsurface influences the rate of bank erosion, as sudden changes in water levels can destabilize streambanks by creating imbalanced forces (e.g., Simon et al., 2000). For example, while surface water levels may fall quickly from a precipitation event, the streambanks may take longer to drain. This causes a difference between the subsurface pore pressure in the streambanks and the adjoining surface water. Without the hydrostatic pressure provided by surface water, the streambanks are more susceptible to failure.

This work builds on an existing modeling framework that captures the integrated surface/subsurface flow system, including water management (e.g., reservoir operations, surface diversions, groundwater pumping). This framework couples the control volume, finite element integrated hydrologic model, HydroGeoSphere (HGS) (Aquanty, 2023), with the optimized linear programming-based surface water operations model, OASIS (Hydrologics, 2009), which was also linked to a module that estimates fluvial erosion in previous research (Brookfield and Layzell, 2019). Here, a streambank stability module is developed to run independently or coupled to the HGS modeling framework to investigate how streambank stability may change due to changes in surface and subsurface hydrologic conditions caused by water management decisions.

All models used in this study require parameterization to represent real-world conditions accurately. Some key parameters such as effective cohesion, internal friction angle, suction angle, and bank geometry are measured through field methods including soil sampling, groundwater monitoring, and Borehole Shear Testing, which are discussed in detail in the following sections.

The primary objective of this study is to investigate how water management decisions, including reservoir operations and groundwater pumping, affect streambank stability. We hypothesize that changes in pore water pressure, surface-subsurface pressure differentials, and hydrologic conditions are critical drivers of streambank instability. Additionally, this work aims to develop a computationally efficient module that simulates these effects using the simplified factor of safety approach. The combination of both scientific inquiry and module development allows for a more comprehensive understanding of streambank erosion processes. Addressing issues such as water availability, sediment loading, and streambank stability is crucial for many regions globally (e.g., Yang et al., 2007), including the demonstration site of the Republican River Basin.

2 Module development and validation

A streambank stability module was developed in Python to work with output from the integrated hydrologic model, HydroGeoSphere (HGS), which was previously coupled with the surface water operations model, OASIS. The module takes the data output from the coupled HGS/OASIS model, such as surface and subsurface hydrologic conditions (e.g., pore water pressure, groundwater levels), and estimates streambank stability using the simplified factor of safety (Fs) approach, through one-way feedback from HGS/OASIS to the module. The HGS and OASIS models are iteratively coupled, providing two-way feedback at each OASIS timestep, ensuring the surface water operations modelled by OASIS are reflected in the hydrologic conditions simulated by HGS (Brookfield et al., 2017).This is consistent with the approach used to simulated fluvial erosion with HGS/OASIS presented by Brookfield and Layzell (2019). In this work, we focus on simulating planar failure, a common type of streambank failure where a section of the bank slides along a nearly flat plane. This occurs when changes in pore water pressure reduce the shear strength of the bank material, leading to instability (Langendoen, 2000).

2.1 Module development

To minimize data and computational requirements, streambank stability is estimated using the Fs approach with the Vertical Slides Method (Langendoen, 2000). In this approach, streambanks are separated into several vertical slices (Figure 1). Fs is estimated using a physically-based approach, based on (1) the forces acting on a streambank, with normal forces, shear forces, and gravity acting in the vertical direction, and (2) hydrostatic forces, with components perpendicular to gravity (Figure 2). An Fs value of less than 1 indicates an unstable streambank, values significantly greater than one are considered stable, and values at or just above 1 indicate streambanks that are at high risk of becoming unstable (Langendoen, 2000).

Forces acting on each slice are calculated separately and combined using the approach outlined in Langendoen (2000), with the Fs value determined by Equation 1 (Simon, 2006).

$$F_s = \frac{\cos\beta \sum_{j=1}^{N} \left(L_j c'_j + S_j \tan \varnothing^b_j + \left(N_j - U_j \right) \tan \phi'_j \right)}{\sin\beta \sum_{j=1}^{N} N_j - F_w} \quad (1)$$

Where c_j' is the effective cohesion (Pa), L_j is the length of the slice base (m), S_j is the shear force mobilized at the base of the slice (N), ϕ^b_j is an angle indicating the increase in shear strength for an increase in matric suction (degree), U_j is the porewater force on the base of the slice (N), ϕ'_j is the effective angle of internal friction (degree), β is the angle of the failure plane (degree), N_j is the normal force (N), and F_w is the hydrostatic force along the whole streambank (N). Full details of the approach, including the development of a user interface, are provided in the Supplementary material.

Due to the limitations of this simplified, computationally-frugal method and the governing equation, this module is suited for scenarios where changes in hydrostatic pressure drive instability, rather than velocity-induced shear stress. Future work aims to include velocity-induced shear stress in the estimation of streambank stability. Full details of the approach, including the development of a user interface, are provided in the Supplementary material.

2.2 Module verification

As no analytical solutions or exact replicates of our approach are available to verify the developed streambank stability module, we utilize results from a similar and commonly used model, the Bank

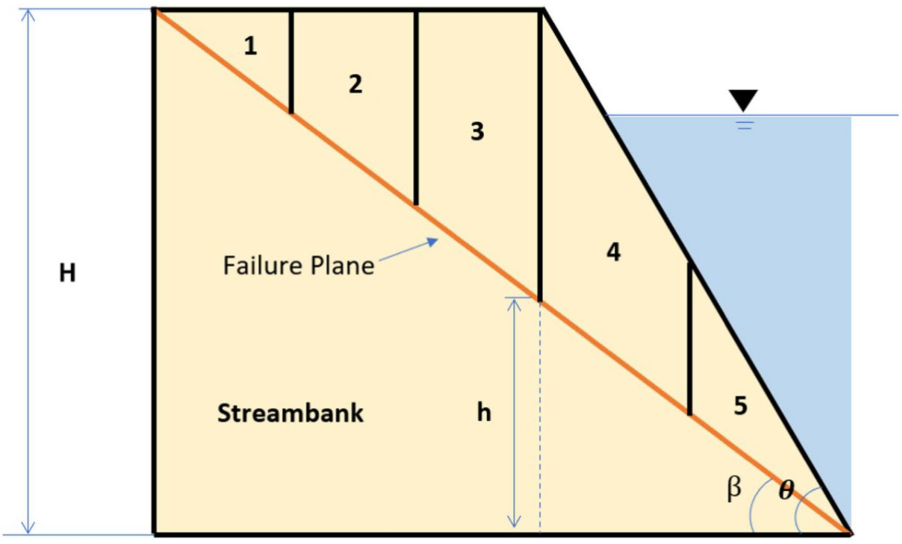


FIGURE 1
A simplified streambank cross-section split into five slices, where H is the height of the streambank, β is the angle of the failure plane, θ is the angle of the streambank, and h is the height of the failure plane at any point along the cross-section.

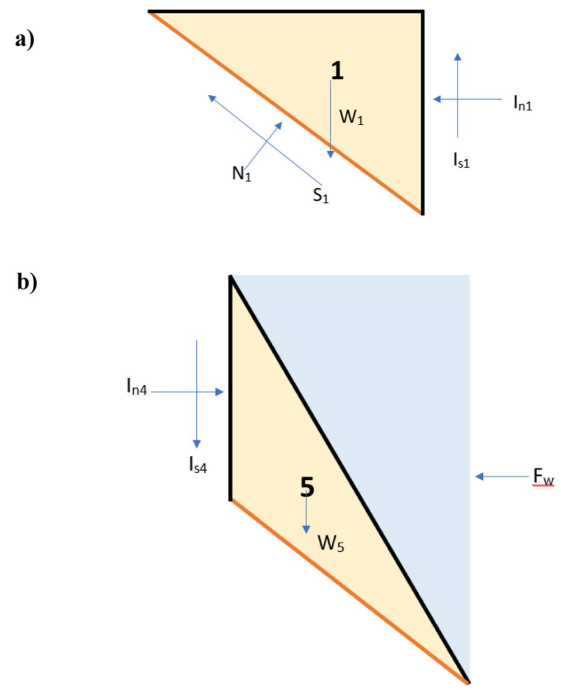


FIGURE 2
Forces acting on streambank cross-section **(A)** slice one and **(B)** slice 5 (*cf.* Figure 1), where N is the normal force, S is the shear force on the slide's base, W is the weight, I_s is the vertical interslice shear force, I_n is the horizontal interslice shear force and F_w is the hydrostatic force.

Stability and Toe Erosion Model (BSTEM) (US Army Corps of Engineers, 2015) for verification and a layer-refining approach to ensure convergence with increased spatial resolution. BSTEM is an Excel-based numerical tool used to assess streambank stability and predict toe erosion by analyzing soil mechanics, hydraulic effects, and vegetation impact, which is widely used by engineers, hydrologists, and environmental scientists. The default parameters used in the verification scenarios are provided in Table 1. The Fs values simulated by BSTEM and the module developed here are compared using different bank angles, bank heights, suction angles, friction angles, and effective cohesion values. The governing equations used by the two models are slightly different, as the streambank stability module uses the slices method (Vertical Slices Method), which is considered a "more classical geotechnical approach to planar failure" (US Army Corps of Engineers, 2015); however, BSTEM uses the layer method. The calculated normal force will be slightly different when using those two methods, and the slices method results are expected to be higher than layer methods without considering tension (US Army Corps of Engineers, 2015). Consistent with this, the module developed in this work has higher Fs values compared to BSTEM under the conditions assessed here (Figure 3). This suggests that our model estimates higher stability under similar conditions. However, the trends in Fs values under changing conditions are similar, and both models show similar increases/decreases in the factor of safety as the tested parameters change (Figure 3 for bank angle; Supplementary Figures S1–S4 for other parameters).

The differences in Fs values between the streambank stability module and BSTEM could be attributed to several factors, including

TABLE 1 Module Verification Data from Sutarto et al. (2014).

Test	Initial	Unit
Gravitational acceleration constant	9.81	N/kg
Void ratio	0.54	–
Specific gravity	3.28	–
Suction angle	17.0	Degree
Saturation	1.0	–
Effective cohesion	6,000	Pa
Internal friction angle	34.88	Degree
The angle of failure surface	32.0	Degree
Water level	2.5	m
Streambank height	4.3	m

differences in the governing equations, simulation methods, approximate calculation methods, assumptions made, and streambank characterization. Each model is underpinned by its unique set of fundamental equations; as mentioned before, the BSTEM uses the layer method, which is different from the stability model, and the choice of numerical simulation methods can notably influence a model's precision and responsiveness to parameter changes. The equations used in these test cases, though not identical, model equivalent physical processes. It is argued that comparing different approaches to the same physical processes can still provide valuable insights. Furthermore, each model employs its own strategies to simplify these complex systems, which can result in discrepancies between their outputs. However, given the consistency between trends in BSTEM and the module developed here, and the limited availability of other verification examples, the new module is considered representative of changing streambank stability conditions.

In the case of the streambank stability module developed in this work, it operates under the premise that the streambank is homogeneous and isotropic for ease of calculation. This simplification process encompasses factors such as soil cohesion, internal friction angles, the weight of the soil, moisture levels, and vegetative impact. However, in reality, streambanks are relatively heterogeneous in all these factors. Such variability can lead to noteworthy differences in model accuracy.

The newly developed module was also tested for spatial convergence. Using consistent input parameters, the number of slices varied from 2 to 100. Results indicate for the test case that the module converges to about 3.69 with 20 slices but approaches a reasonable estimate of 3.70 with about five slices (Figure 4). As a greater number of slices increases the computational burden of the module, it is ideal to balance accuracy with computational demand and therefore, we recommend using at least five slices in future simulations.

3 Module demonstration

The new streambank stability module is applied to a drainage basin with existing coupled HGS/OASIS simulation results (Brookfield and Gnau, 2016). The goal is to demonstrate the module's ability to discern temporal and spatial differences in streambank stability due to changing hydrologic conditions, including those induced by water management practices.

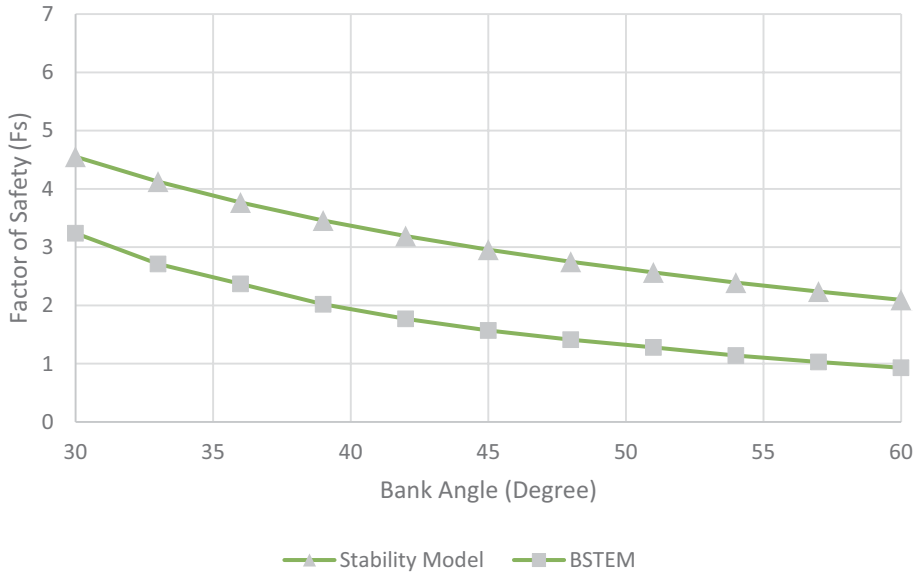


FIGURE 3
Factor of safety results from BSTEM and the streambank stability module across different bank angles.

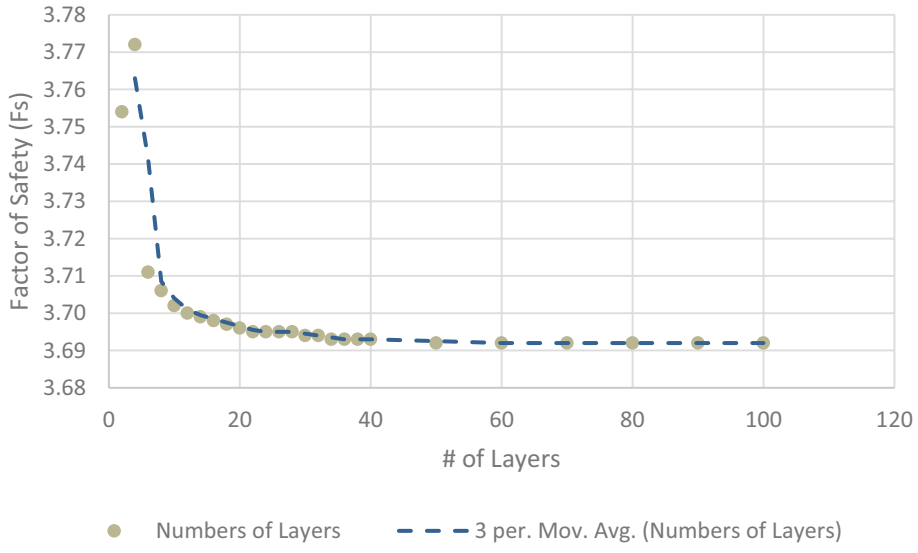


FIGURE 4
The factor of safety results from the streambank stability module vary with the number of vertical layers.

3.1 Site description

The Lower Republican River Basin (LRRB) is located in southern Nebraska and northern Kansas (Figure 5). The entire Republican River basin covers approximately 65,000 km^2, with the rivers originating in northeastern Colorado (U.S. Department of the Interior, 2016). Approximately 31% of the basin is in Colorado, 30% is in Kansas, and 39% is in Nebraska. The dominant use of water in the basin is for agriculture, although water is also used for domestic, industrial, recreational, and wildlife purposes (U.S. Department of the Interior, 2016). Although aquifers underlie most of the basin, the basin is overallocated, and water resources are limited. As such, water management planning in the basin is directly linked to the economic health of the region. Brookfield and Gnau (2016) coupled HGS and OASIS to simulate future water resources in the LRRB (Brookfield and Gnau, 2016), and Brookfield and Layzell (2019) expanded this model to include fluvial erosion to consider some components of sediment transport in the basin.

3.2 Module parameterization

For the streambank stability module, several parameters were estimated from previous literature (e.g., void ratio and specific

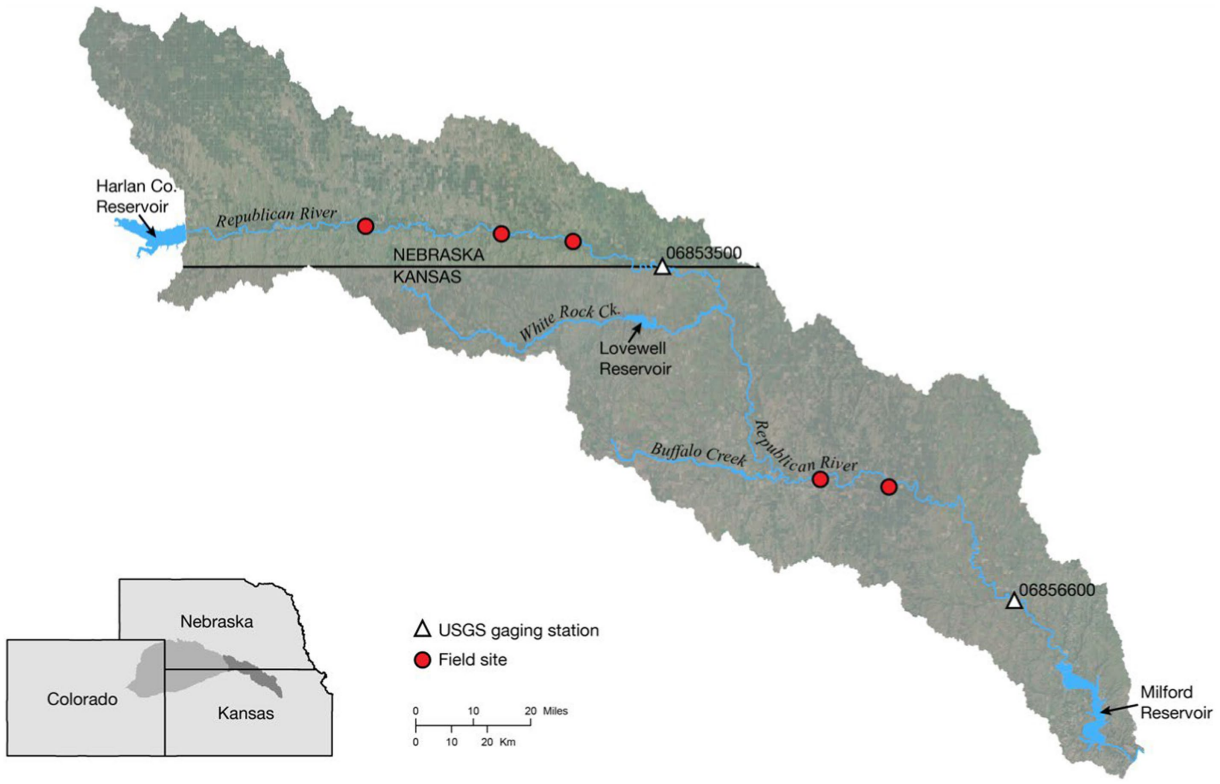


FIGURE 5
Map of Lower Republican River Basin study area.

gravity), and others were measured in the field (e.g., bank height). In order to provide field constraints for the streambank stability module, cohesion (c') and internal friction angle (Φ') were measured at five locations in the LRRB (Figure 5) using a Borehole Shear Test (BST) device (Lutenegger and Hallberg, 1981). For each test, an expandable shear head was lowered into a 3-inch diameter borehole augured to a given depth in the streambank. An initial normal stress was applied to the material by expanding the shear head for 15 min. A vertical force was then applied to the shear head by a hand crank, and the peak shear stress was recorded. The test was repeated with progressively higher normal stresses applied to the material in order to construct a failure envelope and determine the variables c' and Φ.' At each field site, BST tests were performed at different depths, depending on the alluvial stratigraphy and sedimentology. Three tests were repeated to test for variability. Values representative of average site conditions and measurements are used for this module demonstration (Table 2). Results from the coupled HGS/OASIS model for the LRRB are used to characterize hydrologic conditions for this demonstration (Brookfield et al., 2017) specifically surface water depth, groundwater levels, and pore pressure. These hydrologic conditions varied spatially throughout the basin.

To demonstrate how hydrologic conditions resulting from precipitation events and water management decisions (e.g., reservoir releases) affect streambank stability, results from the coupled HGS/OASIS model were selected for two different periods, reflecting "wet" and "dry" conditions. The two periods selected were July 2005 (severe drought) and July 2010 (wet conditions) based on regional climate data from the National Integrated Drought Information System

TABLE 2 Module parameters for LRRB simulations.

Parameters	Values	Unit
Gravitational acceleration constant	9.81	N/kg
Void ratio	0.40	–
Specific gravity	2.75	–
Suction angle	17.0	Degree
Effective cohesion	7,200	Pa
Internal friction angle	32.6	Degree
Angle of failure surface	56.0	Degree
Streambank height	2.70	m

(Drought.Gov, 2022). It should be noted that due to pervasive dry conditions in the basin over the past several decades, the "wet" scenario is characterized by a return to more normal hydrologic conditions (i.e., lack of severe drought) rather than abnormally wet conditions (i.e., extreme flooding). Streamflow was significantly lower in 2005 compared to 2010 at both the USGS gaging station near Hardy, NE (Station 06853500), just downstream of Harlan County Reservoir, and the gaging station at Clay Center, KS (Station 06856600), which is downstream of the inflow from Lovewell Reservoir via White Rock Creek (Figures 5, 6). Average groundwater levels across the basin also were much lower in 2005 compared to 2010, at 415.2 masl and 439.1 masl, respectively. Stream discharge at Hardy averaged 3.98 m^3/s in July 2005 compared to 72.72 m^3/s in July 2010, and downstream at Clay Center averaged 0.94 m^3/s in July 2005

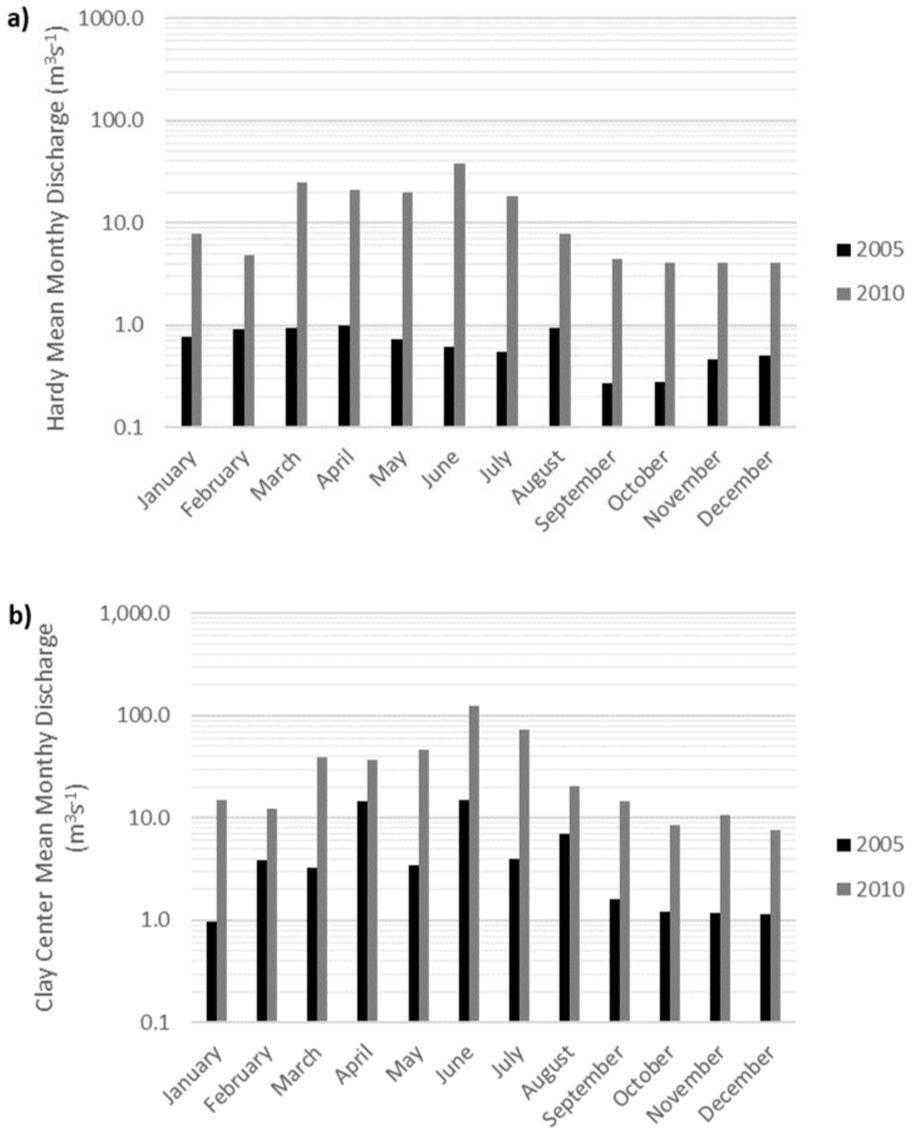


FIGURE 6
Log-scale of average monthly streamflow for the Republican River in 2005 and 2010 **(A)** near Hardy, NE (USGS gage 06853500) and **(B)** at Clay Center, KS (USGS gage 06856600).

compared to 7.73 m^3/s in July 2010. Water management in the LRRB, both between the wet and dry scenarios and between the two reservoirs, was also notably different. For example, in 2005 (dry scenario), Harlan Reservoir released no water and releases from Lovewell Reservoir were limited (Figure 7), likely due to lack of water availability as well as the administration of the Republican River Compact that allocates the waters of the river among the states of Colorado, Nebraska, and Kansas. In 2010 (wet scenario), both reservoirs released significantly more water as water storage demands had likely been met (Figure 7). The pattern of reservoir releases contributed to observed differences in streamflow. For example, in 2005 consistently low flows at the Hardy gage were a product of both drought conditions as well as lack of releases from Harlan Reservoir (Figures 6A, 7A). In contrast, releases from Lovewell Reservoir in the summer of 2005 were able to maintain streamflow at the Clay Center gage (Figures 6B, 7B). These differences allow us to simulate three different hydrologic conditions as a result of reservoir management: (1) wetter conditions when water is released from both reservoirs (2010); (2) dry conditions when water is unavailable for release (2005, upstream of White Rock Creek); and (3) dry conditions where water is available for release (2005, downstream of White Rock Creek).

4 Results

The Fs results are displayed spatially, using ArcGIS, at points that align with nodal coordinates from the HGS/OASIS simulations. A total of 737 HGS nodes along streams are simulated for each scenario. The wet and dry scenario results are presented independently. To facilitate a comparison between results from the two scenarios, a 1:1 plot comparing Fs values under wet and dry conditions is included

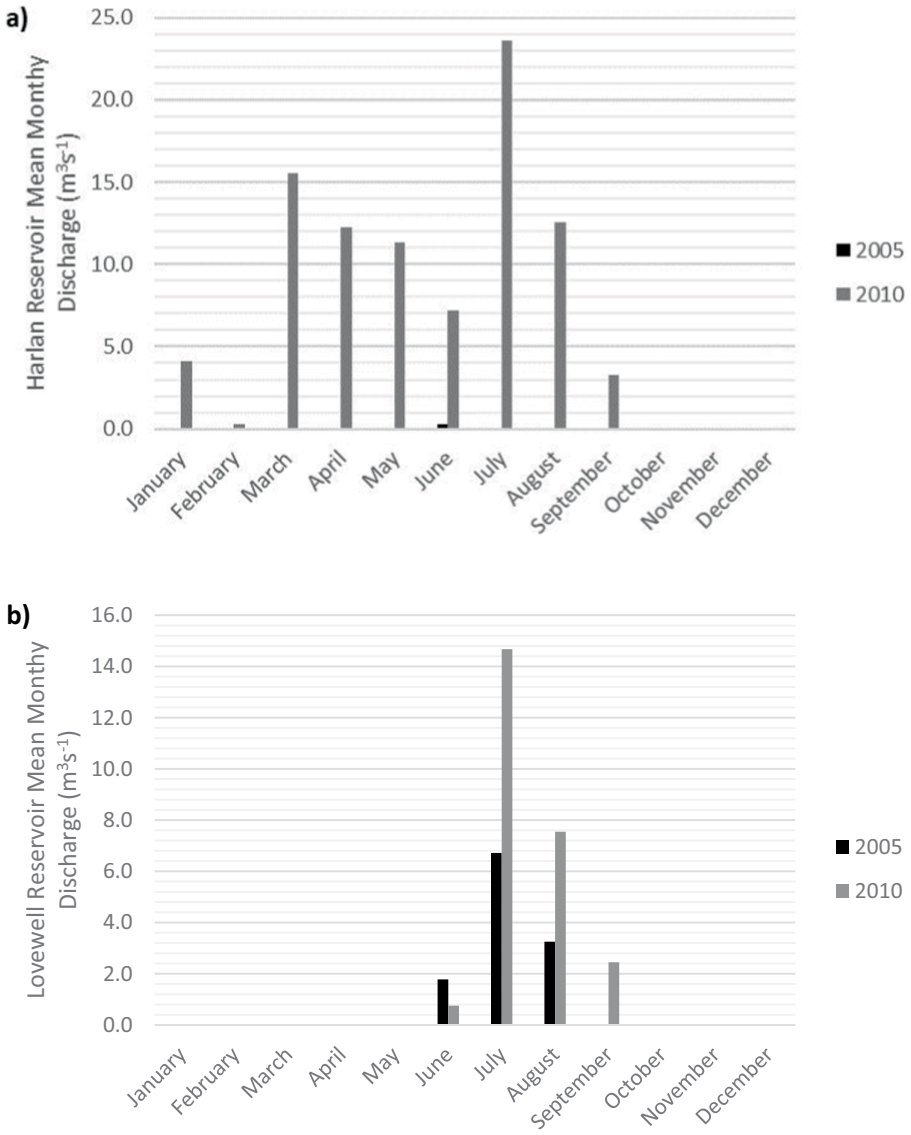


FIGURE 7
Average monthly discharge for 2005 and 2010 from **(A)** Harlan County Reservoir and **(B)** Lovewell Reservoir.

(Figure 8) for each node along the main stem of the Republican River with an unstable streambank in at least one of the scenarios. As shown in the plot, most values fall below the 1:1 line, indicating that the factor of safety is consistently, and sometimes substantially, lower under wet conditions than under dry conditions for the same locations.

When comparing the regions upstream of White Rock Creek (purple) and downstream of White Rock Creek (red) segments, it becomes apparent that upstream nodes show a broader range of factor of safety values between the model scenarios, with several points displaying higher stability under dry conditions (points located in bottom-right portion of Figure 8). This may be due the differences in reservoir releases in the regions upstream and downstream of White Rock Creek. The upstream region is affected by releases from Harlan Reservoir, which had no releases in the dry scenario (2005) but did have releases in the wet scenario (2010, Figure 7). In contrast, downstream segments are influenced by releases from both Harlan and Lovewell Reservoirs, and Lovewell released water more consistently in both scenarios (Figure 7). This would contribute to the factor of safety values clustering more closely together and shows generally lower stability in wet conditions.

4.1 Wet scenario

Simulation results for the wet scenario indicate a total of 92 unstable nodal locations, with 83 unstable locations along the main rivers of the basin (Figure 9). For the 737 HGS nodes along the main stem of the Republican River and major tributaries (White Rock Creek and Buffalo Creek), the average Fs was 2.03, ranging from 0.01 to 2.34. The locations of potential failure are distributed relatively evenly throughout the LRRB, although the number of unstable locations declines slightly in the upper reaches of the basin (Figure 9).

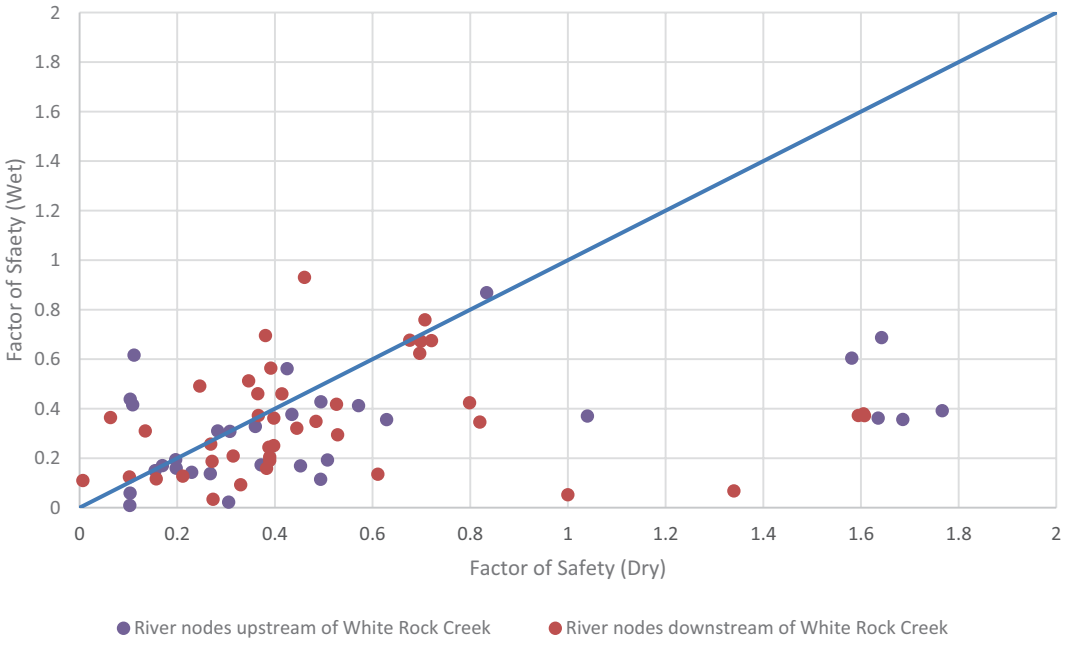


FIGURE 8
Relationship between the factor of safety values under wet (high precipitation and elevated groundwater levels) and dry (low precipitation and reduced groundwater levels) conditions.

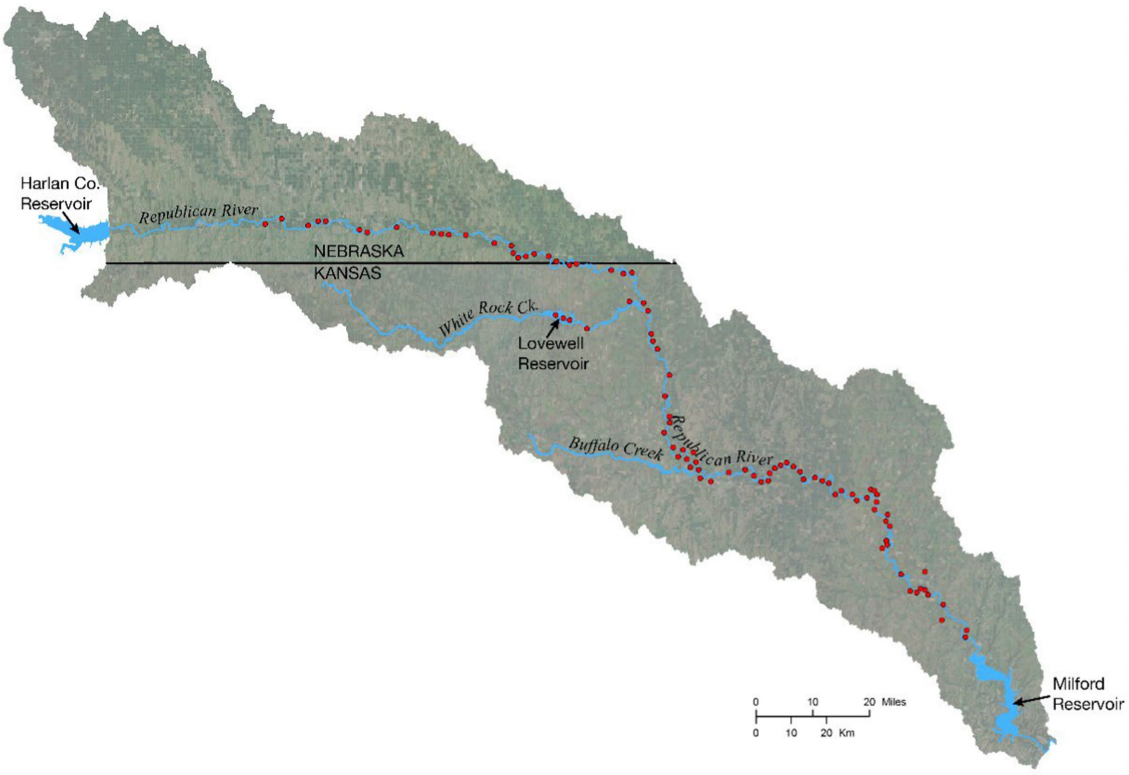


FIGURE 9
Map of streambank instability for the wet scenario. Red circles indicate nodes where streambank failure is predicted to occur (Fs ≤ 1).

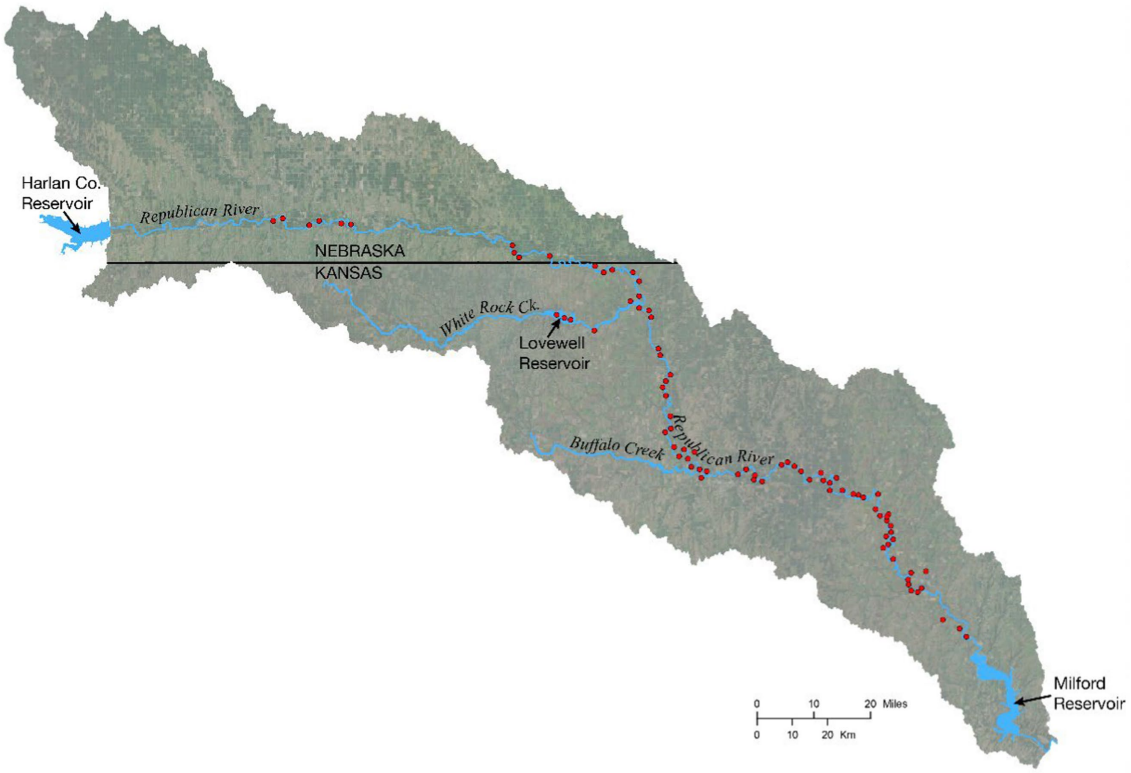


FIGURE 10
Map of streambank instability for the dry scenario. Red circles indicate nodes where streambank failure is predicted to occur (Fs ≤ 1).

4.2 Dry scenario

Results for the dry scenario indicate a total of 84 unstable nodal locations, with 70 along the main stem of the Republican River and major tributaries (Figure 10). For the 737 HGS nodes along these sections, the average Fs was 2.07, with a range from 0.01 to 2.34. In this scenario, the spatial distribution of unstable nodes is similar to the wet scenario for the lower half of the LRRB (i.e., downstream of White Rock Creek). However, there are notably fewer unstable locations located in the upper half of the basin, particularly closer to Harlan County Reservoir, which released no water during this period.

5 Discussion

As shown in Figures 8–10, the stability of streambanks in the LRRB is not the same under wet and dry conditions. There are parts of the basin where streambank stability is similar between scenarios, particularly in the lower half of the basin, which received water from reservoir releases under both wet and dry conditions (Figures 6, 7). However, streambanks in the upstream portion of the basin are more stable under dry conditions with no reservoir releases. This observation is likely due to the equilibrium, or lack thereof, between the surface and groundwater levels. Consistently low streamflow and groundwater levels result in low streambank saturation levels, thereby reducing the effect of gravitational forces (W_1 and W_5 in Figure 2) and inhibiting bank failure. If reservoir releases occur during dry periods, then we would expect downstream surface water levels to be higher relative to natural conditions, which could increase the hydrostatic force (F_w in Figure 2B) on the streambank and further support bank stability. However, prolonged increases in streamflow from reservoir releases can also increase the streambank pore pressure from bank infiltration, Saturation of the bank also increases the weight (i.e., gravitational forces), so that when releases cease and stream stage falls, the bank becomes more unstable (higher W and lower Fs). Moreover, the situation is further complicated under wet conditions by increased soil loading from precipitation events and pore pressures due to high saturation levels. This is likely the reason for increased instability in the wet scenario (precipitation + reservoir releases) and the downstream portion of LRRB during the dry scenario (reservoir releases).

Following this reasoning, we note that in July 2005, streamflow did not respond to precipitation events until the last event of the month (Figure 11A), as opposed to July 2010, where the stream responded to both precipitation events and reservoir releases (Figure 11B). Based on these patterns, we infer that under dry conditions, the soil moisture was so low in July 2005 that infiltrating water from precipitation was retained in the soil rather than transferred through the streambanks to the river by throughflow. However, in July 2010 (wet scenario), we infer that soil moisture levels were high enough to promote the throughflow of infiltrating water to the river, thereby increasing pore pressures in streambanks. While the modelled and observed stream stage at the Hardy gaging station was much lower than at the Clay Center gaging station for both the wet and dry scenario, likely due to differences in reservoir releases between the Harlan and Lovewell reservoirs, it is reasonable to propose that the stream levels were lower than they would have been under natural flow conditions (no reservoirs). This would cause a disequilibrium between the

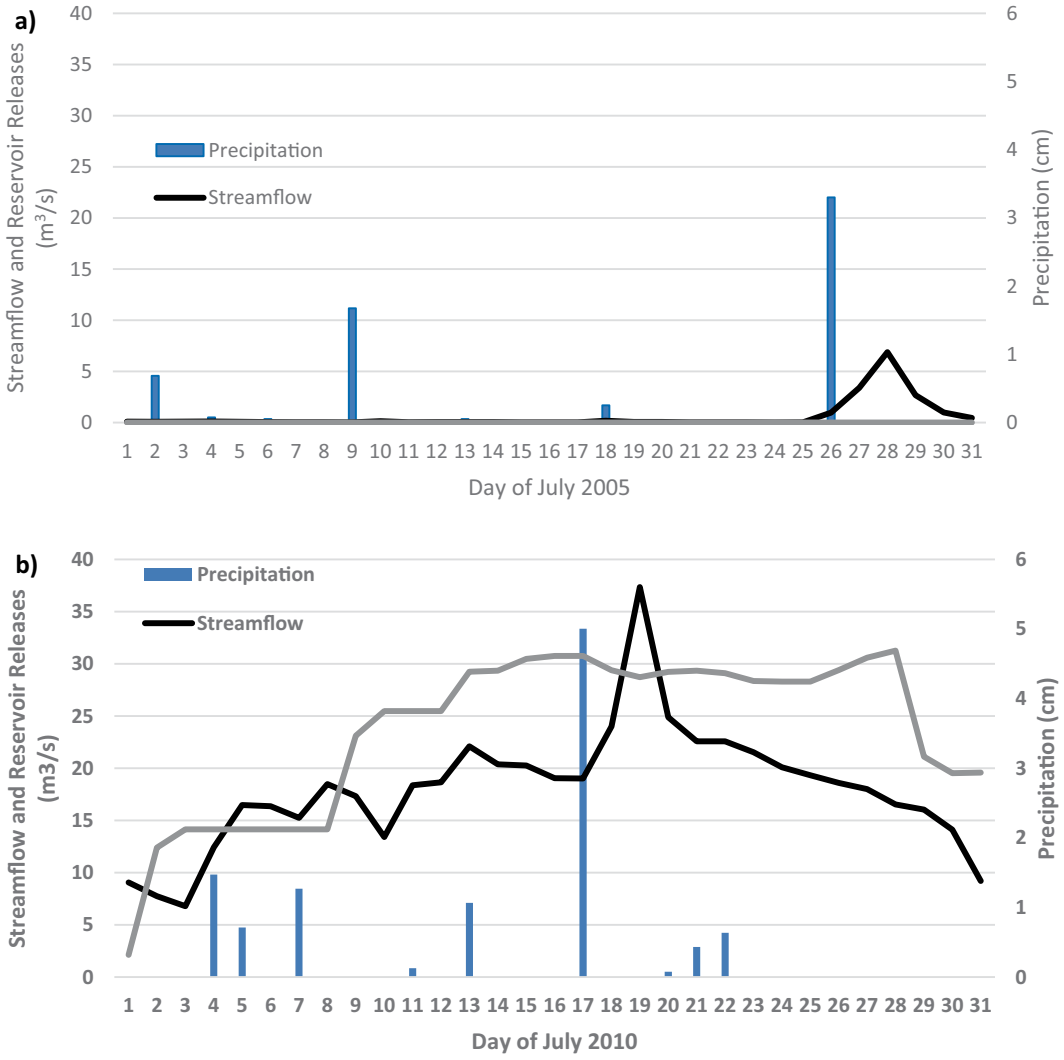


FIGURE 11
Response of streamflow at the USGS gage station near Hardy, NE, to Harlan Reservoir releases and precipitation in **(A)** July 2005 and **(B)** July 2010.

pore pressure of the streambanks (high) and the confining pressure from streamflow (low), contributing to streambank instability under wet conditions. These inferences are supported by simulated and observed groundwater levels, which were much higher in 2010 than in 2005. This higher ambient soil moisture increased the weight of the streambank material, thereby promoting streambank instability and bank failure. While not definitive from the available information, these inferences provide a conceptual idea of how water management could impact streambank stability in the LRRB, and are consistent with other studies (e.g., Kafle et al., 2022).

In this study, we recognize several factors that may contribute to errors and uncertainties in the modeling approach and application. First, the simplified methods used may not fully capture real-world complexities including all of the different streambank failure methods, as discussed in the model development, limiting the model's ability to accurately represent site conditions. Additionally, the HEC-RAS method (For further details, please refer to the supplementary material, Page 5) applied in our analysis may not always yield the minimum factor of safety value under certain conditions, such as when there are monotonic relationships between failure angle and factor of safety (US Army Corps of Engineers, 2015), which could affect the reliability of stability predictions. Another source of uncertainty is the assumption of homogeneous soil properties in the vertical profile. In reality, soil types and properties can vary significantly with depth, impacting both pore pressure distribution and overall stability. Future work could include sensitivity analyses and refinements in parameter selection to address these limitations, enhancing the model's accuracy and applicability.

6 Conclusions

Here, the impact of changing hydrologic conditions, driven in part by water management decisions, on streambank stability is studied by developing a new module that estimates F_S using results from coupled HGS/OASIS simulations. This module was verified using a previously

published, commonly used model (BSTEM) and was tested for spatial convergence. The module was then applied to wet and dry scenarios in the LRRB for demonstration. In applying the verified module to the LRRB, it is evident that water management decisions can have an impact on the stability of streambanks. Results indicated that streambanks were less stable under the wet conditions than dry conditions, likely due to a combination of increased pore water pressures and soil loading. These conditions resulted from increased soil saturation levels and reservoir controlled streamflows that were likely lower than what would have naturally occurred. This was most evident in the upstream portion of the basin studied in this work, where noticeable differences in streambank stability between the wet and dry scenarios were evident. The upstream reservoir (Harlan County Reservoir) released no water during the dry scenario, resulting in very low streamflows in the upper part of the basin. In the lower half of the basin, releases from Lovewell Reservoir augmented downstream streamflow, and, as a result, streambank stability was found to be similar to that of the wet scenario in downstream reaches.

This research intends to provide a tool for the preliminary assessment of streambank stability under hydrologic conditions driven by water management decisions and operations. In future work, we will further explore more spatially variable streambank conditions and simulate longer continuous periods to capture the dynamic interactions and temporal variations in streambank stability. The work presented in this manuscript introduces a modeling framework that is suited for this future work. Additionally, we will investigate how changes in water management practices induced by climate change can affect streambank stability. We believe this future work can provide an even more comprehensive understanding of the factors influencing streambank erosion and stability within the context of water management decisions.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: Federated Research Data Repository (FRDR); Wei, Q., Brookfield, A., Layzell, A. (2024). Quantifying the Effects of Water Management Decisions on Streambank Stability, https://doi.org/10.20383/103.01126.

Author contributions

QW: Validation, Writing – original draft, Writing – review & editing, Software, Visualization. AB: Supervision, Writing – original draft, Writing – review & editing, Formal analysis, Funding acquisition, Validation. AL: Data curation, Investigation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by a US Geological Survey Section 104(b) grant via the Kansas Water Resource Institute (PI Layzell) and the Natural Science and Engineering Research Council of Canada Discovery Grant Program (PI Brookfield).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2024.1430374/full#supplementary-material

References

Abbasov, R., Fahs, M., Younes, A., Nowamooz, H., Jørgen Måløy, K., and Toussaint, R. (2024). Modeling rainfall-induced landslide using the concept of local factor of safety: uncertainty propagation and sensitivity analysis. *Comput. Geotech.* 167:106102. doi: 10.1016/j.compgeo.2024.106102

Aquanty (2023) Aquanty. Available at: https://www.aquanty.com (Accessed February 22, 2023).

Bigham, K. A., Keane, T. D., and Moore, T. L. (2024). Effect of flow regulation on streambank erosion: a perspective downstream of a flood control dam, Kansas, USA. *River Res. Appl.* 40, 14–28. doi: 10.1002/rra.4212

Brookfield, A. E., and Gnau, C. (2016). Optimizing water Management for Irrigation under Climate Uncertainty: evaluating operational and structural alternatives in the lower Republican River basin, Kansas, USA. *Water Resour. Manag.* 30, 607–622. doi: 10.1007/s11269-015-1180-y

Brookfield, A., Gnau, C., and Wilson, B. (2017). Incorporating surface water operations in an integrated hydrologic model: model development and application to the lower Republican River basin, United States. *J. Hydrol. Eng.* 22:04016065. doi: 10.1061/(ASCE)HE.1943-5584.0001486

Brookfield, A. E., and Layzell, A. L. (2019). Simulating the effects of reservoir management strategies on fluvial Erosion. *Water Resour. Manag.* 33, 4983–4995. doi: 10.1007/s11269-019-02380-y

Brunner, P., and Simmons, C. T. (2012). HydroGeoSphere: a fully integrated, physically based hydrological model. *Groundwater* 50, 170–176. doi: 10.1111/j.1745-6584.2011.00882.x

Casagli, N., Rinaldi, M., Gargini, A., and Curini, A. (1999). Pore water pressure and streambank stability: results from a monitoring site on the Sieve River, Italy. *Earth Surf. Process. Landf.* 24, 1095–1114. doi: 10.1002/(SICI)1096-9837(199911)24:12<1095::AID-ESP37>3.0.CO;2-F

Chu-Agor, M. L., Wilson, G. V., and Fox, G. A. (2008). Numerical modeling of Bank instability by seepage Erosion undercutting of layered streambanks. *J. Hydrol. Eng.* 13, 1133–1145. doi: 10.1061/(ASCE)1084-0699(2008)13:12(1133)

Drought.Gov (2022) National Integrated Drought Information System. Available at: https://www.drought.gov/ (Accessed September 30, 2022).

Fox, G. A., Wilson, G. V., Simon, A., Langendoen, E. J., Akay, O., and Fuchs, J. W. (2007). Measuring streambank erosion due to ground water seepage: correlation to bank

pore water pressure, precipitation and stream stage. *Earth Surf. Process. Landf.* 32, 1558–1573. doi: 10.1002/esp.1490

Hydrologics (2009). OASIS with OCL, model version 3.10.8, GUI version 4.6.16. New York, NY, USA: Hydrologics.

Kadhim, J., Waheed, M. Q., Hussein, H. A., and al-Wakel, S. F. A. (2024). Experimental study on the effect of flow velocity and slope on stream Bank stability (part I). *Civil Eng. J.* 10, 2631–2644. doi: 10.28991/CEJ-2024-010-08-013

Kafle, L., Xu, W. J., Zeng, S. Y., and Nagel, T. (2022). A numerical investigation of slope stability influenced by the combined effects of reservoir water level fluctuations and precipitation: a case study of the Bianjiazhai landslide in China. *Eng. Geol.* 297:106508. doi: 10.1016/j.enggeo.2021.106508

Langendoen, E. J. (2000) CONCEPTS - CONservational Channel evolution and pollutant transport system: stream corridor version 1.1. U.S. Department of Agriculture (USDA). Available at: https://www.ars.usda.gov (Accessed November 2022).

Lutenegger, A. J., and Hallberg, G. R. (1981). Borehole shear test in geotechnical investigations. In: *Laboratory Shear Strength of Soil, STP740-EB:0.* Eds. R. N. Yong and F. C. Townsend, ASTM International.

Maxwell, R. M., Condon, L. E., and Kollet, S. J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geosci. Model Dev.* 8, 923–937. doi: 10.5194/gmd-8-923-2015

Midgley, T. L., Fox, G. A., and Heeren, D. M. (2012). Evaluation of the Bank stability and toe Erosion model (BSTEM) for predicting lateral streambank retreat on Ozark streams. *Geomorphology*, 145–146, 107–114. doi: 10.1016/j.geomorph.2011.12.044

Osman, A. M., and Thorne, C. R. (1988). Riverbank stability analysis. I: theory. *J. Hydraul. Eng.* 114, 134–150. doi: 10.1061/(ASCE)0733-9429(1988)114:2(134)

Patsinghasanee, S., Kimura, I., Shimizu, Y., and Nabi, M. (2018). Experiments and modelling of cantilever failures for cohesive riverbanks. *J. Hydraul. Res.* 56, 76–95. doi: 10.1080/00221686.2017.1300194

Qiu, J., Yang, Q., Zhang, X., Huang, M., Adam, J. C., and Malek, K. (2019). Implications of water management representations for watershed hydrologic modeling in the Yakima River basin. *Hydrol. Earth Syst. Sci.* 23, 35–49. doi: 10.5194/hess-23-35-2019

Rinaldi, M., Casagli, N., Dapporto, S., and Gargini, A. (2004). Monitoring and modelling of pore water pressure changes and riverbank stability during flow events. *Earth Surf. Process. Landf.* 29, 237–254. doi: 10.1002/esp.1042

Rinaldi, M., and Darby, S. E. (2007). Modelling river-bank-erosion processes and mass failure mechanisms: progress towards fully coupled simulations. *Dev. Earth Surf. Process.* 11, 213–239. doi: 10.1016/S0928-2025(07)11126-3

Shields, F. D., Simon, A., and Dabney, S. M. (2009). Streambank dewatering for increased stability. *Hydrol. Process.* 23, 1537–1547. doi: 10.1002/hyp.7286

Simon, A. (2006) 'A model of streambank stability incorporating hydraulic erosion and the effects of riparian vegetation', Proceedings of the Eighth Federal Interagency Sedimentation Conference (8thFISC).

Simon, A., Curini, A., Darby, S. E., and Langendoen, E. J. (2000). Bank and near-bank processes in an incised channel. *Geomorphology* 35, 193–217. doi: 10.1016/S0169-555X(00)00036-2

Springston, G. (2007). *Report on streambank stability assessment techniques Vermont geological survey technical Report*. VGTR2007-1. Vermont Geological Survey.

Sutarto, T., Papanicolaou, T., Wilson, C., and Langendoen, E. (2014). Stability analysis of semicohesive streambanks with CONCEPTS: Coupling field and laboratory investigations to quantify the onset of fluvial erosion and mass failure. *J. Hydraul. Eng.* 140:04014041. doi: 10.1061/(ASCE)HY.1943-7900.0000899

Taie Semiromi, M., and Koch, M. (2019). Analysis of spatio-temporal variability of surface–groundwater interactions in the Gharehsoo river basin, Iran, using a coupled SWAT-MODFLOW model. *Environ. Earth Sci.* 78:201. doi: 10.1007/s12665-019-8206-3

U.S. Department of the Interior. (2016). *Republican River Basin Study: Final Executive Summary Report*. Bureau of Reclamation Technical Service Center, Denver, Colorado. Available at: https://www.usbr.gov

US Army Corps of Engineers (2015) HEC-RAS USDA-ARS Bank Stability & toe Erosion Model (BSTEM), Technical Reference and User's manual. Available at: https://www.hec.usace.army.mil/confluence/rasdocs/rassed1d/1d-sediment-transport-technical-reference-manual/bstem-technical-reference-manual/steps-in-a-bank-failure-analysis (Accessed May 30, 2023).

Valerio, A., Rajaram, H., and Zagona, E. (2010). Incorporating groundwater-surface water interaction into river management models. *Groundwater* 48, 661–673. doi: 10.1111/j.1745-6584.2010.00702.x

Yang, S. L., Zhang, J., and Xu, X. J. (2007). Influence of the three gorges dam on downstream delivery of sediment and its environmental implications, Yangtze River. *Geophys. Res. Lett.* 34. doi: 10.1029/2007GL029472

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 23 January 2025
DOI 10.3389/frwa.2024.1520913

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Rana Muhammad Adnan Ikram,
Hohai University, China
Majdi Mansour,
British Geological Survey - The Lyell Centre, United Kingdom

*CORRESPONDENCE
Amelia Peeples
✉ peeples@princeton.edu
Reed M. Maxwell
✉ reedmaxwell@princeton.edu

RECEIVED 31 October 2024
ACCEPTED 18 December 2024
PUBLISHED 23 January 2025

CITATION
Peeples A and Maxwell RM (2025) Subgrid channel formulation in an integrated surface-subsurface hydrologic model. *Front. Water* 6:1520913. doi: 10.3389/frwa.2024.1520913

COPYRIGHT
© 2025 Peeples and Maxwell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Subgrid channel formulation in an integrated surface-subsurface hydrologic model

Amelia Peeples[1]* and Reed M. Maxwell[1,2,3]*

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, United States, [2]High Meadows Environmental Institute, Princeton University, Princeton, NJ, United States, [3]Integrated GroundWater Modeling Center, Princeton University, Princeton, NJ, United States

In hydrologic modeling, the assumption of homogeneity within a cell averages all variability finer than the model resolution. This loss of information can impact a model's ability to accurately represent hydrologic processes, especially in highly heterogeneous domains. This study quantified the impact of this loss of information on surface water fluxes by comparing the outputs of a high-resolution and coarse hydrologic model applied to an idealized domain. This study also presented a framework for including subgrid information in the surface water physics of integrated hydrologic models. Channel width was used as a representative subgrid parameter to better characterize surface water flow in cells containing subgrid channels. A new, nonlinear relationship between flux and calculated flow depth was derived based on assumed bathymetry and known channel width. This flux relationship was incorporated into ParFlow, an integrated 3D subsurface flow and 2D surface flow hydrologic model. In all scenarios tested, the subgrid channel formulation applied to a coarse-resolution model produced peak flows that only differed from the high-resolution model by more than 1% in 11/400 of scenarios and never differed by more than 5%. This is a substantial improvement from the baseline formulation applied to a coarse-resolution model, where peak flow differed by more than 1% in 213/400 scenarios and had a maximum difference of 78%.

KEYWORDS

channel flow, integrated hydrologic model, subgrid formulation, subgrid parameterization, ParFlow

1 Introduction

Large-scale, integrated hydrology models can represent regional hydrology by capturing surface-subsurface interplay and interactions across watershed boundaries. It is becoming increasingly necessary to quantify the terrestrial water cycle across large spatial scales to understand water availability and flood risk under climate stress. To do this, many studies have applied hydrology models to large-scale real world domains (Bauer et al., 2006; Goderniaux et al., 2009; Shen and Phanikumar, 2010; Sutanudjaja et al., 2011; de Paiva et al., 2013; O'Neill et al., 2021; Yang et al., 2023; Delottier et al., 2024). Physics-based modeling at these scales can be very computationally expensive, so other products and models have been developed based on remote sensing, data aggregation, or machine learning of surface (Miller et al., 2018; Mohanasundaram et al., 2021; Durand et al., 2023) and subsurface hydrology (Cools et al., 2006; Ma et al., 2024). These studies are helpful in defining regional hydrology but typically have the disadvantage of producing relatively coarse outputs. Low-resolution is necessary given the computational cost of high-resolution large-scale

hydrologic modeling, but coarse models do not necessarily reflect the same hydrology that high-resolution models would (Foster et al., 2020).

Hydrologic models often assume homogeneity within a cell for both inputs and outputs. Each cell is treated as a representative elementary volume, which is necessary to solve the relevant partial differential equations. This can lead to coarse-resolution model outputs and the loss of natural spatial variability in inputs. Many studies have investigated the impact of subgrid information on the accuracy of hydrologic models and how this lost information can be reincorporated into the modeling process. Specifically, subgrid variability in both land surface processes (Shuttleworth, 1988; Leung et al., 1996; Ghan et al., 1997; Noilhan et al., 1997; Giorgi et al., 2003; Wang and Wang, 2007) and soil characteristics (Wood et al., 1992; Ghan et al., 1997; Kabat et al., 1997; Kreye and Meon, 2016) have been heavily studied. Generally, these studies find that the inclusion of subgrid information is influential on model output and increases model accuracy, yet it is still unclear to what extent subgrid information impacts the large-scale solution.

Subgrid topographic variability ranging in scale from microtopography to subgrid channels can also be impactful on hydrologic fluxes (Jan et al., 2018; Thompson et al., 2010). Subgrid topography has previously been incorporated into overland flow models utilizing a large array of methodologies. This has been done by defining an additional high-resolution grid of topography data within each coarse-grid cell to determine surface storage and flux (Volp et al., 2013; Zhang et al., 2023). The impact of microtopography on surface storage has been represented by developing and modeling storage-outflow curves (Hu et al., 2020). Known anthropogenic impacts on flow paths, such as roads and hedges (Carluer and Marsily, 2004; Gascuel-Odoux et al., 2011) as well as the anisotropic impact of microtopography such as crop rows on overland flow (Viero and Valipour, 2017) have also been incorporated into models. Subgrid topographic information has also been modeled via the use of effective parameters. An effective Manning's n value has been used previously to account for subgrid channels in overland flow calculations (Schalge et al., 2019). Similar to an effective parameter, Moretti and Orlandini (2018) proposed a different methodology to determine the elevation assigned to a cell based on the elevation of the highest Horton-order river within the cell, essentially creating an effective elevation. Additional studies have distilled new subgrid parameters out of subgrid topographic data. These parameters are new values assigned to each grid cell designed to represent lost information and can be used as inputs into the continuity and flux equations (Jan et al., 2018; Li and Hodges, 2019; Ferrari and Viero, 2020).

The issue of accounting for partially inundated cells in hydrologic models has been the subject of many studies (Bates and Hervouet, 1999; Defina, 2000; Casulli, 2009). Recent work has focused on explicitly modeling subgrid channels in 2D hydrologic models to better represent floodplain inundation. Neal et al. (2012) developed a methodology to identify subgrid channel cells and model their fluxes using the dynamic wave equation based on the subgrid parameters of channel width and channel bed elevation in the model LISFLOOD-FP. Viero et al. (2014) used the subgrid method developed by Defina (2000) to model subgrid channel flow in a coupled 2D surface and 2D saturated shallow subsurface model. Other studies have instead modeled channel flow directly as one dimensional stream elements with different geometries. These streams are in one- or two-way communication with other hydrologic processes operating on a regular grid. As an example, David et al. (2011) developed RAPID, an explicit channel network model embedded in a regular grid-based land surface model. Panday and Huyakorn (2004) also modeled channel and reach networks as a separate process from 2D overland flow. The goal of this study is to build on this previous work modeling subgrid channels to further improve the ability to represent surface water accurately across scales. A subgrid approach can increase local accuracy when representing the terrestrial water cycle at large scales without the computational cost of increasing resolution.

The goal of this work was to create a framework to develop and implement subgrid parameterizations in the integrated surface-subsurface hydrologic model ParFlow. The newly implemented subgrid parametrization was evaluated by quantifying the discrepancy in representing channels between coarse-resolution and high-resolution models across an array of input parameters. We then determined when a parameterization of subgrid channels would be most impactful on model accuracy. Channel width was used as a representative topographic subgrid parameter, which is used as an input in the flux equation of the kinematic wave approximation. This formulation is mathematically similar to the effective parameter implementation done by Schalge et al. (2019) but was implemented directly within the surface flux equations in an integrated hydrologic model. This new methodology of directly including subgrid information in the flux formulation allows for nonlinear depth-dependent corrections to the overland flow formulation, which was not possible with effective parameters alone. For example, this subgrid channel formulation was able to account for side-wall friction of subgrid channels while the effective parameter methodology proposed by Schalge et al. (2019) is not. While this is a simple example, this framework paves the way for complex subgrid formulations necessary for regional accuracy in large-scale integrated hydrologic models.

2 Methods

We developed and applied a subgrid parameterization of the kinematic wave equation in the overland flow module of ParFlow, a 3D subsurface and 2D surface flow integrated model which can efficiently be applied at large scales (Ashby and Falgout, 1996; Jones and Woodward, 2001; Kollet and Maxwell, 2006; Maxwell, 2013). This subgrid channel formulation and its derivative are embedded directly into ParFlow's nonlinear solver. We applied this subgrid formulation and the baseline ParFlow overland flow formulation to a coarse idealized model domain. We then compared their outputs to the baseline ParFlow overland flow formulation applied to a high-resolution model domain to determine if using a subgrid formulation resulted in flows that better matched a high-resolution model. Additionally, we compared these model formulation and domain combinations across an array of parameter and precipitation scenarios. This is to determine under what conditions the subgrid formulation increased the accuracy of the coarse-resolution model. Since this study is conducted on an idealized domain, accuracy is defined as how closely a coarse model's outflow matches the high-resolution

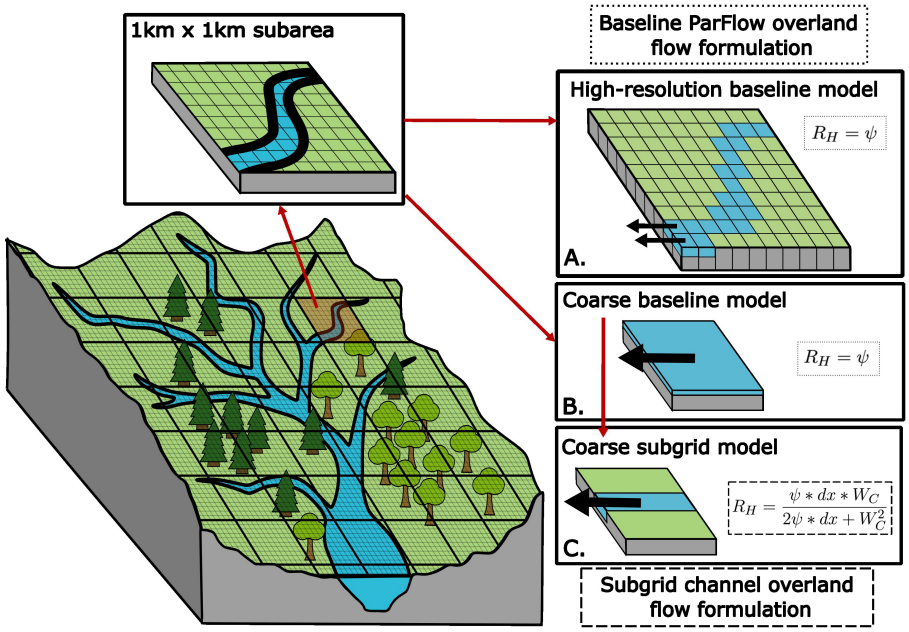


FIGURE 1
Conceptual overview of modeling domain and the characterization of subgrid channels in the **(A)** high-resolution baseline model, **(B)** coarse baseline model, and **(C)** coarse subgrid model.

TABLE 1 Description of all models, defined as a combination of an overland flow formulation and a model domain.

Model name	Description
High-resolution baseline model	• Baseline ParFlow kinematic wave formulation. • Applied to variable-resolution domain where dx is set equal to channel width and dy is set to 1,000 m. • Subfigure A in Figure 1.
Coarse baseline model	• Baseline ParFlow kinematic wave formulation. • Applied to coarse-resolution domain (dx = dy = 1,000 m). • Subfigure B in Figure 1.
Coarse subgrid model	• ParFlow kinematic wave formulation altered to include channel width in the hydraulic radius equation. • Applied to coarse-resolution domain (dx = dy = 1,000 m). • Subfigure C in Figure 1.

model's outflow. This allows the accuracy of the baseline coarse model to be assessed as well as the improvements of the subgrid model to be quantified. All models, here defined as a combination of a model domain and an overland flow formulation, used in this study are described in Figure 1 and Table 1.

2.1 Overland flow formulations

2.1.1 Baseline ParFlow formulation

ParFlow can calculate overland flux using either the diffusive wave or kinematic wave approximations. The baseline ParFlow overland flow formulation used in this study is based on the cell-centered Kinematic Wave Approximation using Manning's equation with the additional assumption that hydraulic radius is equal to pressure head (Kollet and Maxwell, 2006). This is executed in ParFlow as follows:

$$Q = \frac{\sqrt{S}}{n} * \psi^{5/3} * dx \tag{1}$$

where Q is discharge (L^3/t), S is bottom slope (L/L), n is Manning's coefficient ($t/L^{1/3}$), ψ is surface pressure head (L), and dx is cell size or resolution (L). Setting hydraulic radius equal to pressure head is a common assumption in hydrologic models which holds true when flow depth is much smaller than width, slopes are small and pressure is hydrostatic. Since homogeneity is assumed within each cell, all overland flow is represented as sheet flow across each cell. In this representation, flow depth is equivalent to pressure head within a cell and flow width is equivalent to cell size so the assumption that flow depth is much smaller than flow width holds true (Figure 2). However, in scenarios where flow within a cell is confined to a smaller area, such as a subgrid channel, representing flow as sheet flow can lead to an underestimation of the hydraulic radius and therefore an underestimation of flux.

2.1.2 Subgrid channel formulation

The subgrid formulation methodology incorporates channel width as an additional parameter within the kinematic wave equation. This new, nonlinear relationship between flux and calculated flow depth replaces the current flux equation within

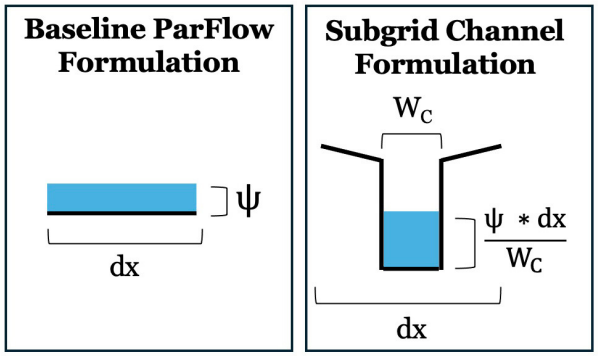


FIGURE 2
Conceptual representation of baseline **(left)** and subgrid channel **(right)** overland flow formulations.

ParFlow. Channel width in both horizontal directions are used as new inputs into ParFlow and are the basis of this new flux calculation. Flux direction is still determined based on cell slope, which is derived from topography.

Overland fluxes are determined using the same kinematic wave approximation as default ParFlow, but we no longer assume that the hydraulic radius is equal to pressure head. Instead, it is assumed that all flow is confined to a rectangular channel of known width (Figure 2). The hydraulic radius can then be calculated based on the depth and width of flow in the rectangular channel:

$$R_H = \frac{\psi * dx * W_C}{2\psi * dx + W_C^2} \tag{2}$$

where W_C is channel width (L), a new subgrid parameter. Depth of flow was calculated based on confining the known volume of surface water to the channel. Substituting this hydraulic radius into the kinematic wave equation results in an overland flux formulation of:

$$Q = \frac{\sqrt{S}}{n} * \psi^{5/3} * dx * \left(\frac{dx * W_C}{2\psi * dx + W_C^2}\right)^{2/3} \tag{3}$$

Given the implicit solution in ParFlow, the derivative of this new flux equation with respect to ψ needs to be included in the Jacobian as well. That derivative is:

$$\frac{dQ}{d\psi} = \frac{\sqrt{S}}{n} * dx^{5/3} * W_C^{2/3} * \left(\frac{5}{3}\left(\frac{\psi}{2\psi * dx + W_C^2}\right)^{2/3} - \frac{4dx}{3}\left(\frac{\psi}{2\psi * dx + W_C^2}\right)^{5/3}\right) \tag{4}$$

This new overland flow formulation and its derivative are added directly into ParFlow's nonlinear solver. As described in Kollet and Maxwell (2006), the surface and subsurface are directly integrated through a global variable of ψ, or pressure, at the surface-subsurface interface. This proposed overland flow formulation replaces the previous overland flow equation and is used in the overland flow boundary condition to simulate integrated surface-subsurface flows.

This formulation is executed in the positive and negative X and Y directions with channel width in both directions defined independently. This is so that channel width only impacts fluxes through a cell in the direction of a channel and not overland fluxes perpendicular to the channel. All overland fluxes from non-channel cells and from channel cells not in the direction of channel flow set W_C equal to cell size. Since ψ is much smaller than dx in these scenarios, this collapses back down to the baseline ParFlow flux formulation.

2.2 Idealized test case

An idealized domain was created to compare all models across a large array of parameters in a controlled environment. This domain is a one cell by five cell rectangle with a channel running down the center of the domain (Figure 3). All models have a cell-size parallel to the channel equal to 1 km and all models except the high-resolution baseline model also have a cell size perpendicular to the channel equal to 1 km. The cell size perpendicular to the channel of the high-resolution baseline model is equal to the width of the rectangular channel. This is so the high-resolution baseline model domain only includes the channel without any overland cells to remove the additional variability associated with the parameterization of those overland cells. Rainfall is scaled for the high-resolution baseline model to account of the lower surface area of each cell. This high-resolution domain will be used as a benchmark to assess performance of the other formulations.

All cells have equal unidirectional slopes downstream. Four hours of spatially invariable rainfall are applied at the beginning of each simulation and then the simulation continues with 0.1-h timesteps until outflow is approaching zero. To simplify the hydrology to allow better comparison of overland flow between models, the impacts of infiltration and groundwater interactions are neglected by setting permeability to nearly zero.

2.3 Parameter scenarios

To determine under what conditions the current ParFlow formulation differs from the subgrid channel formulation as well as when the coarse baseline model diverges from the high-resolution baseline model we ran all three ParFlow models across a large array of different input parameters. The Manning's n and bottom slope parameters are varied across ranges of realistic values (Yang et al., 2023). Four values of channel width ranging from 100 m to 1 km were tested to assess performance when channel width is similar to cell size as well as much smaller than cell size. Four values of rainfall intensity are also applied to assess performance across various water depths. Channel width is also used to determine cell size perpendicular to the channel in the high resolution model. All parameter values are listed in Table 2.

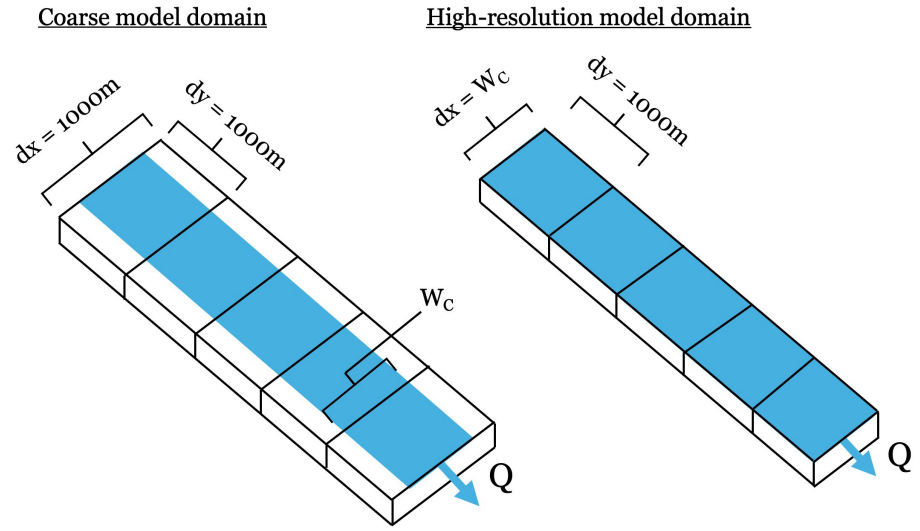


FIGURE 3
Idealized model domains.

TABLE 2 All input parameters varied and their corresponding values.

Channel width (m)	Rainfall intensity (cm/hr)	Manning's n (s/m$^{1/3}$)	Bottom slope (m/m)
100	0.5	3.6e-3	1e-4
200	1	1.8e-2	1e-3
500	5	3.6e-2	1e-2
1,000	10	1.8e-1	1e-1
-	-	3.6e-1	4e-1

3 Results

3.1 Coarse baseline model performance

Parameter scenarios in which the coarse baseline model differs from the high-resolution baseline model must first be identified to assess when subgrid formulations would be most applicable. Figure 4 shows how changes in channel width and rainfall intensity impact hydrograph characteristics at the outlet when comparing the coarse baseline model and the high-resolution baseline model. When channel width is equal to cell size, there is no difference between the low- and high-resolution baseline models, which is to be expected since the models are equivalent in this scenario. As channel width decreases, discrepancies in the peak flow volume begin to emerge between the coarse baseline and high-resolution baseline model.

The difference in peak flow between the coarse baseline model and the high-resolution baseline model across all parameter and rainfall scenarios is shown in Figure 5. As was seen in Figure 4, there is essentially no difference between the two models when channel width is equal to cell size. At the lowest channel width of 100 m, 1/10 the cell size, 35% of the scenarios result in a difference in peak flow of greater than 50% with 20% of scenarios resulting in a difference of greater than 75%. When channel width is 200 m, 1/5 the cell size, 35% of scenarios still result in a peak difference of greater than 50% and no scenarios have a difference of greater than 75%. When channel width is 500 m, 1/2 the cell size, no scenarios have a peak difference exceeding 50% with the greatest difference being 37.02%. Across all channel widths, peak flow differs by greater than 1% in 213 out of 400 scenarios. As channel width decreases, the number of scenarios which differ and the magnitude of those differences increase.

Within each rainfall-channel width scenario, bottom slope and Manning's n are also varied. These two overland flow parameters are influential on the discrepancy in peak flow between the two models when channel width is not equal to cell size. The difference in peak is greatest when Manning's n is high and bottom slope is low. This difference is further exacerbated in lower rainfall conditions, which are representative of when there is lower flow volume in the channel. Overall, the largest discrepancy in peak flow of 78.40% is seen in the scenario where channel width is 100 m, rainfall intensity is 0.5 cm/hr, Manning's n is 3.6e-1 s/m$^{1/3}$, and bottom slope is 1e-4 m/m.

Despite the significant differences in peak flow, peak timing is consistent between the coarse baseline and high-resolution baseline models (Figure 4). While there are some scenarios that result in a single time step difference in the peak timing, these scenarios don't follow any trend relating to the parameters varied in this study. The shape of the rising and falling limb also varied significantly between the coarse baseline and high-resolution baseline models (Figure 4). Qualitatively, hydrograph shape of the coarse baseline and the high-resolution baseline models was most different in the low rainfall and small channel width scenarios.

3.2 Coarse subgrid model performance

The coarse subgrid model performed nearly identically to the high-resolution baseline model across nearly all scenarios

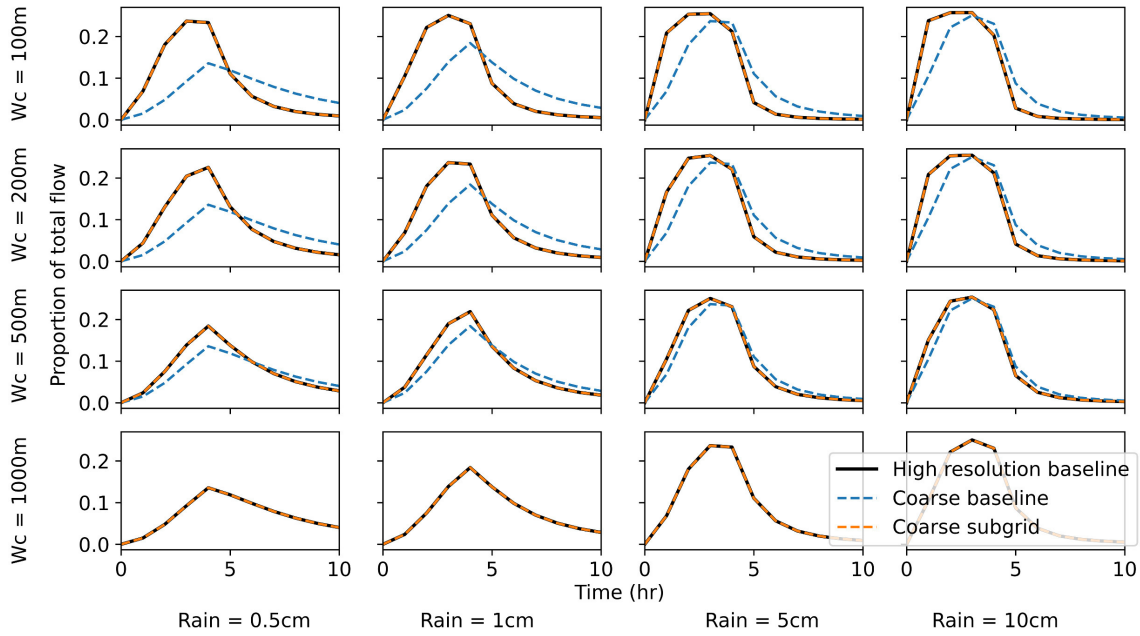


FIGURE 4
Difference in outflow between all three models based on rainfall and channel width.

(Figure 6). In only 11 of 400 total scenarios are there differences in peak flow greater than 1% and no scenarios exceed 5%. These differences are relatively small when compared to the differences seen between the coarse baseline and high-resolution baseline models of up to 78.40%. The differences in peak flow occur when channel width is small, rainfall is high, and flow is slow. Overall, the coarse subgrid model performed more similarly to the high-resolution model than the coarse baseline model across all parameter scenarios tested.

Small differences in peak flow are expected, as the high-resolution baseline model and the coarse subgrid model use different formulations for the kinematic wave equation. The subgrid formulation is based on the assumption of a rectangular channel while the coarse baseline formulation is based on the assumption that flow depth is negligible in comparison to flow width. These scenarios where differences arise highlight when the addition of side-wall friction is impactful on model accuracy, as it is included in the subgrid formulation but not the baseline formulation.

4 Discussion

4.1 Impact of subgrid formulation on coarse model accuracy

The coarse baseline model consistently underestimated peak flow across all rainfall scenarios when channel width was less than cell size. The magnitude of this underestimation was dependent on all parameters evaluated. The addition of the subgrid formulation to the coarse model resulted in the coarse model to perform much more similarly to the high resolution model. These results align with what was found by Schalge et al. (2019).

Rainfall volume was very influential on the accuracy of the coarse baseline model compared to the high resolution model. Lower rainfall scenarios resulted in a higher peak difference between the coarse baseline and high-resolution baseline models across all channel widths, slopes, and n values. Since lower rainfall corresponds with lower flow volume, this implies that the difference between coarse and high-resolution baseline models are most pronounced in low-flow conditions. In high rainfall conditions (when flow depth was high), channel width was much less impactful on peak flow than in low rainfall scenarios (Figure 7). Therefore, in these high flow scenarios the addition of a subgrid formulation only resulted in small improvements from the coarse baseline.

Combinations of high n values and low slopes, resulting in slow flow, also resulted in the greatest differences in peak flow across all rainfall and channel width scenarios. Collectively, this implies that smaller fluxes, such as those from tributaries, in areas with topography and vegetation which result in slow flow would be the most misrepresented by coarse models. These tributaries are likely also small, potentially with widths much smaller than the cell size, which further exacerbates the underestimation of these fluxes. Systematically underestimating tributary fluxes could have a large impact on the overall hydrology of a region. This could compound when there are multiple tributaries with underestimated flows feeding into the same river. The addition of channel width as a subgrid variable was able to mitigate this underestimation and when applied to the coarse model domain resulted in fluxes nearly identical to the high-resolution baseline model.

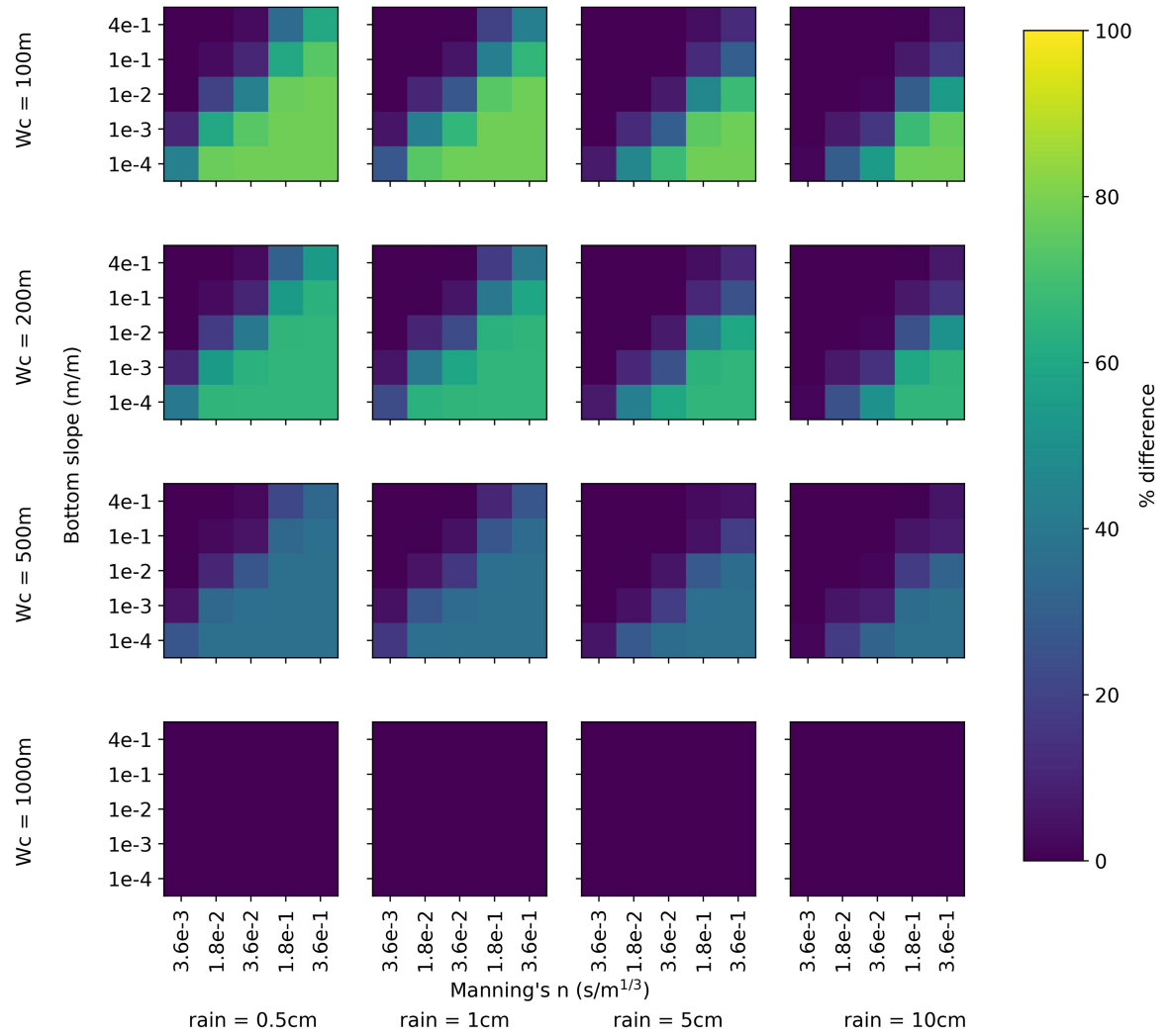


FIGURE 5
Percentage difference in peak flow between the coarse baseline and high-resolution baseline models. Here channel width is not an input in the coarse baseline model but instead is only used to define the domain resolution of the high-resolution model.

4.2 Impact of channel length on coarse model accuracy

All three models across all parameter and rainfall scenarios were also applied to a thirty-cell idealized domain to determine if a longer flow length would result in a larger discrepancy in peak flow or peak flow timing between the coarse baseline and high-resolution baseline model. The outflow of the coarse baseline, high-resolution, and coarse subgrid models were all compared at the outlet of the 30 km channel model.

In the 30 km channel length model, there is still very little difference between the high resolution model and the coarse subgrid model. The few scenarios where these two models differed after 5 km of channel still differed after 30 km, but in peak timing as opposed to peak flow. This is because both the high-resolution baseline model and the subgrid model reach the same constant outflow where inflow from upstream cells is equal to discharge, the high-resolution baseline model just reaches this outflow slightly earlier. This illustrates that over longer channels and time frames, the subgrid model output approaches that of the high-resolution baseline model.

The same trends in peak flow difference between the coarse baseline and high-resolution baseline models observed in the five-cell domain were also observed at all locations in the thirty-cell domain but were further exacerbated at longer distances from the origin. Figure 8 shows the percent difference in peak between the coarse baseline and coarse subgrid model at the end of the 30 km channel. When compared to Figure 5, there are more scenarios which have higher peak flow difference between the two models. The increase in peak discrepancy between the coarse baseline model and the high-resolution baseline model as channel length increases shows that the impact of underestimating flux between subgrid channel cells compounds over distance. Therefore longer channels will be most misrepresented when not accounting for subgrid channels. Despite more scenarios having a larger peak difference between the coarse baseline model and the high-resolution baseline model, the maximum percentage difference between peaks does not change and it can be clearly seen that

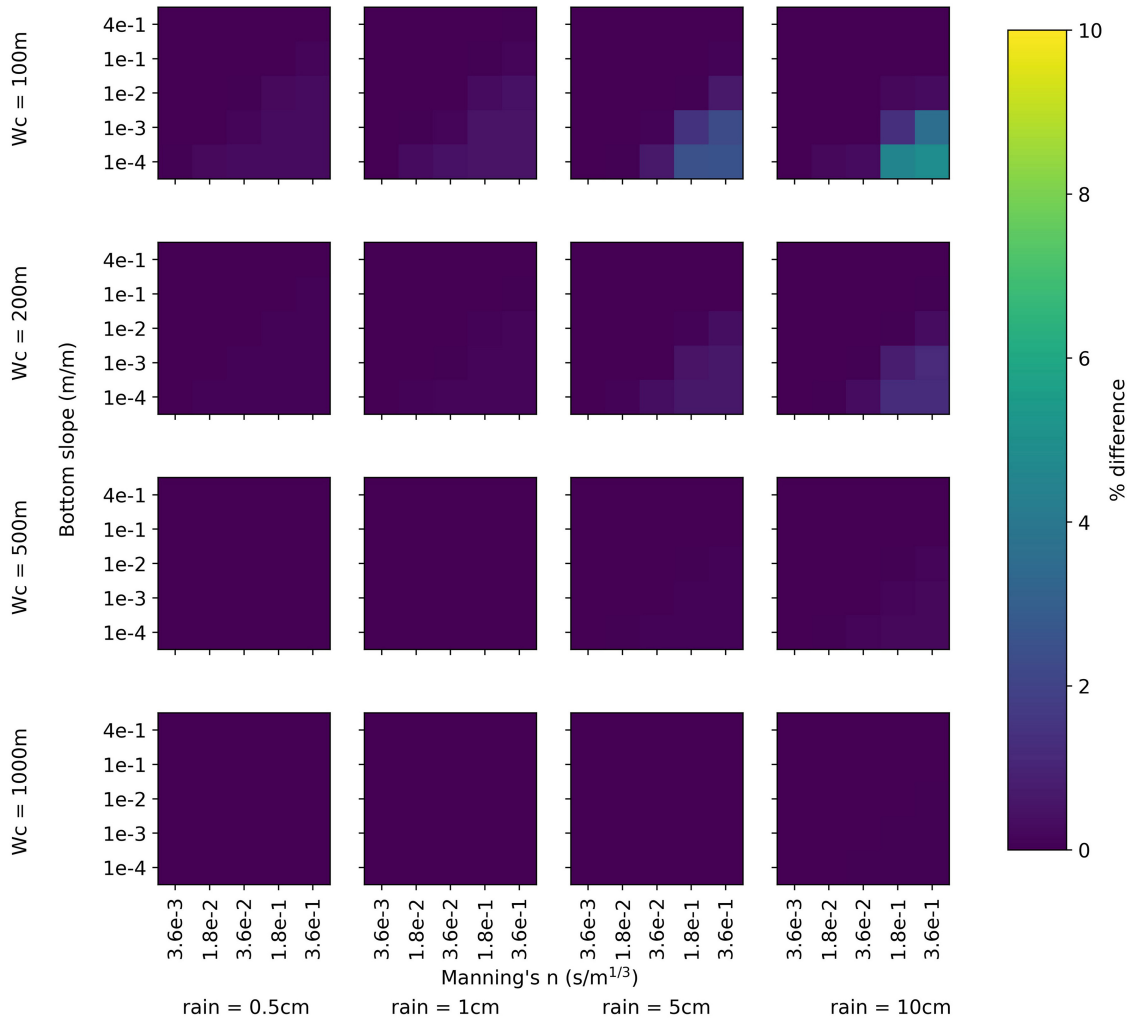


FIGURE 6
Percentage difference in peak flow between the coarse subgrid formulation and high-resolution baseline models. Here channel width is an input in the coarse subgrid model as well as used to define the domain resolution of the high-resolution model.

the maximum peak difference is dependent on channel width (Figure 8).

4.3 Using nondimensional parameters to define applicability subspace

The Froude Number and the Kinematic Wave Number (KWN) are non-dimensional values based on flow parameters and have been used to define ranges of applicability of the Kinematic Wave Approximation and the Diffusive Wave Approximation (Vieira, 1983). These dimensionless values were used in this study to investigate the relationship between domain parameters and accuracy of a coarse-scale model compared to a high-resolution baseline model. Under the assumption of uniform flow and substituting in the Chezy equation, the Froude number can be calculated by:

$$F_0 = C\left(\frac{tan(\theta)}{g}\right)^{1/2} \quad (5)$$

where C is Chezy roughness which is equivalent to $1/n^*(R^{1/6})$, θ is the bottom slope angle, and g is gravitational acceleration. The Kinematic Wave Number is defined as:

$$k = \left(\frac{g^3 L sin(\theta)}{C^4 q^2}\right)^{1/3} \quad (6)$$

where L is defined in Vieira (1983) as the length of the channel at which values are measured and q is lateral inflow into the channel. In this study, we define L as cell length parallel to the channel and q as flux through the cell. This definition allows the kinematic wave number to be calculated for any cell in a hydrologic model and will serve as a non-dimensional value which can be used in conjunction with the Froude number to classify conditions in any given cell to determine if a coarse model output will closely match a higher-resolution model. It is useful to classify these scenarios using non-dimensional parameters so this analysis can be applied to other model resolutions or subgrid parameters of interest.

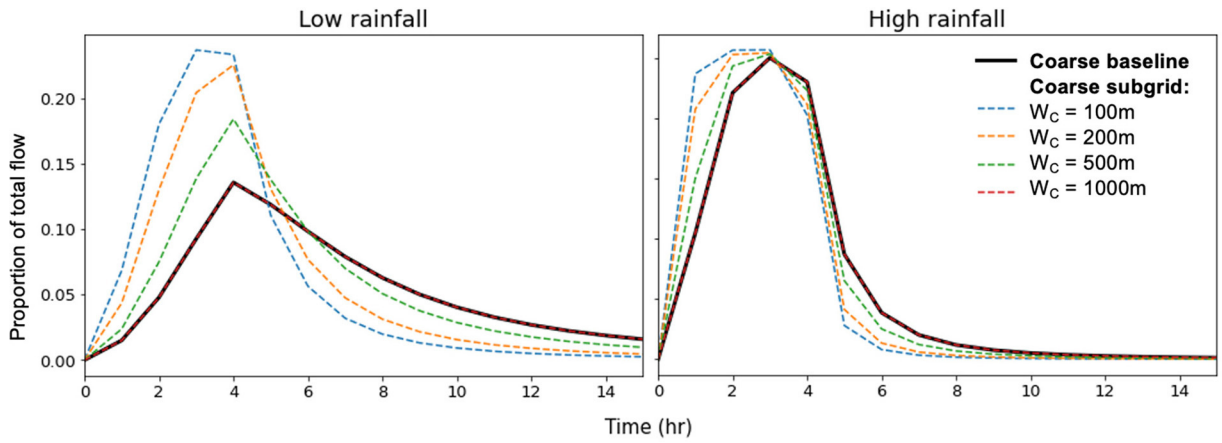


FIGURE 7
Impact of rainfall and channel width on difference between coarse baseline and subgrid formulation models.

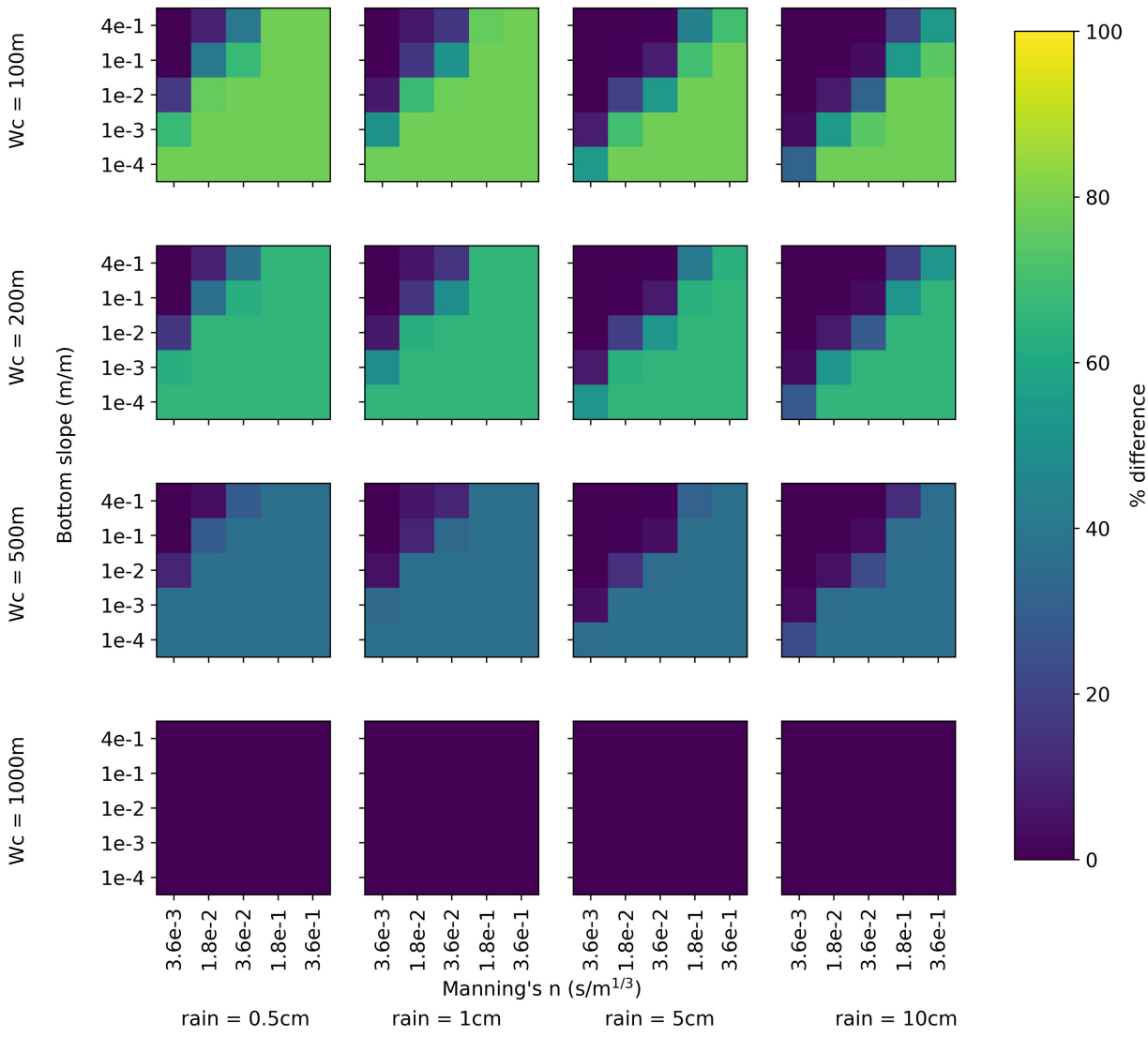


FIGURE 8
Percentage difference in peak flow between the coarse baseline formulation and high-resolution baseline models at the outlet of the 30 km channel.

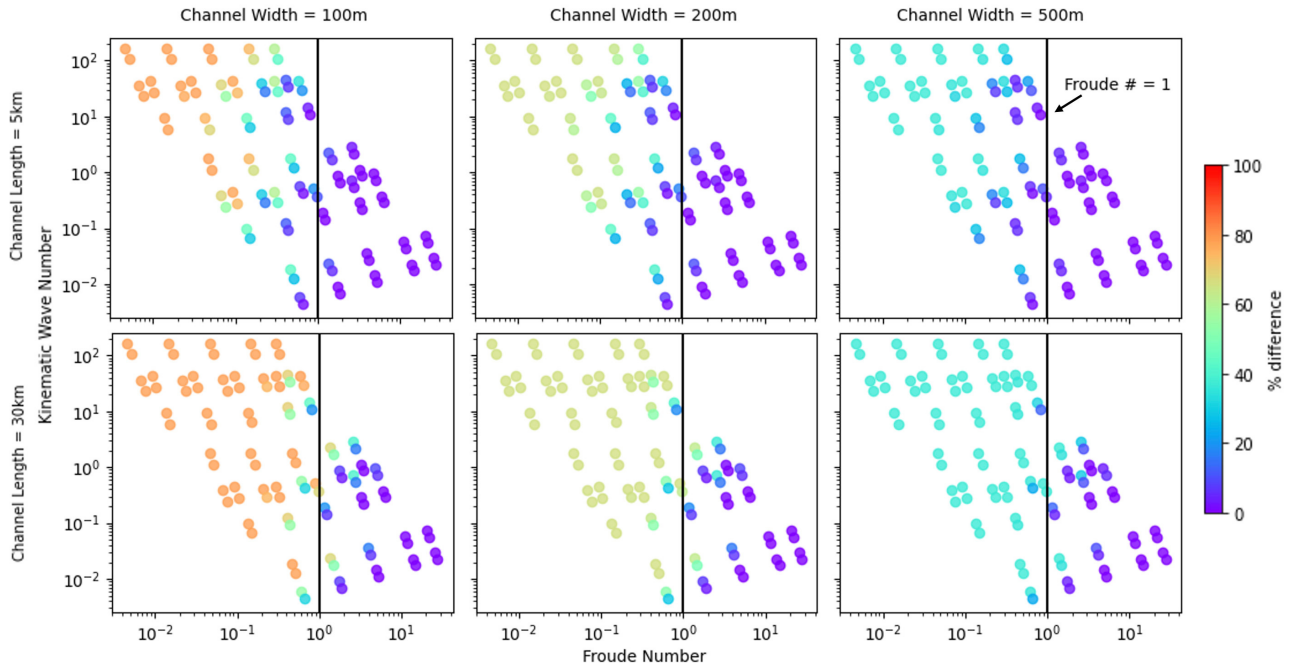


FIGURE 9
Kinematic Wave Number and Froude Number of coarse model across channel width and channel length scenarios. Color represents the difference in peak between the coarse baseline and high-resolution baseline models.

The Froude Number and Kinematic Wave Number were used to determine if there is a non-dimensional pattern where the coarse model performed most differently than the high-resolution baseline model. These cases where differences in peak flow are highest are cases where the subgrid formulation would be most impactful. Figure 9 shows where all 400 parameter combinations fall in the Froude Number and KWN space. The color bar represents the same values as shown in Figure 5, which is the percentage difference between the coarse and high-resolution baseline models for the same parameter values. As seen before, smaller channel width and longer channel length have higher percent differences between the two models. The upper left portion of the sample space, high KWN and low Froude Number, also has higher percent difference between the two models. Low Froude Number and high KWN also corresponds with high Manning's n and low bottom slope across all three models, as was also seen in Figure 5. While there isn't a specific cutoff where a subgrid formulation is necessary, when the Froude Number is less than one is when most large differences between the coarse baseline model and the high-resolution baseline model arise. Analysis of non-dimensional values such as the Froude Number and KWN can be used as guidance regarding when coarse models may not be accurately capturing overland flow in subgrid channel cells and when a subgrid formulation would most improve overland flow accuracy.

4.4 Future work to improve accuracy of subgrid formulations

We used an idealized model domain to benchmark the performance of the subgrid channel formulation, but the only subgrid variability this captures is subgrid channels. This was important to verify this formulation against a high-resolution model without additional sources of variability, but doesn't evaluate the accuracy of this formulation when applied to a real domain which also contains subgrid microtopographic variation outside of channels. To understand how this subgrid channel formulation, as well as future subgrid formulations, are able to improve the accuracy of modeled fluxes, they should be tested on real-world domains. This will also help identify which subgrid parameters are most impactful on the accuracy of overland flow.

An additional challenge in applying subgrid formulations to real-world models is determining the values of new subgrid parameters. There are many ways how channel width has been determined in previous studies including using empirical relationships (Schalge et al., 2019) or satellite data (Neal et al., 2012). When new subgrid formulations are derived, methodologies to assign effective values to the real-world parameters values will also need to be developed so they can be applied to real domains.

5 Conclusions

Subgrid formulations and parameterizations are important next steps in increasing accuracy while maintaining efficiency within large-scale hydrologic models. This study has shown that an additional parameter, channel width, can be incorporated into flux formulations in an integrated surface-groundwater model to increase model accuracy without increasing resolution or runtime. This coarse subgrid model performed nearly identically to the high-resolution baseline model in all scenarios tested in an idealized domain.

The scenarios in which the coarse baseline model differed most from the high-resolution baseline model were identified to

determine under what conditions a subgrid formulation would be most impactful at improving accuracy. The largest difference between the coarse baseline and high-resolution baseline model occurs when flow velocity is low (i.e. when Manning's n is high and channel slope is low), when flow volume is low, and when channel width is much smaller than cell size. When these three flow conditions are combined, there was up to a 78.4% difference in the peak flow between the coarse baseline model and the high-resolution baseline model. This difference in peak was completely mitigated via the use of a subgrid formulation in the coarse model.

Future studies should build on this framework of implementing subgrid formulations into integrated surface-groundwater models to further increase accuracy in more complex systems. This study focused on assessing the accuracy of a subgrid formulation in an idealized domain to isolate the impact on channel flow, but in real-world domains with additional heterogenities, additional subgrid parameters incorporated into flux formulations could further improve accuracy. Large-scale hydrologic modeling is an important next step in answering pertinent hydrology questions and subgrid formulations are necessary to increase their accuracy at coarse resolutions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/miapeeples/subgrid_channel_parflow.

Author contributions

AP: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. RMM: Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This material was based upon work supported by the National Science Foundation Graduate Research Fellowship Program (grant no. DGE-2039656) and HydroGEN, National Science Foundation Convergence Accelerator (grant no. 2134892).

Acknowledgments

The authors would like to thank Georgios Artavanis and Kenadi Waymire for their contributions to the software development of this work in ParFlow. We also thank the editor, Oliver Schilling, and the three reviewers for their constructive feedback during the review process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Ashby, S. F., and Falgout, R. D. (1996). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear Sci. Eng.* 124, 145–159. doi: 10.13182/NSE96-A24230

Bates, P. D., and Hervouet, J. (1999). A new method for moving boundary hydrodynamic problems in shallow water. *Proc. Royal Soc. London. Series A: Mathem. Phys. Eng. Sci.* 455, 3107–3128. doi: 10.1098/rspa.1999.0442

Bauer, P., Gumbricht, T., and Kinzelbach, W. (2006). A regional coupled surface water/groundwater model of the okavango delta, botswana. *Water Resour. Res.* 42:4. doi: 10.1029/2005WR004234

Carluer, N., and Marsily, G. (2004). Assessment and modelling of the influence of man-made networks on the hydrology of a small watershed: implications for fast flow components, water quality and landscape management. *J. Hydrol.* 285, 76–95. doi: 10.1016/j.jhydrol.2003.08.008

Casulli, V. (2009). A high-resolution wetting and drying algorithm for free-surface hydrodynamics. *Int. J. Numer. Methods Fluids* 60, 391–408. doi: 10.1002/fld.1896

Cools, J., Meyus, Y., Woldeamlak, S. T., Batelaan, O., and De Smedt, F. (2006). Large-scale gis-based hydrogeological modeling of flanders: a tool for groundwater management. *Environ. Geol.* 50, 1201–1209. doi: 10.1007/s00254-006-0292-3

David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., and Eijkhout, V. (2011). River network routing on the nhdplus dataset. *J. Hydrometeorol.* 12, 913–934. doi: 10.1175/2011JHM1345.1

Defina, A. (2000). Two-dimensional shallow flow equations for partially dry areas. *Water Resour. Res.* 36, 3251–3264. doi: 10.1029/2000WR900167

Delottier, H., Schilling, O. S., and Therrien, R. (2024). Assessing the impact of surface water and groundwater interactions for regional-scale simulations of water table elevation. *J. Hydrol.* 639:131641. doi: 10.1016/j.jhydrol.2024.131641

Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., et al. (2023). A framework for estimating global river discharge from the surface water and ocean topography satellite mission. *Water Resour. Res.* 59:e2021WR031614. doi: 10.1029/2021WR031614

de Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., et al. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon river basin. *Water Resour. Res. 49*, 1226–1243. doi: 10.1002/wrcr.20067

Ferrari, A., and Viero, D. P. (2020). Floodwater pathways in urban areas: a method to compute porosity fields for anisotropic subgrid models in differential form. *J. Hydrol.* 589:125193. doi: 10.1016/j.jhydrol.2020.125193

Foster, L. M., Williams, K. H., and Maxwell, R. M. (2020). Resolution matters when modeling climate change in headwaters of the colorado river. *Environm. Res. Letters* 15:104031. doi: 10.1088/1748-9326/aba77f

Gascuel-Odoux, C., Aurousseau, P., Doray, T., Squividant, H., Macary, F., Uny, D., et al. (2011). Incorporating landscape features to obtain an object?oriented landscape drainage network representing the connectivity of surface flow pathways over rural catchments. *Hydrol. Process.* 25, 3625–3636. doi: 10.1002/hyp.8089

Ghan, S. J., Liljegren, J. C., Shaw, W. J., Hubbe, J. H., and Doran, J. C. (1997). Influence of subgrid variability on surface hydrology. *J. Clim.* 10, 3157–3166.

Giorgi, F., Francisco, R., and Pal, J. (2003). Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology.: Part i: Effects of temperature and water vapor disaggregation. *J. Hydrometeorol.* 4, 317–333. doi: 10.1175/1525-7541(2003)4<317:EOASTA>2.0.CO;2

Goderniaux, P., Brouyère, S., Fowler, H. J., Blenkinsop, S., Therrien, R., Orban, P., et al. (2009). Large scale surface-subsurface hydrological model to assess climate change impacts on groundwater reserves. *J. Hydrol.* 373, 122–138. doi: 10.1016/j.jhydrol.2009.04.017

Hu, L., Bao, W., Shi, P., Wang, J., and Lu, M. (2020). Simulation of overland flow considering the influence of topographic depressions. *Sci. Rep*. 10:6128. doi: 10.1038/s41598-020-63001-y

Jan, A., Coon, E. T., Graham, J. D., and Painter, S. L. (2018). A subgrid approach for modeling microtopography effects on overland flow. *Water Resour. Res.* 54, 6153–6167. doi: 10.1029/2017WR021898

Jones, J. E., and Woodward, C. S. (2001). Newton krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Adv. Water Resour.* 24, 763–774. doi: 10.1016/S0309-1708(00)00075-0

Kabat, P., Hutjes, R. W. A., and Feddes, R. A. (1997). The scaling characteristics of soil parameters: from plot scale heterogeneity to subgrid parameterization. *J. Hydrol.* 190, 363–396. doi: 10.1016/S0022-1694(96)03134-4

Kollet, S. J., and Maxwell, R. M. (2006). Integrated surface groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.* 29, 945–958. doi: 10.1016/j.advwatres.2005.08.006

Kreye, P., and Meon, G. (2016). Subgrid spatial variability of soil hydraulic functions for hydrological modelling. *Hydrol. Earth Syst. Sci.* 20, 2557–2571. doi: 10.5194/hess-20-2557-2016

Leung, L. R., Wigmosta, M. S., Ghan, S. J., Epstein, D. J., and Vail, L. W. (1996). Application of a subgrid orographic precipitation/surface hydrology scheme to a mountain watershed. *J. Geophys. Res.: Atmospheres* 101, 12803–12817. doi: 10.1029/96JD00441

Li, Z., and Hodges, B. R. (2019). Modeling subgrid-scale topographic effects on shallow marsh hydrodynamics and salinity transport. *Adv. Water Resour.* 129, 1–15. doi: 10.1016/j.advwatres.2019.05.004

Ma, Y., Leonarduzzi, E., Defnet, A., Melchior, P., Condon, L. E., and Maxwell, R. M. (2024). Water table depth estimates over the contiguous united states using a random forest model. *Groundwater* 62, 34–43. doi: 10.1111/gwat.13362

Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Adv. Water Resour.* 53, 109–117. doi: 10.1016/j.advwatres.2012.10.001

Miller, M. P., Carlisle, D. M., Wolock, D. M., and Wieczorek, M. (2018). A database of natural monthly streamflow estimates from 1950 to 2015 for the conterminous united states. *JAWRA J. Am. Water Res. Assoc.* 54, 1258–1269. doi: 10.1111/1752-1688.12685

Mohanasundaram, S., Mekonnen, M. M., Haacker, E., Ray, C., Lim, S., and Shrestha, S. (2021). An application of grace mission datasets for streamflow and baseflow estimation in the conterminous united states basins. *J. Hydrol.* 601:126622. doi: 10.1016/j.jhydrol.2021.126622

Moretti, G., and Orlandini, S. (2018). Hydrography-driven coarsening of grid digital elevation models. *Water Resour. Res.* 54:3654–3672. doi: 10.1029/2017WR021206

Neal, J., Schumann, G., and Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resour. Res.* 48:12514. doi: 10.1029/2012WR012514

Noilhan, J., Lacarrere, P., Dolman, A. J., and Blyth, E. M. (1997). Defining area-average parameters in meteorological models for land surfaces with mesoscale heterogeneity. *J. Hydrol.* 190, 302-316. doi: 10.1016/S0022-1694(96)03131-9

O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M. (2021). Assessment of the parflow-clm conus 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous united states. *Geosci. Model Dev.* 14, 7223–7254. doi: 10.5194/gmd-14-7223-2021

Panday, S., and Huyakorn, P. S. (2004). A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Adv. Water Resour.* 27, 361–382. doi: 10.1016/j.advwatres.2004.02.016

Schalge, B., Haefliger, V., Kollet, S., and Simmer, C. (2019). Improvement of surface run-off in the hydrological model parflow by a scale-consistent river parameterization. *Hydrol. Process.* 33, 2006–2019. doi: 10.1002/hyp.13448

Shen, C., and Phanikumar, M. S. (2010). A process-based, distributed hydrologic model based on a large-scale method for surface-subsurface coupling. Adva. *Water Resour.* 33, 1524–1541. doi: 10.1016/j.advwatres.2010.09.002

Shuttleworth, W. J. (1988). Macrohydrology the new challenge for process hydrology. *J. Hydrol.* 100, 31–56. doi: 10.1016/0022-1694(88)90180-1

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P. (2011). Large-scale groundwater modeling using global datasets: a test case for the rhine-meuse basin. *Hydrol. Earth System Sci.* 15, 2913–2935. doi: 10.5194/hess-15-2913-2011

Thompson, S. E., Katul, G. G., and Porporato, A. (2010). Role of microtopography in rainfall-runoff partitioning: an analysis using idealized geometry. *Water Resour. Res.* 46:8835. doi: 10.1029/2009WR008835

Vieira, J. H. D. (1983). Conditions governing the use of approximations for the saint-vnant equations for shallow surface water flow. *J. Hydrol.* 60, 43–58. doi: 10.1016/0022-1694(83)90013-6

Viero, D. P., Peruzzo, P., Carniello, L., and Defina, A. (2014). Integrated mathematical modeling of hydrological and hydrodynamic response to rainfall events in rural lowland catchments. *Water Resour. Res.* 50, 5941–5957. doi: 10.1002/2013WR014293

Viero, D. P., and Valipour, M. (2017). Modeling anisotropy in free-surface overland and shallow inundation flows. *Adv. Water Resour.* 104, 1–14. doi: 10.1016/j.advwatres.2017.03.007

Volp, N. D., van Prooijen, B. C., and Stelling, G. S. (2013). A finite volume approach for shallow water flow accounting for high-resolution bathymetry and roughness data. *Water Resour. Res.* 49, 4126–4135. doi: 10.1002/wrcr.20324

Wang, D., and Wang, G. (2007). Toward a robust canopy hydrology scheme with precipitation subgrid variability. *J. Hydrometeorol.* 8, 439–446. doi: 10.1175/JHM585.1

Wood, E. F., Lettenmaier, D. P., and Zartarian, V. G. (1992). A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res.: Atmosph.* 97, 2717–2728. doi: 10.1029/91JD01786

Yang, C., Tijerina-Kreuzer, D. T., Tran, H. V., Condon, L. E., and Maxwell, R. M. (2023). A high-resolution, 3d groundwater-surface water simulation of the contiguous us: Advances in the integrated parflow conus 2.0 modeling platform. *J. Hydrol.* 626:130294. doi: 10.1016/j.jhydrol.2023.130294

Zhang, J., Wu, G., Liang, B., and Chen, Y. (2023). Subgrid modeling of urban flooding incorporating buildings effects. *Phys. Fluids* 35:127108. doi: 10.1063/5.0178816

frontiers | Frontiers in Water

TYPE Correction
PUBLISHED 08 May 2025
DOI 10.3389/frwa.2025.1610405

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Oliver S. Schilling,
University of Basel, Switzerland

*CORRESPONDENCE
Amelia Peeples
✉ peeples@princeton.edu
Reed M. Maxwell
✉ reedmaxwell@princeton.edu

RECEIVED 11 April 2025
ACCEPTED 24 April 2025
PUBLISHED 08 May 2025

CITATION
Peeples A and Maxwell RM (2025) Corrigendum: Subgrid channel formulation in an integrated surface-subsurface hydrologic model. *Front. Water* 7:1610405. doi: 10.3389/frwa.2025.1610405

COPYRIGHT
© 2025 Peeples and Maxwell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Corrigendum: Subgrid channel formulation in an integrated surface-subsurface hydrologic model

Amelia Peeples[1]* and Reed M. Maxwell[1,2,3]*

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, United States, [2]High Meadows Environmental Institute, Princeton University, Princeton, NJ, United States, [3]Integrated GroundWater Modeling Center, Princeton University, Princeton, NJ, United States

KEYWORDS

channel flow, integrated hydrologic model, subgrid formulation, subgrid parameterization, ParFlow

A Corrigendum on
Subgrid channel formulation in an integrated surface-subsurface hydrologic model

by Peeples, A., and Maxwell, R. M. (2025). *Front. Water* 6:1520913. doi: 10.3389/frwa.2024.1520913

In the published article, there was an error. A correction has been made to **Methods**, *Idealized test case*, paragraph 2. The model timestep was incorrectly stated to be 1 hour when it is was 0.1 hours. This sentence previously stated: "Four hours of spatially invariable rainfall are applied at the beginning of each simulation and then the simulation continues with hourly timesteps until outflow is approaching zero."

The corrected sentence appears below:

"Four hours of spatially invariable rainfall are applied at the beginning of each simulation and then the simulation continues with 0.1-h timesteps until outflow is approaching zero."

A correction has been made to **Results,** *Coarse baseline model performance*, paragraph 3. The units for Manning's n were incorrectly reported as being s/m$^{1/3}$ when the values given were in min/m$^{1/3}$. This sentence previously stated: "Overall, the largest discrepancy in peak flow of 78.40% is seen in the scenario where channel width is 100 m, rainfall intensity is 0.5 cm/hr, Manning's n is 6e-3 s/m$^{1/3}$, and bottom slope is 1e-4 m/m."

The corrected sentence appears below:

"Overall, the largest discrepancy in peak flow of 78.40% is seen in the scenario where channel width is 100 m, rainfall intensity is 0.5 cm/hr, Manning's n is 3.6e-1 s/m$^{1/3}$, and bottom slope is 1e-4 m/m."

There was an error in Table 2 as published. The units for Manning's n were incorrectly reported as being s/m$^{1/3}$ when the values given were in min/m$^{1/3}$. The corrected Table 2 and its caption appear below.

There was an error in Figure 5 as published. The units for Manning's n were incorrectly reported as being s/m$^{1/3}$ when the values given were in min/m$^{1/3}$. The corrected Figure 5 and its caption appear below.

There was an error in Figure 6 as published. The units for Manning's n were incorrectly reported as being $s/m^{1/3}$ when the values given were in $min/m^{1/3}$. The corrected Figure 6 and its caption appear below.

There was an error in Figure 8 as published. The units for Manning's n were incorrectly reported as being $s/m^{1/3}$ when the values given were in $min/m^{1/3}$. The corrected Figure 8 and its caption appear below.

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

TABLE 2 All input parameters varied and their corresponding values.

Channel width (m)	Rainfall intensity (cm/hr)	Manning's n (s/m$^{1/3}$)	Bottom slope (m/m)
100	0.5	3.6e-3	1e-4
200	1	1.8e-2	1e-3
500	5	3.6e-2	1e-2
1,000	10	1.8e-1	1e-1
-	-	3.6e-1	4e-1

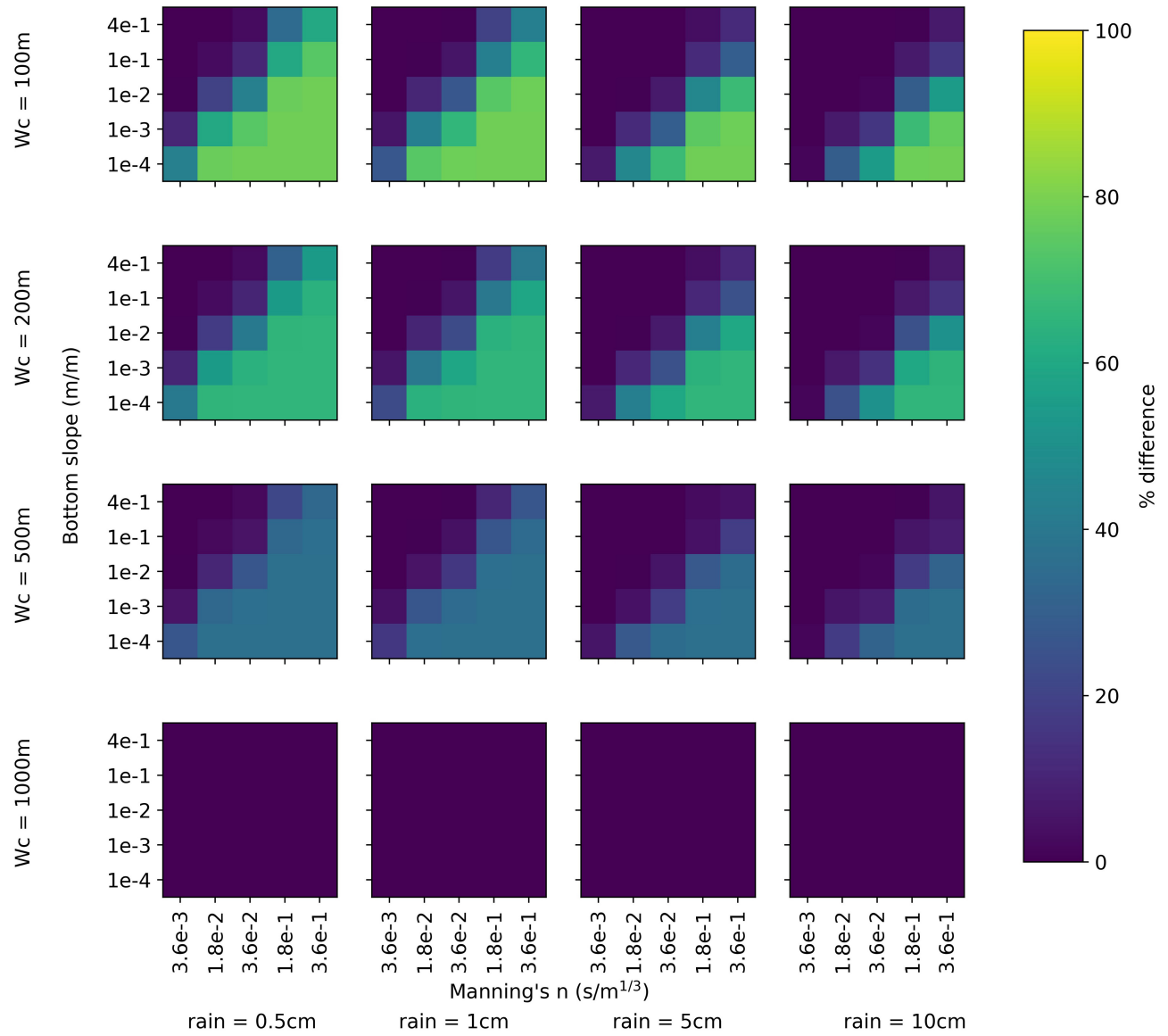


FIGURE 5
Percentage difference in peak flow between the coarse baseline and high-resolution baseline models. Here channel width is not an input in the coarse baseline model but instead is only used to define the domain resolution of the high-resolution model.

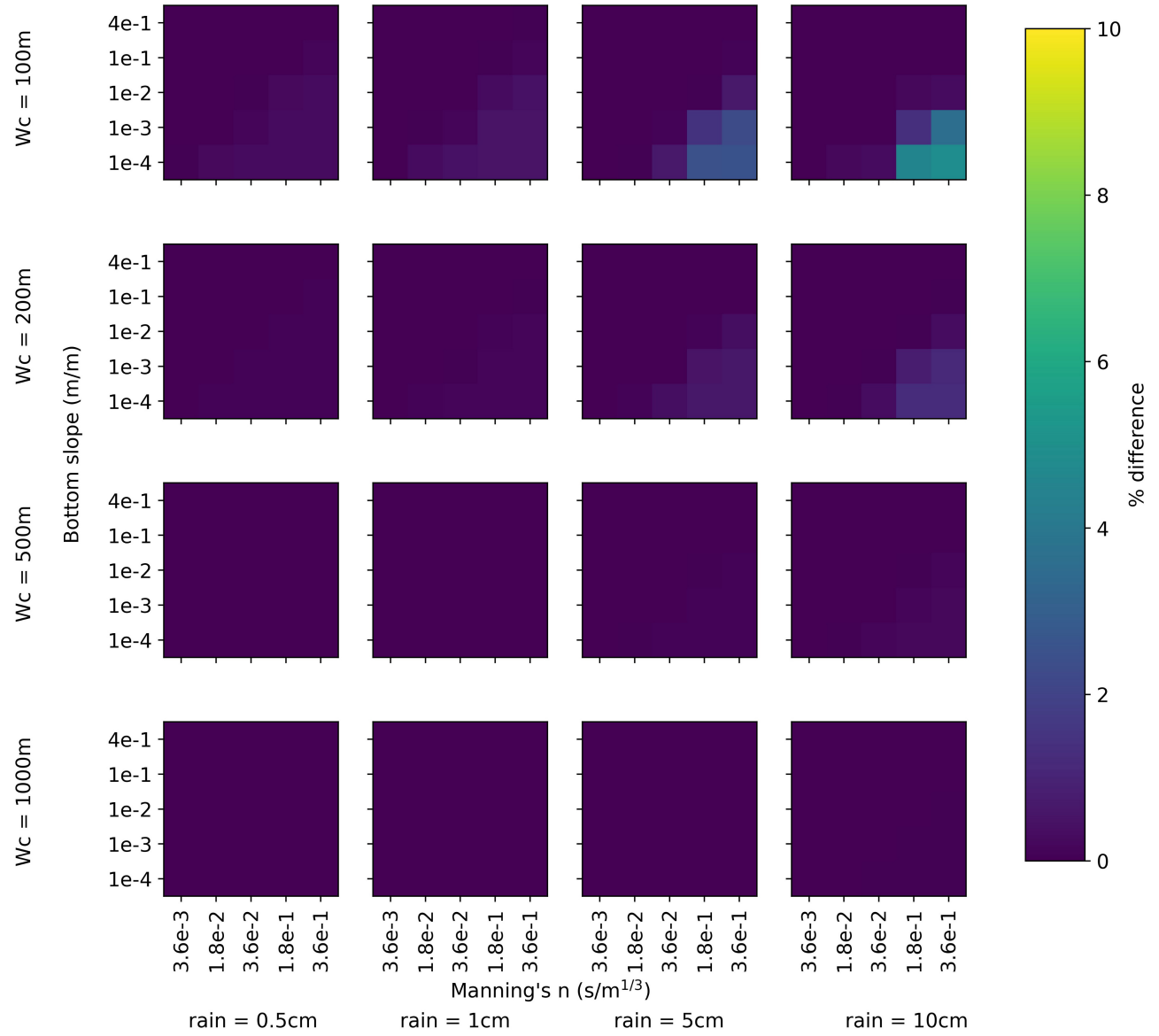


FIGURE 6
Percentage difference in peak flow between the coarse subgrid formulation and high-resolution baseline models. Here channel width is an input in the coarse subgrid model as well as used to define the domain resolution of the high-resolution model.

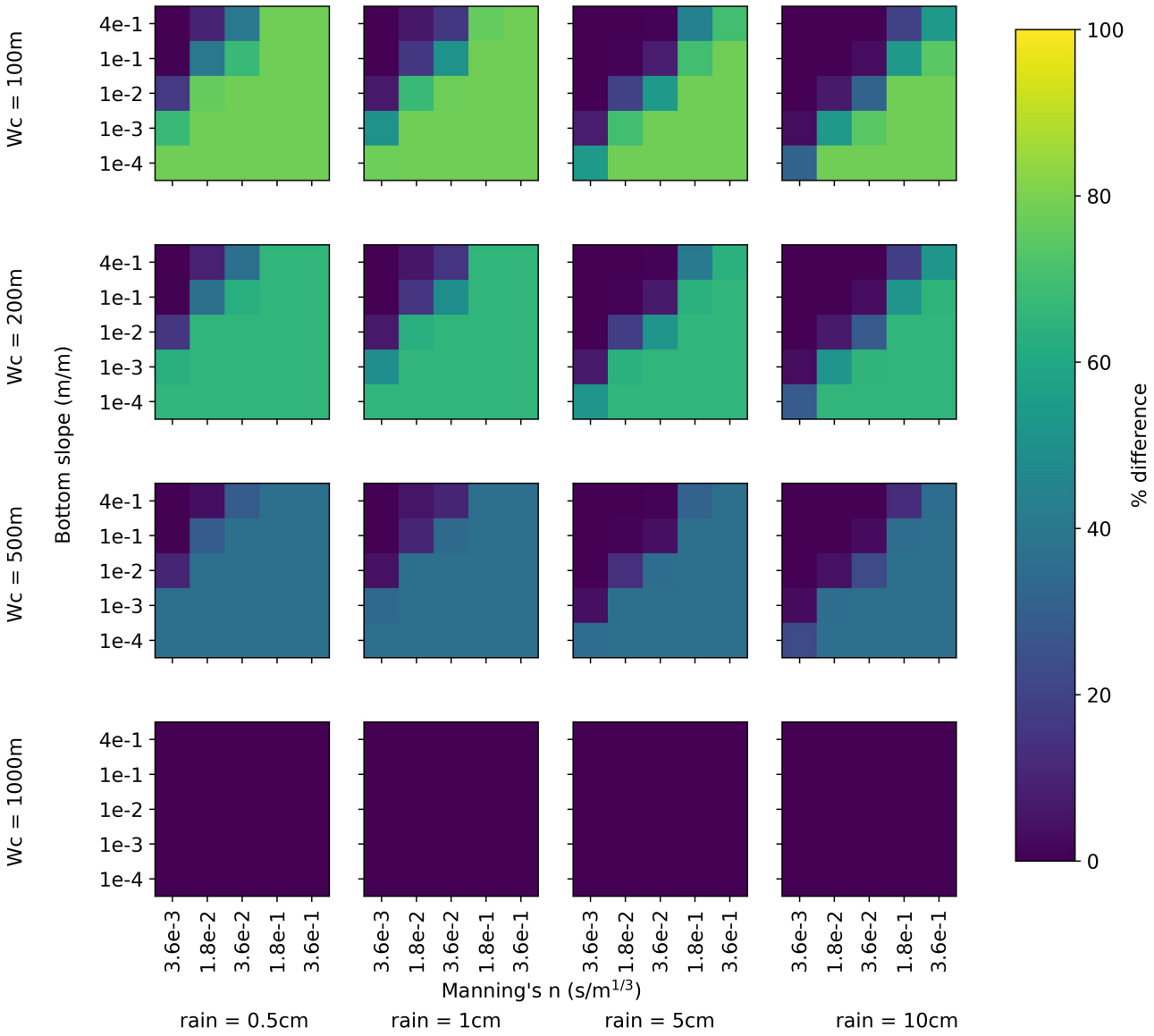


FIGURE 8
Percentage difference in peak flow between the coarse baseline formulation and high-resolution baseline models at the outlet of the 30 km channel.

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 09 May 2025
DOI 10.3389/frwa.2025.1553578

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Padam Jee Omar,
Babasaheb Bhimrao Ambedkar University, India
Xiaofan Yang,
Beijing Normal University, China

*CORRESPONDENCE
Claudio Paniconi
✉ claudio.paniconi@inrs.ca

RECEIVED 30 December 2024
ACCEPTED 10 April 2025
PUBLISHED 09 May 2025

CITATION
Paniconi C, Lauvernet C and Rivard C (2025) Exploration of coupled surface–subsurface hydrological model responses and challenges through catchment- and hillslope-scale examples.
Front. Water 7:1553578.
doi: 10.3389/frwa.2025.1553578

COPYRIGHT
© 2025 Paniconi, Lauvernet and Rivard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploration of coupled surface–subsurface hydrological model responses and challenges through catchment- and hillslope-scale examples

Claudio Paniconi[1]*, Claire Lauvernet[2] and Christine Rivard[3]

[1]INRS-ETE, Université du Québec, Quebec City, QC, Canada, [2]INRAE-RiverLy, Lyon, France, [3]Geological Survey of Canada, Natural Resources Canada, Quebec City, QC, Canada

Selected runs with a physics-based model of surface water–groundwater interactions are used to examine in detail some numerical challenges and surprising behaviors that result from discretization, nested solution schemes, coupling, boundary condition, and other factors. Regardless of the spatial scale of the model domain (field, hillslope, catchment, ...), the processes that are simulated by this class of integrated models can exhibit widely varying dynamics within and across the different subsystems comprising the land surface, the unsaturated zone, and deep groundwater formations. The presence of heterogeneities, nonlinearities, and complex boundary conditions can exacerbate numerical difficulties in resolving exchange fluxes across subsystems and lead to unexpected or undesired results, including localized numerical oscillations and an upper bound on adaptive time stepping. The need for accurate tracking of surface–subsurface exchanges and for better control of aspect ratio and mesh distortion can also influence and constrain spatial and temporal discretization choices. Finally, model performance assessments can be highly sensitive to the response variables of interest. We will illustrate some of these issues via test case simulations at large (13.66 km catchment transect) and small (450 m^2 hillslope) spatial scales, run at time scales from 10 days to hundreds of years.

KEYWORDS

numerical modeling, catchment hydrology model, surface/surface interactions, unsaturated zone, boundary conditions

1 Introduction

Interactions between groundwater and surface water systems play a critical role in regulating important hydrological, ecological, and biogeochemical processes such as baseflow, flooding, riparian zone vegetation growth, and hyporheic zone denitrification and oxygen redistribution (Irvine et al., 2024). Recent advances in integrated groundwater–surface water models make it possible to explore in detail the interactions between the flow of water and the transfer and transport of solutes (which can include contaminants, nutrients, microorganisms, and gasses) within and across surface (overland, streams, lakes) and subsurface (aquifer and soil) domains (Paniconi and Putti, 2015). There are however persistent challenges in the development of reliable and efficient numerical models for these phenomena. When passing from subsurface-only (or surface-only) models to integrated models, customary difficulties associated with heterogeneity and variability in parameters and state variables, nonlinearities and scale effects in process dynamics, and poorly known boundary conditions (BCs) and initial system states (e.g., Sanchez-Vila et al., 2006; Vereecken et al., 2016) are compounded (more complex BCs, nested levels of iteration and time stepping, etc.) and new hurdles are

introduced, namely the need to use appropriate and robust coupling schemes and representations of process interactions and feedbacks (e.g., Discacciati et al., 2002; Dawson et al., 2004; Furman, 2008).

In this study we will restrict our attention to physics-based hydrological models, i.e., models built from the governing equations for surface and subsurface water flow and solute transport, specifically the Saint-Venant or shallow water equations for surface routine, the Richards equation for variably saturated subsurface flow, and the advection–dispersion equation for mass transport. The equations are resolved via appropriate discretization techniques that allow for a detailed representation of parameter variability and boundary condition complexity. These models require reliable numerical techniques for their resolution (e.g., Miller et al., 2013; Farthing and Ogden, 2017). There is by now a quite substantial literature on the seemingly successful application of physics-based integrated surface–subsurface hydrological models (ISSHMs), including a number of model intercomparison and case studies (e.g., Sulis et al., 2010; Maxwell et al., 2014; Kollet et al., 2017; Omar et al., 2021).

While the proliferation of integrated models reflects the considerable need for these tools in a diversity of fields and applications, it is also true that hydrology-focused research studies will not typically report in detail on negative aspects of numerical performance, such as convergence difficulties, localized inaccuracies, and other anomalous behaviors. Indeed, continuing research on any already accepted model will generally focus more on adding process complexity to the model rather than on resolving legacy flaws and limitations (which workarounds – such as grid refinement – can on occasion alleviate). At the same time, studies in the numerics literature do not generally use models that are at the same level of detail (in representing heterogeneities and boundary conditions, for instance) as the current generation of ISSHMs. Thus there can be a long gap before a state-of-the-art numerical scheme demonstrated for an ideal configuration can be adapted for more complex, general-purpose models. As an example of this, the theoretical framework for a mass-conservative scheme for surface–subsurface coupling based on boundary condition switching under infiltration scenarios (Sochala et al., 2009) has not yet been extended to evaporation and mixed (storm–interstorm) scenarios. As another example, theoretically advantageous alternatives to standard Picard or Newton iteration (Paniconi et al., 1991; Kelley, 1995) for linearizing Richards' equation, such as the L–scheme (List and Radu, 2016; Stokke et al., 2023), exist but have not yet been fully vetted for ISSHMs, where factors such as highly varied BC types and strong subsurface heterogeneity can introduce implementation and performance challenges.

Given the complexities of physics-based ISSHMs and their continuing evolution, it is important to highlight some of the undesirable or unexpected numerical results that can occur when running these models, as this may in some measure help guide the development of new or improved equation solvers and discretization schemes for strongly coupled and highly nonlinear systems, and ultimately contribute to future improvements to ISSHMs. Moreover, the coupling itself, manifested across an interface between disparate flow domains and through intricate boundary conditions or exchange terms, can raise new issues for both established and emerging numerical schemes for solving the flow equations, and test simulations with an ISSHM can help to identify these.

Within this context, we seek to illustrate in this study, via example simulations, some of the numerical challenges and surprising behaviors that arise in the use of a physically based, integrated model of surface water and groundwater flow. The approach here is empirical rather than systematic, with the only strictly organizational element being in the selection of the two test cases: one at a quite large spatial scale (a 13.66 km catchment transect) run at very long time scales (hundreds of years) and the other at a small spatial scale (a 450 m^2 hillslope) run at a time scale of days. We will use the CATHY ("catchment hydrology") ISSHM (Camporese et al., 2010; Weill et al., 2011) for all the simulations. This is a widely applied and fully tested model that uses numerical schemes that are standard for this class of hydrological model. In the hillslope example we will examine how activating the coupling scheme and surface routing module in CATHY impacts rainfall–runoff partitioning, including a contribution to ponding that cannot occur when surface–subsurface interactions are not fully resolved. In the transect example we will explore different land surface boundary condition settings, temporal and spatial discretization challenges, and the meaning of a steady state when flow conditions are strongly nonlinear.

2 Methods

2.1 CATHY model

CATHY is a coupled surface water–groundwater model with subsurface flow represented by the 3D Richards equation, solved by Galerkin finite elements, and surface flow represented by a path-based (quasi-2D) Saint-Venant equation, solved by a Muskingum-Cunge finite difference scheme. Analogous numerical schemes are used for the 3D advection–dispersion and quasi-2D advection–diffusion solute transport equations. A preprocessing analysis of the digital elevation model (DEM) of the topographic data identifies the drainage network and distinguishes overland (hillslope) and channel (stream) flow cells, which can then be parameterized separately (Orlandini et al., 2003; Orlandini and Moretti, 2009). This novel path-based representation of surface routing allows for a unified treatment of overland and channel flow and transport (same governing equation, different parameterizations).

Coupling between the subsurface and surface flow modules is through a boundary condition switching approach (Putti and Paniconi, 2004), whereby Neumann (specified flux) or Dirichlet (specified head) conditions are dynamically imposed according to the saturation status of a surface node (ponded, saturated, unsaturated, and air dry). At each time step, once the subsurface equation is solved, a water balance is calculated at each surface node between the atmospheric water supply (precipitation) or demand (potential evapotranspiration), the capacity of the soil to infiltrate or exfiltrate this water, and any ponded water already present at the surface. Any excess of water accumulated at the surface (ponding) becomes available for routing by the surface solver if it exceeds a threshold parameter (minimum water depth before surface routing can occur). Through this procedure, CATHY automatically tracks both infiltration excess (Hortonian) and saturation excess (Dunnian) overland flow generation mechanisms. This boundary condition-based coupling algorithm is one of three common approaches used in ISSHMs (e.g., Haque et al., 2021), and it has been theoretically shown to ensure pressure and flux continuity at the land surface interface (Sochala et al., 2009), at least under the infiltration case that was analyzed. Note that transpiration is not included in the version of CATHY used in this study, thus potential evapotranspiration data input to the model is resolved (converted into an exfiltration flux from the soil) according

to a BC switching scheme for actual evaporation that is analogous to the scheme used for rainfall–runoff partitioning.

2.2 Test cases

The hillslope-scale test case (Figure 1) is from the Biosphere 2 Landscape Evolution Observatory (LEO). See Niu et al. (2014a) and Scudeler et al. (2016a) for a description of LEO and of the CATHY experiments on this hillslope. For the runs reported in this study we used a uniform discretization at the surface ($\Delta x = \Delta y = 1$ m; 30 m × 15 m total) and 20 layers of equal thickness vertically ($\Delta z = 7.5$ cm; 1.5 m total), for a grid of 10,416 nodes. The average terrain slope is 17.6% (10°), with a maximum slope of 30.6% (17°) around the outlet or convergence zone.

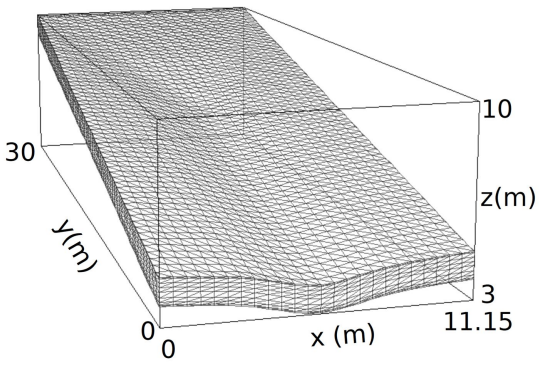


FIGURE 1
The LEO hillslope test case showing the uniform mesh discretization.

The transect test case (Figure 2) is from the 425 km^2 Tony Creek subcatchment near the town of Fox Creek in west-central Alberta, Canada. See Meneses-Vega et al. (under review)[1] for a detailed presentation of the 700 km^2 Fox Creek study area and of the CATHY applications in both two-dimensional (2D transect) and fully 3D configurations. For the runs reported in this study we used a uniform discretization along the transect ($\Delta x = 20$ m; 13.66 km total) and 15 layers of varying thickness vertically (from 15 cm to 60 m; 300 m total), for a grid of 32,832 nodes (this includes a 3-node discretization in the no-flow transverse direction since CATHY is a 3D model). A finer layering was used near the surface and toward the base of the domain, in order to more accurately resolve the atmospheric forcing and free drainage boundary conditions. The average terrain slope is 1.7% (1.0°) on the transect segment to the right of the creek (valley bottom) and 1.6% (0.9°) to the left, with a maximum slope of 6.8% (3.9°) along the right streambank.

The soil and aquifer parameter values for these two test cases are summarized in Table 1.

3 Results

3.1 LEO hillslope example

For the CATHY simulations on LEO, the hillslope was initially completely unsaturated, with a pressure head distribution obtained from a long-term drainage experiment. Drainage occurs across a

1 Meneses-Vega, B. J., Paniconi, C., Rivard, C., and Guarin-Martinez, L. I. (under review). A multi-model study of the subsurface and surface hydrodynamics of a 700 km^2 watershed in western Canada (Fox Creek area, Alberta). *J. Hydrol. Reg. Stu.*

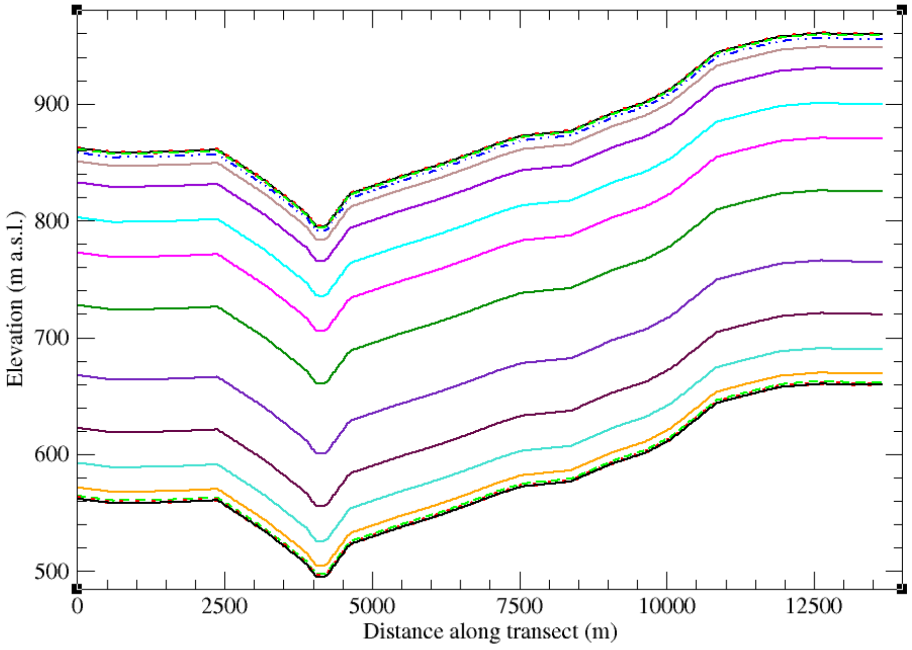


FIGURE 2
The Tony Creek transect test case showing the vertical mesh discretization.

TABLE 1 Parameter values for the two test cases.

Test case	LEO hillslope	Tony Creek transect
Saturated hydraulic conductivity K_s (m s^{-1})	4.05×10^{-5}	4.5×10^{-6} lateral; 4.5×10^{-7} vertical except bottom layer 2.0×10^{-9}
Porosity θ_s (m^3 m^{-3})	0.41	0.35
Residual water content θ_r (m^3 m^{-3})	0.057	0.08
van Genuchten n (−)	2.28	2.2
van Genuchten ψ_a (m)	−0.0806	−0.15
Specific storage S_s (m^{-1})	5.0×10^{-4}	8.1×10^{-5} for the 50-year simulations; 1.0×10^{-2} for the 729-year simulations

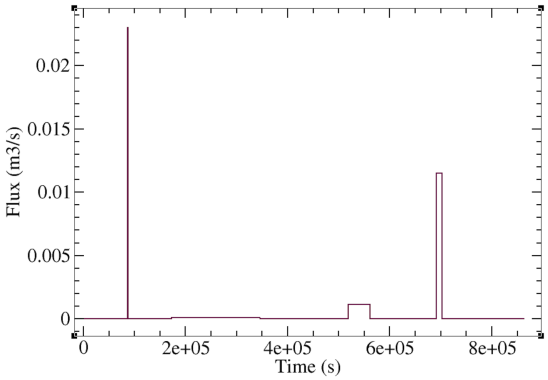


FIGURE 3
Atmospheric forcing for the 10-day (8.64×10^5 s) LEO experiment showing the four rain events.

Dirichlet (zero pressure head) boundary condition imposed at the 16 nodes along the bottom of the downslope face of the hillslope, with the highest outflows in the middle of this face, in correspondence with the hillslope's planform. The soil for this experiment was considered homogeneous and isotropic (these assumptions were relaxed in subsequent LEO trials with CATHY, reported in Niu et al., 2014a; Pasetto et al., 2015; Scudeler et al., 2016a), with saturated hydraulic conductivity $K_s = 4.05 \times 10^{-5}$ m/s, porosity $\theta_s = 0.41$, specific storage $S_s = 5 \times 10^{-4}$ m^{-1}, and van Genuchten (1980) soil water retention curve parameter values of: fitting exponent $n = 2.28$, residual moisture content $\theta_r = 0.057$, and air entry pressure head $\psi_a = -0.0806$ m (see Table 1). The experiment was of a 10-day (8.64×10^5 s) duration and comprised four precipitation events generated by the rainfall generator at LEO: 180 mm/h (5.0×10^{-5} m/s) for a duration of 20 min, 0.9 mm/h (2.5×10^{-7} m/s) for 2 d, 9 mm/h (2.5×10^{-6} m/s) for 12 h, and 90 mm/h (2.5×10^{-5} m/s) for 3 h (Figure 3). Since K_s is smaller than the rainfall rate of the first event, infiltration excess runoff is generated during this event. For the middle two (low rainfall) events, no surface runoff is produced. For the last event, notwithstanding the initially quite dry soil conditions and the presence of a Dirichlet BC at the base of the hillslope, Dunne saturation excess runoff is produced over a portion of the hillslope. We will examine more closely the surface saturation response of the hillslope in the next section. Here we will look instead at the fluxes that are generated.

3.1.1 Flux partitioning

The impact of surface–subsurface interactions can be seen in Figure 4 were fluxes imposed (atmospheric BC) and generated by the model (actual, overland, and return flow) are plotted for the first (top graph) and last (bottom graph) rain events for CATHY simulations with and without coupling (in no-coupling mode, only the subsurface module is run). In the first (Horton runoff-generating) event, when the rainfall period ends the actual flux (infiltration rate) falls to zero in the subsurface-only run, whereas in the coupled run infiltration persists because there is some water that has accumulated at the surface from the runoff generation and routing processes, seen as the green overland flow curve (negative after the rain event because it is subtracted from the rainfall rate). Even during the rainfall period, the actual flux is slightly higher in the coupled case because any ponded water amount is "added" to the rainfall rate, creating a stronger head gradient for infiltration after the "time to ponding" is reached. There is no return flow (exfiltration across a ponded or saturated surface) for the first event because below the land surface the soil profile is still unsaturated.

For the last (Dunne runoff-generating) event, we again see post-event persistence of infiltration for the coupled run, of a smaller magnitude and duration here because the rainfall rate is much lower than the first event. During the rainfall period, the actual flux is now slightly lower in the coupled case than in the subsurface-only case. This is because there is now return flow occurring in the fully saturated portion of the hillslope (downslope around the convergence zone), making the net (downward) infiltration lower. Later in the simulation (and not shown here), some of this exfiltration eventually reinfiltrates into the soil. The dashed green curve in the plot for the last rain event is from a CATHY run with no coupling and no solute transport (so subsurface flow only). We observe that overland flow is higher when transport is active, indicating an additional dispersive or mixing layer impact on surface–subsurface interactions (Gatel et al., 2020; Gatto et al., 2021) that is highly complex and possibly conditioned by how boundary conditions are treated. This effect requires further study, as indeed do many aspects of solute transport handling in ISSHMs, which has received much less attention in the literature compared to flow modeling. Solute transport will not be considered in the Tony Creek example.

3.1.2 Surface saturation response

For the same two events for which we examined flux partitioning across the land surface interface in Figure 4, we now look at the saturation response of this boundary. The top graph in Figure 5 shows the fraction of the hillslope surface that is saturated during the first and most intense of the four rain events. Since the rainfall rate exceeds the K_s of the surface, and since this event occurs early in the 10-day simulation period when the soil is still quite dry throughout the hillslope, only Horton runoff occurs here. Moreover, since K_s is homogeneous, when saturation does occur, it occurs over the entire hillslope, and thus the saturation fraction reaches 1. The time to ponding is quite rapid (about 500 s after the start of the rain event), and is roughly but not exactly equal over the entire hillslope, due to a nonuniform distribution of soil moisture at the start of the event. As in the preceding section, the inclusion or not of solute transport

10.3389/frwa.2025.1553578

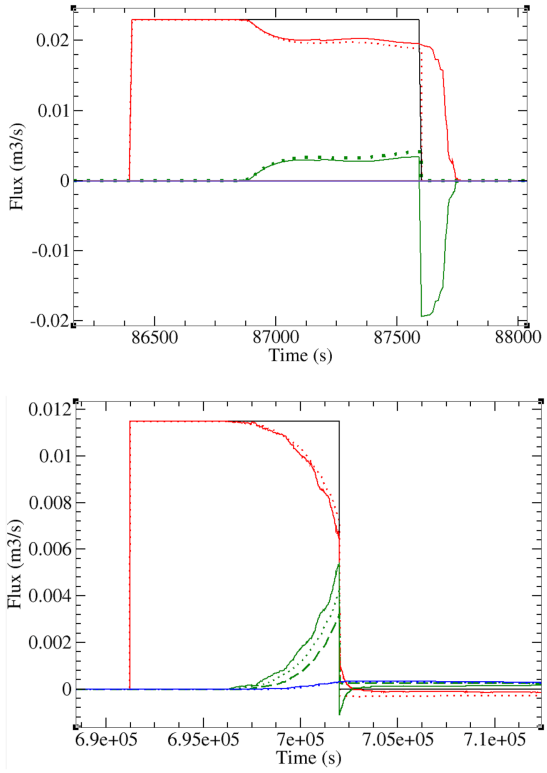


FIGURE 4
Atmospheric input (rainfall) flux (black), actual flux across the land surface (red), overland flow rate (green), and return flow rate (blue) for CATHY in coupled (surface–subsurface flow and transport) mode (solid lines) and uncoupled (subsurface flow and transport only) mode (dotted lines) for the first (top graph) and last (bottom graph) rain events of the LEO hillslope test case. The dashed green line is for a subsurface flow only (i.e., no transport) run.

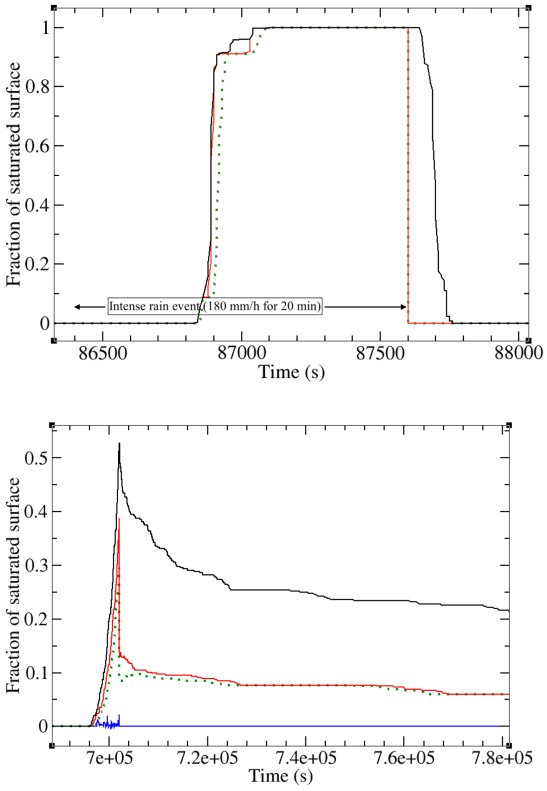


FIGURE 5
Coupled (black), subsurface only (red), and subsurface flow only (i.e., no transport) (green) saturation responses for the first (top) and last (bottom) rain events of the LEO hillslope test case. The blue curve in the bottom graph is the Horton contribution to surface saturation that is produced in the coupled case.

appears to influence the flow responses produced by the model. Also noteworthy is the difference between the coupled and uncoupled model at the end of the rain event. In the coupled case, even though the atmospheric flux is now zero, Horton saturation persists for a short period after the rain has ceased (and the degree of saturation gradually rather than abruptly drops to zero) because of the spatial distribution of ponded water that provides a supplemental input flux. This ponding-assisted Horton saturation is an interesting counterpoint to Horton-assisted upstream expansion of a catchment's Dunne saturated areas reported in Zanetti et al. (2024). These phenomena are a direct result of using an integrated model, and they should be explored in more detail.

For the last rain event (bottom graph), Dunne runoff occurs for all model runs (and notwithstanding the Dirichlet BC at the base of the hillslope, as mentioned previously), and it is clearly spatially nonuniform due to topography, slope, and planform effects. The influence of surface–subsurface interactions is very marked here, with a higher peak in surface saturation (over 50% for the coupled run compared to under 40% for the uncoupled runs) and a more persistent tail in its spatial distribution as the runoff event recedes. The blue curve in this graph at near-zero shows the Horton saturation fraction for the coupled run. There is no Horton saturation for the uncoupled runs, as expected since the rainfall rate is lower than K_s for this event. The surprising occurrence of a small amount of Horton saturation in the coupled case is again due to the presence of ponded water.

3.2 Tony Creek transect example

For the CATHY simulations on Tony Creek, the transect was initially in vertically hydrostatic equilibrium with the water table 2 m below the surface throughout the domain. Different combinations of boundary condition settings and specific storage values were used, as described in the following sections. The soil and aquifer for this experiment were considered homogeneous (this assumption was relaxed in subsequent Tony Creek trials with CATHY, reported in see footnote 1) but not isotropic, with lateral $K_s = 4.5 \times 10^{-6}$ m/s, vertical $K_s = 4.5 \times 10^{-7}$ m/s (except for the bottom layer), porosity $\theta_s = 0.35$, and van Genuchten soil water retention curve parameter values of: fitting exponent $n = 2.2$, residual moisture content $\theta_r = 0.08$, and air entry pressure head $\psi_a = -0.15$ m (see Table 1). The atmospheric forcing consisted of a constant rainfall rate of 70 mm/y (= 2.22×10^{-9} m/s = 6.065×10^{-5} m^3/s), and free drainage was imposed at the bottom of the domain, at a rate of 2×10^{-9} m/s (= 5.464×10^{-5} m^3/s), or just slightly lower than the rainfall rate. To impose this rate in CATHY, the vertical saturated hydraulic conductivity of the bottom layer was set to the desired free drainage value.

3.2.1 50-year simulations with low specific storage

Figure 6 presents the results from several 50-year simulations using a specific storage (S_s) value of 8.1×10^{-5} m^{-1}. The response variables shown, together with the imposed rainfall and free drainage fluxes, are the actual flux across the land surface, the return flow rate, and the overland flow rate. Five simulations were performed: Run 1, with no coupling (i.e., subsurface flow only) and with no boundary conditions imposed besides the atmospheric forcing; Run 2, with no coupling and with Dirichlet BC imposed at the stream node (valley bottom in Figure 2); Run 3, with no coupling and with seepage face BCs imposed along 4 nodes on either side of the valley and including the stream node; Run 4, no coupling and with seepage face BCs imposed along 25 nodes on either side of the valley and including the stream node; and Run 5, coupled with no BCs on the surface but with a designated outlet cell for the surface routing module. In addition to these response variables, Figure 6 also shows the fluxes across the seepage face boundary for Runs 3 and 4 and the outlet hydrograph for Run 5.

In Figure 7 the land surface saturation responses for the coupled simulation (Run 5) and three of the uncoupled cases (Runs 1, 2, and 3) are shown. The responses for these latter 3 runs are visually identical. All of the responses in this graph are from the saturation excess runoff mechanism, except for a small occurrence of infiltration excess runoff for Run 2, discussed later in this section.

Figure 8 shows the ponding dynamics along a 200 m section across the stream channel valley, and centered at the stream node, for the coupled simulation (Run 5). In the time snapshots we see that noticeable ponding first occurs between the first and sixth month of simulation (at time 0 the water table is uniformly 2 m below the surface, and the rainfall rate is more than two orders of magnitude smaller than the vertical K_s), that the ponding levels increase until 10 years, and that they recede thereafter. It is also evident that the recession is very slow, and is perhaps reaching a steady state by the end of the simulation. Note that the spatial resolution for the output extracted for this graph is rather coarse, as only three nodes spaced 100 m apart were sampled out of the 11 nodes spaced 20 m apart.

Several points emerge from the five 50-year Tony Creek transect simulations shown in Figures 6–8 (some of these remarks apply also

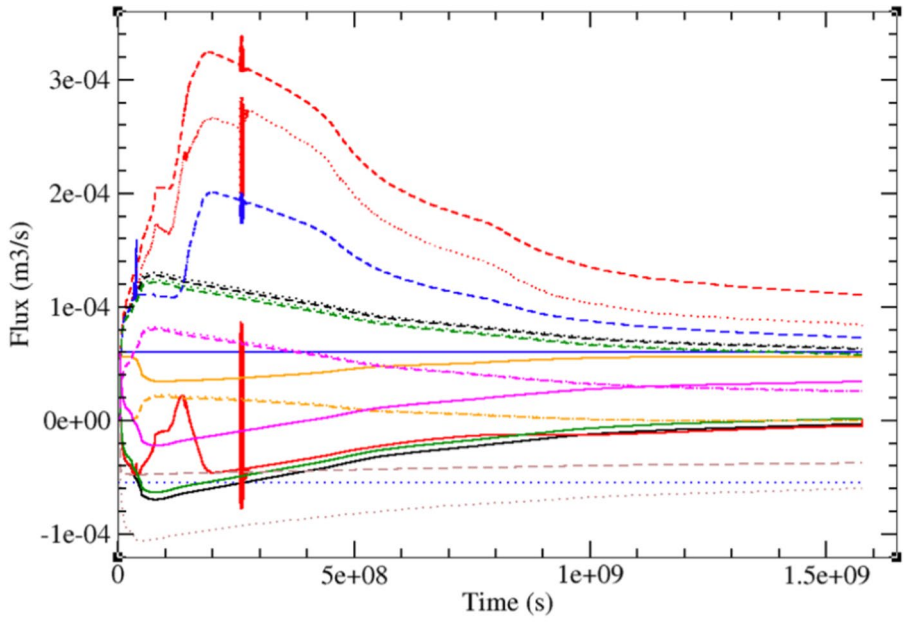


FIGURE 6
Results for five 50-year (1.5768×10^9 s) simulations of the Tony Creek transect. Imposed incoming (rainfall = 70 mm/y = 2.22×10^{-9} m/s = 6.065×10^{-5} m^3/s, solid blue line) and outgoing (free drainage = −63 mm/y = -2×10^{-9} m/s = -5.464×10^{-5} m^3/s, dotted blue line) fluxes are shown, together with the actual flux across the land surface, the return flow rate, and the overland flow rate (respectively solid, dashed, and dotted curves) for: Run 1 (no coupling, i.e., subsurface flow only, and no BCs imposed on the surface besides the rainfall) in black; Run 2 (no coupling and Dirichlet BC imposed at the stream node) in green; Run 3 (no coupling and seepage face BCs imposed along 4 nodes on either side of the valley and including the stream node) in magenta; Run 4 (no coupling and seepage face BCs imposed along 25 nodes on either side of the valley and including the stream node) in orange; and Run 5 (coupled and no BCs on the surface but with a designated outlet cell for the surface routing module) in red. In addition, the brown curves are the fluxes across the seepage face boundary for Runs 3 (dashed) and 4 (dotted), and the dashed blue curve is the outlet hydrograph for Run 5.

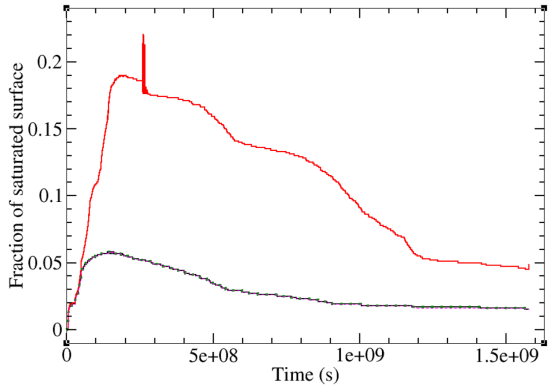


FIGURE 7
Land surface saturation response for the coupled simulation (Run 5) in red and for three of the no coupling runs: no BCs, Run 1 (solid black curve); Dirichlet BC at the stream node, Run 2 (dotted green curve); and seepage face BC along 8 nodes, Run 3 (dashed magenta curve). The curves for Runs 1, 2, and 3 are visually identical.

to the 729-year simulations of the next section). In uncoupled (subsurface only) mode (Runs 1–4), the model is able to reproduce "expected" (for an uncoupled run) dynamics at the land surface even without imposing any of the standard boundary conditions that are typically applied in simulations of this sort. Indeed, the generated fluxes at the land surface (infiltration, surface runoff, return flow, etc.) are generally very similar between the no BC, Dirichlet BC, and seepage face cases. This underscores the role of topography in driving the hydrologic responses for transects (and watersheds) such as the one analyzed here (climate and soil parameters are uniform – except for vertical K_s at the bottom of the domain – and thus do not exert significant control over the spatiotemporal dynamics observed).

In coupled mode (Run 5) all the responses are very different from the uncoupled responses, underscoring the importance of including a proper representation of surface–subsurface interactions when these phenomena are of interest. This means not just capturing instantaneous interactions (e.g., rainfall exceeding infiltration capacity at a given point on the land surface), but also allowing for continuous interactions (e.g., overland flow generated at one point and reinfiltrating further downslope). This is apparent both in the land surface fluxes shown in Figure 6 and in the land surface saturation response shown in Figure 7. It is clear from Figure 7 that accounting for surface–subsurface interactions leads to a much greater fraction of the land surface actively contributing to the overall dynamics.

For the coupled run, we observe in Figure 6 a double hump in the rising limb of the outlet response (and also to some degree in the other response variables for this run). This response likely reflects different characteristic time scales of the shorter and steeper transect to the left of the stream and the longer right side. We also see for this run that the return and overland flow hydrographs are significantly higher than the outlet hydrograph. This is another manifestation of continuous or recycled interactions (as runoff is routed downslope), and of the fact that not all surface runoff reaches the outlet.

In the coupled run we also observe spurious numerical oscillations that may be indicative of time stepping or coupling algorithm constraints in the discretization schemes (Dagès et al., 2012; Fiorentini et al., 2015). This is a complex issue as there are numerous time stepping and iteration nestings in the numerical solution procedure (linearization of Richards' equation; different characteristic time scales and time step constraints between the saturated zone, the vadose zone, overland flow, and channel flow; BCs that can switch from iteration to iteration and from time step to time step). Since the oscillations were only observed for the coupled run, it is not thought to be an issue connected to resolution of the subsurface flow system. It also does not appear to be an issue with the substepping scheme of the surface flow solver. This scheme is based on a Courant number criterion wherein the number of surface solver time steps is calculated at each subsurface solver time step (Camporese et al., 2010). Many tests for the Tony Creek transect were run using both tighter and looser Courant targets (the target is normally set to a value of 1.0). Increasing the target value allows for larger surface routing time step sizes, while decreasing it means that more substeps are taken within each subsurface time step. Neither increasing nor decreasing the Courant target had an impact on the oscillations. Dismissing the surface or subsurface solvers, taken separately, as possible causes, the explanation for the observed oscillations is thought to lie in the coupling scheme. For instance, the boundary condition switching algorithm may need to be improved to allow a smoother buildup of ponding heads when Dirichlet BCs are activated, or the convergence criterion on the subsurface solver needs to be based not only on convergence of the nodal pressure head solution (as is currently the case), but also on convergence of the BC switching procedure (to Dirichlet or Neumann type) at any given surface node. The oscillations are concerning and these conjectures need to be further investigated, but it should also be noted that this behavior is localized and does not persist, intensify, or otherwise affect the overall solution. In other words, this erratic behavior of an episodic nature does not point to numerical instabilities in the schemes used to resolve the surface and subsurface model equations and their coupling.

For the uncoupled runs, it is clear that by allocating a greater portion of the land surface to be a potential seepage face, the outflow across this BC increases while the overland and return flow rates are diminished, as is the amplitude and variation of the actual flux across the surface. The 4-node (Run 3) and 25-node (Run 4) seepage face BC configurations were chosen to approximately mimic, respectively, the Dirichlet BC case and the maximum extent of the variably saturated area that develops around the stream. A more detailed discussion of seepage face BCs in the context of coupled and uncoupled hydrological modeling can be found in Scudeler et al. (2017). For the Dirichlet BC simulation (Run 2), not visible in the hydrographs in Figure 6 is that initially, and until the water table reaches the surface at the stream location, this BC acts as a source of water, meaning that an important mass balance error is being committed since this water does not actually exist. This even causes a small amount of Horton runoff (to a maximum land surface saturation fraction of 0.1% in the first 20.4 d of the simulation). Therefore, care is needed when applying Dirichlet BCs at the land surface, as they may not serve their intended purpose (a drain or sink of water, for instance) during the entire course of a simulation.

It is worth noting that ponding (and thus also land surface saturation) develops at other points along the transect besides along the stream channel, in particular where there are small topographic dips (see Figure 2). However, the ponding levels observed are very small (<1 mm, compared to a peak of almost 2 m at the stream, as

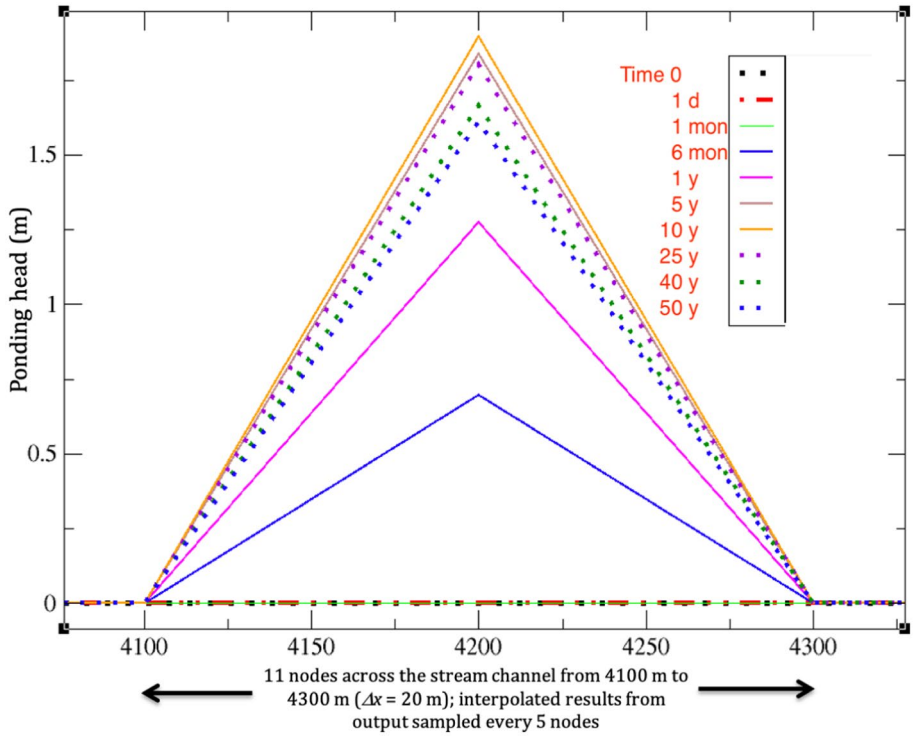


FIGURE 8
Ponding dynamics along a 200 m section across the stream channel valley, and centered at the stream node, for the coupled simulation (Run 5).

shown in Figure 8), and they drop to zero well before the end of the 50-year simulation.

3.2.2 729-year simulations with high specific storage

Very many simulations of the Tony Creek transect were run before arriving at the 50-year scenario described in the previous section. Three recurring, and to some degree surprising, outcomes in these trials were (a) that, although driven by steady and uniform atmospheric forcing and bottom leakage, it was not altogether clear that the flow system was approaching a steady state; (b) that any emergence of a steady state, and the eventual rate at which such a state was being approached, appeared to depend on the response variable being examined; and (c) that there was evidently a very strict and persistent upper limit to the time step size that could be used during the simulation. With regard to point (c) it was moreover observed that altering the value of the specific storage coefficient had a direct impact on the apparent maximum time step size that could be used during a simulation. Since in the governing subsurface flow equation for a saturated system, in both its continuous and discretized forms, S_s also has a direct impact on the system's evolution to steady state (larger S_s implies slower dynamics), it was decided to further pursue issues such as the three enumerated above by running a longer simulation with a larger value of S_s. Note that any hypothesized direct link between a system's storage coefficient, its rate of progression toward a steady state, and eventual upper constraints on time step size in a numerical discretization of the system's governing equation may be quite tenuous for a nonlinear system (e.g., unsaturated or variably saturated zone), as indeed we are dealing with here, as opposed to a fully saturated system (saturated domain).

Following the above reasoning, in this section we present the results from simulating the Tony Creek transect for a 729-year period and with a specific storage value of $S_s = 0.01\ m^{-1}$. The time series of the actual flux, return flow, overland flow, and groundwater recharge response variables are shown in Figure 9. The results are plotted on a linear time axis (top graph) as well as on a logarithmic time axis (bottom graph). The latter gives a better picture of the early-time behavior of the response variables, including the onset of rainfall–runoff partitioning and overland flow. Moreover, while in the top graph it might appear that the response variables are all asymptotically approaching a steady state value, in the bottom graph it is not at all evident, except for the groundwater recharge, that a steady state is imminent.

Another response variable, the water table level, is shown in Figure 10 and suggests as well that a steady state is still elusive after 729 years. It seems from these results that the water table will keep dropping at the highest elevations, and perhaps even in the valley around the stream node, eventually leading to the creek becoming a losing stream (i.e., the valley becoming a point of groundwater recharge) rather than a gaining stream. Much longer simulations would be needed to establish this.

Moreover, as the water table drops and the depth of the unsaturated zone increases in the upslope regions of the transect, the amount of return flow decreases (less groundwater is contributing to it), and as a result the actual flux across the land surface (see Figure 9) gradually decreases in magnitude. In fact it appears to approach zero (from negative values) and does not return to positive values, meaning

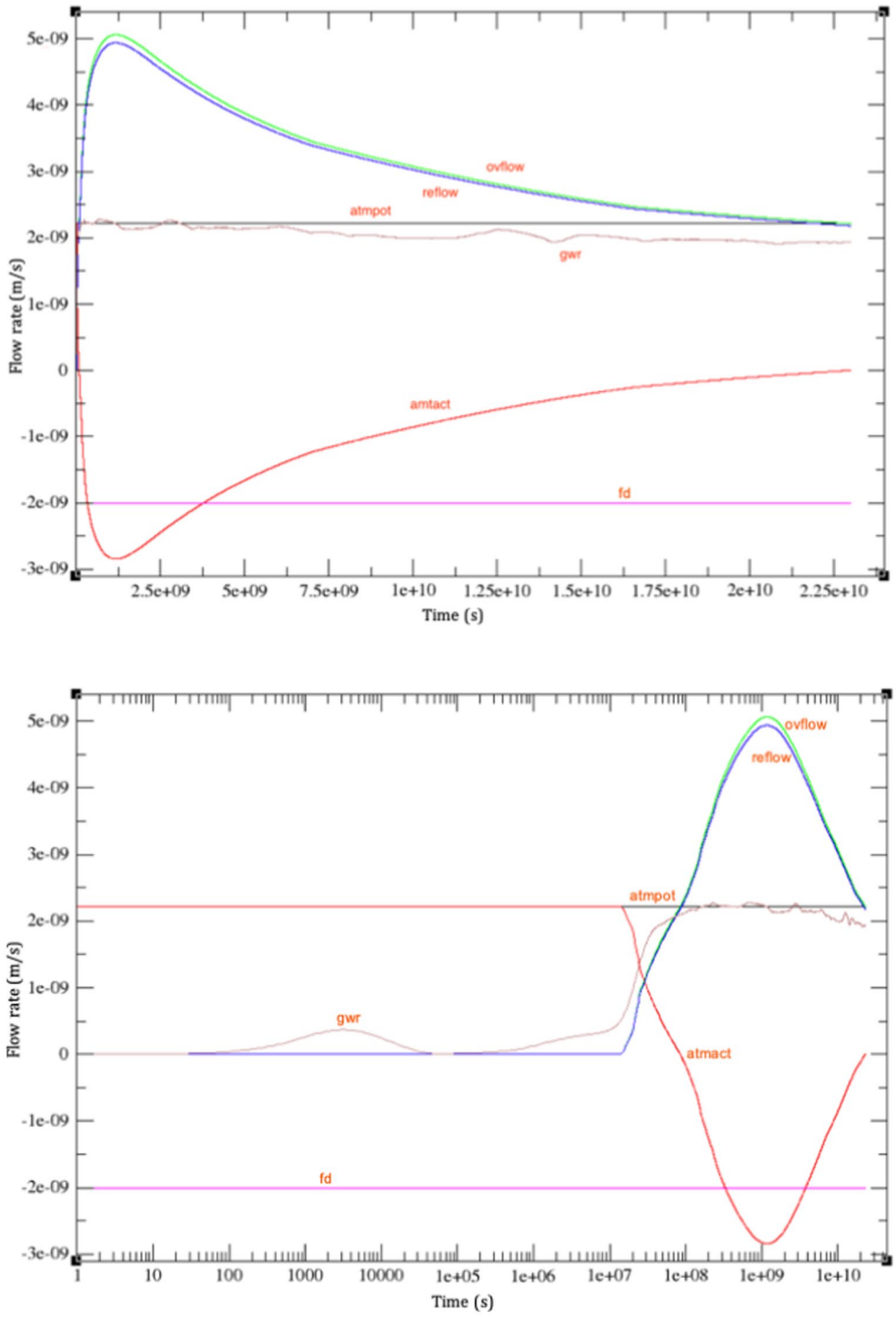


FIGURE 9
Results for a 729-year (2.3×10^{10} s) simulation of the Tony Creek transect. The run is driven by a constant rainfall (black line; labeled "atmpot") of 70 mm/y (2.22×10^{-9} m/s) and a free drainage flux of -2×10^{-9} m/s (magenta; "fd"). The response variables shown are the actual flux across the land surface (red; "atmact"), the return flow rate (blue; "reflow"), the overland flow rate (turquoise; "ovflow"), and the groundwater recharge flux (brown; "gwr"). The results are plotted on a linear time axis (top graph), as well as on a logarithmic time axis (bottom graph).

that in this later stage of the simulation, practically all of the incoming rain exits the transect (there may be some small contribution from mixing with groundwater, i.e., pre-event water, around the variable source areas). This runoff is mostly through return flow (i.e., subsurface runoff), although there is also a small component of direct runoff on the variable source areas. Therefore, except around these topographic depressions or variable source areas, the unsaturated and saturated zones become essentially decoupled.

As a general remark on the results shown in Figures 9, 10, it should be noted that the correct tracking (in a mass balance sense) of

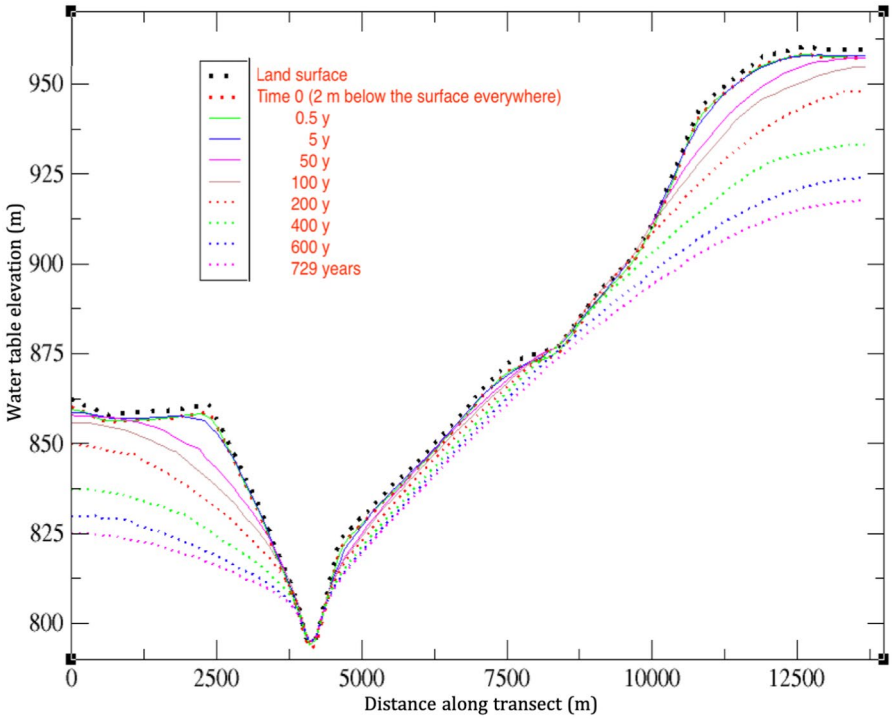


FIGURE 10
Water table dynamics for the 729-year simulation.

the large variety of incoming and outgoing fluxes, across the land surface in particular, is not a simple matter, and becomes even more complex for conventional scenarios where atmospheric forcing alternates between rainfall and evaporation. One approach for tracking the dynamics of surface–subsurface interactions is through analysis of the patterns of streamflow intermittency and of disconnected clusters of ponded water that are formed and dissipated in storm–interstorm scenarios (e.g., Ward et al., 2018; Senatore et al., 2021; Özgen-Xian et al., 2023; Zanetti et al., 2024). Another strategy that makes direct use of ISSHMs is to perform a meticulous dissection of fluxes and stores of water across each node or element face comprising the discretized land surface, adapting boundary conditions as needed (Putti and Paniconi, 2004; Sochala et al., 2009). This latter approach underscores the important and yet often neglected role that boundary conditions play in controlling surface–subsurface dynamics.

There is a strong interest in using ISSHMs as one of the many tools for estimating groundwater recharge (e.g., Waldowski et al., 2023), an important variable in water resources management. When computed in CATHY, we observe very high spatial and temporal variability in recharge (Figure 11; see also the "gwr" curve in Figure 9). Notwithstanding this variability, when taken cumulatively (over the entire transect and the entire simulation), as shown in Figure 12, the behavior of groundwater recharge is very smooth, and even quite linear. Moreover, the yearly average derived from the cumulative recharge is 64 mm (47.2 m / 729 y), which is quite consistent with several other estimates for the Tony Creek subcatchment and Fox Creek region (see footnote 1). It should be noted that there are numerous thorny issues surrounding groundwater recharge estimation, both in its conceptualization and in its calculation via numerical models and other techniques (e.g., Camporese et al. under review).[2] The recharge results presented here will thus likely be revisited in future studies.

3.2.3 Spatial and temporal discretization challenges

Much attention in subsurface and coupled surface–subsurface flow modeling is devoted to issues of spatial grid resolution and temporal discretization (e.g., Dawson et al., 2004; Sulis et al., 2011; Liggett et al., 2012; Lipnikov et al., 2016), which is understandable given the complexities of resolving such highly nonlinear and strongly coupled systems whose processes also typically evolve at widely differing characteristic scales. There is thus much scope in CATHY and other ISSHMs for improving discretization schemes, linearization and coupling algorithms, and the accuracy of groundwater velocity and other fluxes that are important derivatives of the primary variable calculations, essential for incorporating solute transport, soil–plant interactions, and other phenomena (e.g., Keyes et al., 2013; Scudeler et al., 2016b).

On this note, we will conclude the Tony Creek transect example with two illustrations of the numerical issues that can arise from spatial and temporal discretizations. As alluded to at the beginning of the previous section (point c), there was indeed an apparent maximum time step size, Δt, that could be achieved for these runs, and this limit or threshold was quite different for the two values of specific storage coefficient: a Δt on the order of 50 min was the upper limit for the run with $S_s = 8.1 \times 10^{-5}$ m^{-1},

2 Camporese, M., Paniconi, C., and Putti, M. (under review). Groundwater recharge is not the whole story: saturated storage dynamics provides a complete picture of subsurface water availability. *Water Resources Res.*

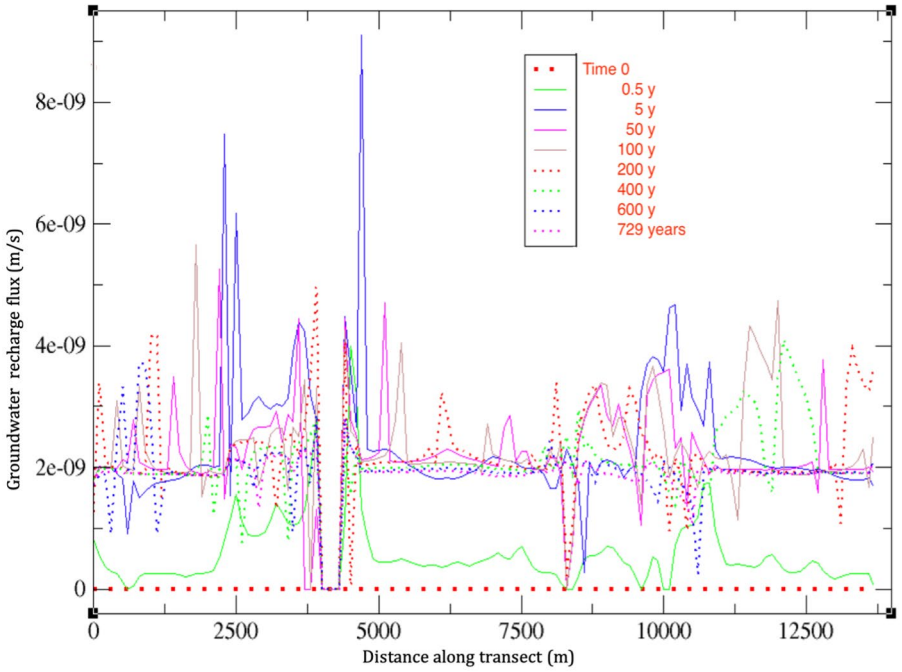


FIGURE 11
Spatial distribution of groundwater recharge flux computed by CATHY for the 729-year simulation.

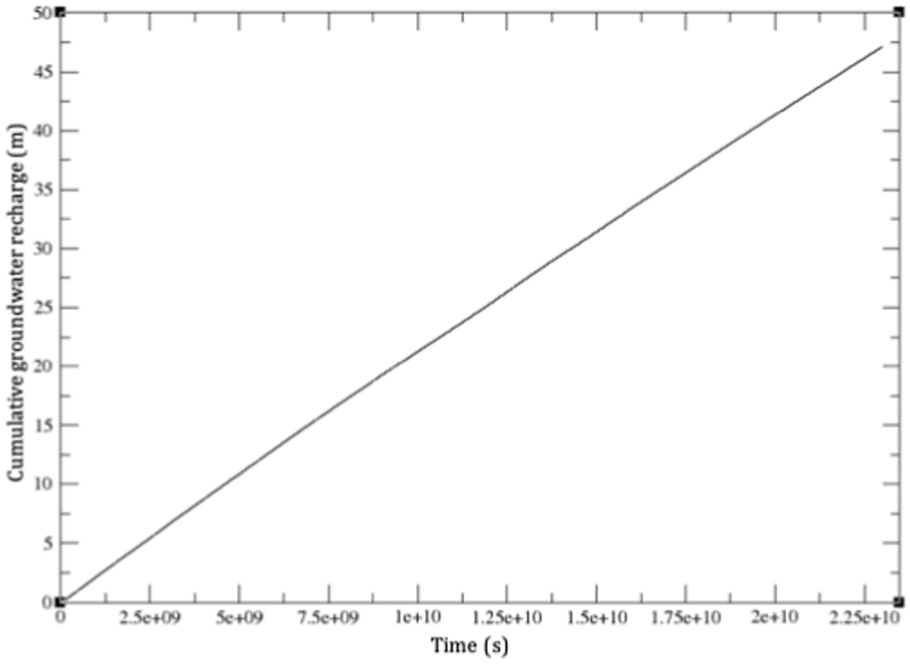


FIGURE 12
Cumulative (over space and time) groundwater recharge computed by CATHY for the 729-year simulation.

while for the $S_s = 0.01\ m^{-1}$ case it was approximately 3.5 d. Similar outcomes were obtained using other S_s values. Interestingly, the Δt and S_s ratios for these two cases are both around 100, indicating a possible linear scaling relationship.

Figure 13 shows the behavior of the adaptive time stepping scheme used in CATHY (D'Haese et al., 2007) during a very narrow time interval of one of the transect simulations. In CATHY, as well as many other Richards equation-based models, the time step size is allowed to gradually increase so long as the iterative procedure used to solve (linearize) the equation converges rapidly. When convergence difficulties are encountered, on the other hand, the solution procedure performs a "back step," wherein the time step is repeated with a much

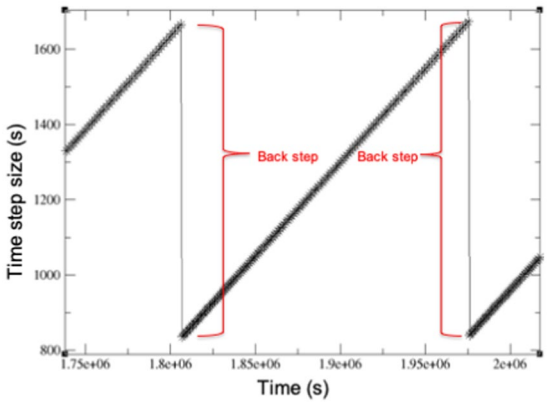


FIGURE 13
Adaptive time stepping (each "*" is a time step) with two back step occurrences shown for a short time interval extracted from a Tony Creek transect simulation.

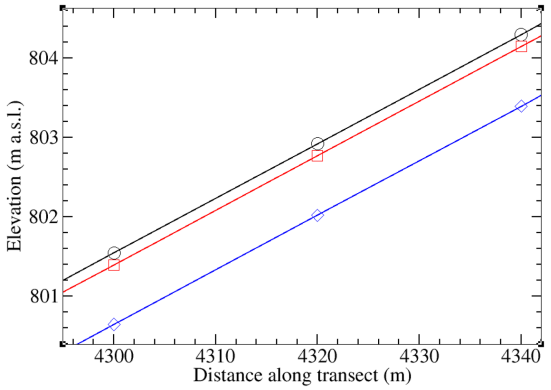


FIGURE 14
Zoom on two cells (Δx = 20 m for both) and 2 layers (Δz = 15 cm between the black and red lines that demarcate the topmost layer; Δz = 75 cm between the red and blue lines demarcating the second layer) of the Tony Creek transect discretization.

smaller Δt. The pattern shown in Figure 13 occurs repeatedly over the course of a simulation, and adds an additional constraining factor to the model's computational efficiency, on top of the more familiar control on Δt exerted by the temporal resolution and degree of variability in the atmospheric forcing inputs.

There are additional considerations on this issue that warrant attention. When variably saturated conditions prevail, as in the cases examined here, where the water table drops significantly over the course of the simulation (see Figure 10), S_s is not the only storage parameter involved. The general storage term in variably saturated flow models based on Richards' equation contains also the specific soil moisture capacity ($d\theta/d\psi$, where θ is moisture content and ψ is pressure head) and the porosity, and these parameters would presumably also affect time step behavior. Furthermore, as mentioned earlier, the intuitive connection between larger S_s and slower dynamics is predicated on a linear equation (saturated groundwater flow); the behavior under (strongly) nonlinear conditions (e.g., a continually dropping water table) is not as intuitive. Heterogeneity, not addressed in the tests presented here (see Gauthier et al., 2009 for a CATHY-related example), would likely complicate matters further if, in addition to hydraulic conductivity, storage parameters are also spatially variable.

With regard to spatial discretization, it is generally advisable, when dealing with 2D and 3D domains, to avoid highly distorted elements within a computational grid, as these can lead to solver convergence problems and solution inaccuracies. In integrated surface water–groundwater modeling, mesh skewness typically arises in discretizing steep terrain, but it can occur also over relatively flat terrain where distortion is accentuated by very high aspect ratios between the horizontal and vertical discretizations, especially when applying ISSHMs at large spatial scales. This is illustrated in Figure 14 for the Tony Creek transect. As mentioned in Section 2.2, the average terrain slope on each side of the Tony Creek transect is quite mild (~1.7%). Even locally, the steepest slope, along the right streambank, reaches only 6.8%. (This is in contrast, for instance, to the much steeper terrain gradient for the LEO hillslope, of 17.6%.) However, the 20 m horizontal grid size used for the Tony Creek transect, combined with a 15 cm vertical grid size for the critical topmost layer, results in an aspect ratio $\Delta x/\Delta z$ of 133. For an average slope of 1.7%, this gives a vertical drop of 34 cm over a 20 m lateral distance, which is more than double the layer thickness. Where the terrain is locally steeper the drop is much greater, as shown in Figure 14.

In the simulations presented for this study, we did not conduct a detailed analysis of aspect ratio impacts, to the extent for example of attributing specific numerical problems observed to mesh distortion versus other factors, but it is certainly an issue that warrants further attention in the context of integrated hydrological modeling. An aspect ratio of 133 was not considered unreasonable in view of a prior study (Paniconi and Wood, 1993) where satisfactory results were obtained with aspect ratios as high as 150, provided the convergence criterion on the iterative solver for Richards' equation was not too stringent. Calver and Wood (1989), on the other hand, recommended stricter limits on aspect ratio ($\Delta x/\Delta z \leq 20$) for subsurface flow models. There can be significant trade-offs between keeping computational costs low (large aspect ratios) and maintaining accurate numerical solutions (small aspect ratios), as investigated for instance by Badrot-Nico et al. (2007) for the advection–diffusion equation. In addition to imposing limits on the spatial grid, the fine vertical grid spacing typically used for near-surface layers in order to resolve flux partitioning also restricts the temporal discretization, adding another facet to the accuracy–cost trade-off. There is a need to extend analysis of these issues to ISSHMs, as coupling may introduce additional factors that constrain mesh aspect ratios. In the specific case of CATHY and other DEM-based ISSHMs, another limitation arises from the use of a nonuniform surface grid. Mesh refinement where the terrain is steeper would allow for smaller aspect ratios without unduly amplifying computational costs. This is an important research topic for this class of ISSHMs.

4 Conclusion

Through the small-scale hillslope and large-scale transect examples examined in this paper, we have highlighted some numerical issues that can occur in integrated hydrological modeling, and we have illustrated some of the unexpected or anomalous responses that can arise in simulating such strongly coupled, nonlinear systems. In the LEO hillslope test case, we showed that although qualitatively similar results can be obtained in the capture of flux partitioning across the land surface

boundary between coupled and uncoupled runs of the model, the fine details (for, e.g., timing and duration of overland flow) can be quite different. Moreover, there are phenomena that cannot be reproduced without properly resolving surface–subsurface interactions, such as enhanced ponding due to the interplay between the infiltration excess and saturation excess mechanisms of runoff generation. A general point that was emphasized through both the hillslope and transect examples is that correct tracking, including in a mass balance sense, of surface water–groundwater interactions is very challenging, and there is much scope for future research on this central constituent of ISSHMs.

In the Tony Creek transect test case, we showed that, in subsurface-only mode, similar responses for key fluxes at the land surface are reproduced in simulations with and without "guideposts" such as seepage face or Dirichlet boundary conditions placed at or along the stream. However, the responses obtained in coupled mode are significantly different, illustrating the important contribution made by interactions across the land surface that can occur in a continuous or recycled manner, such as downslope reinfiltration. Moreover, it was pointed out that care must be taken when assigning Dirichlet BCs at the surface, in order to avoid unintended (non-physical) sources or sinks of water. In further simulations of the Tony Creek transect, we explored the nature of steady state flow under the flow conditions that prevailed in these runs, featuring strong coupling (exemplified for instance by persistent land surface saturation and high levels of localized ponding) and an expanding unsaturated zone (exemplified by a continually dropping water table). It was observed that if any semblance of a steady state did emerge in this nonlinear flow system, it was in any case highly dependent on the response variable being examined. In these steady state trials, it was moreover found that there was a sustained upper limit to the allowable time step size for any run, and that this upper limit depended strongly on the value of the specific storage coefficient parameter. Thus time stepping is constrained not just by factors such as the temporal resolution and variability of atmospheric (and any other) forcing terms and the strength of coupling and nonlinearity, but also by parameters controlling the internal dynamics of the system such as storage coefficients. Finally, in the transect simulations we suggested that grid aspect ratio issues can arise not just in modeling very steep terrain, but also gently sloping terrain, in particular for large-scale applications.

In addition to challenges associated with complex boundary conditions, mesh irregularities (skewness and aspect ratio), and the resolution and tracking of surface water–groundwater interactions, attention to improving linearization schemes (variants and adaptations of standard Picard and Newton iteration) and refining the substepping scheme in the surface routing module (for, e.g., using more flexible and adaptive criteria linked to ponding levels) are required to ensure robust and accurate results from ISSHM simulations. For models like CATHY, further theoretical development and testing of the boundary condition switching algorithm is needed, and the idea of adding another level of nesting to the model, namely iterative coupling between the surface and subsurface solvers, should be explored.

Overall, it is hoped that the results from the test cases presented in this study have provided some guidelines for further improvements to the coupling and discretization schemes used in ISSHMs. There is no doubt moreover that new numerical issues will emerge as physics-based integrated hydrological models continue to evolve and expand, to more fully include vegetation processes (e.g., Manoli et al., 2014; Brunetti et al., 2019) and ecohydrology in a broad sense (e.g., Niu et al., 2014b; Guswa et al., 2020), and as these models are pushed to ever larger scale applications (Lemieux et al., 2008; Ala-Aho et al., 2015).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CP: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. CL: Conceptualization, Investigation, Writing – review & editing. CR: Conceptualization, Investigation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ala-Aho, P., Rossi, P. M., Isokangas, E., and Kløve, B. (2015). Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: model verification with stable isotopes and airborne thermal imaging. *J. Hydrol.* 522, 391–406. doi: 10.1016/j.jhydrol.2014.12.054

Badrot-Nico, F., Brissaud, F., and Guinot, V. (2007). A finite volume upwind scheme for the solution of the linear advection–diffusion equation with sharp gradients in multiple dimensions. *Adv. Water Resour.* 30, 2002–2025. doi: 10.1016/j.advwatres.2007.04.003

Brunetti, G., Kodešová, R., and Šimůnek, J. (2019). Modeling the translocation and transformation of chemicals in the soil-plant continuum: a dynamic plant uptake module for the HYDRUS model. *Water Resour. Res.* 55, 8967–8989. doi: 10.1029/2019WR025432

Calver, A., and Wood, W. L. (1989). On the discretization and cost-effectiveness of a finite element solution for hillslope subsurface flow. *J. Hydrol.* 110, 165–179. doi: 10.1016/0022-1694(89)90242-4

Camporese, M., Paniconi, C., Putti, M., and Orlandini, S. (2010). Surface–subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resour. Res.* 46:W02512. doi: 10.1029/2008WR007536

D'Haese, C. M. F., Putti, M., Paniconi, C., and Verhoest, N. E. C. (2007). Assessment of adaptive and heuristic time stepping for variably saturated flow. *Int. J. Numer. Methods Fluids* 53, 1173–1193. doi: 10.1002/fld.1369

Dagès, C., Paniconi, C., and Sulis, M. (2012). Analysis of coupling errors in a physically-based integrated surface water–groundwater model. *Adv. Water Resour.* 49, 86–96. doi: 10.1016/j.advwatres.2012.07.019

Dawson, C., Sun, S., and Wheeler, M. F. (2004). Compatible algorithms for coupled flow and transport. *Comput. Methods Appl. Mech. Eng.* 193, 2565–2580. doi: 10.1016/j.cma.2003.12.059

Discacciati, M., Miglio, E., and Quarteroni, A. (2002). Mathematical and numerical models for coupling surface and groundwater flows. *Appl. Numer. Math.* 43, 57–74. doi: 10.1016/S0168-9274(02)00125-3

Farthing, M. W., and Ogden, F. L. (2017). Numerical solution of Richards' equation: a review of advances and challenges. *Soil Sci. Soc. Am. J.* 81, 1257–1269. doi: 10.2136/sssaj2017.02.0058

Fiorentini, M., Orlandini, S., and Paniconi, C. (2015). Control of coupling mass balance error in a process-based numerical model of surface–subsurface flow interaction. *Water Resour. Res.* 51, 5698–5716. doi: 10.1002/2014WR016816

Furman, A. (2008). Modeling coupled surface–subsurface flow processes: a review. *Vadose Zone J.* 7, 741–756. doi: 10.2136/vzj2007.0065

Gatel, L., Lauvernet, C., Carluer, N., Weill, S., and Paniconi, C. (2020). Sobol global sensitivity analysis of a coupled surface/subsurface water flow and reactive solute transfer model on a real hillslope. *Water* 12:121. doi: 10.3390/w12010121

Gatto, B., Paniconi, C., Salandin, P., and Camporese, M. (2021). Numerical dispersion of solute transport in an integrated surface–subsurface hydrological model. *Adv. Water Resour.* 158:104060. doi: 10.1016/j.advwatres.2021.104060

Gauthier, M.-J., Camporese, M., Rivard, C., Paniconi, C., and Larocque, M. (2009). A modeling study of heterogeneity and surface water–groundwater interactions in the Thomas brook catchment, Annapolis Valley (Nova Scotia, Canada). *Hydrol. Earth Syst. Sci.* 13, 1583–1596. doi: 10.5194/hess-13-1583-2009

Guswa, A. J., Tetzlaff, D., Selker, J. S., Carlyle-Moses, D. E., Boyer, E. W., Bruen, M., et al. (2020). Advancing ecohydrology in the 21st century: a convergence of opportunities. *Ecohydrology* 13:e2208. doi: 10.1002/eco.2208

Haque, A., Salama, A., Lo, K., and Wu, P. (2021). Surface and groundwater interactions: a review of coupling strategies in detailed domain models. *Hydrology* 8:35. doi: 10.3390/hydrology8010035

Irvine, D. J., Singha, K., Kurylyk, B. L., Briggs, M. A., Sebastian, Y., Tait, D. R., et al. (2024). Groundwater-surface water interactions research: past trends and future directions. *J. Hydrol.* 644:132061. doi: 10.1016/j.jhydrol.2024.132061

Kelley, C. T. (1995). Iterative methods for linear and nonlinear equations. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Keyes, D. E., McInnes, L. C., Woodward, C., Gropp, W., Myra, E., Pernice, M., et al. (2013). Multiphysics simulations: challenges and opportunities. *Int. J. High Perf. Comput. Appl.* 27, 4–83. doi: 10.1177/1094342012468181

Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., et al. (2017). The integrated hydrologic model intercomparison project, IH-MIP2: a second set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resour. Res.* 53, 867–890. doi: 10.1002/2016WR019191

Lemieux, J. M., Sudicky, E. A., Peltier, W. R., and Tarasov, L. (2008). Dynamics of groundwater recharge and seepage over the Canadian landscape during the Wisconsinian glaciation. *J. Geophys. Res.* 113:F01011. doi: 10.1029/2007JF000838

Liggett, J. E., Werner, A. D., and Simmons, C. T. (2012). Influence of the first-order exchange coefficient on simulation of coupled surface–subsurface flow. *J. Hydrol.* 415, 503–515. doi: 10.1016/j.jhydrol.2011.11.028

Lipnikov, K., Konstantin, L., David, M., and Daniil, S. (2016). New preconditioning strategy for Jacobian-free solvers for variably saturated flows with Richards' equation. *Adv. Water Resour.* 94, 11–22. doi: 10.1016/j.advwatres.2016.04.016

List, F., and Radu, F. A. (2016). A study on iterative methods for solving Richards' equation. *Comput. Geosci.* 20, 341–353. doi: 10.1007/s10596-016-9566-3

Manoli, G., Bonetti, S., Domec, J.-C., Putti, M., Katul, G., and Marani, M. (2014). Tree root systems competing for soil moisture in a 3D soil–plant model. *Adv. Water Resour.* 66, 32–42. doi: 10.1016/j.advwatres.2014.01.006

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., et al. (2014). Surface–subsurface model intercomparison: a first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resour. Res.* 50, 1531–1549. doi: 10.1002/2013WR013725

Miller, C. T., Dawson, C. N., Farthing, M. W., Hou, T. Y., Huang, J., Kees, C. E., et al. (2013). Numerical simulation of water resources problems: models, methods, and trends. *Adv. Water Resour.* 51, 405–437. doi: 10.1016/j.advwatres.2012.05.008

Niu, G.-Y., Paniconi, C., Troch, P. A., Scott, R. L., Durcik, M., Zeng, X., et al. (2014b). An integrated modelling framework of catchment-scale ecohydrological processes: 1. Model description and tests over an energy-limited watershed. *Ecohydrology* 7, 427–439. doi: 10.1002/eco.1362

Niu, G.-Y., Pasetto, D., Scudeler, C., Paniconi, C., Putti, M., Troch, P. A., et al. (2014a). Incipient subsurface heterogeneity and its effect on overland flow generation – insight from a modeling study of the first experiment at the biosphere 2 landscape evolution observatory. *Hydrol. Earth Syst. Sci.* 18, 1873–1883. doi: 10.5194/hess-18-1873-2014

Omar, P. J., Shivhare, N., Dwivedi, S. B., Gaur, S., and Dikshit, P. K. S. (2021). "Study of methods available for groundwater and surface water interaction: a case study on Varanasi, India" in The Ganga River basin: A Hydrometeorological approach. eds. M. S. Chauhan and C. S. P. Ojha (Cham: Springer), 67–83.

Orlandini, S., and Moretti, G. (2009). Determination of surface flow paths from gridded elevation data. *Water Resour. Res.* 45:W03417. doi: 10.1029/2008WR007099

Orlandini, S., Moretti, G., Franchini, M., Aldighieri, B., and Testa, B. (2003). Path-based methods for the determination of nondispersive drainage directions in grid-based digital elevation models. *Water Resour. Res.* 39:1144. doi: 10.1029/2002WR001639

Özgen-Xian, I., Molins, S., Johnson, R. M., Xu, Z., Dwivedi, D., Loritz, R., et al. (2023). Understanding the hydrological response of a headwater-dominated catchment by analysis of distributed surface–subsurface interactions. *Sci. Rep.* 13:4669. doi: 10.1038/s41598-023-31925-w

Paniconi, C., Aldama, A. A., and Wood, E. F. (1991). Numerical evaluation of iterative and noniterative methods for the solution of the nonlinear Richards equation. *Water Resour. Res.* 27, 1147–1163. doi: 10.1029/91WR00334

Paniconi, C., and Putti, M. (2015). Physically based modeling in catchment hydrology at 50: survey and outlook. *Water Resour. Res.* 51, 7090–7129. doi: 10.1002/2015WR017780

Paniconi, C., and Wood, E. F. (1993). A detailed model for simulation of catchment scale subsurface hydrologic processes. *Water Resour. Res.* 29, 1601–1620. doi: 10.1029/92WR02333

Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., and Troch, P. A. (2015). Impact of sensor failure on the observability of flow dynamics at the biosphere 2 LEO hillslopes. *Adv. Water Resour.* 86, 327–339. doi: 10.1016/j.advwatres.2015.04.014

Putti, M., and Paniconi, C. (2004). Time step and stability control for a coupled model of surface and subsurface flow. *Dev. Water Sci.* 55, 1391–1402. doi: 10.1016/S0167-5648(04)80152-7

Sanchez-Vila, X., Guadagnini, A., and Carrera, J. (2006). Representative hydraulic conductivities in saturated groundwater flow. *Rev. Geophys.* 44:RG3002. doi: 10.1029/2005RG000169

Scudeler, C., Pangle, L., Pasetto, D., Niu, G.-Y., Volkmann, T., Paniconi, C., et al. (2016a). Multiresponse modeling of variably saturated flow and isotope tracer transport for a hillslope experiment at the landscape evolution observatory. *Hydrol. Earth Syst. Sci.* 20, 4061–4078. doi: 10.5194/hess-20-4061-2016

Scudeler, C., Paniconi, C., Pasetto, D., and Putti, M. (2017). Examination of the seepage face boundary condition in subsurface and coupled surface/subsurface hydrological models. *Water Resour. Res.* 53, 1799–1819. doi: 10.1002/2016WR019277

Scudeler, C., Putti, M., and Paniconi, C. (2016b). Mass-conservative reconstruction of Galerkin velocity fields for transport simulations. *Adv. Water Resour.* 94, 470–485. doi: 10.1016/j.advwatres.2016.06.011

Senatore, A., Micieli, M., Liotti, A., Durighetto, N., Mendicino, G., and Botter, G. (2021). Monitoring and modeling drainage network contraction and dry down in Mediterranean headwater catchments. *Water Resour. Res.* 57:e2020WR028741. doi: 10.1029/2020WR028741

Sochala, P., Ern, A., and Piperno, S. (2009). Mass conservative BDF-discontinuous Galerkin/explicit finite volume schemes for coupling subsurface and overland flows. *Comput. Methods Appl. Mech. Eng.* 198, 2122–2136. doi: 10.1016/j.cma.2009.02.024

Stokke, J. S., Mitra, K., Storvik, E., Both, J. W., and Radu, F. A. (2023). An adaptive solution strategy for Richards' equation. *Comput. Mathematics Appl.* 152, 155–167. doi: 10.1016/j.camwa.2023.10.020

Sulis, M., Meyerhoff, S. B., Paniconi, C., Maxwell, R. M., Putti, M., and Kollet, S. J. (2010). A comparison of two physics-based numerical models for simulating surface water–groundwater interactions. *Adv. Water Resour.* 33, 456–467. doi: 10.1016/j.advwatres.2010.01.010

Sulis, M., Paniconi, C., and Camporese, M. (2011). Impact of grid resolution on the integrated and distributed response of a coupled surface–subsurface hydrological model for the des Anglais catchment, Quebec. *Hydrol. Proc.* 25, 1853–1865. doi: 10.1002/hyp.7941

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44, 892–898. doi: 10.2136/sssaj1980.03615995004400050002x

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., et al. (2016). Modeling soil processes: review, key challenges, and new perspectives. *Vadose Zone J.* 15, 1–57. doi: 10.2136/vzj2015.09.0131

Waldowski, B., Sánchez-León, E., Cirpka, O. A., Brandhorst, N., Hendricks Franssen, H.-J., and Neuweiler, I. (2023). Estimating groundwater recharge in fully integrated pde-based hydrological models. *Water Resour. Res.* 59:e2022WR032430. doi: 10.1029/2022WR032430

Ward, A. S., Schmadel, N. M., and Wondzell, S. M. (2018). Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Adv. Water Resour.* 114, 64–82. doi: 10.1016/j.advwatres.2018.01.018

Weill, S., Mazzia, A., Putti, M., and Paniconi, C. (2011). Coupling water flow and solute transport into a physically-based surface–subsurface hydrological model. *Adv. Water Resour.* 34, 128–136. doi: 10.1016/j.advwatres.2010.10.001

Zanetti, F., Camporese, M., and Botter, G. (2024). How do different runoff generation mechanisms drive stream network dynamics? Insights from physics-based modelling. *Hydrol. Proc.* 38:e15234. doi: 10.1002/hyp.15234

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 05 June 2025
DOI 10.3389/frwa.2025.1571704

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Michael Tso,
UK Centre for Ecology and Hydrology (UKCEH), United Kingdom
Maria Staudinger,
University of Zurich, Switzerland

*CORRESPONDENCE
Klaus Goergen
k.goergen@fz-juelich.de

RECEIVED 05 February 2025
ACCEPTED 19 May 2025
PUBLISHED 05 June 2025

CITATION
Goergen K, Belleflamme A, Hammoudeh S, Vanderborght J and Kollet S (2025) The July 2021 flood event in the Eifel-Ardennes mountains as simulated by the high-resolution integrated hydrologic model ParFlow.
Front. Water 7:1571704.
doi: 10.3389/frwa.2025.1571704

COPYRIGHT
© 2025 Goergen, Belleflamme, Hammoudeh, Vanderborght and Kollet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The July 2021 flood event in the Eifel-Ardennes mountains as simulated by the high-resolution integrated hydrologic model ParFlow

Klaus Goergen[1,2]*, Alexandre Belleflamme[1,2], Suad Hammoudeh[1,2], Jan Vanderborght[1,2,3] and Stefan Kollet[1,2]

[1]Institute of Bio- and Geosciences (IBG-3, Agrosphere), Forschungszentrum Jülich, Jülich, Germany, [2]Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Jülich, Germany, [3]Department of Earth and Environmental Sciences, Soil and Water Management, KU Leuven, Leuven, Belgium

In mid-July 2021, a quasi-stationary extratropical cyclone over parts of western Germany and eastern Belgium led to unprecedented sustained widespread precipitation, nearly doubling climatological monthly rainfall amounts in less than 72 h. This resulted in extreme flooding in many of the Eifel-Ardennes low mountain range river catchments with loss of lives, and substantial damage and destruction. Despite many reconstructions of the event, open issues on the underlying physical mechanisms remain. In a numerical laboratory approach based on a 52-member spatially and temporally consistent high-resolution hindcast reconstruction of the event with the integrated hydrological surface-subsurface model ParFlow, this study shows the prognostic capabilities of ParFlow and further explores the physical mechanisms of the event. Within the range of the ensemble, ParFlow simulations can reproduce the timing and the order of magnitude of the flood event without additional calibration or tuning. What stands out is the large and effective buffer capacity of the soil. In the simulations, the upper soil in the highly affected Ahr, Erft, and Kyll river catchments are able to buffer between about one third to half of the precipitation that does not contribute immediately to the streamflow response and leading eventually to widespread, very high soil moisture saturation levels. In case of the Vesdre river catchment, due to its initially higher soil water saturation levels, the buffering capacity is lower; hence more precipitation is transferred into discharge.

KEYWORDS

integrated hydrological model, ParFlow, July 2021 flood, Eifel-Ardennes, terrestrial water cycle

1 Introduction

During 14 and 15 July 2021, extreme flooding affected parts of western Germany and eastern Belgium, as well as parts of Luxembourg and the Netherlands (Cornwall, 2021; Davies, 2021a; Davies, 2021b; Mohr et al., 2023). In Germany and Belgium, the event caused more than 200 fatalities (EM-DAT, 2021). Widespread substantial damage and destruction of infrastructure and properties occurred, with aggregated losses of about US$ 58bn associated with the event (Munich Re, 2025). For Europe, these are extreme numbers associated with flooding in recent history (Paprotny et al., 2018).

An area most extensively affected was the Eifel-Ardennes low mountain range and parts of the Lower Rhine Embayment in western Germany, with Federal States North Rhine-Westphalia and Rhineland-Palatinate, and eastern Belgium, including some of the Meuse River tributaries (Lehmkuhl et al., 2022; Mohr et al., 2023) (Figure 1).

Several studies investigated the 2021 flood event, e.g., in the realm of meteorological conditions, precipitation-runoff prediction, precipitation uncertainty or counterfactuals. Synoptic conditions, the hydrologic event description, including collected observations, the hydro-morphodynamics, the valley morphology, and changes thereof due to erosion, sediment and debris transport, that jointly contributed to the flood generation and evolution, are described, analyzed, and summarized comprehensively in the companion papers of Mohr et al. (2023) and Ludwig et al. (2023).

The first order driver of the extreme floods was widespread, persistent, and intense rain, associated with a quasi-stationary, blocked extratropical cyclone with an occluded front, orographic lifting at the Eifel-Ardennes low mountain range, and unstable moisture-laden air masses. These synoptic conditions were analyzed extensively, e.g., in reports by Junghänel et al. (2021) or Schäfer et al. (2021) with a focus on Germany and by Journée et al. (2023) for Belgium. In an extensive hindsight analysis of the event, the July heavy precipitation event was ranked by Ludwig et al. (2023) as one of the five heaviest precipitation events in Germany within the past 70 years. Figure 1 shows the observed 72 h precipitation sums preceding and during the built-up of the main flood events from 12 to 15 July based on the RADOLAN radar precipitation product (Weigl and Winterrath, 2009; DWD, 2025). Widespread precipitation totals exceeding 100 mm 72 h^{-1}, locally up to 150 mm 72 h^{-1}, and more than 75 mm 24 h^{-1} during 14 July, with local amounts up to 100 mm 12 h^{-1} during the afternoon of 14 July affected most of the drainage and river networks of the Eifel-Ardennes area north of the Moselle river, west of the Rhine river and east of the Meuse river on 14 and 15 July during the main flood events. Strongly impacted regions were in the catchments of the rivers Ahr, Erft, Kyll, Vesdre (see the catchments highlighted in Figure 1). Whether the extreme July 2021 precipitation can be attributed to climate change remains inconclusive in a study by Tradowsky et al. (2023); there is, however, an increased likelihood and severity of extreme precipitation events for the larger Western European region in a future climate.

Hydrologic observations of the event are either incomplete, as most gaging stations were destroyed, or existing rating curves cannot be applied to the stage observations that exceeded high water marks (McMillan et al., 2012; Roggenkamp et al., 2024). Mohr et al. (2023) contains hydrograph and water level reconstructions for some gages. Based on field surveys Roggenkamp et al. (2024) derive a peak discharge between 1,000 m^3 s^{-1} and 1,250 m^3 s^{-1} for the Ahr river at gage Dernau, which is five times larger than the previous peak discharge on record.

The 2021 event is usually characterized as a pluvially triggered flash flood (e.g., Lehmkuhl et al., 2022). A high antecedent precipitation, as an important preconditioning of the subsurface water storage, in combination with shallow soils, in many cases steep valley slopes, high gradient valleys, and heavy, widespread, persistent precipitation, that supposedly caused infiltration as well as saturation excess overland flow, are often considered as the root causes for the flood, with a further intensification and modulation through morphodynamic changes and debris transport in the river valleys (Lehmkuhl et al., 2022; Mohr et al., 2023; Roggenkamp et al., 2024). Another common explanation in this context, with a focus on soil water retention, assumes that shallow soils on top of impermeable bedrock have high infiltration rates, especially in forested areas through macropores, but only little buffering capacity. Once soils are saturated, precipitation water propagates to the streams through fast interflow along preferential flowpaths on top of impermeable bedrock, with little overland flow outside the riverbeds.

In Ludwig et al. (2023) the operational flood forecasting model LARSIM is able to reproduce the reconstructed Ahr flood discharge for a selected gage, based on observed precipitation and a storyline approach with pseudo global warming atmospheric simulations. Also, the lumped conceptual GR4H model and the integrated surface-subsurface hydrological model (ISSHM) ParFlow in the DE06 setup and configuration, as described below, can realistically reproduce some hydrographs, based on improved radar precipitation products and adjusted Manning's roughness coefficients (Saadi et al., 2023). However, given the complexity and the morphodynamic changes during the event, streamflow parametrization tuning seems unfeasible. Also, the hydrodynamics of the event could be simulated for the lower Ahr river downstream of Altenahr at spatial resolutions down to 1 m with the RIM2D and SERGHEI models by Khosh Bin Ghomash et al. (2024). In a counterfactual precipitation scenarios study, Voit and Heistermann (2024) can generate peak flows, which are on average up to a factor 2 higher than the reconstructed discharge.

Still knowledge gaps exist, and debates continue on the physical mechanisms that generated the 2021 flood event. For example, many of the analyses see preferential flow and surface runoff as drivers of the rapid streamflow response. What to our knowledge has not been considered so far is that a rapid subsurface stormflow in hillslopes in the (shallow) vadose zone and saturated zone can be triggered by an intense rainfall event, which introduces a significant amount of water into the subsurface. This can produce a pressure wave. The celerity, i.e., the speed at which this pressure wave propagates, is much faster than the Darcy flux velocity. Such a pressure wave can lead to rapid subsurface hydrologic response processes, e.g., groundwater flooding in the river valley convergence zones with shallow water tables, increased river baseflow, or saturation excess overland flow (Rasmussen et al., 2000; McDonnell and Beven, 2014; Zhao et al., 2017). We hypothesize that this might also be one of the drivers for the magnitude and the rapid rise of the hydrographs during the 2021 event.

In this context, we reconstruct the event with a 52-member ensemble of the ISSHM ParFlow (Kollet and Maxwell, 2006, 2008; Kuffour et al., 2020), including the Common Land Model. ParFlow is run at a spatial resolution of 611 m with 15 vertical subsurface model

Abbreviations: ASTER, Advanced Spaceborne Thermal Emission and Reflection Radiometer; CLC, Corine Land Cover; DE06, ParFlow Setup for Hydrological Germany at 611 m Horizontal Grid Spacing; GDEM, Global Digital Dlevation Model; GRDC, Global Runoff Data Center; IGBP, International Geosphere-Biosphere Programme; ISSHM, Integrated Surface-Subsurface Hydrologic Model; IHME1500, International Hydrogeological Map of Europe 1:1,500,000; LARSIM, Large Area Runoff Simulation Model; MERIT, Multi-Error-Removed Improved-Terrain; PFT, Plant Functional Types; RADOLAN, Radar-Online-Aneichung; RADKLIM, Radarklimatologie; USDA, U. S. Department of Agriculture.

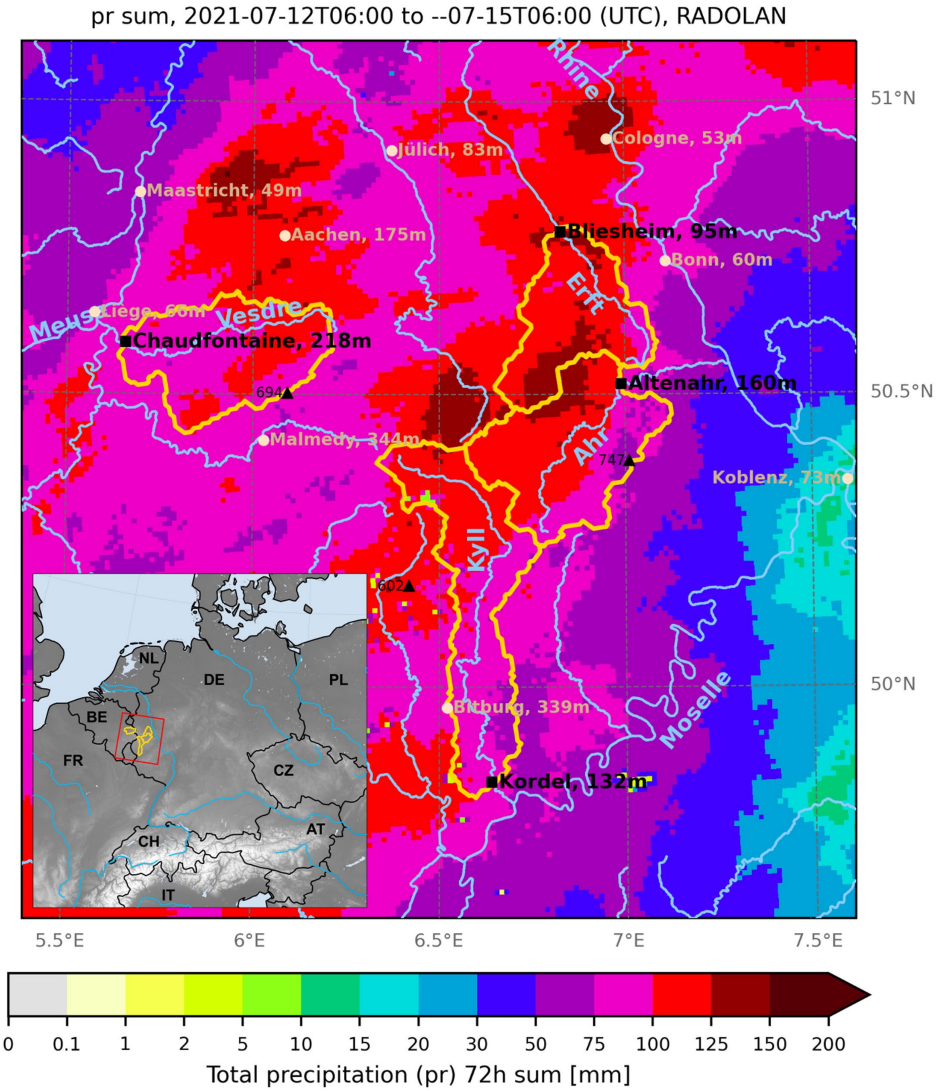


FIGURE 1
Geographic overview of the study region and 72 h accumulated hourly precipitation from Monday, 2021-07-12 06UTC, to Thursday, 2021-07-15 06UTC [mm 72 h^{-1}]; please note the non-linear colorbar scale. The Eifel-Ardennes low mountain range lies between the rivers Moselle, Rhine, and Meuse; the triangle Jülich, Bliesheim, Cologne roughly borders the Lower Rhine Embayment. Yellow: ParFlow model catchment boundaries as used in the analysis. Black squares: gaging stations under consideration with place names. Bright blue: major rivers and their tributaries (those rivers under special consideration are labeled: Ahr, Erft, Kyll, Vesdre). Off white dots: city names for improved orientation. Low, interspersed precipitation values are artifacts in the radar-based data. Numbers behind place names give the mean altitude of the place or gage. The inset shows the topography of the complete 2,000 × 2,000 grid point ParFlow hydrologic model DE06 domain with the study region (red) and analysis catchments (yellow) and country borders (black). Data sources: DWD RADOLAN RW (precipitation) (Weigl and Winterrath, 2009; DWD, 2025); River and Catchment Characterization and Model database v2.1 (river network) (de Jager and Vogt, 2003).

levels down to 60 m depth for a mid-European model domain and provides a spatially and temporally consistent high-resolution reconstruction of the flood event. These ParFlow simulations allow studying the flood and ponding height generating processes in a numerical laboratory approach complementing existing and ongoing discussions and analyses of the 2021 flood event.

Physics-based ISSHMs such as ParFlow, where 2D/3D surface and subsurface hydrodynamics are treated in continuum approach allow for an explicit simulation of transport processes and feedbacks, toward more realistic process and terrestrial water cycle representations. ISSHMs show, e.g., added value due to redistribution of surface and groundwater, streamflow aquifer interactions, ponding and flowing water in river convergence zones, evolution of river networks, km-scale heterogeneity, hill-slope processes, including human interventions, etc. (Clark et al., 2017; Kollet et al., 2017; Brookfield et al., 2023).

Based on the uncalibrated physics-based ISSHM ParFlow, through a process-based analysis, the goals of this study are: (i) To demonstrate the prognostic capabilities of the ParFlow/CLM DE06 ensemble—originally set up for water resources applications—to show whether the dynamics and magnitude of the July 2021 flood event are captured; and (ii) to further explore the physical mechanisms that generated the horrendous flood wave and water ponding heights of more than 7 m on the valley floor over extremely short time scales of only a few hours. However, as the modeling approach is limited in the detailed

representation of the complex small-scale topography and hydrogeology of the catchments, does neither resolve nor include any morphodynamic changes during the event, and cannot consider impacts due to infrastructures such as dams or obstruction at bridges, an exact reproduction of the reconstructed events is neither intended nor expected. We also do not touch upon any forecast aspect backed by model skill scores, despite the fact that the ParFlow ensemble we use is based on a forecast simulation. While we focus our analyses on four strongly affected catchments of the rivers Ahr, Erft, Kyll, Vesdre (see the catchments highlighted in Figure 1), the ISSHM provides a hydrologic reconstruction of the event at the local scale over the complete Eifel-Ardennes region. The results encompass the complete state of the hydrologic system from groundwater across the variably saturated soil zone to the land surface including streams and vegetation and allow to interrogate the evaluation of the hydrologic states that led up to the event.

After a brief description of the study region (Section 2.1), the model setup is presented (Section 2.2). In the ensuing sections, we focus on the hydrometeorological evolution of the July 2021 floods (Section 3.1), before studying the discharge, subsurface storage, and precipitation evolution in four severely affected, meso-scale catchments (Sections 3.2 and 3.3). Section 4 discusses the findings and concludes with a summary and outlook.

2 Materials and methods

2.1 Geographic setting and scope of the study

The Eifel-Ardennes low mountain region ranges between 200 and 500 m altitude. Mountain peaks reach up to 750 m (Figure 1). The river network is characterized by narrow valleys. Soils are generally shallow and rocky, prevailing are Cambisols and Luvisols according to the SoilGrids database (Hengl et al., 2017). The region is characterized by an oceanic, moist climate. Long-term mean annual precipitation sums in the area range between 600 mm per year in rain-shadowed regions of the Lower Rhine Embayment to 1,400 mm in the elevated parts of the Ardennes-Eifel. The four catchments under consideration are the Ahr (753 km^2), Erft (549 km^2), and Kyll (835 km^2) in Germany, and the Vesdre (669 km^2) in Belgium. Typical long-term mean average river discharge for the gages in Figure 1 is about 7 m^3s^{-1} at Altenahr (Ahr), 2.5 m^3s^{-1} at Bliesheim (Erft), 9 m^3s^{-1} at Kordel (Kyll) and 10.6 m^3s^{-1} at Chaudfontaine (Vesdre) (DGJ, 2025; SPW-MI, 2025).

Extreme rainfall happened on 13 and 14 July (see Figure 1; Section 3.1) with main flood events taking place on 14 and 15 July 2021 with discharge peaks occurring in the afternoon of 14 July and the morning hours of 15 July in the most affected catchments (Mohr et al., 2023). Our study encompasses a timespan from Tuesday, 13 July 2021 12UTC, to Friday, 23 July 2021 12UTC. All analyses are done for Coordinated Universal Time; this 2 h time delay with respect to local time needs to be considered when comparing to in-situ reconstructions of the events.

Many drainage and river networks in the Eifel-Ardennes region and part of the Lower Rhine Embayment were affected, i.e., damage was wide-spread. Other areas in North Rhine-Westphalia, Belgium, and the Netherlands also experienced severe flooding, as well as parts of Southern Germany, Switzerland, and Austria where flooding, flash floods, and mudflows happened (Mohr et al., 2023). In the following, we delineate the effective river catchments of the applied hydrologic model with reference to some main gaging stations; these model-defined catchments coincide with the real catchments. The selected gages are close to the river mouths. Except for the Vesdre river, gages and analysis catchments are also selected to avoid highly managed catchments.

2.2 Hydrologic model and experimental forecasting setup

The hydrologic model used in this study is ParFlow v3.8 in its GPU version (Hokkanen et al., 2021), a parallel, fully coupled, physically based ISSHM. ParFlow simulates surface flow with a 2D kinematic wave equation which acts as the upper boundary condition for the subsurface flow with the 3D Richards equation in a continuum approach (Kollet and Maxwell, 2006, 2008; Kuffour et al., 2020). The ParFlow built-in Common Land Model (CLM) parameterizes the moisture, energy, and momentum balances and determines the flux partitioning at the land surface as a function of the time-variable external atmospheric forcing, vegetation properties, and external parameter fields, such as land cover or rooting depth (Maxwell and Miller, 2005; Kollet and Maxwell, 2008). Through the land surface energy budget CLM resolves net precipitation from evapotranspiration, and interception over vegetated surfaces, and infiltration.

The ParFlow simulations of this study are taken from an experimental, quasi-operational forecasting system for a hydrologic Germany domain with 2,000 × 2,000 model grid points, centered over Germany at 611 m spatial resolution (DE06) (see inset in Figure 1 for the complete model domain). The DE06 setup, configuration, external parameter fields and their generation, as well as the experiment design are introduced for the first time in detail in Belleflamme et al. (2023). In the DE06 setup, ParFlow has 15 vertical layers with a variable grid spacing ranging from 0.02 m at the land surface to 18 m near the bedrock resulting in a depth of 60 m using a terrain-following grid. The land sea mask and the D4 slopes are calculated from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) at about 30 m resolution (Abrams et al., 2020). A hydrologically adjusted DEM from the MERIT database is used to adjust the slopes to follow the river trajectories (Yamazaki et al., 2019). The static land cover information is from a 100 m resolution 2018 Corine v20 version (CLC2018) (EEA, 2018), which is transferred to 18 International Geosphere-Biosphere Programme (IGBP) land cover types, that are translated to Plant Functional Types (PFTs) of CLM. Soil hydraulic parameters (porosity, permeability, Van Genuchten (1980) parameters saturated and residual water content, alpha, and N values) are derived through a remapping to USDA soil texture classes from SoilGrids v2017 at 250 m resolution (Hengl et al., 2017) and for hydrogeology the International Hydrogeological Map of Europe 1:1,500,000 (IHME) hydrogeological types are used (Duscher et al., 2015). Supplementary Figure S1 gives a spatial overview of some of the most relevant external parameters. The model timestep as well as the output interval is 1 h. ParFlow simulates the complete 3D subsurface water budget as well as the interactions at the surface. It must be emphasized that the DE06 simulations have undergone a careful setup and configuration (see Belleflamme et al., 2023) combined with a multi-decadal spinup, but other than that, the model is not calibrated or tuned at all, e.g., for soil moisture distributions or discharge.

Our flood event reconstruction makes use of three types of simulations taken from the stream of ParFlow DE06 hydrologic forecast simulations. The analysis timespan ranges from Tuesday, 13 July 2021 12UTC, to Friday, 23 July 2021 12UTC, encompassing the 14 and 15 July main event days and a single forecast initialization and time frame. The baseline is the 10-day deterministic DE06 ParFlow forecast driven by the ECMWF HRES deterministic weather forecast (Owens and Hewson, 2018) at 1 h temporal and 0.1 deg. spatial resolution (resampled to the 611 m ParFlow grid) as atmospheric forcing (ParFlow-HRES). To sample a large range of atmospheric conditions and precipitation forecasts, additional 50 ParFlow DE06 10-day forecast ensemble members are considered (ParFlow-ENS), driven by the 50-member ECMWF ENS probabilistic weather prediction ensemble. ParFlow-HRES and -ENS use the same initial conditions of 2021-07-13 12UTC, based on the ParFlow-CLIM climatology run. ParFlow-CLIM consists of a sequence of seamlessly concatenated 24 h deterministic ParFlow-HRES forecasts from 12UTC to 12UTC of the following day. ParFlow-CLIM simulations are initialized at 2007-01-01 from a multi-decadal spinup. Except for the atmospheric forcing, all simulations of the study use the same configuration and explicitly calculate overland flow.

The ParFlow DE06 forecasts are not associated with any official hydrologic forecasting or early warning system, the production is research-driven without any mandate or data delivery obligation. However, as the ISSHM is run daily since the beginning of 2021, is based on an automatic workflow, with a strict forecast clock, covers a national level, uses operational weather forecasts as a driver, and has been run with only few interruptions, we term the system "quasi-operational."

The ParFlow model has been extensively tested, validated, and used in many different types of applications, Kuffour et al. (2020) provide an overview. Related to the flood event simulation skills of ParFlow needed for this study, Kollet et al. (2017) demonstrate in their ISSHM intercomparison study the physical models' capability to simulate, e.g., hill-slope processes in an idealized tilted v-catchment setup. Lapides et al. (2020) find a good agreement of ParFlow's overland flow simulation with an analytical solution to the kinematic wave equation and lab experiments. In a real data parameter and input uncertainty study, Pomeon et al. (2020) use a high resolution ParFlow for flash flood hindcasting with NSE values of up to 0.85.

The ParFlow DE06 setup and configuration underlying this study was evaluated by Belleflamme et al. (2023) with respect to important water cycle components: monthly volumetric soil moisture anomalies from 2011 to 2020 show a spatial mean root mean square error (RMSE) of 0.02m^3 m^{-3} when compared to ESA CCI satellite data, and a Pearson correlation coefficient (r) of >0.5 for 80% of the grid elements; evapotranspiration is compared to the GLEAM dataset with an RSME of 0.15 mm day^{-1} and $r > 0.5$ for 86% grid elements; water table depth yields in a comparison with 5,799 wells from 2011 to 2016 an RSME of 0.32 m and $r > 0.5$ for 49% of the wells; river discharge at 231 GRDC gages shows from 2011 to 2019 a $r > 0.5$ at 96% of the gages and at 70% of the gages the Nash–Sutcliffe efficiency score (NSE) is improved compared to the mean flow benchmark. The ParFlow DE06 setup and configuration is also used in Saadi et al. (2023) who can reproduce reconstructed Eifel rivers' hydrographs in model parameter and atmospheric forcing sensitivity study. Thus, we are confident that the relevant processes are adequately represented in the model and provide a consistent and interpretable spatiotemporal view of the event.

3 Results

3.1 Atmospheric drivers and soil moisture evolution

The synoptic situation characterizing the low-pressure system which caused the wide-spread large rainfall amounts (see Figure 1) has been extensively analyzed, e.g., by the DWD (Junghänel et al., 2021), the RMI (2021), or the ECMWF (2021) and comprehensive analysis in Ludwig et al. (2023). The ECMWF HRES and ENS41 weather forecasts, Figures 2A–D,I–K taken from the ParFlow-HRES and ParFlow-ENS41 atmospheric forcing, show that the flood-causing rain events were mainly restricted to July 13 and July 14 along a stationary front. We show the ENS41 ECMWF weather forecast ensemble member as it yields the highest discharge in our simulations (see Section 3.2). According to the meteorological analysis by the DWD (Junghänel et al., 2021), based on the observational HYRAS dataset (Razafimaharo et al., 2020), the 30-year July spatio-temporal average precipitation for the Ahr catchment is 70 mm vs. 115 mm 72 h^{-1} from July 12 to 14; for the Erft catchment numbers are 68 mm $month^{-1}$ vs. 130 mm 72 h^{-1}, and for the Kyll catchment 73 mm $month^{-1}$ vs. 126 mm 72 h^{-1}. Precipitation rates during July 13 were on average below 25 mm day^{-1} for the German catchments. The maximum hourly precipitation rates anywhere in the catchment from RADOLAN observations were 24.3 mm h^{-1} for the Ahr, 31.8 mm h^{-1} for the Erft, 24.4 mm h^{-1} for the Kyll, and 17.0 mm h^{-1} for the Vesdre throughout July 13 and 14. This suggests that sustained, wide-spread intense rainfall, instead of convectively driven rain bursts, triggered the flood, which resulted ultimately in saturation excess instead of infiltration excess runoff production. Most rain gage precipitation amounts reached record highs beyond 100-year return periods; many measured more than 150 mm in 24 h (Junghänel et al., 2021). Considering the HYRAS numbers above it should be noted that the RADOLAN radar observations are exceeded in the radar climatology RADKLIM dataset by DWD (Winterrath et al., 2017) with more widespread > 100 mm 72 h^{-1} precipitation, or quantitative precipitation estimates products in Saadi et al. (2023) with > 150 mm 24 h^{-1}.

The ECMWF forecasted daily precipitation sums in Figure 2 resemble the order of magnitude of the aforementioned precipitation amounts from different observational products. The smooth precipitation fields result from the bicubic resampling of the atmospheric forecasts to the 611 m ParFlow grid. As outlined in the introduction, our study does not depend on an exact reproduction of the observed hydrometeorological extreme events in terms of the spatial distribution, timing, and precipitation rates, as our interest is in the dynamics and the physical mechanisms of the event. The ECMWF HRES 24 h accumulated precipitation field of 2021-07-15 00UTC (Figure 2C) does not resemble the radar-based observation of Figure 1, irrespective of the difference in the aggregation time span. Using the same color scale, ECMWF ENS41 more closely resembles the peak precipitation in the center of the analysis domain and the vicinity of the affected rivers' source areas; for many grid elements more than 150 mm total precipitation were forecasted on for July 14 (Figure 2J).

In both ParFlow simulations shown, the intense precipitation, especially during 14 July (Figures 2A–D,I–K), results in a sharp increase in instantaneous soil water saturations of the upper 0.3 m in

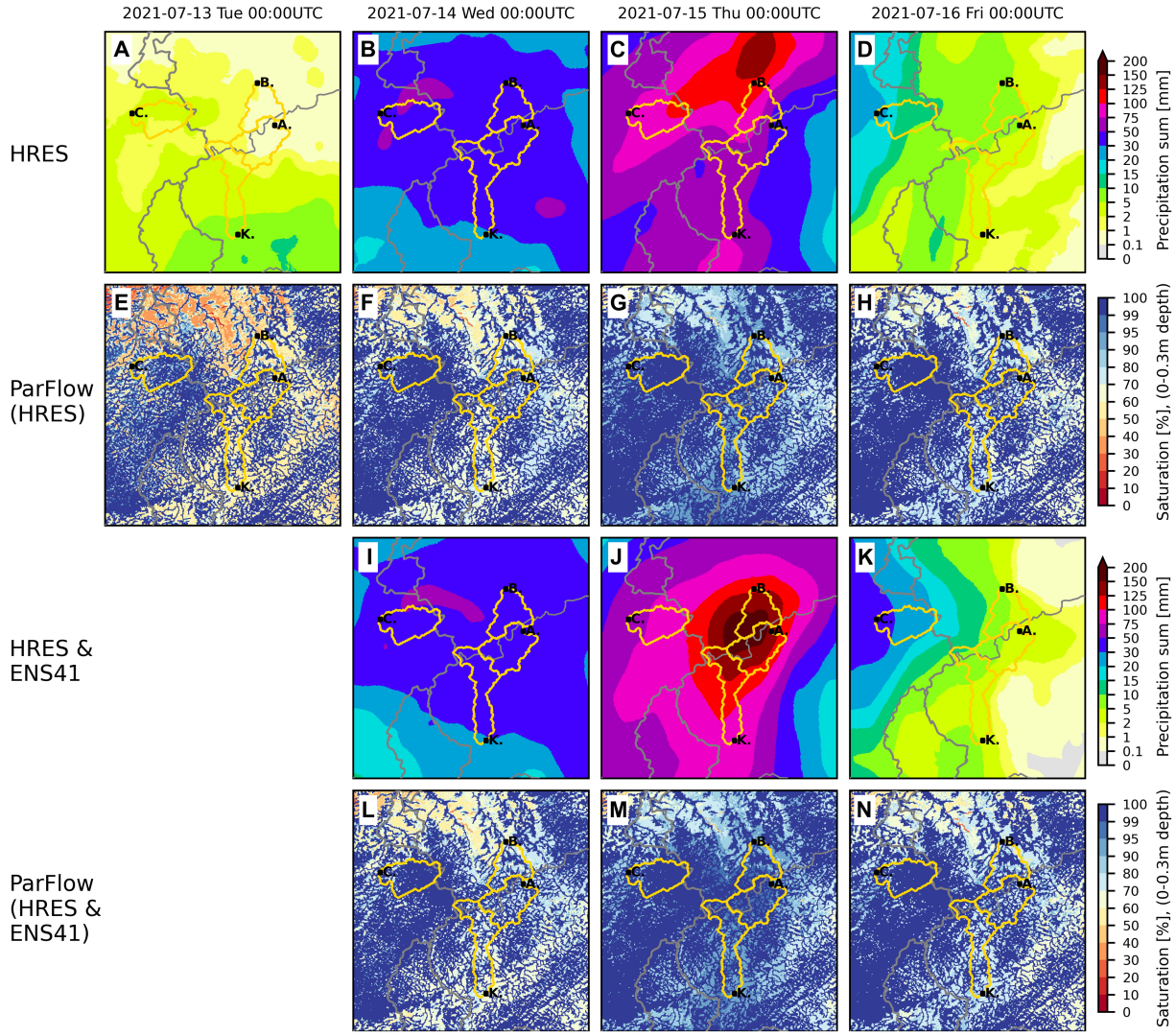


FIGURE 2
Temporal evolution of precipitation and soil moisture over the study area. First row **(A–D)**: 24 h precipitation sums [mm d–1] from ECMWF HRES deterministic forecasts at the time indicated above the plot. Second row **(E–H)**: Soil water saturation [%], instantaneous field at 00UTC, mean of the upper 0.3 m (5 model layers) from ParFlow-HRES. Third and fourth row **(I–N)**: As **(A–H)**, but for ECMWF ENS41 forecast and ParFlow-ENS41 simulation. The instantaneous soil water saturation always results from the field of 24 h precipitation sums shown above. Because the ParFlow-HRES and -ENS41 runs are started at 2021-07-13 12UTC with the same initial conditions, the HRES and ENS41-based fields in column 1 are the same; the precipitation sums at 2021-07-14 00UTC **(B)** are a combination of HRES (00 to 12UTC) and ENS41 (12 to 24UTC). Please note the non-linear scaling of the color bars.

Figures 2E–H,L–N which coincides with the spatial distribution of precipitation. In ParFlow-HRES the spatial average soil water saturation over the entire analysis region (Figure 1) increases from 70% on 13 July at 00UTC to 75% saturation on 14 July and to 88% on 15 July. Hence, the soils in large parts of the catchments still have some capacity to store parts of the precipitation at the onset of the event. Spatial averages over the four analysis catchments show maximum saturation values between 91% for the Erft and 97% for the Vesdre catchment. From the four analysis catchments, the Vesdre catchment in ParFlow-CLIM shows a notably different preconditioning Figure 2E with saturations greater than 95% for most grid elements and nearly full saturation at the onset of the event (Figures 2F,L). Section S2 of the supplement contains a comparison of the ParFlow volumetric soil moisture to the ESA-CCI satellite observations.

3.2 Flood characteristics

Figure 3 shows the hydrographs at Altenahr, Bliesheim, Kordel, and Chaudfontaine gage locations (see Figures 1, 2). Discharge is derived from ParFlow's overland flow calculation at these locations in the river convergence zones.

The ParFlow-HRES baseline Altenahr and Chaudfontaine hydrographs (Figures 3A,D) are characterized by steep rises during 14 July afternoon and peak discharge during the night from 14 to 15 July

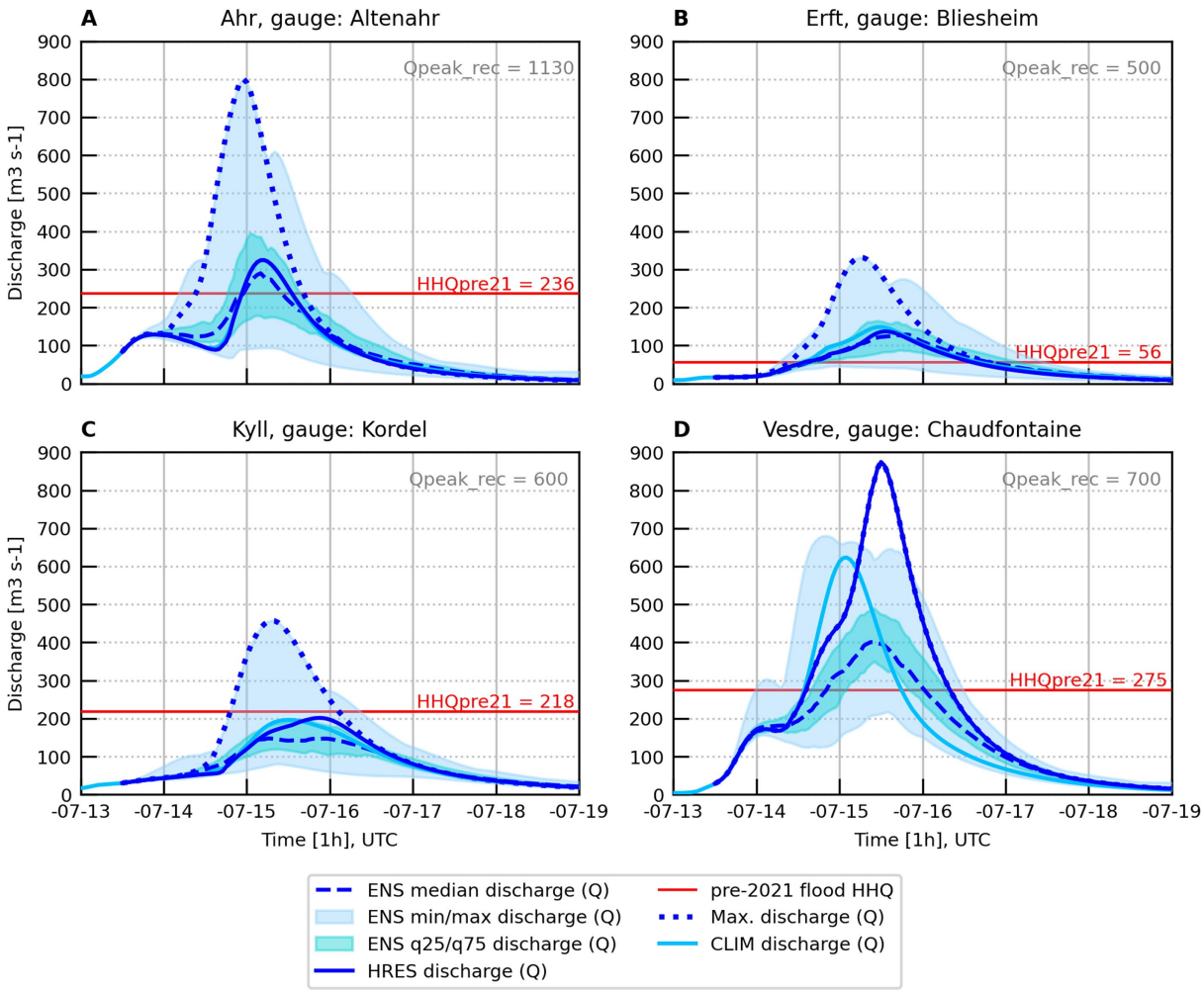


FIGURE 3
Discharge (Q) [m^3s^{-1}] for four gages of the analysis catchments. **(A)** Ahr river catchment, gage Altenahr, **(B)** Erft river catchment, gage Bliesheim, **(C)** Kyll river catchment, gage Kordel, **(D)** Vesdre river catchment, gage Chaudfontaine. Note the initialisation time at 2021-07-13 12UTC from the ParFlow-CLIM simulation from where the different model realizations diverge (ParFlow-CLIM, -HRES, -ENS[1...50]). Blue solid line: ParFlow-HRES; bright blue solid line: ParFlow-CLIM; blue/bright blue shaded area: minimum and maximum / 25th and 75th quantile of ParFlow-ENS[1...50]; blue dashed line: median of ParFlow-ENS; thick blue dotted line: hydrograph of peak discharge ParFlow run (HRES and ENS-41) (HRES causes peak discharge for Chaudfontaine); red solid line: pre-July 2021 observed peak discharge in [m^3s^{-1}] from DGJ (2025) and SPW-MI (2025); gray numbers (Qpeak_rec), upper right corner: estimated and reconstructed 2021 flood discharge in [m^3s^{-1}] from Dewals et al. (2021) and Mohr et al. (2023).

and during the day of 15 July, as an immediate response to the precipitation forcing (Figures 2, 4). The ParFlow-HRES Kordel and Bliesheim hydrographs show a similar behavior, albeit with a much lower peak discharge (Figures 3B,C). In all cases the historical pre-July 2021 maximum discharge values for the investigated gages Altenahr (HHQ = 236m^3s^{-1}), Bliesheim (HHQ = 56m^3s^{-1}), Chaudfontaine (HHQ = 275m^3s^{-1}), Kordel (HQQ = 218m^3s^{-1}) (DGJ, 2025; SPW-MI, 2025) are clearly exceeded (except for Kordel) in the ParFlow simulations (Figure 3), which agrees with observed event exceedances of water levels as compiled soon after the event in Kreienkamp et al. (2021). Except for Chaudfontaine the daily 12UTC reinitialized ParFlow-CLIM discharge matches ParFlow-HRES which hints to a good HRES forecast skill and indicates that the frequent atmospheric forcing reinitialization in ParFlow-CLIM does not affect the simulation's temporal consistency.

The large precipitation bandwidth in the ECMWF driving data ensemble with reference to timing, total amounts, intensity, and spatial distribution (data not shown) leads to highly differing hydrographs and discharge peaks simulated by ParFlow, as indicated by the shading in Figure 3. In a study by Saadi et al. (2023), that investigated the effect of different quantitative precipitation estimate datasets from meteorological radar observations, ParFlow peak discharge with a DE06-derived configuration also shows a large bandwidth.

The ParFlow-ENS41 for Altenahr, Bliesheim, and Kordel and ParFlow-HRES for Chaudfontaine yield the highest peak-flows. Although the study's primary goal is not to exactly reproduce the event, the ISSHM ParFlow can capture the timing, evolution, and the order of magnitude of the reconstructed event remarkably well, if only those ParFlow simulations out of the 52 simulations are considered

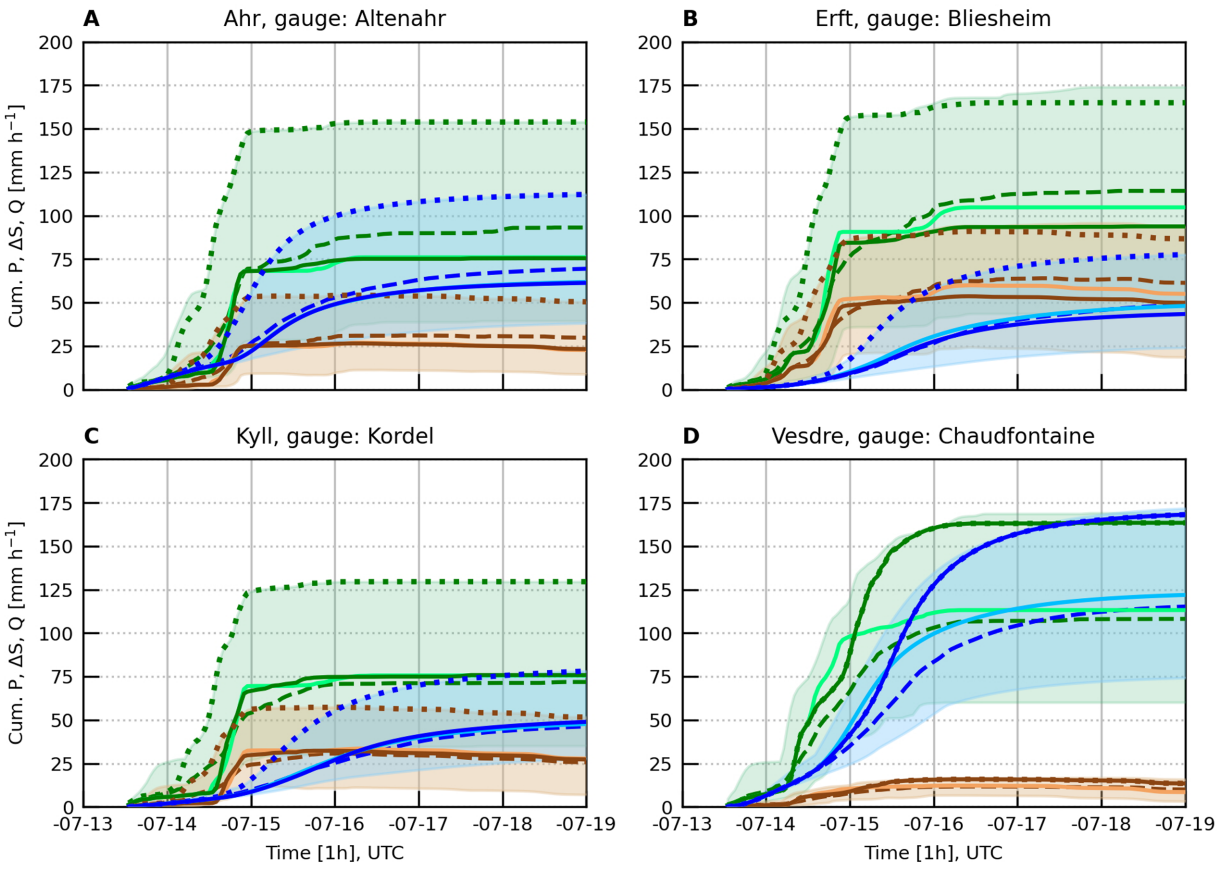


FIGURE 4
For the gaged catchment areas as defined in Figures 1, 2 of the **(A)** Ahr, **(B)** Erft, **(C)** Kyll, **(D)** Vesdre catchments: Cumulative precipitation (P, green colors), discharge (Q, blue colors), and total subsurface storage change (ΔS, brown colors) per unit catchment area in [mm h^{-1}] with respect to 2021-07-13 12UTC. Light green: ECMWF-HRES precipitation for ParFlow-CLIM; green: ECMWF-HRES precipitation for ParFlow-HRES; green shading: minimum and maximum of ECMWF-ENS[1...50]; green dashed line: median of ECMWF-ENS[1...50]; thick green dotted: precipitation of peak discharge ParFlow runs (HRES and ENS-41). Light blue/brown solid line: ParFlow-CLIM discharge/storage; blue/brown solid line: ParFlow-HRES discharge/storage; blue/brown shaded area: ParFlow-ENS[1...50] minimum and maximum discharge/storage; blue/brown dashed line: ParFlow-ENS[1...50] median discharge/storage; thick blue/brown dotted line: discharge/storage of peak discharge ParFlow runs (HRES and ENS-41). For Chaudfontaine peak discharge is caused by HRES and curves are overlain.

which yield the highest discharge (thick blue dotted line in Figure 3). The behavior of this hindcast ensemble clearly underlines the importance of using large atmospheric forcing data ensembles and many hydrologic ensemble members, as emphasized also, e.g., by Cloke and Pappenberger (2009) or Harrigan et al. (2023).

Mohr et al. (2023), their Table 1 and Figure 6 for Germany, and Dewals et al. (2021) for Belgium give overviews of the event's approximated hydrographs and water stages or peak discharge values, that could be estimated or reconstructed (Figure 3, gray numbers in the plots), despite largely incomplete hydrologic observations due to many destroyed gaging stations, and the fact that empirical water stage-discharge relationships are not applicable under such extreme flood conditions.

The event's extremeness becomes even more obvious when considering the detailed reconstructions and event descriptions in DGJ (2025), Mohr et al. (2023) or Roggenkamp et al. (2024). The long-term mean discharge of the Ahr river at gage Altenahr is 6.3m^{3}s^{-1}; the daily mean discharge was approximated to about 600 m^{3}s^{-1} (±50 m^{3}s^{-1}) on July 15, as opposed to the previously highest daily mean of 162 m^{3}s^{-1} on 21 December 1993. On 14 July the water level was about 1 m at 10UTC, which equals about 60m^{3}s^{-1} discharge, reaching about 550 m^{3}s^{-1} between 19 and 20 CEST before the measurements ceased.

In the Roggenkamp et al. (2024) reconstruction a peak discharge is quantified between about 1,000 m^3s^{-1} and 1,250 m^3s^{-1} with a water level above the riverbed of about 6 m for the place of Dernau, shortly downstream of gage Altenahr. These values exceed the previous highest peak on instrumental record of about 236 m^3s^{-1} with water levels of about 3.7 m on 2 June 2016 by up to a factor of 5 (DGJ, 2025; Roggenkamp et al., 2024).

One of the main results of this study is the proven capability of the ParFlow ISSHM, in its DE06 setup and configuration designed for water resources forecasts, to capture the extreme July 2021 flood event without any additional calibration or tuning.

3.3 Storage buffer characteristics

To better understand the ambient hydrologic conditions that led to the event and its evolution, we estimate and inspect the storage buffer of the different watersheds, which strongly impacts the onset and peak of flooding. Figure 4 juxtaposes the cumulative precipitation forcing and discharge responses at the gages per unit catchment area of the different ParFlow simulations, and in addition also the total subsurface water storage change since the start of the simulation. The storage is calculated from the relative saturation output of ParFlow, based on the model porosities and layer thicknesses.

In our purely model-based assessment, the July 2021 hydrometeorological extreme event is constrained to 14 and 15 July. The large precipitation bandwidth of the ECMWF-ENS ensemble, ranging for example in the Ahr catchment from precipitation totals from 2021-07-13 12UTC to -07-15 00UTC from about 15 mm 36 h^{-1} to about 150 mm 36 h^{-1} is characteristic for all catchments. For the actual event, the ECMWF-HRES sums and the median of the ECMWF-ENS show a similar behavior. The steep and steady rise of ECMWF-HRES and ECMWF-ENS median cumulative precipitation in the forecasts agrees with the assessment from Section 3.1. Precipitation fell nearly entirely during the 2nd half of 14 July and caused the extreme runoff generation and flooding in the Kyll and Ahr river valleys and along the Erft in the Lower Rhine Embayment region late on 14 July and on 15 July. The Vesdre catchment received about the same amount of precipitation during 14 July, but with an earlier onset and a slower temporal evolution of the flooding.

The order of magnitude of ECMWF-HRES precipitation totals matches with radar observations (see Figure 1) and analyses in Ludwig et al. (2023) or Saadi et al. (2023), however with an underestimation of about 15-20 mm 36 h^{-1} for the Ahr, Erft, and Kyll catchments (data not shown). The coarser resolution atmospheric model runs, such as the 9 km resolution ECMWF-HRES and 18 km -ENS datasets are typically associated with a lower extreme event precipitation intensity (e.g., Ban et al., 2021). This also shows with the highest precipitation sums (which also cause the highest discharge). Precipitation rates between the ECMWF-ENS41 and ECMWF-HRES are very similar, however the ECMWF-ENS41 precipitation onset is about 12 h earlier in case of Ahr, Erft and Kyll catchments and yields eventually about double the precipitation amount than ECMWF-HRES. In general, many ECMWF-ENS ensemble members produce plausible precipitation amounts.

The lower slopes of the cumulative runoff as opposed to the cumulative precipitation indicate that water continues to infiltrate. The Ahr and Kyll catchments show a behavior where on average about one third of the total precipitation is buffered by the subsurface. For the Erft catchment buffering capacity is about half of the precipitation. In the Vesdre catchment with higher initial saturation levels (Figure 2) and less subsurface buffer, about all precipitation is transformed into discharge, with a time lag of about 24 h. The ratios of storage change to the precipitation sum ratios remain similar for all simulations per catchment, starting off from the same initial conditions at 2021-07-13 12UTC, which indicates that infiltration continued and storage capacity is not yet exceeded.

Per unit catchment area, the cumulative discharge-precipitation ratios at 2021-07-17 00UTC for the highest discharge simulations are 71% (ParFlow-ENS41) (110 mm discharge vs. 155 mm precipitation) for the Ahr, 46% (73 mm discharge vs. 160 mm precipitation) for the Erft, 57% (74 mm discharge vs. 129 mm precipitation) for the Kyll, and 96% (ParFlow-HRES) (157 mm discharge vs. 163 mm precipitation) for the Vesdre.

This, as well as the fact that the ENS precipitation ensemble and runoff ensemble do only partly overlap, indicates a substantial buffering capacity in the modeled system, which is reflected also in the estimation of the cumulative subsurface water storage change (Figure 4). The modeled buffering capacity also reduces the spread in the hydrological response with respect to the atmospheric forcing in almost all catchments.

The 2nd main result of the study is that for the Ahr, Erft, and Kyll catchments between about 30 and 50% of the precipitation can infiltrate, avoiding even higher discharge peaks, despite already high antecedent soil moisture saturation. For the Vesdre catchment with less subsurface buffer, about all precipitation transforms into discharge.

4 Discussion

In this study, we used the ISSHM ParFlow in a high-resolution experimental, quasi-operational forecasting setup for water resources assessments, to provide a spatially and temporally consistent event reconstruction and an analysis of the hydrologic characteristics and physical mechanisms that led to the extreme July 2021 flood event and its evolution in the Eifel-Ardennes low mountain range in western Germany and eastern Belgium.

In a purely simulation-driven numerical laboratory approach, without specific additional calibration or tuning, ParFlow with the DE06 setup and configuration can reproduce the extreme July 2021 flood event in a hindcast ensemble with a lead time of 24 h in terms of magnitude, timing, and location when compared to available information from detailed reconstructions of the floods. This is also due to the driving ECMWF-HRES and -ENS precipitation forecasts, whose large variance encompasses the observed precipitation amounts.

In terms of the hydrologic modeling, the study shows that an uncalibrated physically based ISSHM can simulate extreme events that are beyond the currently available observational data records, which arguably renders calibration meaningless. This demonstrates the prognostic capabilities of the ParFlow DE06 ensemble. Thus, the proposed modeling approach is an applied, highly versatile monitoring and forecasting system especially useful in ungaged or data-scarce regions, and when trans-regional or transnational exchange of data is limited. The high-resolution model approach, that can be run efficiently on state-of-the-art GPU-clusters, cannot only be used for

water resources applications, but is able to capture the sub-daily dynamics of extreme flooding and provide useful information at the local scale.

Despite the successful reproduction of the July 2021 flood event, the simulation-driven numerical laboratory approach has uncertainties and limitations which affect the results. For example, external surface and subsurface parameters are simplified. The ParFlow DE06 setup uses a very recent, but static land cover dataset. The soil and subsurface hydraulic properties, i.e., the description of the complex hydrogeology of the Eifel-Ardennes region, are based on SoilGrids soil texture classes, which are reclassified into U. S. Department of Agriculture (USDA) soil types to be used with the USDA ROSETTA pedotransfer function model (Schaap et al., 2001) for Van Genuchten parameters. Below the SoilGrids depth-to-bedrock the permeability, porosity, Van Genuchten parameters saturated and residual water content, n and alpha, are assigned to IHME hydrogeological types. In the DE06 setup, the Manning's surface roughness coefficient is constant; in their study on the July 2021 flood, Saadi et al. (2023) used spatially varying Manning's parameters to obtain reconstructed peak discharge, which may help to adjust the model to changing morphodynamics in the river valleys during the event. The 611 m horizontal grid spacing does not resolve the very steep and narrow river valleys of the Eifel-Ardennes mid mountain ranges; the scale-consistent river parameterization as developed by Schalge et al. (2019) for ParFlow is not used. A first order uncertainty additionally exists through the atmospheric forcing data. In our study this is accounted for by using the 50-member ECMWF-ENS forecast in conjunction with the ECMWF-HRES forecast, which introduces a considerable resolution gap (18 km and 9 km versus 0.611 km). Even for high-resolution meteorological radar quantitative precipitation estimates, Saadi et al. (2023) documented underestimations of the RADOLAN data of up to 20%. Because the hindcast ensemble is based on quasi-operational forecasting simulations and the goal is to demonstrate the capabilities of this quasi-operational setup, we only consider precipitation induced uncertainties.

Note, ParFlow ISSHM in the DE06 setup reaches reconstructed discharge magnitudes (Figure 3) at the four analysis gages, when those hydrologic model ensemble members (ParFlow-ENS41, ParFlow-HRES) are considered that are driven by the highest cumulative precipitation amount from the atmospheric forecasts. By tuning (reducing) Manning's coefficient with historical flood events and landscape and soil heterogeneity in the DE06 setup, Saadi et al. (2023) simulate the reconstructed peak discharge also with a precipitation forcing equivalent to ECMWF-HRES. In this study, because DE06 is always forced with an atmospheric ensemble, the event is albeit captured. Also, ECMWF-HRES is close to RADOLAN precipitation sums, which are arguably too low when compared to quantitative precipitation estimates based on different radar retrieval algorithms in Saadi et al. (2023) or with reprocessed RADKLIM data (Winterrath et al., 2017; DWD, 2025). Furthermore, the extensive DE06 evaluation by Belleflamme et al. (2023) yields good agreement with observations and the DE06 setup and configuration needs to be suitable for a wide range of applications and catchments without additional tuning.

Because of the physical consistency of the ISSHM ParFlow simulations, the DE06 ensemble can be used to further explore the physical mechanisms that generated the widespread extreme flood and water ponding heights in Eifel-Ardennes valleys during the 2021 event. Due to the surface and subsurface water interaction, the dynamics of the 2021 flood event in our modeling study are different than those of flash floods, that develop rapidly and are dominated by infiltration excess overland flow.

An important driving mechanism for the extreme July 2021 flood event was sustained, intense rainfall over a timespan of about 12 h over the almost complete area of the catchments under consideration. The high average amounts of catchment-wide precipitation might have caused fast pressure responses of the drainage system in the center of the valleys, leading to rapidly rising hydrographs and eventually high discharge peaks along the river networks. Considering the ParFlow-ENS41 simulation in the Ahr, Erft, and Kyll catchments, we hypothesize that the slope of the cumulative discharge per unit catchment area in Figure 4, which is smaller than the slope for cumulative precipitation and subsurface storage change, might be indicative of a relatively slow interflow, i.e., the contribution to the river discharge at the gage is delayed due to slow subsurface water movement. The fast reaction of the hydrograph at the gages (Figure 3, dotted lines) with only little temporal delay to the precipitation signal (Figure 4, dotted lines) might however be an indication of a strong and fast pressure response in the subsurface, i.e., a pressure wave, induced by heavy precipitation loading, that propagated quickly through saturated soils eventually leading to high exfiltration in the river valleys, i.e., the convergence zones, with high water tables, accompanied by little sheet flow outside the river channels.

The initial soil moisture state is highly relevant in the initiation and evolution of flooding, especially with widespread intense precipitation events that affect the complete catchment, because soil moisture storage acts as an effective precipitation buffer (Merz et al., 2020). However, contrary to common beliefs, a considerable soil moisture buffer exists in our study during the onset of the event. Also, given the complex topography, and varying valley cross sections with funneling effects and blocking situations due to debris, accurate discharge forecasts are arguably impossible under these conditions. Thus, instead of focusing only on discharge, the study additionally interrogated the spatiotemporal evolution of the storage buffer characteristics, which are a key factor in the generation and magnitude of the flood event.

In the simulations, the upper soil in the Ahr, Erft and Kyll catchments can buffer between about one third to one half of the precipitation that does not contribute immediately to the streamflow response. This in turn suggests that under higher ambient soil moisture conditions at the beginning of the precipitation event, discharge and peak flows may be even larger, which is disconcerting in view of the experienced catastrophic destruction and loss of lives. In case of the Vesdre catchment, due to its initial moisture state with higher soil water saturation levels, the buffering capacity is lower; hence more runoff is simulated in comparison to the other catchments.

This study adds a process-based analysis to the body of existing literature on the 2021 Eifel-Ardennes flood. With the ParFlow DE06 setup and configuration the capability of a transferable, uncalibrated ISSHMs is demonstrated to identify a combination of critical hydrometeorological system states in advance of an extreme event and to simulate and reconstruct the flood event based on simulated initial conditions. Given the large ensemble precipitation spread with even higher plausible daily precipitation sums, and the importance of the soil water buffer, sensitivity studies could be designed to systematically quantify the discharge behavior of flood-prone mesoscale catchments

with similar topographic conditions under different initial soil moisture conditions in combination with large-scale, extreme precipitation ensembles.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://doi.org/10.26165/JUELICH-DATA/GROHKP.

Author contributions

KG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. AB: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft. SH: Data curation, Resources, Software, Writing – original draft. JV: Conceptualization, Methodology, Supervision, Writing – original draft. SK: Conceptualization, Formal analysis, Methodology, Resources, Software, Supervision, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Helmholtz Association Knowledge Transfer project ADAPTER (WT-0104) and through the BMBF BioökonomieREVIER funding scheme with its "BioRevierPlus" project (funding ID: 031B1137D/031B1137DX) with its Innovation Laboratory Digital Geosystem Rhineland Region. The authors gratefully acknowledge the Earth System Modeling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Acknowledgments

The RADOLAN precipitation radar data are from the German Weather Service (DWD). Acknowledgement is made also for the use of ECMWF's computing, archive facilities and data in this research. We acknowledge the supercomputing support as well as computational and storage resources provided to us by the Jülich Supercomputing Centre (JSC) through the Simulation and Data Laboratory Terrestrial Systems of the Centre for High-Performance Scientific Computing in Terrestrial Systems (Geoverbund ABC/J) and the JSC, Germany.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2025.1571704/full#supplementary-material

References

Abrams, M., Crippen, R., and Fujisada, H. (2020). ASTER global digital elevation model (GDEM) and ASTER global water body dataset (ASTWBD). *Remote Sens.* 12:1156. doi: 10.3390/rs12071156

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., et al. (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Clim. Dyn.* 57, 275–302. doi: 10.1007/s00382-021-05708-w

Belleflamme, A., Goergen, K., Wagner, N., Kollet, S., Bathiany, S., El Zohbi, J., et al. (2023). Hydrological forecasting at impact scale: the integrated ParFlow hydrological model at 0.6 km for climate resilient water resource management over Germany. *Front. Water* 5:1183642. doi: 10.3389/frwa.2023.1183642

Brookfield, A. E., Ajami, H., Carroll, R. W. H., Tague, C., Sullivan, P. L., and Condon, L. E. (2023). Recent advances in integrated hydrologic models: integration of new domains. *J. Hydrol.* 620:129515. doi: 10.1016/j.jhydrol.2023.129515

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., et al. (2017). The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. *Hydrol. Earth Syst. Sci.* 21, 3427–3440. doi: 10.5194/hess-21-3427-2017

Cloke, H. L., and Pappenberger, F. (2009). Ensemble flood forecasting: a review. *J. Hydrol.* 375, 613–626. doi: 10.1016/j.jhydrol.2009.06.005

Cornwall, W. (2021). Europe's deadly floods leave scientists stunned. *Science* 373, 372–373. doi: 10.1126/science.373.6553.372

Davies, R. (2021a). Germany – Fatalities, Evacuations, Dozens Feared Missing After Severe Flooding. Available online at: https://floodlist.com/europe/germany-floods-july-2021 (Accessed February 2, 2025).

Davies, R. (2021b). Belgium – Thousands Evacuate Floods as Death Toll Rises. Available online at: https://floodlist.com/europe/floods-belgium-july-2021 (Accessed February 2, 2025).

de Jager, A., and Vogt, J. (2003). Rivers and Catchments of Europe - Catchment Characterisation Model (CCM). Available online at: http://data.europa.eu/89h/8c681046-726b-413d-aff8-b1afebd73c0a (Accessed February 2, 2025).

Dewals, B., Erpicum, S., Pirotton, M., and Archambeau, P. (2021). "Extreme floods in Belgium. The July 2021 extreme floods in the Belgian part of the Meuse basin", in: Hydrolink 2021/4. International Association for Hydro-Environment Engineering and Research (IAHR). 104–107. Available online at: https://www.iahr.org/library/hydrolink?hid=412 (Accessed February 2, 2025).

Duscher, K., Günther, A., Richts, A., Clos, P., Philipp, U., and Struckmeier, W. (2015). The GIS layers of the "international hydrogeological map of Europe 1:1,500,000" in a vector format. *Hydrogeol. J.* 23, 1867–1875. doi: 10.1007/s10040-015-1296-4

DWD (2025). RADOLAN: Analysen radarbasierter stündlicher (RW) und täglicher (SF) Niederschlagshöhen. Available online at: https://www.dwd.de/DE/leistungen/radolan/radolan.html (Accessed February 5, 2025).

ECMWF (2021). 202107 - Rainfall - Germany and Belgium. Available online at: https://confluence.ecmwf.int/display/FCST/202107+-+Rainfall+-+Germany+and+Belgium (Accessed February 3, 2025).

EEA (2018). Copernicus Land Monitoring Service, Corine Land Cover 2018. Available online at: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018 (Accessed February 3, 2025).

EM-DAT (2021). 2021-0411, severe floods; Germany, Belgium. In: The International Disaster Database. Available online at: https://doi.org/10.17616/R3QQ1X (Accessed February 3, 2025).

Harrigan, S., Zsoter, E., Cloke, H., Salamon, P., and Prudhomme, C. (2023). Daily ensemble river discharge reforecasts and real-time forecasts from the operational global flood awareness system. *Hydrol. Earth Syst. Sci.* 27, 1–19. doi: 10.5194/hess-27-1-2023

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12:e0169748. doi: 10.1371/journal.pone.0169748

Hokkanen, J., Kollet, S., Kraus, J., Herten, A., Hrywniak, M., and Pleiter, D. (2021). Leveraging HPC accelerator architectures with modern techniques — hydrologic modeling on GPUs with ParFlow. *Comput. Geosci.* 25, 1579–1590. doi: 10.1007/s10596-021-10051-4

Journée, M., Goudenhoofdt, E., Vannitsem, S., and Delobbe, L. (2023). Quantitative rainfall analysis of the 2021 mid-July flood event in Belgium. *Hydrol. Earth Syst. Sci.* 27, 3169–3189. doi: 10.5194/hess-27-3169-2023

Junghänel, T., Bissolli, P., Daßler, J., Fleckenstein, R., Imbery, F., Janssen, W., et al. (2021). Hydro- klimatologische Einordnung der Stark- und Dauerniederschläge in Teilen Deutschlands im Zusammenhang mit dem Tiefdruckgebiet "Bernd" vom 12. bis 19. Juli 2021. Deutscher Wetterdienst, Offenbach am Main. Available online at: https://www.dwd.de/DE/leistungen/besondereereignisse/niederschlag/20210721_bericht_starkniederschlaege_tief_bernd.html (Accessed February 2, 2025).

Kollet, S. J., and Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.* 29, 945–958. doi: 10.1016/j.advwatres.2005.08.006

Kollet, S. J., and Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resour. Res.* 44:W02402. doi: 10.1029/2007WR006004

Kollet, S. J., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., et al. (2017). The integrated hydrologic model Intercomparison project, IH-MIP2: a second set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resour. Res.* 53, 867–890. doi: 10.1002/2016WR019191

Khosh Bin Ghomash, S., Apel, H., and Caviedes-Voullième, D. (2024). Are 2D shallow-water solvers fast enough for early flood warning? A comparative assessment on the 2021 Ahr valley flood event. *Nat. Hazards Earth Syst. Sci.* 24, 2857–2874. doi: 10.5194/nhess-24-2857-2024

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., et al. (2021). Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021. World Weather Attribution. Available online at: https://www.worldweatherattribution.org/heavy-rainfall-which-led-to-severe-flooding-in-western-europe-made-more-likely-by-climate-change/ (Accessed February 3, 2025).

Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S., and Maxwell, R. M. (2020). Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geosci. Model Dev.* 13, 1373–1397. doi: 10.5194/gmd-13-1373-2020

Lapides, D. A., David, C., Sytsma, A., Dralle, D., and Thompson, S. (2020). Analytical solutions to runoff on hillslopes with curvature: numerical and laboratory verification. *Hydrol. Process.* 34, 4640–4659. doi: 10.1002/hyp.13879

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., et al. (2022). Assessment of the 2021 summer flood in Central Europe. *Environ. Sci. Eur.* 34:107. doi: 10.1186/s12302-022-00685-1

DGJ (2025). Deutsches Gewässerkundliches Jahrbuch im Internet. Available online at: https://dgj.de (Accessed February 3, 2025).

Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., et al. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in Central Europe – part 2: historical context and relation to climate change. *Nat. Hazards Earth Syst. Sci.* 23, 1287–1311. doi: 10.5194/nhess-23-1287-2023

Maxwell, R. M., and Miller, N. L. (2005). Development of a coupled land surface and groundwater model. *J. Hydrometeorol.* 6, 233–247. doi: 10.1175/JHM422.1

McDonnell, J. J., and Beven, K. (2014). Debates - the future of hydrological sciences: a (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resour. Res.* 50, 5342–5350. doi: 10.1002/2013WR015141

McMillan, H., Krueger, T., and Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol. Process.* 26, 4078–4111. doi: 10.1002/hyp.9384

Merz, R., Tarasova, L., and Basso, S. (2020). The flood cooking book: ingredients and regional flavors of floods across Germany. *Environ. Res. Lett.* 15:114024. doi: 10.1088/1748-9326/abb9dd

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., et al. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in Central Europe – part 1: event description and analysis. *Nat. Hazards Earth Syst. Sci.* 23, 525–551. doi: 10.5194/nhess-23-525-2023

Munich Re (2025). Flood risks on the rise - greater loss prevention is needed. Available online at: https://www.munichre.com/en/risks/natural-disasters/floods.html (assessed February 3, 2025).

Owens, R. G., and Hewson, T. D. (2018). ECMWF forecast user guide. Available online at: https://doi.org/10.21957/m1cs7h (Assessed February 3, 2025).

Paprotny, D., Sebastian, A., Morales-Nápoles, O., and Jonkman, S. N. (2018). Trends in flood losses in Europe over the past 150 years. *Nat. Commun.* 9:1985. doi: 10.1038/s41467-018-04253-1

Pomeon, T., Wagner, N., Furusho, C., Kollet, S., and Reinoso-Rondinel, R. (2020). Performance of a PDE-based hydrologic model in a flash flood modeling framework in sparsely-gauged catchments. *Water* 12:2157. doi: 10.3390/w12082157

Rasmussen, T. C. Jr., Baldwin, R. H., Dowd, J. F., and Williams, A. G. (2000). Tracer vs. pressure wave velocities through unsaturated Saprolite. *Soil Sci. Soc. Am. J.* 64, 75–85. doi: 10.2136/sssaj2000.64175x

Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M., and Deutschländer, T. (2020). New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS). *Theor. Appl. Climatol.* 142, 1531–1553. doi: 10.1007/s00704-020-03388-w

RMI (2021). Ce que l'on sait sur les pluies exceptionnelles des 14 et 15 juillet 2021 [online]. Available online at: https://www.meteo.be/fr/infos/actualite/ce-que-lon-sait-sur-les-pluies-exceptionnelles-des-14-et-15-juillet-2021 (Accessed February 3, 2025).

Roggenkamp, T., Herget, J., and Roggenkamp, T. (2024). Flood reconstruction – the unexpected rather frequent event at river Ahr in July 2021. *Glob. Planet. Chang.* 240:104541. doi: 10.1016/j.gloplacha.2024.104541

Saadi, M., Furusho-Percot, C., Belleflamme, A., Chen, J.-Y., Trömel, S., and Kollet, S. (2023). How uncertain are precipitation and peak flow estimates for the July 2021 flooding event? *Nat. Hazards Earth Syst. Sci.* 23, 159–177. doi: 10.5194/nhess-23-159-2023

Schaap, M. G., Leij, F. J., and van Genuchten, M. T. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* 251, 163–176. doi: 10.1016/S0022-1694(01)00466-8

Schäfer, A., Mühr, B., Daniell, J. E., Ehret, U., Ehmele, F., Küpfer, K., et al. (2021). Hochwasser Mitteleuropa, Juli 2021 (Deutschland), Bericht Nr. 1 "Nordrhein-Westfalen & Rheinland-Pfalz". Karlsruhe: Center for Disaster Management and Risk Reduction Technology (CEDIM), Forensic Disaster Analysis (FDA) Group.

Schalge, B., Haefliger, V., Kollet, S., and Simmer, C. (2019). Improvement of surface run-off in the hydrological model ParFlow by a scale-consistent river parameterization. *Hydrol. Process.* 33, 2006–2019. doi: 10.1002/hyp.13448

SPW-MI (2025). Hydrological statistical yearbook. Available online at: http://voies-hydrauliques.wallonie.be/opencms/opencms/fr/hydro/Archive/annuaires/index.html (Accessed February 3, 2025).

Tradowsky, J. S., Philip, S. Y., Kreienkamp, F., Kew, S. F., Lorenz, P., Arrighi, J., et al. (2023). Attribution of the heavy rainfall events leading to severe flooding in Western Europe during July 2021. *Clim. Chang.* 176:90. doi: 10.1007/s10584-023-03502-7

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44, 892–898. doi: 10.2136/sssaj1980.03615995004400050002x

Voit, P., and Heistermann, M. (2024). A downward-counterfactual analysis of flash floods in Germany. *Nat. Hazards Earth Syst. Sci.* 24, 2147–2164. doi: 10.5194/nhess-24-2147-2024

Weigl, E., and Winterrath, T. (2009). Radargestützte Niederschlagsanalyse und -vorhersage (RADOLAN, RADVOR-OP). Promet 35, 78–86. Available online at: http://nbn-resolving.de/urn:nbn:de:101:1-2017012510513.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Walawender, E., et al. (2017). Erstellung einer radargestützten Niederschlagsklimatologie. Berichte des Deutschen Wetterdienstes 251, Deutscher Wetterdienst, Offenbach am Main. Available online at: https://opendata.dwd.de/climate_environment/GPCC/radarklimatologie/Dokumente/Endbericht_Radarklimatologie_final.pdf (Accessed February 5, 2025).

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M. (2019). MERIT hydro: a high-resolution global hydrography map based on latest topography datasets. *Water Resour. Res.* 55, 5053–5073. doi: 10.1029/2019WR024873

Zhao, P., Zhao, P., Liang, C., Li, T., and Zhou, B. (2017). Understanding the rapidity of subsurface storm flow response from a fracture-oriented shallow vadose through a new perspective. *J. Hydrol.* 544, 628–639. doi: 10.1016/j.jhydrol.2016.12.007

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 15 July 2025
DOI 10.3389/frwa.2025.1539177

Check for updates

OPEN ACCESS

EDITED BY
Matteo Camporese,
University of Padua, Italy

REVIEWED BY
Ilja H. J. van Meerveld,
University of Zurich, Switzerland
Luisa Hopp,
University of Bayreuth, Germany

*CORRESPONDENCE
W. Payton Gardner
payton.gardner@umontana.edu

RECEIVED 03 December 2024
ACCEPTED 09 June 2025
PUBLISHED 15 July 2025

CITATION
von Trapp I, Jencso KG, Hoylman ZH, Livesay R and Gardner WP (2025) On the emergent scale of bedrock groundwater contribution to headwater mountain streams. *Front. Water* 7:1539177. doi: 10.3389/frwa.2025.1539177

COPYRIGHT
© 2025 von Trapp, Jencso, Hoylman, Livesay and Gardner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

On the emergent scale of bedrock groundwater contribution to headwater mountain streams

Isabellah von Trapp[1], Kelsey G. Jencso[2], Zachary H. Hoylman[2], Robert Livesay[2] and W. Payton Gardner[1]*

[1]Department of Geosciences, University of Montana, Missoula, MT, United States, [2]Department of Forest Management, Franke College of Forestry and Conservation, Missoula, MT, United States

We investigated the contribution of bedrock groundwater to streamflow as a function of catchment scale in a headwater stream. Synoptic surveys were conducted during hydrologically important periods of the year using multiple environmental tracers in stream water, soil water, and bedrock groundwater, along a first-order montane stream, in west-central Montana. Sampled analytes included ^{222}Rn, used to constrain total subsurface flux, and major and minor elements, used in end-member mixing analysis (EMMA) to identify the contributions of soil and bedrock groundwater to the stream. Partitioning between soil-derived and bedrock-derived groundwater was then analyzed as a function of the incremental and accumulated sub-catchment sizes. Radon results indicated that subsurface water contributions accounted for the majority of streamflow at all surveyed times. EMMA results revealed that the bedrock groundwater contribution to streamflow varied between 26% during peak snowmelt and 44% during late summer. Streamflow generation was dominated by soil groundwater contribution along the entire reach, but the bedrock groundwater contribution increased consistently with accumulated sub-catchment size. However, groundwater contributions were not well-correlated with incremental sub-catchment size. The scale at which increased bedrock groundwater discharge can be correlated with sub-catchment size appears to be >1 km^2 for our study. Our results are consistent with a conceptual model where streamflow is predominantly generated by a 3D subsurface nested flow system. Local subsurface heterogeneities control the stream source at local scales but begin to average out at scales >2 km^2. Our study indicates that, while soil groundwater is the dominant source, bedrock groundwater remains an important and predictable contributor to streamflow throughout the year, even in a snow-dominated, mountainous headwater catchment.

KEYWORDS

groundwater—surface water interactions, streamflow generation, mountain hydrology, environmental tracers, mountain aquifer

Introduction

Upland mountain watersheds host active hydrologic systems that play important roles in the storage and release of water to the larger stream network (Hale et al., 2016; Hale and McDonnell, 2016; Liu et al., 2004; Williams et al., 2015). These groundwater systems are locally important sources of streamflow (Asano et al., 2022; Gabrielli et al., 2012; Uchida et al., 2003) and recharge regional groundwater systems, which discharge to larger streams

in the lowlands (Beisner et al., 2018; Fan et al., 2013; Frisbee et al., 2011; Gardner et al., 2011; Gleeson and Manning, 2008; Smerdon et al., 2012). The amount of bedrock groundwater discharge in montane streamflow is important for conceptualizing watershed function and predicting hydrologic resource availability in the future (Condon and Maxwell, 2019; Meixner et al., 2016; Williams et al., 2015). However, groundwater flow systems are marked by extreme heterogeneity in hydraulic characteristics and the resulting flow path length and rate; as a result, groundwater discharge volumes and locations are all highly variable (Beisner et al., 2018; Gardner et al., 2011; Hale et al., 2016; Harrington et al., 2013; Smerdon and Gardner, 2021). Neither the spatial scale at which 3D groundwater flow becomes an important and predictable source of streamflow nor the seasonal variation in these processes are well-documented at present.

Subsurface water is an important source of streamflow, even in snow-dominated watersheds (Liu et al., 2004). Subsurface sources are commonly conceptualized as originating from both bedrock (weathered bedrock and saprolite) and soil reservoirs (Anderson et al., 1997a; Carroll and Williams, 2018; Hale et al., 2016; Hale and McDonnell, 2016; Salve et al., 2012; Torres et al., 1998). In this study, we refer to shallow groundwater in the soil zone as soil groundwater and deeper, fractured-bedrock groundwater as bedrock groundwater. The relative importance of soil and bedrock reservoirs depends on the soil and bedrock hydraulic characteristics (Flint et al., 2008; Gardner et al., 2020; Katsuyama et al., 2005; Pfister et al., 2017; Uchida et al., 2002, 2006). The amount of groundwater in a watershed significantly influences dynamic properties, such as hydrograph variability (Asano et al., 2022; Gardner et al., 2010; Pfister et al., 2017) and mean transit time (Asano et al., 2002; Asano and Uchida, 2012; Hale et al., 2016; Pfister et al., 2017; Uchida et al., 2006).

Soil water lateral flow and bedrock groundwater circulation are tightly coupled (Appels et al., 2015; Carroll et al., 2019; Gardner et al., 2020; Tromp-van Meerveld et al., 2007; Uchida et al., 2002). As a result of this coupling, the location and amount of bedrock recharge are controlled by subsurface permeability distribution, hillslope position, and bedrock topography, which affect the lateral soil moisture distribution and location of vertical drainage (Appels et al., 2015; Gardner et al., 2020). The relative contributions of soil and bedrock groundwater to stream water are expected to vary temporally (Spencer et al., 2021). The celerity of subsurface response depends on the soil and bedrock hydraulic properties (Anderson et al., 1997a,b).

Groundwater discharge location, rate, and flow path source can be highly localized and variable because they depend upon the local geology and surface characteristics (Beisner et al., 2018; Bergstrom et al., 2016; Gardner et al., 2011; Harrington et al., 2013; Smerdon et al., 2012). In addition, the localized bedrock and soil water contributions from hillslopes are buffered by the local alluvial groundwater system, and they can be mixed and transported within this system before reaching the stream (Jencso et al., 2010). Despite the spatial heterogeneity at local scales, it is clear that the contribution from deeper bedrock groundwater systematically increases with increasing drainage area at regional scales (Asano et al., 2022; Frisbee et al., 2011, 2012). Since there is observed heterogeneity in the amount and location of groundwater discharge occurring locally, which contrasts with predictable increases in groundwater discharge at larger, integrated scales, it appears that groundwater discharge in streams is an emergent property.

Is there, in fact, a scale at which groundwater discharge volume becomes more predictable? If so, does this scale change temporally? In this article, we investigate how the amount and location of bedrock and soil groundwater discharge vary as a function of space and time in a montane, first-order watershed in west-central Montana. The study objectives are to constrain the amount of bedrock vs. soil groundwater discharge along a stream reach as a function of seasonal changes in catchment wetness and increasing spatial scale. Leveraging a high-resolution network of gauging stations, we conducted a multi-tracer interrogation of the incremental sources of streamflow generation. We estimate the total volume of subsurface discharge and distinguish the contributions of soil and bedrock groundwater discharge along the stream. We then explore relationships of the bedrock groundwater contribution as a function of the incremental and integrated upland characteristics.

Study area

The Cap Wallace Watershed (CWW) is located in west-central Montana, ~56 km northeast of Missoula on the north slope of the Garnet Range (Figure 1). The catchment drains 6 km^2 of forested land and is a tributary of the Blackfoot River. The CWW ranges in elevation from 1,163 m at the confluence of Cap Wallace Creek (CWC) and Elk Creek to 1,918 m at the ridge tops. Vegetation consists primarily of second-growth western larch (*Larix occidentalis*) on the north-slope mid and upper elevations, Douglas-fir (*Pseudotsuga menziesii*) on north-facing, low-elevation slopes, and Ponderosa Pine (*Pinus ponderosa*) on south-aspect slopes. North-aspect slopes have well-developed soils, whereas south-aspect slopes lack soil coverage and are generally covered by steep talus fields. Due to topographic variation in the radiation balance and resulting microclimatic effects (Holden et al., 2011a,b), south-aspect slopes remain significantly drier throughout the year (Hoylman et al., 2019).

At the Lubrecht Flume SNOTEL station (site #604; 1,425 m), the average annual precipitation over the record period of 10 years is 514 mm. Nearly 50% of precipitation falls as snow, classifying the Cap Wallace as a snowmelt-dominated watershed. In 2017, temperatures ranged from −27.6°C to 34.2°C with an annual average temperature of 6.1°C (Figure 2). Peak snow water equivalent typically occurs in late March, with snow melt taking place during April, May, and early June. May and June are typically cool and rainy, while July, August, and September are generally hot and dry, with precipitation beginning again in October.

The Garnet Range is part of the Northern Rocky Mountains and was formed through folding and thrusting associated with the Jurassic Sevier and Cretaceous Laramide Orogenies (140–55 Ma). Subsequent extensional normal faulting was initiated in the early Eocene as a result of Basin and Range tectonics, triggering the uplift of metamorphic core complexes in the Northern Rockies (Portner et al., 2011). The CWW is situated on the flank of a large granitic batholith called the Garnet Stock, which intruded the area during the Late Cretaceous (Lonn et al., 2010). CWC follows an east–west trending normal fault. Quaternary glacial

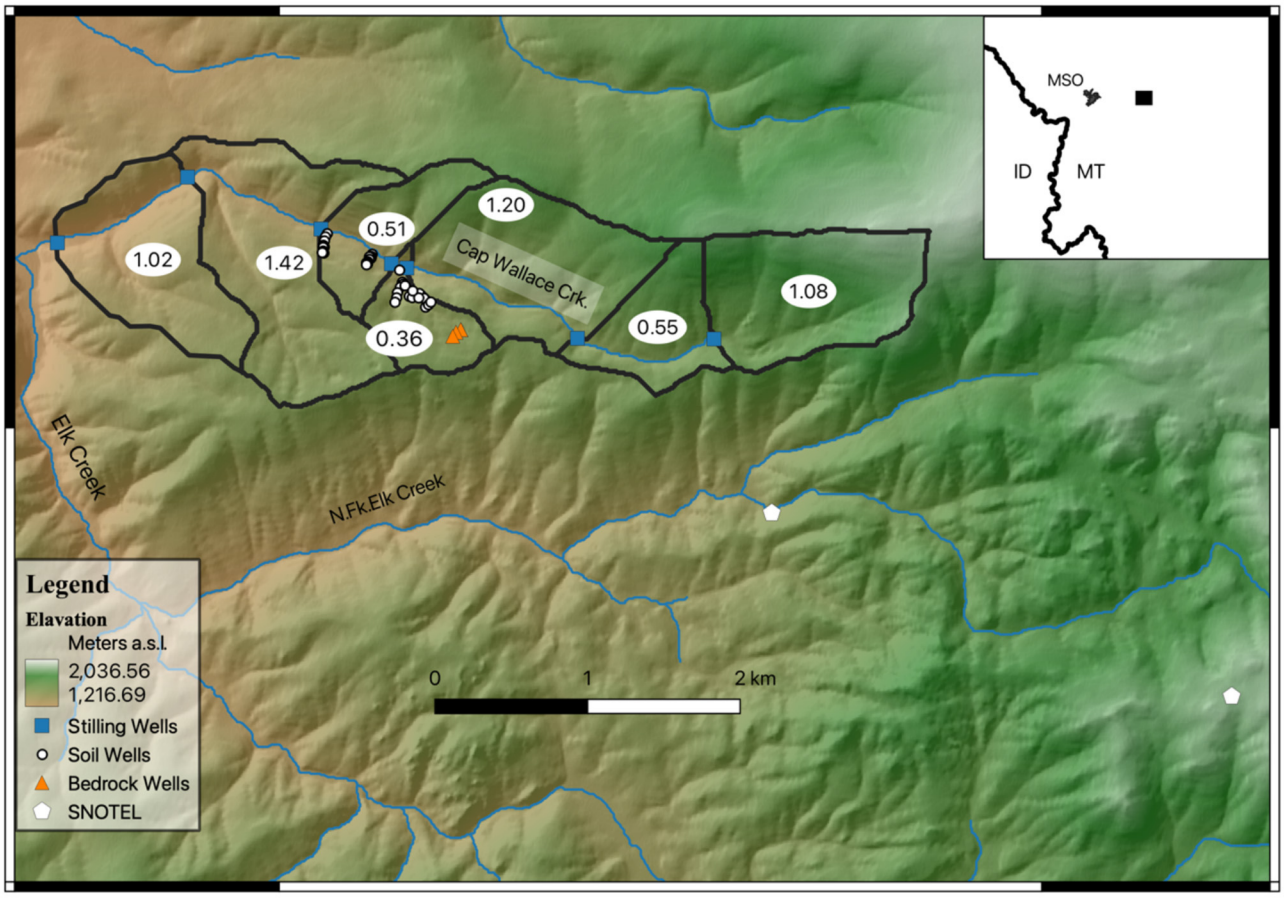


FIGURE 1
Map of the Cap Wallace watershed on the north slope of the Garnet Range. Outlines indicate sub-watersheds, with drainage areas in km^2 given in white ellipses. The inset map shows east-central Idaho (ID) and west-central Montana (MT) with the relative location of Missoula (MSO).

and alluvial deposits cover the narrow valley bottom and stream floor, while the hillslopes are composed of gravelly, silt-loam soils (Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, n.d.) that overlie several kilometers of argillites, siltites, and quartzites of the Precambrian metasedimentary Belt Supergroup. Soils are classified as Typic Haplustalfs (well-draining with a lower clay horizon), with depths ranging from 0.5 to 2 m that are thicker in hollows and thinner in upper elevation ridges (Hoylman et al., 2018). The north-aspect slope is composed of the Garnet Range Formation, a siltstone with hummocky sand lenses, which lies unconformably across the valley from the stratigraphically lower Bonner Quartzite, a massive, slightly metamorphosed, arkosic sandstone that makes up the south-aspect slopes (Lonn et al., 2010). These formations have been subject to tectonism associated with compression from major mountain-building events and subsequent active extension due to gravitational relaxation, leading to a dense fracture network within the bedrock (Brenner, 1968; Gardner et al., 2020).

Methods

We measured multiple environmental tracers in soil, groundwater, and stream water to perform hydrograph separation along CWC. We conducted synoptic sampling along CWC for discharge of the stream and dissolved ^{222}Rn and major ions at four hydrologically important times: (1) early-spring baseflow (March), (2) the rising limb of spring snowmelt freshet (April), (3) the falling limb of freshet (May–early June), and (4) summer baseflow (late June–October).

Field methods

Cap Wallace Watershed is instrumented with 36 shallow soil wells, three groundwater wells, and seven stilling wells for measuring stream discharge. Shallow soil wells are 1–2 m deep, fully screened, and pounded with a steel rod until refusal. Bedrock groundwater wells were drilled to depths of 10–20 m using a man-portable drilling rig. Bedrock wells were fully screened, and the annulus was filled with the sand pack from the bottom to 7 m below the land surface and then cased and the annulus sealed with bentonite to the land surface. Stilling wells were spaced at ~1 km intervals, strategically located downstream of converging sub-watersheds with large contributing areas to estimate the hillslope contribution from sub-watersheds along the stream (Figure 1). Field data collection began during the winter of 2017. Solinst pressure transducers were installed in stilling, soil,

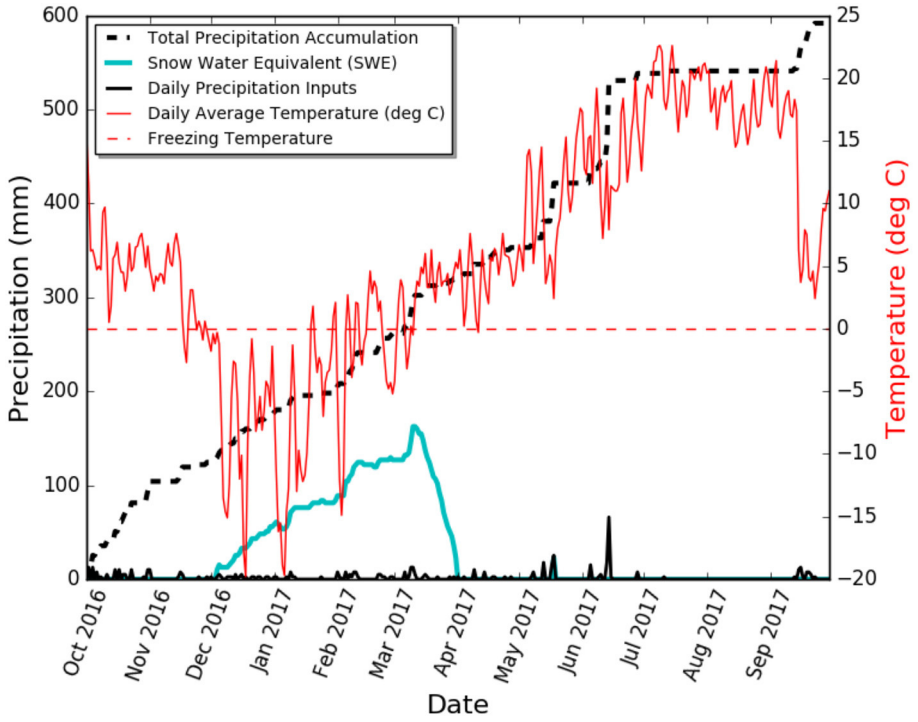


FIGURE 2
Lubrecht climate data for the 2017 water year. Data source: Lubrecht SNOTEL (site #604). 2017 was a heavy precipitation year with many days below freezing.

and groundwater wells to continually monitor and record stage measurements along the stream and contributing hillslopes.

Throughout the spring, summer, and fall of 2017, synoptic discharge measurements were made repeatedly at each stilling well using dilution gauging (Gooseff and McGlynn, 2005). Replicate measurements were collected at multiple stilling well locations during each sampling period. On average, replicates of stream discharge differed by ±4% of one another, but during baseflow, one replicate recorded a discharge value of 54% lower than the original measurement (28 August 2017). There are several potential reasons for this discrepancy. In particular, this measurement was collected at one of the lowest recorded flows. Low flows have been shown to complicate the dilution gauging technique when hyporheic exchange occurs (Moore, 2004). Because this replicate was identified as an outlier, this baseflow measurement was not used in our determination of the average uncertainty in replicate measurements (5%); however, it is clear that as we approach low flows, our ability to accurately measure the flow is reduced.

During synoptic discharge measurement campaigns, each stilling well location was sampled for ^{222}Rn and major ions (Ca^{2+}, K^{+}, Mg^{2+}, Na^{+}, F^{-}, Cl^{-}, NO_3^{-}, SO_4^{2-}). Sample bottles were rinsed three times with sample water. Samples for ^{222}Rn analysis were collected underwater in airtight 250 mL glass bottles to avoid exchange with the atmosphere. Major ion samples were filtered using a 0.45-μm disposable filter and stored in acid-washed polyethylene bottles. Cation samples were acidified with 0.2 mL of nitric acid. Field parameters, including pH, temperature, oxidation-reduction potential, and electric conductivity, were collected using an *In-Situ* Aqua TROLL 600 sonde. Precipitation samples were collected at the Jones Pond Montana Mesonet site ~10 km away and were measured throughout the year. Integrated snowmelt samples were collected during the melt season using a snow core to collect complete snowpack samples and melted in the laboratory refrigerator. Water samples were collected from soil and groundwater wells, using a peristaltic pump in April 2017. Prior to collection, wells were purged until field parameters remained constant (~2–3× the volume of the well).

Laboratory methods

^{222}Rn samples were analyzed within 48 h of collection using a Durridge Instruments RAD7 spectral alpha-decay detector. A sample of 250 ml was analyzed using the WAT-250 protocol, with 5 min of degassing (bubbling) followed by five rounds of counting, where each counting round was 5-min long. The RAD7 was purged for >20 min between each sample. Concentration was determined from the last four counting rounds (20 min total). The obtained concentrations were corrected to account for decay during the period between sample collection and analysis using the formula:

$$C_{tc} = C_i e^{\left(\frac{t}{132.4}\right)} \quad (1)$$

where C_{tc} is the time-corrected radon concentration in Bq/L, C_i is the initial radon concentration prior to correction (Bq/L),

and t is the time elapsed between collection and analysis in hours. Samples for the analysis of anions were run on a DIONEX DX500 (IC) and cations on a Perkin-Elmer OPTIMA 5300 (ICP-OES). Detection limits and accuracy for IC and ICP-OES analyses are presented in Table 1. All ^{222}Rn and major ion samples were analyzed at the University of Montana's Environmental Biogeochemical Laboratory.

Modeling approach and data analysis

Two different modeling techniques were used to identify and estimate the sources of streamflow. First, in-stream radon concentrations were used to estimate total subsurface discharge. End-member mixing analysis (EMMA) was then conducted on major ions in stream, soil, snowmelt, and groundwater to estimate the appropriate number and type of end members. The obtained end members were then used in a mass-balance mixing model to quantify end-member contributions to the stream. Mixing fractions were compared with landscape topographic characteristics. The effects of different internal landscape characteristics on spatial heterogeneity and duration of subsurface flow were quantified seasonally at spatial scales ranging from 0.4 to 6 km^2.

Radon modeling

^{222}Rn is advantageous as an analyte because it is not present in the atmosphere and is non-existent in overland flow, direct snowmelt, and precipitation. Meanwhile, all subsurface waters with a residence time of over a few hours have measurable ^{222}Rn, and all waters with a residence time over ~2 weeks achieve secular equilibrium and have relatively constant ^{222}Rn concentration for a given lithology. Thus, ^{222}Rn concentrations can be used to isolate subsurface streamflow sources. Estimating subsurface inflow from ^{222}Rn requires a model that accounts for subsurface discharge, gas exchange with the atmosphere, and radioactive decay. A 1D stream transport model was created to simulate longitudinal radon activity (Cook et al., 2003) using the equations:

$$\frac{\partial Q}{\partial x} = I(x) - L(x) - E(x) \tag{2}$$

$$Q\frac{\partial c}{\partial x} = I(c_i - c) + wEc - kwc - dw\lambda c \tag{3}$$

where c is the concentration of radon in the stream (Bq/L), c_i is the concentration in bedrock or soil groundwater (Bq/L), I is the groundwater inflow rate (m^3/m/day), w is the stream width (m), d is the mean stream depth (cross-sectional area/width) (m), k is the gas exchange velocity (m/day), λ is the radon decay coefficient (per day), Q is the stream discharge (m^3/day), E is the evaporation rate (m/day), and L is the stream extraction rate from pumping or diversions (m^3/m/day). Equation 2 is the mass balance equation for water in the stream, and Equation 3 is the 1D steady-state equation for advective solute transport in the stream with groundwater inflow, atmospheric gas exchange, and in-stream radioactive decay. We assumed an atmospheric ^{222}Rn concentration of zero, no production of ^{222}Rn within the stream, and steady-state flow conditions.

The ultimate goal of the ^{222}Rn fitting was to assess the amount of subsurface-derived discharge vs. overland discharge, assuming that soil and groundwater with residence times over 2 weeks have similar radon concentrations and overland flow has zero radon concentration. The distributed total groundwater discharge (soil plus bedrock) was estimated by fitting the observed ^{222}Rn stream discharge. Radon and discharge values were fit by varying stepwise groundwater discharge (Gardner et al., 2011). Groundwater inflow was assumed to be constant over 500 m steps, which resulted in 10 unknown groundwater inflow steps along the stream reach. The number and location of groundwater steps were held constant at all sampling times. Steps in the radon concentration correspond to boundaries of the stepwise constant groundwater inflow. The total number of observations for inversion is 14, resulting in an overdetermined system.

To quantify incremental gains from subsurface sources at each stilling well, we calculated the ratio of subsurface discharge to total streamflow gain (F_{ss}):

$$F_{ss} = \frac{Q_{ss_1} - Q_{ss_2}}{\Delta Q_t} \tag{4}$$

where Q_{ss_1} is the upstream modeled subsurface discharge (L/s), Q_{ss_2} is the downstream modeled subsurface discharge (L/s), and ΔQ_t is the measured difference in streamflow from upstream to downstream (L/s).

The parameters used for radon modeling are shown in Table 2. Those that were not directly measured in the field were estimated based on a series of equations also detailed in Table 2. Following the approach of Cook et al. (2003), a sensitivity analysis was conducted by varying each individual model parameter by ±50%, while holding the other parameters constant to observe the associated effect on estimated subsurface inflow. Particular interest was given to the parameters not measured in the field since they were the largest source of uncertainty in the model.

As the gas transfer velocity (k) is difficult to constrain in low-order mountainous streams due to their variable geometries, flow velocities, and temperatures, special attention was given to the effect of k on estimated subsurface inflow rates. Four common equations that depend on slope, velocity, and depth (Raymond et al., 2012) were used to approximate the gas transfer velocity range.

End-member mixing analysis

End-member mixing analysis was performed following the approach of Hooper (2003). Details of the specific application of EMMA in this article can be found in Supplementary material. Here, we used four tracers (calcium, magnesium, sodium, and chloride, see Results) to constrain a two-end-member mixing model. The end members chosen were the reprojected average bedrock and soil groundwater compositions from all samples throughout the year. End-member concentrations were held constant over time. However, most bedrock and all soil groundwater samples were collected in spring when water tables

TABLE 1 Limits of detection and precision of field and laboratory duplicates for ion chromatography (IC) and inductively coupled plasma emission spectroscopy (ICAP-OES) analyses.

Parameter	Ca	K	Mg	Na	F	Cl	SO_4	NO_3
Limit of detection (mg/L)	0.102	0.51	0.102	0.51	0.015	0.100	0.100	0.100
Error % (lab)	±1.35	±1.40	±1.20	±1.12	±3.86	±2.31	±1.78	±0.21
Error % (field)	±3.20	±1.80	±3.90	±3.30	±6.57	±1.86	±0.19	±5.14

Errors associated with field and laboratory duplicates are within acceptable range for geochemical interpretation.

TABLE 2 Details of the parameters used in ^{222}Rn modeling.

Parameter	Units	Value	Method of collection or estimation
In-stream radon concentration (c)	Bq/L	Avg. 0.907; 0.074–4.64 ±15%	Measured in field and analyzed using the RAD7 Alpha Decay Spectrometer at the University of Montana; spatially and temporally variable
Total river length (x)	km	5	Measured in field
Evaporation rate (E)	mm/day	0	Assumed negligible
Initial radon concentration (c)	Bq/L	Avg. 0.37; 0.12–0.60 ±15%	Concentration of most upstream sampling location (variable); measured in field
Atmospheric radon concentration	Bq/L	0	
Radon decay coefficient (λ)	/day	0.18	Constant
Gas transfer velocity (k)	m/day	Avg. 16.6; 5.5–27.7	Estimated from the study of Raymond et al. (2012)
Groundwater inflow Concentrations (c_i)	Bq/L	32.0	Highest measured concentration in groundwater samples; measured in field and analyzed using the RAD7 Alpha Decay Spectrometer at the University of Montana
Stream width (w)	m	0.75	Mean value of widths measured in field
Stream depth (d)	m	0.15	Mean value of depths measured in field

were shallow enough to be sampled. Our end members are referred to here as soil groundwater and bedrock groundwater, where soil water represents shallow groundwater in the soil zone, and groundwater represents bedrock groundwater in fractured bedrock. We estimated uncertainty for the EMMA results by propagating the error due to EMMA misfit through the mixing calculation details, which can be found in Supplementary material. Error bars on mixing plots are given as the mean relative error $\frac{\bar{\delta}Q}{\bar{Q}}$ for all samples, which was 13%.

Terrain analysis

To explore the topographic characteristics that influence streamflow generation, the catchment was delineated into "sub-catchments." All-terrain analyses were performed using the SAGA GIS software (Conrad et al., 2015). A 10-m resolution digital elevation model of the study area had sinks filled, and the stream network burned in. The CWW boundary was delineated by creating flow direction and accumulation grids to determine the contributing area for each stilling well. The watershed was then divided into seven sub-catchments based on those areas. In the process, each sequential downstream sub-catchment integrated all upstream sub-catchments; we refer to these sub-catchments as "accumulated" sub-catchments. "Incremental" sub-segments contributing between stilling wells were determined by subtracting the upstream sub-watershed from the downstream sub-watershed (Figure 1). Mean elevation, Topographic Wetness Index (TWI), and upslope accumulated catchment area (UAA) were calculated for each accumulated and incremental sub-catchment. Linear regression was then performed to assess the modeled end-member discharge correlation to terrain attributes for both the incremental and accumulated areas.

To analyze the relationship between landscape and end-member discharge in an incremental sense, fractions of end-member discharge gained in each sub-catchment were calculated by quantifying the solute concentration of water entering the incremental reach (ΔC) at each stilling well along the reach using the following formula:

$$\Delta C = \frac{Q_d C_d - Q_u C_u}{\Delta Q} \tag{5}$$

where Q_d is downstream discharge (L/s), C_d is downstream solute concentration (mg/L), Q_u is upstream discharge (L/s), C_u is the upstream solute concentration (mg/L), and ΔQ is the net change in discharge from the upstream stilling well to the downstream stilling well (L/s). Incremental solute concentrations, such as calcium, magnesium, sodium, and chloride, were calculated for each tracer used in mixing models. These solute concentrations were applied in mass-balance mixing models (Supplementary material) to quantify fractional gains in groundwater and soil water from one sub-catchment to the next. Modeled end-member fractions were compared against the mean elevation and TWI values as well as the area for each incremental sub-catchment to evaluate the relationships between internal catchment characteristics and discharge. To evaluate these

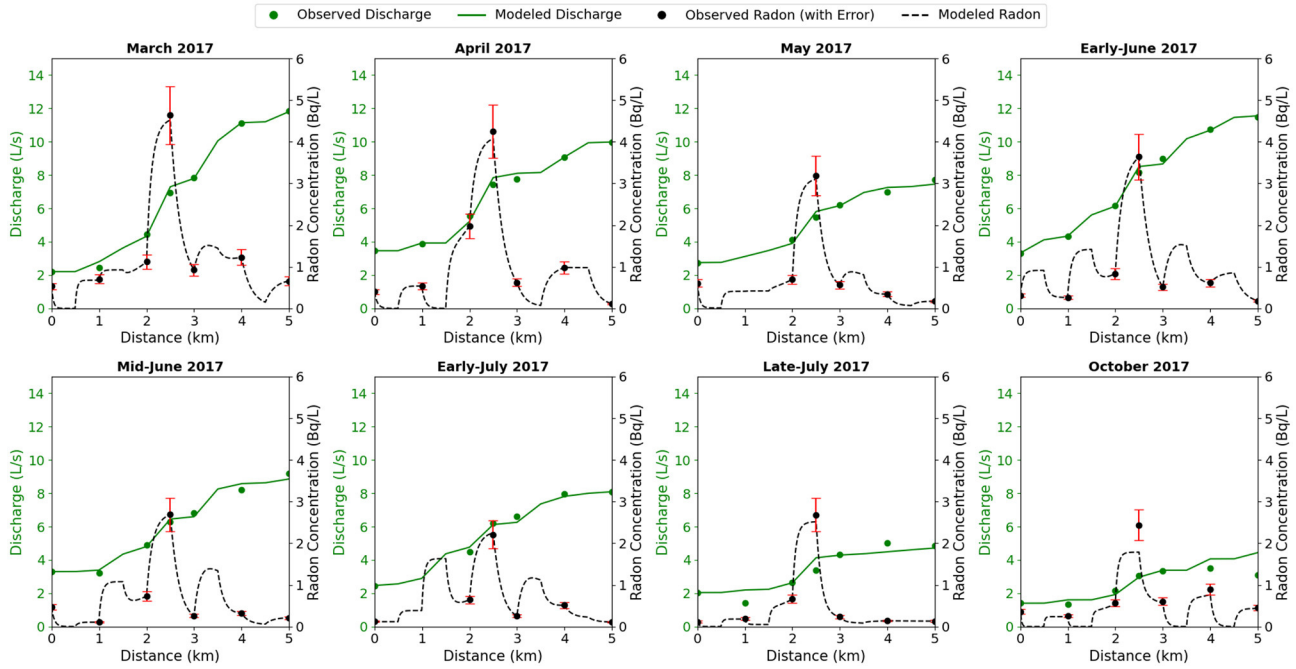


FIGURE 3
Longitudinal profiles of measured and modeled ^{222}Rn concentration and stream discharge. Black and green dots represent measured in-stream ^{222}Rn concentration and discharge, respectively. Dashed black and green lines represent modeled ^{222}Rn concentration and streamflow. Error bars of 15% added as an average field of duplicate errors.

relationships for the accumulated area, the elevation, TWI, and UAA values for the sub-watershed draining to each stilling well were compared to the original EMMA-modeled groundwater and soil water contributions at each stilling well location. Both accumulated and incremental analyses were conducted for each synoptic sampling campaign in order to evaluate seasonal change.

Hydrometric data from soil wells at various landscape positions were used as a proxy for determining when the catchment was "wet," as the data indicated the interconnectedness of the water table over time. Full hydrographs for soil and groundwater over the study period can be found in Gardner et al. (2020). Soil water connection was evaluated through binary plots of saturated vs. unsaturated conditions.

Results

Radon concentrations ranged from 0.074 to 4.25 Bq/L and showed the same general spatial pattern across varying discharges (Figure 3). During each sampling period, a significant increase in radon activity was observed approximately halfway down the observed reach (~2.5 km). This increase coincides with the location of a large contributing hillslope area from the south (Figure 1). The difference between duplicate samples ranged from 0 to 40%, with low-concentration samples exhibiting larger variability, consistent with Poisson statistics of radioactive decay. For example, at a concentration of 0.074 Bq/L (corresponding duplicate 0.111 Bq/L), there is a 40% variability, while at a concentration of 2.527 Bq/L (duplicate 2.597 Bq/L), there is a 2.7% variability.

Figure 4 shows total subsurface discharge estimates along with modeled and observed total stream discharge. The longitudinal profile of the accumulated subsurface discharge closely follows the increase in total stream discharge. For each interval, subsurface discharge accounts for 19–100% of the incremental gains in streamflow, with an average of 81% (Figure 5). In May and mid-June, the months of highest rainfall, the influence of non-subsurface discharge is strongest, with subsurface discharge accounting for, on average, 68% and 59%, respectively, of the incremental gains in streamflow. Non-subsurface discharge generation estimates predominantly occur near the center of the reach (2.5–3 km downstream of the headwaters) at the convergence of two large hillslopes.

Reprojected mean groundwater and soil water end-member concentrations are provided in Table 3. The two end members chosen, spanned the sample variance and the residual error in modeled vs. observed concentration showed no structure. The fractions of soil and bedrock groundwater in stream water samples are shown in Figure 6. The bedrock groundwater discharge fraction ranged from 0 to 0.45, with an average of ~0.2. In general, bedrock groundwater becomes a larger component of streamflow as the distance downstream increases, with the highest fraction of groundwater discharge occurring at the farthest downstream gauging location. Figure 7 shows the percentage of bedrock groundwater at the outlet as a function of time. Groundwater at the outlet fluctuated between 26 and 44% and averaged 38% of total streamflow. The fraction of bedrock

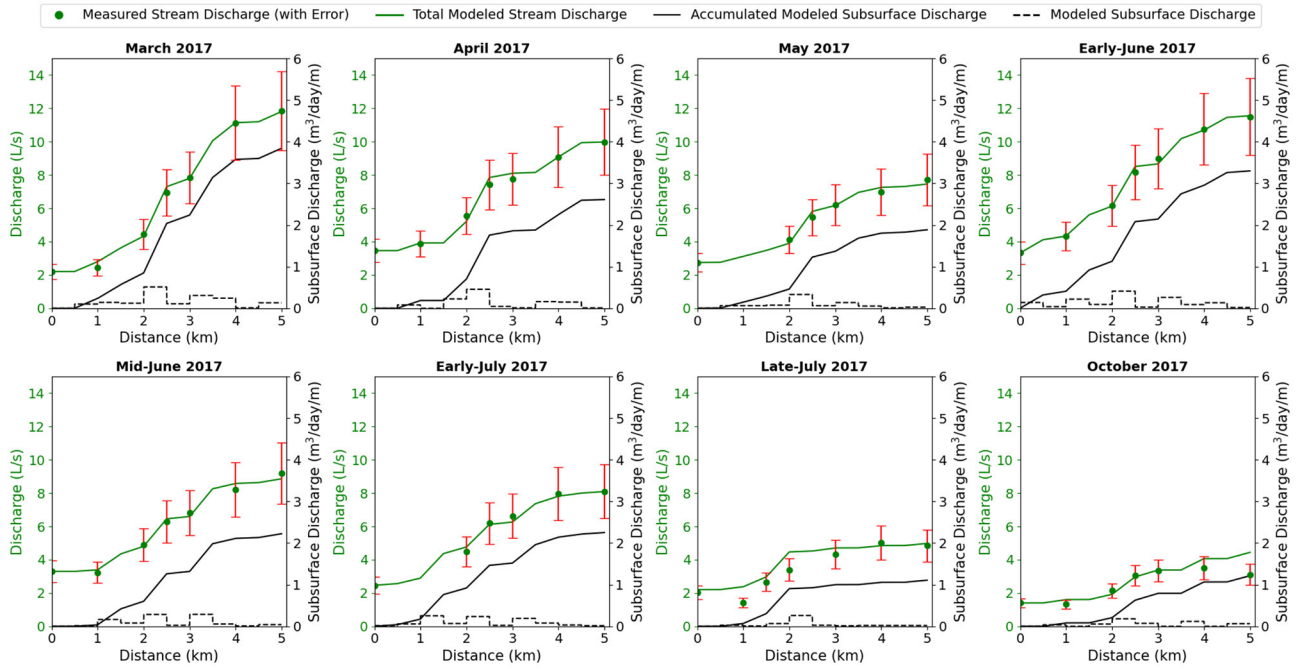


FIGURE 4
Longitudinal profiles of stream discharge. The green line, green dots, and black line correspond to the primary y-axis and represent accumulated modeled total stream discharge, measured stream discharge at each stilling well location, and accumulated modeled subsurface discharge along the reach. The dashed black line corresponds to the secondary y-axis and shows modeled subsurface inflow in $m^3/m/d$. Error bars of 20% added as an average field duplicate error.

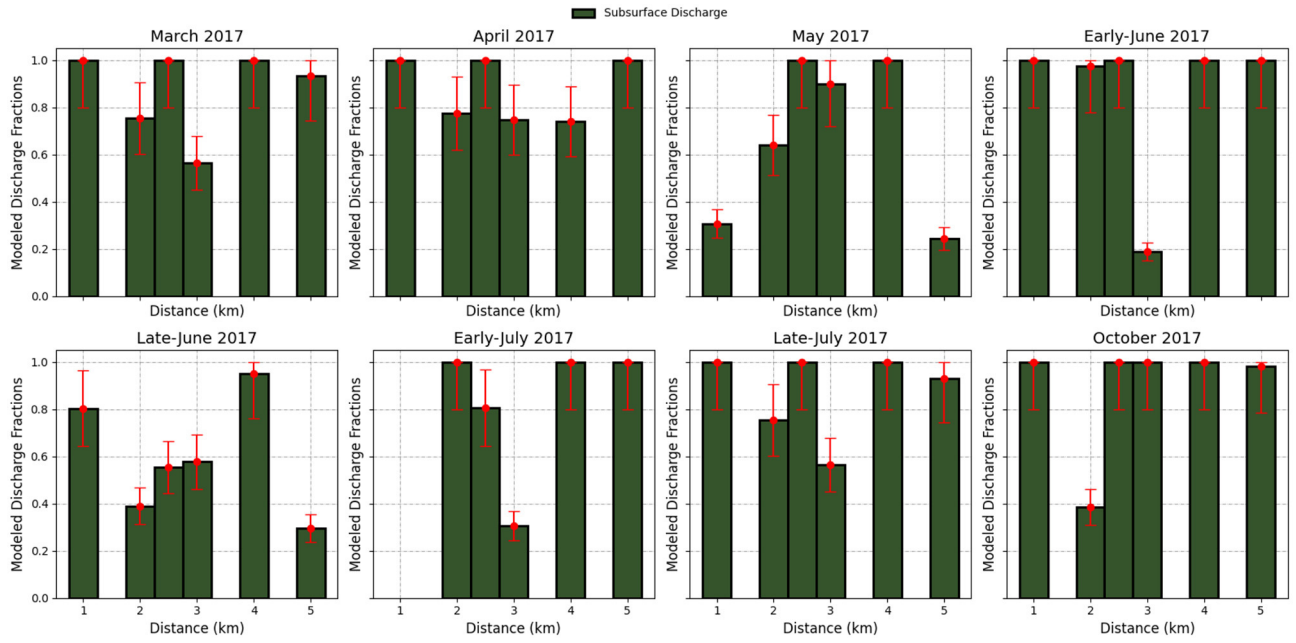


FIGURE 5
Incremental gains in modeled streamflow as a fraction of subsurface to total streamflow at each gauging location along the reach. The darker green color indicates the fraction of the accumulated discharge accounted for by a subsurface discharge between gauging station locations. Error bars estimated from 20% uncertainty in groundwater discharge estimate using our Rn transport model (Goble, 2018).

groundwater to total streamflow at the outlet was highest during the late spring (2 June 2017) and lowest during mid-spring (19 May 2017).

Table 4 summarizes linear regression results for incremental and accumulated terrain characteristics with the bedrock groundwater mixing fraction. Linear regression of cumulative

TABLE 3 Solute concentrations used in EMMA mass-balance mixing models.

Tracer concentration (mg/L)	Groundwater	Soil water
Calcium	7.94	2.72
Chloride	1.22	1.30
Magnesium	3.93	1.84
Sodium	8.47	2.39

mean terrain indices and soil and bedrock groundwater fractions revealed strong relationships with both cumulative average elevation and cumulative upslope accumulated area (Figure 8, Table 4). R^2 values ranged from 0.52 to 0.98 over the 7-month study period. While the R^2 values ranged from 0.05 to 0.81 for the incremental terrain characteristics, they tended to switch drastically from high to low values throughout the year. No systematic trend was found relating incremental end-member discharges with incremental TWI and area (Figure 9, Table 4).

Figure 10 plots the r^2 values from cumulative and incremental regression with binary soil well responses as a function of time. Soil wells with high TWI values tended to stay continually saturated, whereas those with lower TWI values exhibited more transient-like behavior in response to climatic forcing. In both cases, the r^2 value was highest (0.98 and 0.81, respectively) in early July, ~3 weeks after the largest catchment "wet-up."

Discussion

Radon results indicate the majority of runoff in CWC is derived from subsurface sources (soil and bedrock groundwater) throughout the year, even during the snowmelt cycle, consistent with prior observations in alpine catchments (Liu et al., 2004). These results may be expected in later parts of the season (late summer and fall) when the overland flow is expected to be zero in our catchment. However, the fact that subsurface flow dominates during snowmelt (March, April, and May) indicates that direct runoff through the snowpack is not a major source of water for our stream. Radon results are consistent with soil hydrometric data. Soil wells with low TWI values (generally located at higher elevations with less accumulated area) exhibit transient behavior to catchment wet-ups (Figure 10) (Gardner et al., 2020). Some wet-ups had durations on the order of hours to days, suggesting that soil water in these landscape positions is rapidly lost through lateral inter-flow and/or infiltration to the bedrock aquifer.

While incremental discharge source plots show that subsurface sources account for the majority of streamflow gains, there are some instances where non-subsurface sources are estimated to be more important in streamflow generation (Figure 5). Non-subsurface streamflow generation occurred throughout the year but had the strongest influence on stream chemistry during the late spring to mid-summer (May through late July). These instances occurred primarily in the middle section of the reach, where multiple convergent zones from large hillslopes converge (Figure 3). Hydrometric data from soil wells indicate that the catchment was at its most saturated state during the May and June periods (Figure 10). Thus, during May and June, we hypothesize that non-subsurface sources could be due to saturation excess and overland flow in these large contributing drainage bottoms or soil and/or preferential/piping flow in the unsaturated zone. The fact that our radon modeling indicates some non-subsurface flow during the baseflow periods (late July and October) is likely due to model error. A probable source of error is the discharge of short residence time flow paths that have low Rn.

End-member mixing analysis provided an independent test of the discharge source findings, along with the ability to further divide stream-flow composition. The results indicate that the stream is generally composed of shallow saturated soil water at the headwaters and that shallow soil water is the dominant source of streamflow generation, but bedrock groundwater discharge increases with downstream distance, even during periods of snowmelt and heavy rainfall. Our bedrock groundwater discharge fractions are generally comparable with those of other mountain aquifer studies (Somers and McKenzie, 2020). The increasing influence of bedrock groundwater with downstream distance argues for a 3D watershed conceptual model consistent with those proposed by previous researchers evaluating groundwater flow in mountainous catchments in the western U.S. (Carroll et al., 2018; Frisbee et al., 2011, 2012). However, this study demonstrates that this emergent relationship exists in headwater catchments >2 km^2 in scale.

Temporal analysis indicates surprisingly low variability in modeled end-member proportions throughout the observed time period; however, early summer months have the highest proportion of bedrock groundwater discharge. Several large precipitation events (~60 mm/day) occurred in early June. These precipitation events could have rapidly filled the subsurface reservoirs, which could, in turn, drive increased discharge through a variety of mechanisms, including groundwater ridging, preferential recharge, and high fractured bedrock celerity. Rapid groundwater reactions to precipitation inputs are consistent with the experimental observations of Anderson et al. (1997a). However, our observed late-season bedrock groundwater ratios of streamflow are surprisingly low. Several previous hydrograph separation studies have assumed that baseflow is entirely composed of groundwater (Genereux et al., 1993; Pinder and Jones, 1969). Indeed, our results indicate that shallow soil groundwater is a substantial portion of streamflow during the late season. These results could be explained by several different mechanisms. For example, the substantial difference in the hydraulic diffusivity of soil and fractured bedrock can lead to situations where the fractured bedrock discharge exhibits a rapid response to infiltration, which is then followed by a long-term storage release from porous soil (Anderson et al., 1997b). Additionally, soil interflow and groundwater recharge and discharge have been shown to be tightly coupled at the scales of hillslopes (Gardner et al., 2020) and catchments (Carroll et al., 2019), and this interaction is capable of sustaining groundwater recharge and soil water throughflow throughout the year (Gardner et al., 2020).

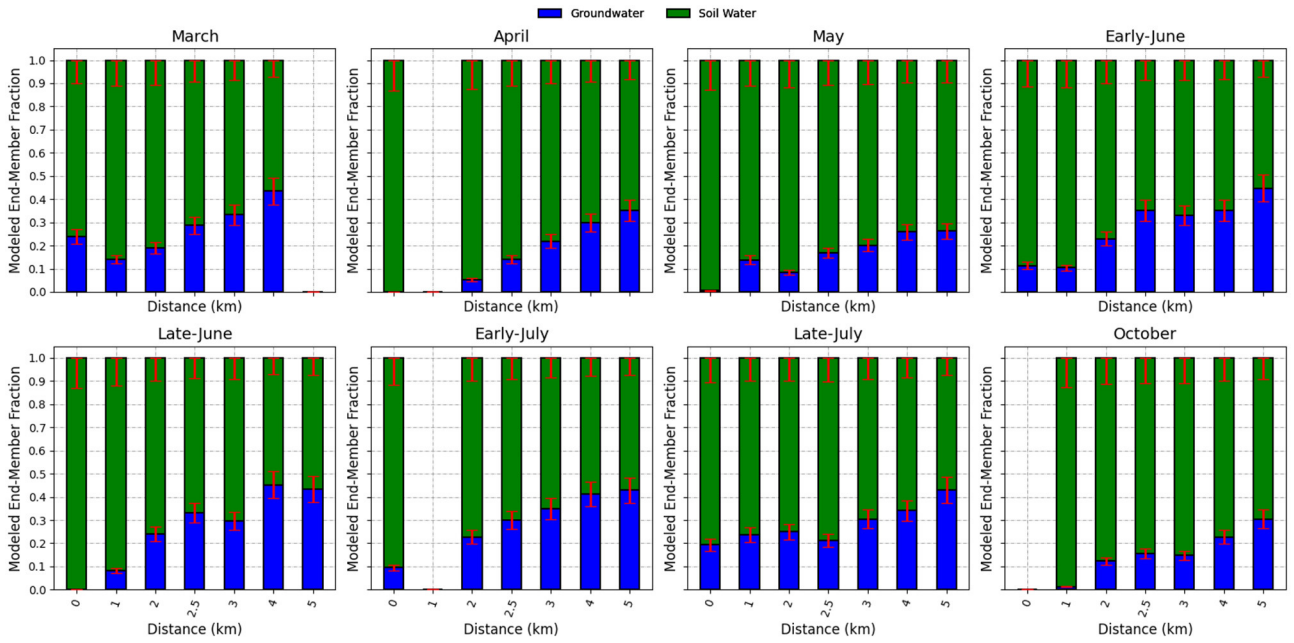


FIGURE 6
Longitudinal profiles of mixing model results showing how soil and bedrock groundwater were partitioned along the reach for each sampling campaign. Stream chemistry and discharge measurements were occasionally unavailable in circumstances where it would have been dangerous to collect them (i.e., the presence of bears), if samples became contaminated or if sample bottles were broken prior to chemical analysis.

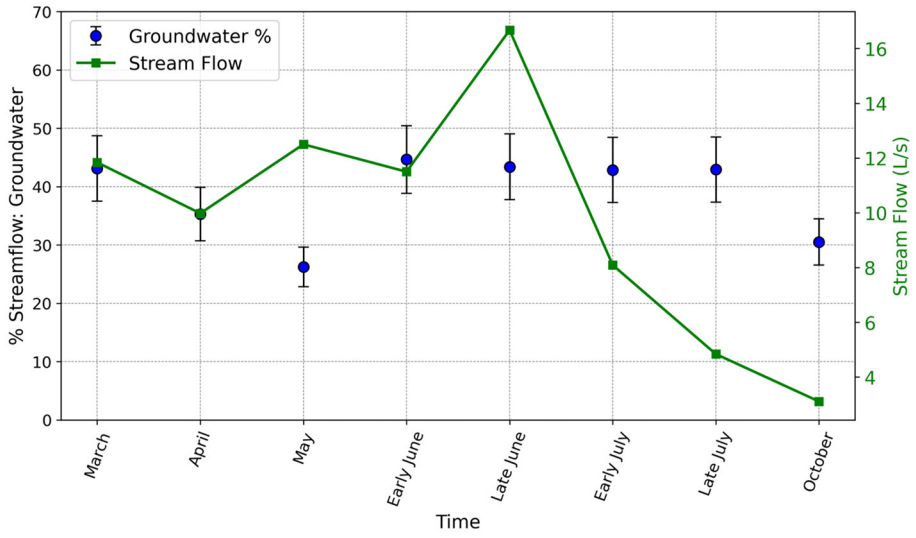


FIGURE 7
Seasonal fluctuation of the groundwater fraction of total streamflow at the outlet.

Soil groundwater remains the dominant source of subsurface discharge over our entire reach throughout the year. However, bedrock groundwater discharge becomes an increasingly higher fraction of discharge as the amount of accumulated area increases. These observed trends of high soil groundwater discharge and increasing bedrock groundwater contribution with increased catchment scale imply that streamflow represents an integration of individual hillslope responses and larger-scale deeper groundwater flow paths (Figure 11). We attribute the increasing contribution of bedrock groundwater discharge with increasing accumulated area due to the integration of more heterogeneity and higher groundwater discharge zones (Asano et al., 2022; Frisbee et al., 2011). However, we demonstrate this relationship at a landscape scale that is orders of magnitude smaller than what has been observed in previous research. The relationship was generally consistent across variable streamflow states, with small temporal inconsistencies that are likely associated with dynamic catchment properties, such

TABLE 4 Summary of r^2 values from cumulative and incremental regression against mean elevation, topographic wetness index, and sub-watershed area.

Type	Attribute	Mar	Apr	May	Early June	Late June	Early July	Late July	Oct
Incremental	Elev.	0.45	0.93	0.89	0.87	0.92	0.96	0.73	0.96
	TWI	0.47	0.22	0.59	0.78	0.93	0.74	0.35	0.69
	Area	0.09	0.05	0.00	0.00	0.04	0.00	0.11	0.23
Accumulated	Elev.	0.52	0.99	0.89	0.89	0.94	0.98	0.71	0.93
	TWI	0.66	0.96	0.80	0.95	0.97	0.98	0.62	0.90
	Area	0.71	0.91	0.80	0.86	0.89	0.92	0.85	0.96

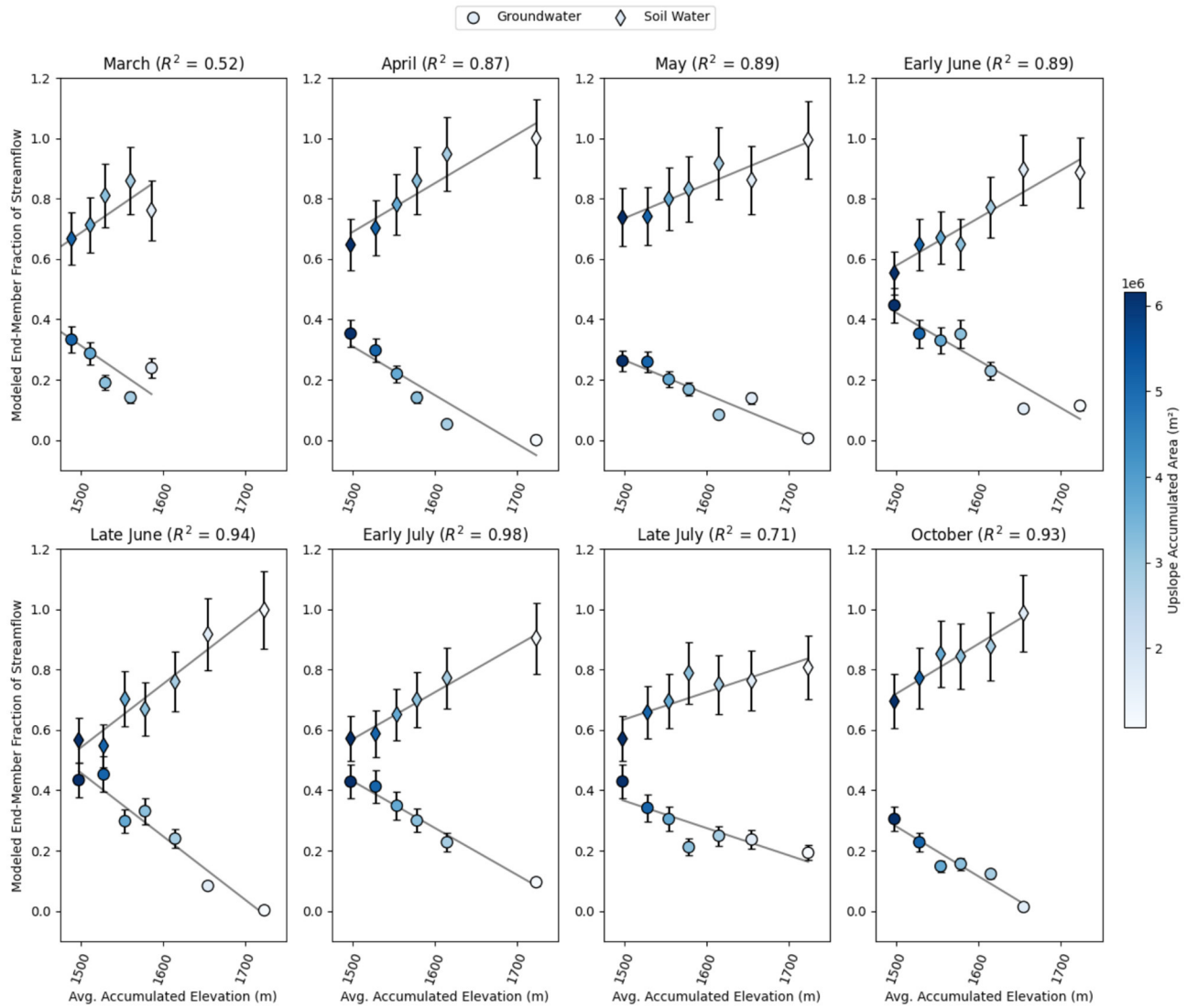


FIGURE 8
Regression results from a cumulative analysis of end-member discharge against elevation and UAA. Blue dots and diamonds are color-coordinated with cumulative mean UAA to show relative fractions of modeled groundwater and soil water discharge along the reach.

as climate and soil moisture, highlighting the importance of antecedent conditions.

The lack of trend in bedrock groundwater discharge with incremental sub-watershed area suggests that a minimum scale exists below which it is difficult to observe trends in bedrock groundwater discharge with terrain metrics. This lack of trend could result from two proposed processes: (1) regional groundwater flow paths could contribute to streamflow in a non-linear fashion

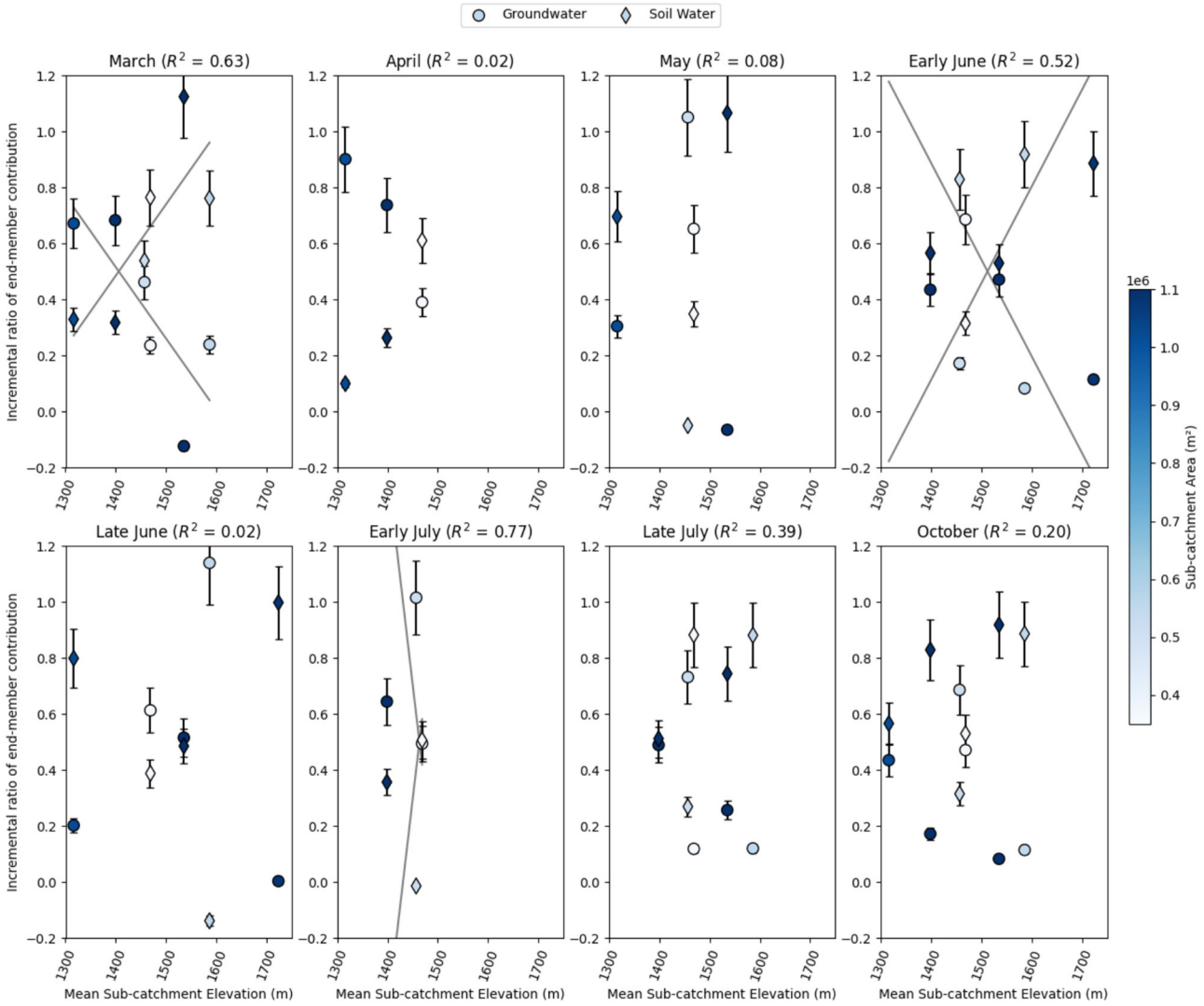


FIGURE 9
Regression results from the incremental analysis of end-member discharge against elevation and UAA. Blue dots and diamonds are color-coordinated with sub-catchment areas to show the relative incremental fractions of modeled groundwater and soil water discharge.

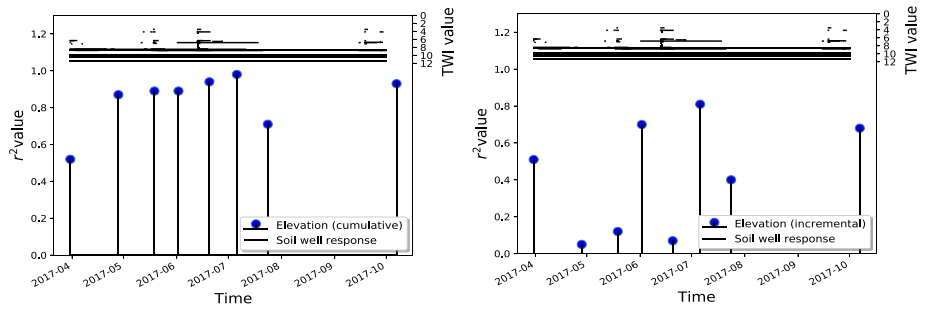


FIGURE 10
Blue dots show the r^2 values from cumulative **(Left)** and incremental **(Right)** regression of elevation values against modeled end-member fractions through time. Black lines near the top of the plot are binary soil well responses with corresponding TWI values in which the length of the line corresponds to the amount of time a soil well was saturated.

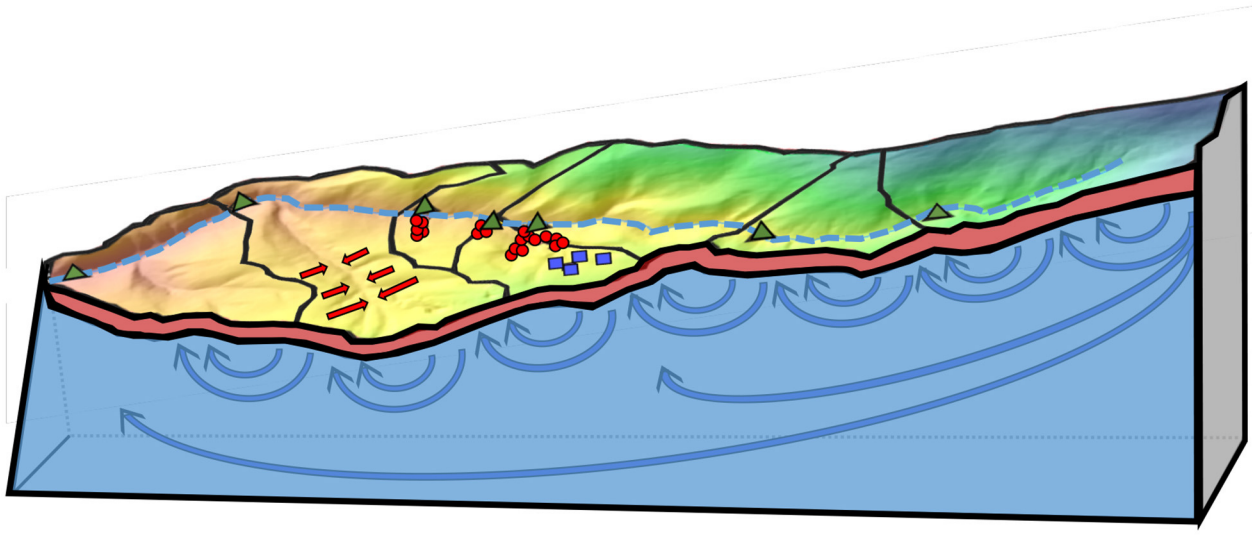


FIGURE 11
A 3D conceptual model of CWW showing the creek as an integration of hillslope surface and unsaturated slow (red arrows) and local (short blue arrow) regional groundwater (long blue arrow) flow paths that contribute to streamflow at lower elevations as well as individual hillslope responses. Stilling well locations are shown as triangles, groundwater wells as blue squares, and soil wells as red circles.

due to geologic heterogeneities and (2) valley-bottom exchange of water. Linear trends in modeled soil and bedrock groundwater contributions against elevation and upslope accumulated area should not necessarily be expected, as a sub-catchment high in the watershed could export its regional groundwater to locations lower in the watershed. Thus, in the case of active regional groundwater flow, low-elevation sites would receive disproportionally more groundwater than high-elevation sites, potentially explaining the observed correlations with accumulated catchment area.

Caveats and limitations

Our estimates of the bedrock groundwater discharge fraction are affected by the total flow and our error in gauging low flows. For example, on 8 June 2017, a 0.25-L/s increase in streamflow was recorded when going from stilling well "CWSTW6" (located ~2.5 km downstream from headwaters) to stilling well "CWSTW2-upper" (~3 km downstream from headwaters). Over this same reach, a 0.15-L/s increase in modeled subsurface discharge was calculated, implying that 60% of discharge gained was obtained from the subsurface while 40% was obtained from the surface or unsaturated sources. However, the observed 0.25-L/s increase is small compared to the uncertainty in discharge measurements. In addition, estimated subsurface inflows are subject to uncertainty in model parameters. Groundwater inflow estimates derived from ^{222}Rn are sensitive to the gas transfer velocity (k), which is also one of the least constrained parameters in the radon transport model. Varying the gas transfer velocity by ±50% has a marked effect on estimated groundwater inflow: a 50% decrease in ^{222}Rn would result in an estimated gain of 0.0 L/s, implying that 100% of streamflow comes from the surface, while a 50% increase in ^{222}Rn would result in an estimated gain of 0.45 L/s, which was more than what was actually measured. In addition, radon modeling is limited to detecting subsurface sources that have an appreciably high concentration of ^{222}Rn, which requires subsurface residence times of more than a couple of days, with secular equilibrium being reached only after ~2 weeks. It is possible that subsurface sources with brief residence times and thus low ^{222}Rn concentration could go undetected.

Overall, these uncertainties could contribute to uncertainty in the exact volume and proportion of subsurface discharge sources for a given reach, particularly at low flows for any one of our data interpretations. However, if the errors are of similar magnitude for each reach, then the overall pattern across the watershed should remain. In addition, we have the added benefit of combining discharge, soil hydrometric data, and multiple stream tracers, which helps minimize the influence of error in any single data source. Given these factors, we argue that our results are suitable for characterizing the watershed's behavior and developing conceptual models of streamflow generation.

Conclusion

In this study, multiple environmental tracers were used to estimate the discharge of soil and bedrock groundwater along a 5 km reach of a snowmelt-dominated mountainous catchment at a variety of hydrologically important times. ^{222}Rn modeling indicated that streamflow is generated predominantly from subsurface sources for all sampled periods. EMMA results indicated that streamflow chemistry could be explained by discharge from soil and bedrock groundwater. On average, bedrock groundwater comprised 38% of streamflow at the watershed outlet but fluctuated between a low of 26% in spring and a high of 44% in early summer. Bedrock groundwater discharge for individual stream segments was not well-correlated with landscape characteristics for incremental

sub-watersheds draining to the segment (01 km^2); however, it was well-correlated with the aggregated watershed landscape characteristics upstream of that segment ($>$ 2 km^2). EMMA-derived mixing models showed that headwaters are dominantly composed of soil groundwater and that bedrock groundwater becomes an increasingly more important component to streamflow at larger catchment scales with lower elevations and higher UAA values, suggesting 3D topography-driven flow. For the accumulated watershed characteristics, correlations among landscape and end-member discharge were strong across variable states of catchment "wetness" conditions, and accumulated elevation or total upslope area were robust predictors of groundwater discharge during all seasons (r^2 = 0.52–0.98). Groundwater discharge can be expected to constitute a significant source of streamflow generation in upland and montane catchments throughout the year. Bedrock groundwater flow paths follow a 3D nested flow pattern where recharge originating in upper reaches can discharge in further downstream reaches according to the heterogeneous subsurface architecture. Groundwater discharge locations are strongly controlled by local heterogeneities and cannot be well-predicted from the incremental contributing area at scales below 1 km^2. As the watershed scale exceeds 2 km^2, the aggregated average groundwater discharge from these 3D flow paths increases with an increase in drainage area. These findings could be useful to other researchers in developing conceptual models of streamflow generation in other watersheds.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

IV: Writing – original draft, Writing – review & editing. KJ: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. ZH: Conceptualization, Investigation, Software, Writing – review & editing. RL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. WG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This research was funded in part by DOE DE-NE0008786, DE-SC0021088, and the Montana Wyoming Water Center faculty seed grant program.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2025.1539177/full#supplementary-material

References

Anderson, S. P., Dietrich, W. E., Montgomery, D. R., Torres, R., Conrad, M. E., and Logue, K. (1997a). Subsurface flow paths in a steep, unchannelled catchment. *Water Resour. Res.* 33, 2637–2653. doi: 10.1029/97WR02595

Anderson, S. P., Dietrich, W. E., Torres, R., Montgomery, D. R., and Loague, K. (1997b). Concentration-discharge relationships in runoff from a steep, unchanneled catchment. *Water Resour. Res.* 33, 211–225. doi: 10.1029/96WR02715

Appels, W. M., Graham, C. B., Freer, J. E., and Mcdonnell, J. J. (2015). Factors affecting the spatial pattern of bedrock groundwater recharge at the hillslope scale. *Hydrol. Process.* 29, 4594–4610. doi: 10.1002/hyp.10481

Asano, Y., Suzuki, S. N., and Kawasaki, M. (2022). Peak discharges per unit area increase with catchment area in a high-relief mountains with permeable sedimentary bedrock. *J. Hydrol.* 610:127876. doi: 10.1016/j.jhydrol.2022.127876

Asano, Y., and Uchida, T. (2012). Flow path depth is the main controller of mean base flow transit times in a mountainous catchment. *Water Resour. Res.* 48:W03512. doi: 10.1029/2011WR010906

Asano, Y., Uchida, T., and Ohte, N. (2002). Residence times and flow paths of water in steep unchannelled catchments, Tanakami, Japan. *J. Hydrol.* 261, 173–192. doi: 10.1016/S0022-1694(02)00005-7

Beisner, K. R., Gardner, W. P., and Hunt, A. G. (2018). Geochemical characterization and modeling of regional groundwater contributing to the Verde River, Arizona between Mormon Pocket and the USGS Clarkdale gage. *J. Hydrol.* 564, 99–114. doi: 10.1016/j.jhydrol.2018.06.078

Bergstrom, A., Jencso, K., and McGlynn, B. (2016). Spatiotemporal processes that contribute to hydrologic exchange between hillslopes, valley bottoms,

and streams. *Water Resour. Res.* 52, 4628–4645. doi: 10.1002/2015WR017972

Brenner, R. L. (1968). *The Geology of Lubrecht Experimental Forest*. Missoula, MT: The Montana Forest and Conservation Experiment Station, 71.

Carroll, R. W., H., Bearup, L. A., Brown, W., Dong, W., Bill, M., and Willlams, K. H. (2018). Factors controlling seasonal groundwater and solute flux from snow-dominated basins. *Hydrol. Process.* 32, 2187–2202. doi: 10.1002/hyp.13151

Carroll, R. W. H., Deems, J. S., Niswonger, R., Schumer, R., and Williams, K. H. (2019). The Importance of Interflow to Groundwater Recharge in a Snowmelt-Dominated Headwater Basin. *Geophys. Res. Lett.* 46, 5899–5908. doi: 10.1029/2019GL082447

Carroll, R. W. H., and Williams, K. H. (2018). Data from: Discharge data collected within the East River for the Lawrence Berkeley National Laboratory Watershed Function Science Focus Area (Water years 2015-2018). *ESS-DIVE*. doi: 10.21952/WTR/1495380

Condon, L. E., and Maxwell, R. M. (2019). Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion. *Sci. Adv.* 5:eaav4574. doi: 10.1126/sciadv.aav4574

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8:1991–2007. doi: 10.5194/gmd-8-1991-2015

Cook, P. G., Favreau, G., Dighton, J. C., and Tickell, S. (2003). Determining natural groundwater influx to a tropical river using radon, chloroflourocarbons and ionic environmental tracers. *J. Hydrol.* 277, 74–88. doi: 10.1016/S0022-1694(03)00087-8

Fan, Y., Li, H., and Miguez-Macho, G. (2013). Global patterns of groundwater depth. *Science* 339, 940–943. doi: 10.1126/science.1229881

Flint, A. L., Flint, L. E., and Dettinger, M. D. (2008). Modeling Soil moisture processes and recharge under a melting snowpack. *Vadose Zone J.* 7, 350–357. doi: 10.2136/vzj2006.0135

Frisbee, M. D., Phillips, F. M., Campbell, A. R., Liu, F., and Sanchez, S. A. (2011). Streamflow generation in a large, alpine watershed in the southern {Rocky Mountains of Colorado: Is} streamflow generation simply the aggregation of hillslope runoff responses? *Water Resour. Res.* 47:W06512. doi: 10.1029/2010WR009391

Frisbee, M. D., Phillips, F. M., Weissmann, G. S., Brooks, P. D., Wilson, J. L., Campbell, A. R., et al. (2012). Unraveling the mysteries of the large watershed black box: implications for the streamflow response to climate and landscape perturbations. *Geophys. Res. Lett.* 39:L01404. doi: 10.1029/2011GL050416

Gabrielli, C. P., McDonnell, J. J., and Jarvis, W. T. (2012). The role of bedrock groundwater in rainfall-runoff response at hillslope and catchment scales. *J. Hydrol.* 450–451, 117–133. doi: 10.1016/j.jhydrol.2012.05.023

Gardner, W. P., Harrington, G. A., Solomon, D. K., and Cook, P. G. (2011). Using terrigenic 4He to identify and quantify regional groundwater discharge to streams. *Water Resour. Res.* 47, 1–13. doi: 10.1029/2010WR010276

Gardner, W. P., Jensco, K., Hoylman, Z., Livesay, R., and Maneta, M. (2020). A numerical investigation of bedrock groundwater recharge and exfiltration on soil mantled hillslopes. *Hydrol. Process.* 34, 3311–3330. doi: 10.1002/hyp.13799

Gardner, W. P., Susong, D. D., Solomon, D. K., and Heasler, H. P. (2010). Snowmelt hydrograph interpretation: revealing watershed scale hydrologic characteristics of the Yellowstone volcanic plateau. *J. Hydrol.* 383, 209–222. doi: 10.1016/j.jhydrol.2009.12.037

Genereux, D. P., Hemond, H. F., and Mulholland, P. J. (1993). Use of radon-222 and calcium as tracers in a three-end-member mixing model for streamflow generation on the West Fork of Walker Branch Watershed. *J. Hydrol.* 142, 167–211. doi: 10.1016/0022-1694(93)90010-7

Gleeson, T., and Manning, A. H. (2008). Regional groundwater flow in mountainous terrain: three-dimensional simulations of topographic and hydrogeologic controls. *Water Resour. Res.* 44:W10403. doi: 10.1029/2008WR006848

Goble, D. (2018). *Estimating Groundwater Inflow and Age Characteristics in an Alluvial Aquifer Along the Little Wind River, Wyoming* (Masters). University of Montana. Available online at: https://scholarworks.umt.edu/etd/11263

Gooseff, M. N., and McGlynn, B. L. (2005). A stream tracer technique employing ionic tracers and specific conductance data applied to the Maimai catchment, New Zealand. *Hydrol. Process.* 19, 2491–2506. doi: 10.1002/hyp.5685

Hale, V. C., McDonnell, Jeffrey, J., Stewart, M. K., Solomon, D. K., Doolitte, J., Ice, G. G., et al. (2016). Effect of bedrock permeability on stream base flow mean transit time scaling relationships: 2. Process study of storage and release. *Water Resour. Res.* 52, 1375–1397. doi: 10.1002/2015WR017660

Hale, V. C., and McDonnell, J. J. (2016). Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison. *Water Resour. Res.* 52, 1358–1374. doi: 10.1002/2014WR016124

Harrington, G. A., Gardner, W. P., and Munday, T. J. (2013). Tracking groundwater discharge to a large river using tracers and geophysics. *Groundwater* 52, 837–852. doi: 10.1111/gwat.12124

Holden, Z. A., Abatzoglou, J. T., Luce, C. H., and Baggett, L. S. (2011a). Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain. *Agric. For. Meteorol.* 151, 1066–1073. doi: 10.1016/j.agrformet.2011.03.011

Holden, Z. A., Crimmins, M. A., Cushman, S. A., and Littell, J. S. (2011b). Empirical modeling of spatial and temporal variation in warm season nocturnal air temperatures in two North Idaho mountain ranges, USA. *Agric. For. Meteorol.* 151, 261–269. doi: 10.1016/j.agrformet.2010.10.006

Hooper, R. P. (2003). Diagnostic tools for mixing models of stream water chemistry. *Water Resour. Res.* 39. doi: 10.1029/2002WR001528

Hoylman, Z. H., Jencso, K. G., Hu, J., Holden, Z. A., Martin, J. T., and Gardner, W. P. (2019). The climatic water balance and topography control spatial patterns of atmospheric demand, soil moisture and shallow subsurface flow. *Water Resour. Res.* 55, 2370–2389. doi: 10.1029/2018WR023302

Hoylman, Z. H., Jencso, K. G., Hu, J., Martin, J. T., Holden, Z. A., Seielstad, C. A., et al. (2018). Hillslope topography mediates spatial patterns of ecosystem sensitivity to climate. *J. Geophys. Res. Biogeosci.* 123, 353–371. doi: 10.1002/2017JG004108

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Bencala, K. E., and Wondzell, S. M. (2010). Hillslope hydrologic connectivity controls riparian groundwater turnover: implications of catchment structure for riparian buffering and stream water sources. *Water Resour. Res.* 46:W10524. doi: 10.1029/2009WR008818

Katsuyama, M., Ohte, N., and Kabeya, N. (2005). Effects of bedrock permeability on hillslope and riparian groundwater dynamics in a weathered granite catchment. *Water Resour. Res.* 41, 1–11. doi: 10.1029/2004WR003275

Liu, F., Williams, M. W., and Caine, N. (2004). Source waters and flow paths in an alpine catchment, Colorado Front Range, United States. *Water Resour. Res.* 40:W09401. doi: 10.1029/2004WR003076

Lonn, J. D., McDonald, C., Sears, J. W., and Smith, L. N. (2010). *Geologic Map of the Missoula East 30′ x 60′ Quadrangle, Western Montana.* Montana Bureau of Mines and Geology Open-File Report 593.

Meixner, T., Manning, A. H., Stonestrom, D. A., Allen, D. M., Ajami, H., Blasch, K. W., et al. (2016). Implications of projected climate change for groundwater recharge in the western United States. *J. Hydrol.* 534, 124–138. doi: 10.1016/j.jhydrol.2015.12.027

Moore, R. (2004). Slug injection using salt in solution. *Streaml. Watershed Manag. Bull.* 8, 1–6.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., et al. (2017). Bedrock geology controls on catchment storage, mixing, and release: a comparative analysis of 16 nested catchments. *Hydrol. Process.* 31, 1828–1845. doi: 10.1002/hyp.11134

Pinder, G. F., and Jones, J. F. (1969). Determination of the ground-water component of peak discharge from the chemistry of total runoff. *Water Resour. Res.* 5, 438–445. doi: 10.1029/WR005i002p00438

Portner, R. A., Hendrix, M. S., Stalker, J. C., Miggins, D. P., and Sheriff, S. D. (2011). Sedimentary response to orogenic exhumation in the northern rocky mountain basin and range province, flint creek basin, west-central Montana. *Can. J. Earth Sci.* 48, 1131–1153. doi: 10.1139/e10-107

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., et al. (2012). Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnol. Oceanogr. Fluids Environ.* 2, 41–53. doi: 10.1215/21573689-1597669

Salve, R., Rempe, D. M., and Dietrich, W. E. (2012). Rain, rock moisture dynamics, and the rapid response of perched groundwater in weathered, fractured argillite underlying a steep hillslope. *Water Resour. Res.* 48, 1–25. doi: 10.1029/2012WR012583

Smerdon, B. D., and Gardner, W. P. (2021). Characterizing groundwater flow paths in an undeveloped region through synoptic river sampling for environmental tracers. *Hydrol. Process.* 36:e14464. doi: 10.1002/hyp.14464

Smerdon, B. D., Gardner, W. P., Harrington, G. A., and Tickell, S. J. (2012). Identifying the contribution of regional groundwater to the baseflow of a tropical river (Daly River, Australia). *J. Hydrol.* 464–465, 107–115. doi: 10.1016/j.jhydrol.2012.06.058

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture (n.d.). *Soil Survey Geographic (SSURGO) Database for West-Central Montana*. Available online. (Accessed August 1, 2017).

Somers, L. D., and McKenzie, J. M. (2020). A review of groundwater in high mountain environments. *Wires Water*. 7:e1475. doi: 10.1002/wat2.1475

Spencer, S. A., Anderson, A. E., Silins, U., and Collins, A. L. (2021). Hillslope and groundwater contributions to streamflow in a Rocky Mountain watershed underlain by glacial till and fractured sedimentary bedrock. *Hydrol. Earth Syst. Sci.* 25, 237–255. doi: 10.5194/hess-25-237-2021

Torres, R., Dietrich, W. E., Montgomery, D. R., Anderson, S. P., and Loague, K. (1998). Unsaturated zone processes and the hydrologic response of a steep,

unchanneled catchment. *Water Resour. Res.* 34, 1865–1879. doi: 10.1029/98WR01140

Tromp-van Meerveld, H. J., Peters, N. E., and McDonnell, J. J. (2007). Effect of bedrock permeability on subsurface stormflow and the water balance of a trenched hillslope at the Panola Mountain Research Watershed, Georgia, USA. *Hydrol. Process.* 21, 750–769. doi: 10.1002/hyp.6265

Uchida, T., Asano, Y., Ohte, N., and Mizuyama, T. (2003). Seepage area and rate of bedrock groundwater discharge at a granitic unchanneled hillslope. *Water Resour. Res.* 39, 1–12. doi: 10.1029/2002WR001298

Uchida, T., Kosugi, K., and Mizuyama, T. (2002). Effects of pipe flow and bedrock groundwater on runoff generation in a steep headwater catchment in Ashiu, central Japan. *Water Resour. Res.* 38, 1–14. doi: 10.1029/2001WR000261

Uchida, T., McDonnell, J. J., and Asano, Y. (2006). Functional intercomparison of hillslopes and small catchments by examining water source, flowpath and mean residence time. *J. Hydrol.* 327, 627–642. doi: 10.1016/j.jhydrol.2006.02.037

Williams, M. W., Hood, E., Molotch, N. P., Caine, N., Cowie, R., and Liu, F. (2015). The 'teflon basin' myth: hydrology and hydrochemistry of a seasonally snow-covered catchment. *Plant Ecol. Divers.* 8, 639–661. doi: 10.1080/17550874.2015.1123318

frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 16 July 2025
DOI 10.3389/frwa.2025.1584947

Check for updates

OPEN ACCESS

EDITED BY
Oliver S. Schilling,
University of Basel, Switzerland

REVIEWED BY
Balaji Etikala,
Yogi Vemana University, India
Parnian Ghaneei,
University of Alabama, United States

*CORRESPONDENCE
Catherine Moore
✉ c.moore@gns.cri.nz

RECEIVED 28 February 2025
ACCEPTED 16 June 2025
PUBLISHED 16 July 2025

CITATION
Moore C, Oluwunmi P, Hemmings B, Cameron S, Yang J, Taves M, Rajanayaka C, Woodward SJR and Moreau M (2025) The significance of groundwater contributions to New Zealand rivers.
Front. Water 7:1584947.
doi: 10.3389/frwa.2025.1584947

COPYRIGHT
© 2025 Moore, Oluwunmi, Hemmings, Cameron, Yang, Taves, Rajanayaka, Woodward and Moreau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The significance of groundwater contributions to New Zealand rivers

Catherine Moore[1]*, Paul Oluwunmi[1], Brioch Hemmings[2], Stewart Cameron[1], Jing Yang[3], Mike Taves[1], Channa Rajanayaka[3], Simon J. R. Woodward[4] and Magali Moreau[1]

[1]GNS Science, Lower Hutt, New Zealand, [2]INTERA Incorporated, Austin, TX, United States, [3]National Institute of Water and Atmospheric Research (NIWA), Auckland, New Zealand, [4]DairyNZ, Hamilton, New Zealand

Groundwater and surface water are highly interconnected systems, with the connections varying spatially, temporally and by catchment. Representing this connectivity is of key importance for future effective water management, and to address the global decline of surface water flows. Previous studies have used baseflow separation methods to quantify the groundwater contribution to surface flow volumes. However, few studies have analysed the different dynamics of deep and shallower groundwater contributions to surface water flow rates across the flow regime and attempted to quantify this changing contribution. We analysed the distribution of fast (near-surface event flow), medium (seasonal shallow groundwater discharge) and slow (deeper groundwater) pathways into surface water flows for a case study involving 58 river water quality and flow monitoring sites across New Zealand. This involved a novel application of the chemistry assisted baseflow separation method (BACH). We found that shallow and deep groundwater pathways were the most significant contributor (>80% of the daily flow rate) to river flow at most sites at the 75th flow percentile, and for many sites even at the 95th flow percentile. These findings emphasise the need to better integrate groundwater into surface water management strategies, particularly as droughts intensify, floods become more frequent and severe, and legacy nutrient input increases under changing climate and land-use.

KEYWORDS

baseflow separation, Bayesian, chemistry-assisted hydrograph separation, groundwater, surface water-groundwater exchange

1 Introduction

Groundwater has long been known to be a significant contributor to river flows (Beck et al., 2013). Despite this, the dynamics and disposition of surface- and ground-water interactions remain poorly understood. This knowledge gap has been linked to decreasing river flow trends globally (Scanlon et al., 2012, 2023; Seo et al., 2018a,b; Konikow and Leake, 2014; Taylor et al., 2013; Das et al., 2021; Gorelick and Zheng, 2015), as declining river flows are attributed to increased abstractions from inter-connected groundwater systems (de Graaf et al., 2019). These declining flows are impacting the chemistry and ecology of aquatic ecosystems (Rosenberry et al., 2008), the ability to assimilate nutrient and pollutant fluxes into streams and rivers (we refer to both as rivers in this paper), and the provision of drinking water (Fan et al., 2013; Bosch et al., 2017; Singh et al., 2019). A clear understanding of the importance of the groundwater contribution to surface waters is therefore critical for effective catchment

management. This paper advances our understanding of the interconnections of the surface- and ground-water components in the hydrological system (Scanlon et al., 2023; Seo et al., 2018b).

Baseflow separation methods have long been used to provide a rapid assessment of groundwater contributions to stream flow. These methods conceptualise river flow hydrographs as a combination of a lumped groundwater or "baseflow" estimate and precipitation runoff or "quick flow" (Beck et al., 2013). The lumped groundwater "baseflow" component may be variously comprised of regional groundwater flow, or more shallow near river sources such as adjacent perched aquifers, bank storage, the hyporheic zone, and braid plain aquifers (Wilson et al., 2024), and interflow from the unsaturated zone (Howcroft et al., 2019). Baseflows are represented implicitly and estimated through minimising model-to-measurement misfit rather than being based on hydrogeological properties and the simulation of hydrological processes; hence they cannot simulate spatially explicit causal relationships within the upriver catchment.

The key advantage of baseflow separation methods is that they are simple to use and can be applied even where catchment data is limited. However, this simplicity comes with compromises as different baseflow separation methods adopt alternative simple conceptualisations of groundwater inflows, each of which can introduce different biases (e.g., Yang et al., 2021), as occurs with simple representations of complex systems (Doherty and Christensen, 2011), leading to Cartwright (2022) observing that estimating the groundwater inflow to streams is not straightforward.

Typically, hydrograph baseflow separation methods are based on an analytical model (Martinez and Gupta, 2010) or are defined using filtering or smoothing methods (Lott and Stewart, 2016). Digital filters are also used (Eckhardt, 2005), where precipitation runoff and groundwater contributions are estimated based on decomposing the time series, into a series of responses with different frequencies and are similar to Fourier transforms and wavelet filtering methods. Typically, these baseflow separation techniques are based on a conceptualisation of baseflow having a longer wavelength and lower amplitude variation, while quick flow is characterised by shorter wavelengths. In New Zealand, Singh et al. (2019) applied the digital filter method at a national level to separate river flow hydrographs into two components, i.e., precipitation (quick flow) and groundwater (baseflow); note that application of this method does not involve ground-truthing. Traditionally, a stream baseflow index, defined as the ratio of the baseflow volume to total flow volume over a specified time period, is used to characterise this groundwater contribution, and is estimated using baseflow separation methods. These methods are applied to a single river location at a time and are used to reflect the upriver catchment's average hydrological process behaviour.

End-member mixing analysis (EMMA) (Adams et al., 2009; Klaus and McDonnell, 2013) is an alternative mass balance approach to baseflow separation, which makes use of additional groundwater chemistry information to identify the groundwater components of river flow, and using river chemistry data to anchor the baseflow estimation. It uses the assumption that waters flowing through specific compartments of a combined surface and subsurface watershed have unique chemical signatures (e.g., Bugaets et al., 2023). These signatures reflect the average water chemistry sources in the capture zone above the measurement point. While attractive conceptually, this method is challenged by the difficulty of adequately characterising the source water chemistry, which typically requires high resolution sampling, and is generally accompanied by considerable uncertainty (Barthold et al., 2011; Delsman et al., 2013; Woodward et al., 2017). Cartwright (2022) also notes that the long-term variability and spatial heterogeneity of groundwater inflows also severely complicate efforts to calibrate baseflow separation techniques using such chemical mass balance methods.

A variant to both EMMA and hydrograph separation methods, is the Bayesian chemistry assisted hydrograph separation method, "BACH" (Woodward and Stenger, 2018; Woodward and Stenger, 2020; Stenger et al., 2024). BACH combines a recursive digital filter (Su et al., 2016) and EMMA in a Bayesian Framework using chemical or tracer data to inform and reduce the uncertainty of the separated hydrograph components. This combination of the strengths of both EMMA and digital filter methods allows the BACH method to harness more information than either EMMA or digital filters alone, while retaining the minimal data requirements and application simplicity of recursive digital filter methods.

The BACH method relies on river flows conceptualised as the product of time varying combinations of slow, medium and fast flow paths, which are attributed to deeper groundwater, seasonal shallow groundwater discharge, and near-surface event flow, respectively. This results in BACH being able to provide insights into possible shallow and deeper groundwater contributions to low and high stream flows very quickly, using river flow and chemical concentration data. Further, while EMMA typically requires conservative chemistry analytes, BACH is freed from this requirement, because any attenuation in groundwater, the hyporheic zone or instream is implicitly included in the concentrations assigned to each flow path (Woodward et al., 2017). Similarly, depth, area, storage and other hydraulic properties relevant to flow path delineation when using distributed numerical models, are implicitly included in the flow path filter parameters when using BACH. To date, studies using BACH have focussed on identifying nutrient load pathways into surface water ways (Woodward and Stenger, 2018; Woodward and Stenger, 2020; Stenger et al., 2024; and most recently a companion paper to this study in Yang et al., 2025).

In this study we use the advantages of BACH, to present an ambitious application of the BACH method that for the first time explores and quantifies the dynamics of changing slow, medium and fast flow path contributions to rivers across a national river monitoring network. This paper describes the application of the BACH method in a nationwide study of 58 river sites across New Zealand to quantify the dynamic contributions of fast (near-surface event flow), medium (seasonal shallow groundwater discharge) and slow (deeper groundwater) pathways into surface water daily flows. The results of this analysis are discussed followed by some conclusions on the important insights the application of the method reveals. These results provide important information for the application of integrated land and water management policy by contributing to a broader understanding of the magnitude and dynamics of the combination of near-surface and groundwater flow paths to a stream flow varying over time.

2 Methodology

2.1 General

The BACH method, developed by Woodward and Stenger (2018, 2020), assumes that a river flow hydrograph can be separated into fast,

medium and slow flow components, whose relative contributions can be inferred from the temporally varying observations of water chemistry. These flow components are conceptually associated with distinct chemical end members, which have typically been differentiated using two routinely measured water analytes acting as tracers (e.g., nitrogen and phosphorous).

The method as applied to date has generally assumed fast flow components represent event-response, surface or near-surface flow, medium flow components represent seasonal shallow local groundwater flow, and slow flow components represent a persistent deeper regional groundwater pathway (Figure 1). Similar conceptualisations of these flow paths has been established in other catchment studies (e.g., Hesser et al., 2010; Broda et al., 2011; O'Brien et al., 2013; Aubert et al., 2013), and we have adopted the same flow path conceptualisation in this study. Each pathway implicitly connects land to the surface water monitoring site as depicted in Figure 1, however the method is not able to identify the spatial disposition of these pathways.

The BACH method uses a three-component recursive digital filter (RDF) with chemical mixing models to separate the hydrograph into the fast, medium and slow flow components. Each flow component is linked to the physical near surface, shallow groundwater and deep groundwater pathways by assuming they have characteristic concentrations of each analyte, that can be considered time-invariant (Figure 2). The BACH analysis outputs comprise simulated river concentration time-series for: the mixed nutrient components that are used during history matching, the separate nutrient fluxes from the fast, medium, and slow flow components, and the fast, medium and slow contributing flows. Though not the focus of this paper, BACH outputs also include nutrient flux volumes as discussed in Yang et al. (2025).

The BACH signal filtering method conceptually models a flow response to long term recharge and rainfall events, with the incorporation of water chemistry information. The outcome is a hydrograph separation that better reflects slow to fast rates of movement through the catchment system by inferring dominant flow path velocities.

The BACH model is described fully in Woodward and Stenger (2018, 2020), but we give a brief summary here for clarity. The method uses two RDFs, that separate the river flow (y_n) on day n, first, into fast flow (f_n) and remaining flow (r_n) (Equation 1), and then into medium flow (m_n) and slow flow components (s_n) (Equation 2):

$$f_n = y_n - r_n \tag{1}$$

$$m_n = r_n - s_n, \tag{2}$$

where r_n is estimated by the first RDF:

$$r_n = \min\left(y_n, \left(a_{1,m} * r_{n-1} + b_{0,m} * y_n\right)\right), \tag{3}$$

And the second RDF, s_n is estimated by

$$s_n = \min\left(r_n, \left(a_{1,s} * s_{n-1} + b_{0,s} * r_n\right)\right). \tag{4}$$

Here, $a_{1,m}$, $b_{0,m}$, $a_{1,s}$, and $b_{0,s}$ are filter parameters in the interval [0,1], with the constraint that $a_1/(1 - b_0) \leq 1$, this constraint is also used in Eckhardt (2005) and Su et al. (2016). These filter parameters are estimated along with component concentration parameters (for each analyte; c_f, c_m, and c_s for fast, medium, and slow component analyte concentrations, respectively). The mixing model (Equation 5) predicts analyte concentration in the river (on day n) as:

$$y_n * c_n = f_n * c_f + m_n * c_m + s_n * c_s. \tag{5}$$

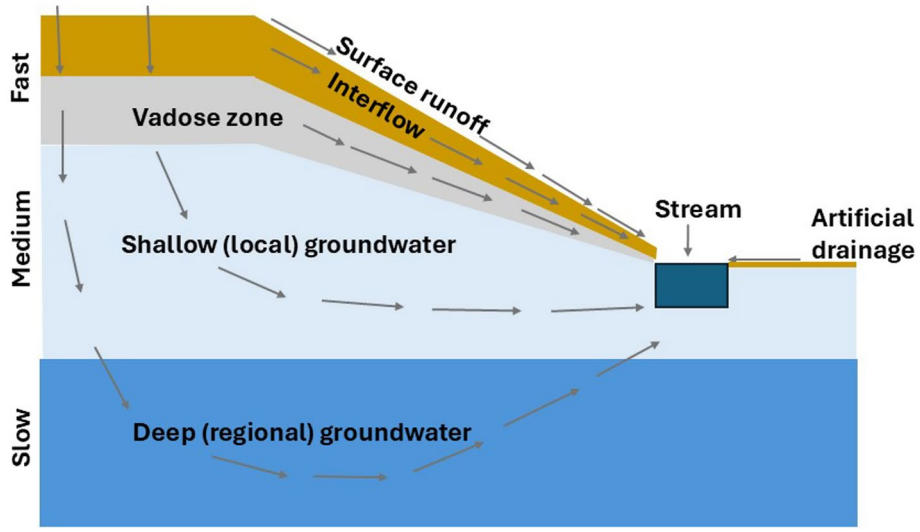


FIGURE 1
Schematic depiction of fast, medium and slow flow paths contributing to a river flow, modified from Stenger (2022). Surface runoff, interflow and near-surface flow contribute to fast flows into the river. Shallow groundwater provides medium flow contributions into the river. Deeper groundwater provides a slower contribution to flows.

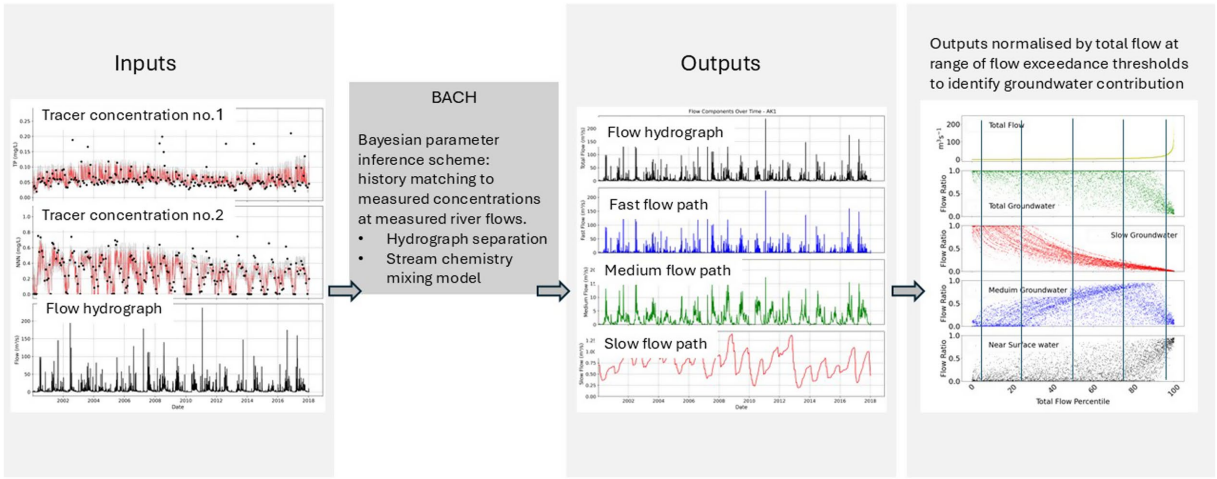


FIGURE 2
Illustration of BACH model inputs, their concurrent processing by the Bayesian parameter inference scheme, and resulting outputs. Modified from Stenger et al. (2024).

In the current study, c_f, c_m, and c_s are assumed to vary nonlinearly through time according to the following harmonic form, as described in Equation 6 (Woodward and Stenger, 2020),

$$c_t = e_0 * (1-x) + e_1 * x + f_1 * \sin(\pi x) + f_2 * \sin(2\pi x) \quad (6)$$

where e_0 is the initial concentration, e_1 is the final concentration, x is the time scaled between 0 and 1, and f_1 and f_2 are harmonic coefficients to be inferred.

If using two analytes, as in this study where we used nitrate-nitrite nitrogen (NNN) and total phosphorus (TP), a total of 28 parameters are estimated when applying the BACH model, i.e., the four filter parameters in Equations 3, 4, and 12 concentration parameters for each analyte (e_0, e_1, f_1, and f_2 for each flow path component).

2.2 Model convergence, goodness of fit and uncertainty quantification

History matching of the model parameters was achieved using the Bayesian Markov Chain Monte Carlo (MCMC) methodology, this was implemented using the Stan algorithm (Stan Development Team, 2024). Goodness of fit metrics, as described below, are used to evaluate the credibility of the model parameterisation.

Detail on the specific computation algorithm and its implementation in BACH can be found in Woodward and Stenger (2018, 2020), and references therein. In our study, we employed the Gelman and Rubin (1992) statistic, often known as the R-hat statistic, to determine convergence of the MCMC algorithm. It is a metric for analysing the variability of samples collected from multiple chains of a Bayesian sampling method. It is defined as the sample size-adjusted ratio of between-chain variance to within-chain variance (Equation 7).

$$R \approx \frac{\text{between} - \text{chain variance}}{\text{within} - \text{chain variance}}. \quad (7)$$

where between-chain variance is the variance of the parameters across all Markov chains, within-chain variance is the average of the variance in the parameters for each chain.

An R-hat near to one, suggests convergence to a stable posterior distribution. Values noticeably higher than 1 indicate that more iterations are required for improved convergence. In our study we aimed for a Gelman-Rubin statistic of 1.1, as suggested by Gelman et al. (2004).

We also used a goodness-of-fit statistic to evaluate model to measurement fit. In this investigation, we used daily flow and available coincident river NNN and TP data to calibrate our BACH models (refer to Supplementary Table S4), initially assuming that the standard uncertainty for each analyte was 0.02 mg L^{-1} for TP and 0.2 mg L^{-1} for NNN as employed by Woodward and Stenger (2018).

The root mean squared error (RMSE) is the most used goodness-of-fit statistic in BACH: RMSE measures the average squared difference between observed and predicted values, providing a sense of the model's accuracy in the original units of the data to assess the goodness-of-fit of model findings and to provide a visual comparison of simulated and modelled data. A generalised form of the RMSE (GRMSE, after Woodward and Stenger, 2018) is presented below. This provides a measure of reliability of the estimated "endmember" concentrations for fast, medium and slow flows components, and the ratio of flow components from low to high flows based on how well they provide outputs that can reproduce the observed phosphorous and nitrate concentrations:

$$\text{GRMSE} = \sqrt{\frac{1}{nD}\sum_{j=1}^{nD} 2\sigma_{Dj}^2 \left[l(0) - l\left(\hat{D}_j - D_j\right) \right]} \quad (8)$$

where, $\hat{D}_j$ is value of the jth observation, with a total number of n_D; D_j is the corresponding model prediction value; $l\left(\hat{D}_j - D_j\right)$ is the log-likelihood of residual $\left(\hat{D}_j - D_j\right)$, while $l(0)$ is the log-likelihood of $\left(\hat{D}_j - D_j\right)$ being zero; σ_{D_j} is the standard deviation of the observation.

The initial 3 years of the time series are used as warm-up data, and hence goodness of fit criteria are only applied after that time.

The goodness of fit metrics do not necessarily provide a reliable indication of the uncertainty of the model predictions of fast, medium and slow flow components (Kitlasten et al., 2022). Instead, the MCMC adopted with BACH is used to quantify the uncertainty of the simulated outputs due to parameter non-uniqueness, model simplification, observation error and an assumed prior parameter distribution.

2.3 Groundwater contributions to river flow dynamics

Previous studies have used the BACH method to explore the fast, medium and slow flow path contributions to estimate nutrient loads to rivers. In this study we instead explore groundwater contributions to flow rates. Fast, medium and slow contributions to river flows can be analysed for any hydrograph, and the patterns of these pathways in river dynamics can be explored across the cumulative flow distribution. We discuss the contribution of different flow pathways to river flow rates, with pathway contributions normalised as percentages of the measured total flow at seven selected flow exceedance thresholds or "percentiles" (i.e., 1st, 5th, 25th, 50th, 75th, 95th, 99th) at each of the selected river monitoring sites.

To explore potential relationships between the BACH separated flow ratios and a representative site characteristic we utilised the flashiness index (FI) as used in Woodward and Stenger (2018) (Table 1). The FI (Equation 9) is a dimentionless measure of the river flow variability at short timescales. It is computed by dividing the pathlength of flow oscillations for that time period (i.e., the sum of the absolute values of daily variations in mean daily flow) by the total discharge for a given time period (Baker et al., 2004; Equation 8). A higher FI number implies a river in which the flow varies quickly and is often associated with reduced baseflow contributions (Deelstra and Iital, 2008).

$$\mathrm{FI} = \frac{\sum_{i=1}^{n} \left| q_i - q_{i-1} \right|}{\sum_{i=1}^{n} q_i} \tag{9}$$

where FI is the Flashiness Index, $\sum_{i=1}^{n} \left| q_i - q_{i-1} \right|$ is the sum of the absolute values of daily changes in mean daily flow (pathlength of flow oscillations), $\sum_{i=1}^{n} q_i$ is the total of the mean daily flows.

2.4 New Zealand case study

New Zealand river catchments occur across multiple geologies, soils, topographies, land uses, and rainfalls, resulting in multiple catchment types. Many of these catchments have headwaters in hilly or mountainous terrain, and outflows at the coast, and are associated with mapped alluvial aquifers. In most of these catchments groundwater flow is known to interact substantially with the surface water flow systems (Yang et al., 2025; Singh et al., 2019).

The data were supplied by the New Zealand Institute of Water and Atmospheric (NIWA) National River Water Quality Network (NRWQN, 2024; Ballantine and Davies-Colley, 2014). Surface water flow and water quality time-series data suitable for BACH analysis were available for 58 sites across our New Zealand case study (Figure 3 and Table 1). Further site details can be found in Supplementary Table S4.

Flow time series data was available at daily intervals at all sites, but with some data gaps that were filled using linear interpolation methods. Generally, the dataset provided monthly measurements of water quality analytes, collected by a combination of automated and grab samples. The time series of flow and water quality generally spanned more than 20 years (e.g., Figure 4).

We elected to use the combination of flow, TP and NNN data for the BACH analysis, as these analytes were most widely available in NRWQN (2024). Variations in the river TP and NNN concentrations at varying flow rates were used to identify distinct water quality signatures associated with slow, medium, and fast surface water flow components. This enabled us to explore the deeper and shallower groundwater pathway contributions to river flow across the full hydrograph range for multiple sites in New Zealand.

The reliability of flow-concentration relationships, and therefore the performance of the BACH method can be compromised by artificial modulation of the flow response to catchment recharge events. This includes the presence of dams, controlled flows, and lakes—the presence of these features is inconsistent with the governing assumptions of the applied methodology. The flows at monitoring site AX1 on the Clutha River, for example, are impacted by the dam on Lake Hawea (as well as by buffering from the lake itself), while flows at site AX2 on the Kawarau River are likely impacted by the buffering effect of Lake Wakatipu and the control gates that influence lake outflow, and at Clutha River site AX4 by the Roxburgh Dam. The sequence of hydroelectric dams on the Waitaki River strongly modulates flow at site TK4, as well as at the downriver, non-convergent site TK6. Sites RO6 and RO1 are proximally downriver of the outflows of Lake Taupō and Lake Tarawera, respectively, and the flow-concentration relationship is likely to be impacted by buffering of the lakes. Site TK3 on the Makaroro River, had a large number of gaps in the site flow record, and Site RO1 had very low nutrient concentration, compromising the ability to use these data time series. These eight sites were removed from the analysis (see grey shading in Table 1), leaving a total of 50 sites used for this analysis.

Observation weights assigned in the BACH analysis were initially based on the assumption of a uniform standard uncertainty across all sites (0.2 mg L^{-1} and 0.02 mg L^{-1} for TP and NNN, respectively), with a Gaussian distribution assumed for observation errors. These weights essentially serve as a fitting tolerance, reflecting the anticipated combined error from measurement error and model structure and parameterisation error (Woodward and Stenger, 2018). For the sites with very low concentrations, this error term was subsequently reduced, to ensure the signal was not obscured by underfitting (refer to Section 3.1).

3 Results and discussion

3.1 Data summary

A range of catchments with various dimensions, topography, hydrogeology and land use patterns are represented across the

10.3389/frwa.2025.1584947

TABLE 1 Statistics of count, mean, and standard deviation for observed annual mean flow, TP, NNN, and FI for sites assessed in this study.

Data metric	River name and site	Flow ($m^3 s^{-1}$)			NNN ($mg m^{-3}$)			TP ($mg m^{-3}$)			FI
		n	Mean	Std.	*n*	Mean	Std.	*n*	Mean	Std.	
AK1	hoteo_@_gubbs	7,993	5.7	11.7	263	293	233	262	60	35.2	0.52
AK2	rangitopuni_@_walkers	7,993	1.4	3.3	263	185	138.6	261	66	41.1	0.66
AX1	clutha_@_luggate_br	8,008	264.8	98.1	259	32	11.5	258	4	16.4	0.06
AX2	kawarau_@_chards_rd	7,980	202.0	90.6	258	24	6.6	255	35	109.9	0.05
AX3	shotover_@_bowens_peak	8,008	36.0	26.3	261	17.7	9.8	261	97	293.6	0.20
AX4	clutha_@_millers_flat	8,007	521.5	198.0	260	35	18.4	257	9	15.8	0.07
CH1	hurunui_@_mandamus	7,994	51.1	43.0	258	18	22.4	261	36	235.7	0.17
CH2	hurunui_@_sh1_br	7,995	71.8	69.1	263	365	182.5	262	44	166.1	0.17
CH3*	waimakariri_@_gorge	7,994	125.3	101.7	262	79	36.2	262	48	177.6	0.24
CH4*	waimakariri_@_old_highway_bge	7,995	111.0	102.6	256	148	119.3	260	63	235.8	0.27
DN2	sutton_stm_@_sh87	7,210	1.2	2.6	237	24	66.3	237	22	19.4	0.40
DN4	clutha_@_balclutha	8,009	567.7	218.6	262	94	104.5	262	14	17.1	0.08
DN5	mataura_@_seaward_downs	8,008	87.8	85.3	262	1,155	352.1	262	54	57.8	0.21
DN9	waiau_@_tuatapere	8,009	132.2	148.9	261	246	125.9	261	18	29.9	0.20
GS1	waipaoa_@_kanakanaia_c/w	8,001	30.5	62.0	262	163	172.6	262	157	324.9	0.42
GS2	waikohu_@_no_1_br	8,001	0.8	1.5	262	216	180.1	262	34	59.1	0.39
GS3*	motu_@_waitangirua	8,000	11.8	17.8	261	177	164.8	260	42	55.0	0.40
GS4	motu_@_houpoto	8,001	85.5	113.9	262	61	53.6	262	59	119.8	0.40
GY1	buller_@_te_kuha	8,001	411.1	440.9	262	72	45.8	260	20	42.0	0.37
GY2	grey_@_dobson	8,002	348.0	321.5	262	138	72.1	259	21	50.7	0.36
GY3	grey_@_waipuna	8,002	55.2	62.0	262	44	41.2	262	16	68.7	0.45
GY4	haast_@_roaring_billy	7,971	187.4	220.3	260	32	11.0	260	27	97.9	0.47
HM1	waipa_@_otewa	8,008	11.8	12.6	263	276	159.1	263	37	63.4	0.25
HM2	waipa_@_sh23_br_whatawhata	8,009	82.6	80.5	263	764	388.9	262	80	42.1	0.16
HM6	ohinemuri_@_karangahake	8,002	11.5	18.4	262	441	267.5	262	16	16.0	0.38
HV1*	Makaroro_@_Burnt_Bridge	10,179	6.5	9.3	334	66	59	332	34	93	0.27
HV2	Tukituki_@_Red_Bridge	11,419	42,0	93.2	374	637	508	373	54	141	0.42
HV3*	Ngaruroro_@_Chesterhope	11,418	40.8	70.6	371	96	94	370	42	95	0.34
HV4	Ngaruroro_@_Kuripapango	11,417	16.9	20.3	374	11	11	371	7	20	0.31
NN1*	motueka_@_woodstock	8,001	51.6	61.6	262	205	149.1	262	15	32.7	0.36
NN2*	motueka_@_gorge	8,001	7.0	10.3	262	26	23.0	261	6	17.8	0.51
NN3*	wairau_@_dip_flat	8,000	26.5	22.8	260	14	12.5	260	16	58.0	0.21
NN5*	buller_@_longford	8,000	69.5	53.0	262	35	26.0	262	12	54.3	0.19
RO1	tarawera_@_lake_outlet_recorder	8,000	6.9	1.4	260	1	1.7	260	8	2.2	0.02
RO2	tarawera_@_awakaponga	8,001	28.9	6.4	263	386	45.5	263	110	22.8	0.03
RO3	rangitaiki_@_murupara	8,001	20.1	7.2	263	804	241.4	262	31	7.5	0.03
RO4	whirinaki_@_galatea	8,001	14.6	12.3	263	111	81.6	263	39	26.4	0.13
RO5	rangitaiki_@_te_teko	8,001	66.2	33.1	262	404	158.0	261	38	11.1	0.13
RO6	waikato_@_reids_farm	8,000	150.9	62.2	261	0.9	1.0	261	5	1.6	0.20
TK1	opihi_@_waipopo	8,001	17.9	40.3	263	572	388.2	263	13	29.4	0.22
TK2	opihi_@_rockwood	8,001	4.8	7.9	263	1,121	663.2	263	24	79.9	0.30
TK3*	Opuha_@_Skipton	8,001	8.2	6.2	263	278	157.7	261	14	17.9	0.07
TK4	waitaki_@_kurow	8,002	348.9	121.5	263	7	8.2	262	8	55.4	0.12

(Continued)

TABLE 1 (Continued)

Data metric	River name and site	Flow ($m^3 s^{-1}$)			NNN ($mg m^{-3}$)			TP ($mg m^{-3}$)			FI
		n	Mean	Std.	*n*	Mean	Std.	*n*	Mean	Std.	
TK5	hakataramea_@_above_mhbr	8,002	5.1	11.3	262	50	131.9	260	9	26.3	0.26
TK6	waitaki_@_sh1_br	8,001	356.8	124.1	263	67	71.3	263	11	16.0	0.11
TU1*	whanganui_@_te_maire	8,002	70.0	69.9	259	263	200.9	259	42	91.1	0.25
TU2	tongariro_@_turangi	8,002	30.7	21.7	258	40	30.1	258	22	21.4	0.16
WA1	waitara_@_bertrand_rd	8,001	54.0	75.6	262	323	196.5	262	93	151.5	0.45
WA2	manganui_@_sh3	8,001	1.6	2.3	262	114	89.6	262	32	190.5	0.61
WA4	whanganui_@_paetawa	8,002	206.8	247.9	259	218	185.6	258	81	126.9	0.32
WA5*	rangitikei_@_mangaweka	8,001	63.2	63.1	261	85	84.1	260	53	181.0	0.27
WA7	manawatu_@_weber_rd	7,988	13.5	28.8	263	470	338.1	263	100	299.4	0.46
WH1	waipapa_@_forest_ranger	7,995	4.8	10.9	263	24	27.5	263	13	21.5	0.62
WH2	waitangi_@_wakelins	7,993	7.9	14.5	263	238	188.6	262	30.0	29.6	0.46
WH3	mangakahia_@_titoki_br	7,978	23.3	42.0	263	120	124.2	263	34	38.4	0.49
WH4	wairua_@_purua	8,009	17.5	28.8	263	391	339.2	263	75	53.9	0.36
WN2	hutt_@_kaitoke	8,008	8.0	11.1	263	34	17.5	263	9	29.0	0.60
WN5*	ruamahanga_@_mt_bruce	8,009	9.6	13.2	263	36	23.9	263	11	31.6	0.69

Grey background are sites where the BACH method could not be reliably applied.

monitored sites. Summary statistics for the flow, NNN, and TP datasets used in the analysis, for each site, are presented in Table 1. Across these sites the mean flows and concentrations vary widely. For example, the mean flow across all sites ranges from a minimum of 0.8 $m^3 s^{-1}$ (site GS2) up to 567.7 $m^3 s^{-1}$ (site DN4), the mean NNN concentration varies from 0.9 (RO6) to 1,155 mg m^{-3} (site DN5) and the mean TP concentrations varies from 4 (site AX1) and 157 mg m^{-3} (site GS1).

At a number of sites with low nutrient concentrations, some adjustment of the observation weights was required to enhance the signal to noise ratio, and to facilitate the extraction of the flow partitioning information used by the BACH model (see Section 3.2). These sites are denoted by an "*" symbol in Table 1, if TP or NNN was a concern. Weights adjustment generally involved an iterative increase in observation weight (i.e., seeking a tighter fit to the data) for sites with low TP and NNN concentrations. However, in a few cases the weights were decreased to avoid fitting to noise. The final weights and parameters applied to the data at each site are listed in Supplementary Table S1.

3.2 History matching

History matching with the BACH model was able to establish fast, medium and slow flow components of river flow, based on acceptably good fits to the measured nutrient concentrations for all 50 sites used in the analysis. Figures 4, 5 depict the model to measurement fits to the observed instream concentration values (depicting the AK1 and WA4 sites). These 50 sites all converged to an acceptable level as defined by the Gelman R statistic, with the R-hat values all within the acceptable 1–1.1 range. The performance of the model is slightly challenged however in reproducing high observed TP concentrations, while the model was able to fit intermediate and low TP concentrations reasonably well. This reflects a weakness of the BACH model, which does not represent the non-linear increases in TP entering near-surface pathways during storm events as particulates.

The BACH method relies on the presence of a clear (and simple) relationship between river flow and analyte concentrations to effectively and reliably separate flow components. At the same time BACH allows the hysteresis between the flow components to be represented, allowing the mix of flow path components for any flow rate to vary on the rising and falling limb of a hydrograph. At other sites the relationship between these variables appears even more complex. Despite a converging solution, it was sometimes challenging to reproduce (relatively infrequent) high TP concentrations, suggesting that there is a stream flow and phosphorous loading behaviour that the BACH model conceptualisation is unable to represent as depicted in Figures 4, 5 for sites AK1 and WA4, with mean flows of 5.7 $m^3 s^{-1}$ and 206.8 $m^3 s^{-1}$, respectively. At other sites it was difficult to achieve an acceptable converged solution without careful attention to observation weights, and this reflects the lack of a unique parameter solution potentially caused by similar flow path tracer concentrations. Application of BACH at these sites is presumably challenged by the absence of strong relationship between river flow and analyte concentrations and potentially contamination of the relationships by erroneous data points.

The fits to the NNN observations in Figure 4 indicate some bias in the simulated concentration values, with high NNN concentrations biased low. Woodward and Stenger (2020) suggested that similar biased trends could be related to structural deficiency within the model inhibiting its ability to represent flashy system behaviour and its relationship to Total Nitrogen concentration.

Many sites demonstrate strong seasonal signal in NNN concentration data. For example, in the case of AK1 and WA4 (Figure 4) NNN concentrations peak in winter. Typically, where such seasonal variations are present the BACH model is able to reproduce them. However, there is still variation in the fit of the NNN predictions,

with greater misfit occurring for more extreme observed values (e.g., near zero concentrations).

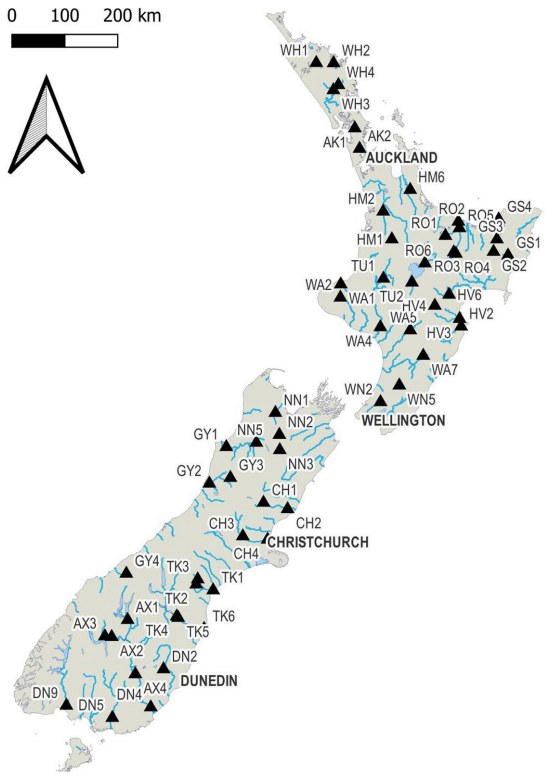


FIGURE 3
Location of the 58 New Zealand River Water Quality Network monitoring sites used across the New Zealand case study.

TP observations are not dominated by seasonal processes to the same extent. The modelled TP time series typically has greater variability, with a higher frequency and amplitude of variations, than observations. Peak observed concentrations, especially for TP, are often not well reproduced by the BACH model. These model to observation misfits can be attributed to system dynamics that are not well represented by the underlying simplifying assumptions of the BACH model.

The magnitude of the prediction uncertainty varies from site to site. Generally, for most catchments the GRMSE was around or below the assumed standard deviation for each analyte (e.g., 0.02 mg/L for TP and 0.2 mg/L for NNN, Supplementary Figure S1). The GRMSE for a number of sites is well below the assumed standard deviation (especially for NNN). This may be an indication that the model is over-fitting. However, it is noted that the sites with particularly low GRMSE generally have low concentrations, and hence a lower GRMSE is appropriate.

We note that despite some poor fits to the concentration time series, Stenger et al. (2024) observed that in general the relative pathway flow contributions had small errors. Our study is consistent with this, with all sites estimating the proportion of flow path contributions to stream flow with a small uncertainty quantified from the MCMC process (e.g., Figure 5). Slow (deep groundwater) flow path contributions typically have the narrowest confidence interval, with medium (shallow groundwater) and fast (near surface) flow contributions having slightly larger confidence intervals, indicating greater correlation between the medium and fast flow paths.

As mentioned above, field validation of chemistry assisted baseflow separation techniques using such chemical mass balance methods are challenged by long-term variability and spatial heterogeneity of groundwater inflows. However, two of the study

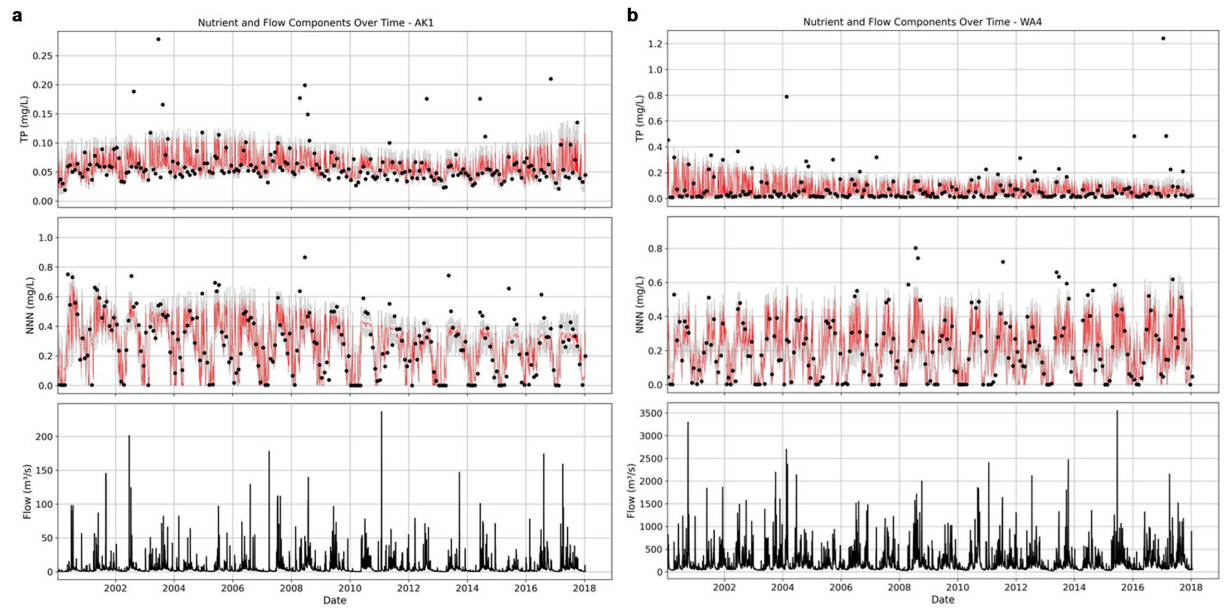


FIGURE 4
Simulated and observed TP and NNN concentrations and flows at site; **(a)** AK1 and **(b)** WA4. The model prediction for TP and NNN are displayed in red for the median model prediction, and grey bands include the 95% credible interval (which reflects parameter uncertainty).

catchments (HM6 and HV2) have relatively hydrogeologically simple catchment flow systems with some long-term groundwater nitrate concentration data (Moreau et al., 2025). We compared stream NNN concentrations for these sites with the groundwater nitrate concentrations in the surrounding catchment; we focussed only on measured nitrate concentrations, as in groundwater NNN is almost solely comprised of nitrate. Assuming a simplifying assumption of a medium and slow flow path separation depth of 80 m and 90 m below ground surface in the HM6 and HV2 stream catchments respectively, a reasonable correspondence between average groundwater measured concentrations and BACH flow-path concentration estimates was observed (refer to Supplementary Figure S2 and Supplementary Table S5).

However, this result does not verify the BACH flow-path insights. If independent groundwater chemistry data was more widely available, within delineated groundwater capture zones above the stream monitoring site, such a field validation of the chemistry end members could be possible, as noted by Cartwright (2022). However, it is worth emphasising that the strength of baseflow separation methods, including BACH, lies in the insights it provides quickly with little data. These insights can be used to guide future more detailed sampling and modelling work.

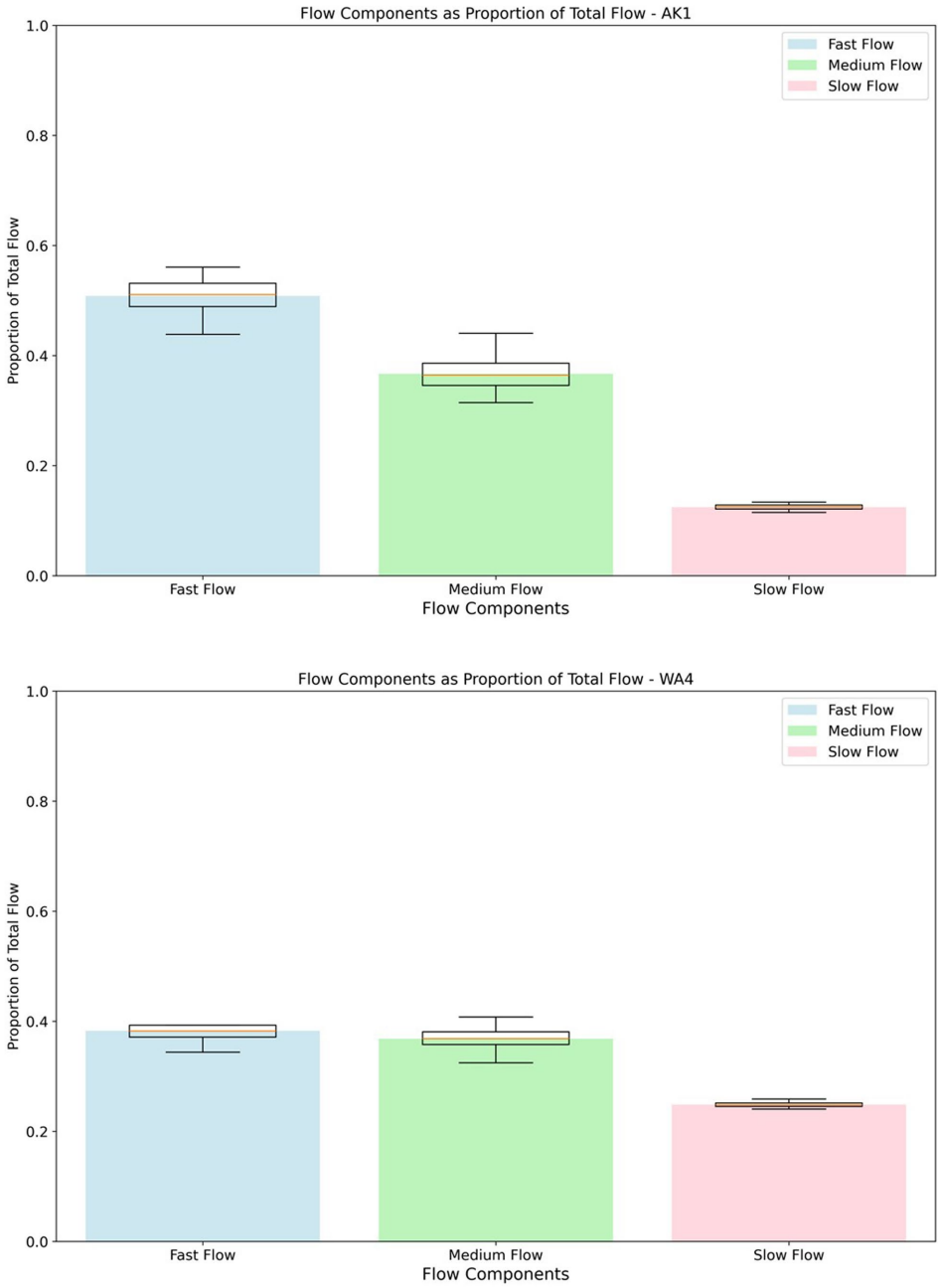


FIGURE 5
Predicted flow contributions as a fraction of total flow volume over the 20 year period for sites AK1 and WA4. Boxplots indicate the 0.025, 0.25, 0.5, 0.75, and 0.975 uncertainty quantiles for the inferred value.

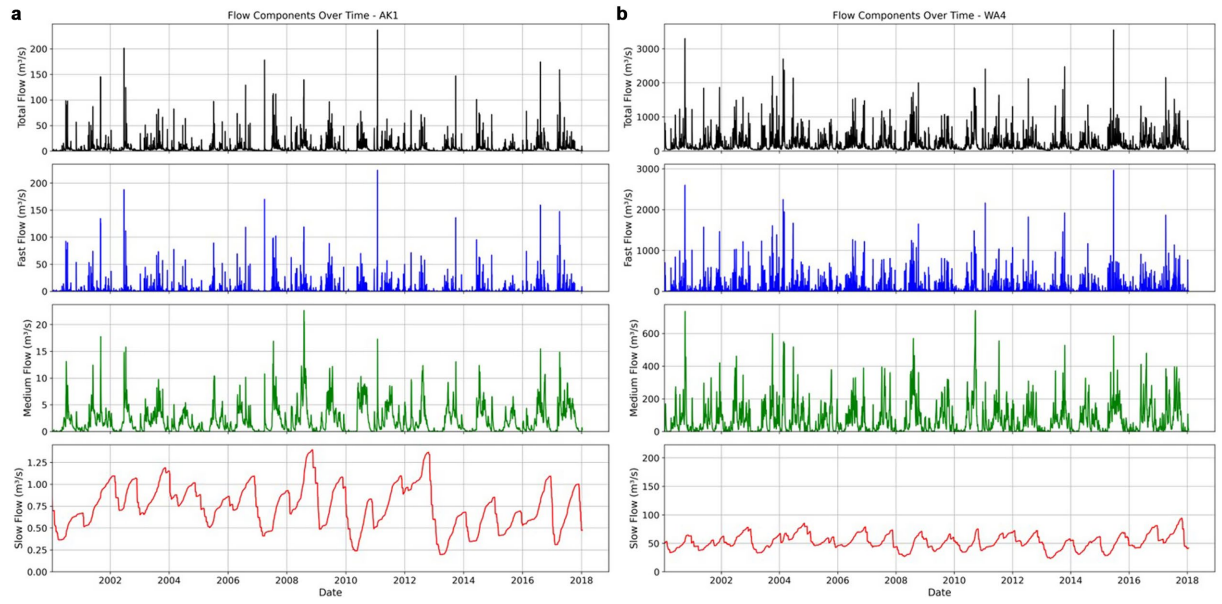


FIGURE 6
Separation of flow at **(a)** AK1 and **(b)** WA4 into slow, medium and fast flow components.

3.3 Contribution of groundwater to surface water flows

While previous studies using BACH analysis have focussed on nutrient loads entering rivers, also implicit in the BACH analyses is a characterisation of the surface- and ground-water interactions above the measurement site over time.

Figures 6a,b show the median of the posterior predicted timeseries for the flow pathway components for sites AK1 and WA4, respectively. From this figure it is clear that the variability of the slow flow component is significantly dampened relative to the total river flow, showing only low frequency signals. The medium flow component has a dampened frequency and amplitude of the total flow time series, while the fast flow component tends to have a frequency and amplitude commensurate with the total flow time series.

Also implied in Figure 6 is that groundwater input to rivers is irregular over short timescales (days to months). A similar observation was noted by Cartwright (2022), who comments that it is not clear whether baseflow as a whole varies smoothly, as is typically assumed in hydrograph baseflow separation techniques.

Due to the relative magnitude of the three flow components (e.g., slow flow, median flow and quick flow) it is often difficult to identify the changes in the dominant flow component during temporally varying river flow conditions. To address this, we explored the relative magnitude of each flow path component for the flow duration (exceedance) curve at all sites. When the three flow components are normalised by the total flow, the relative contributions of slow, medium and fast components across the flow range become more apparent. Figures 7a,b provides an example of these results for sites AK1 (Hoteo River, Auckland) and WA4 (Whanganui River) which correspond to the time series plots shown in Figure 6.

As expected, at all sites, at low river flows (highest probability of flow exceedance) the slow flow components dominate, reflecting deeper groundwater inflow. Fast flows (reflecting near surface flows) dominate at the highest river flows. The median flow component contributions (shallow groundwater) also increase with increasing total river flow.

Different flow paths are activated during the rising limb compared to the falling limb of a flow event and seasonal patterns, with a resulting hysteresis of flow pathway characteristics (Woodward and Stenger, 2020; Bowes et al., 2005). Therefore, hysteresis in the flow path contribution to total flow ratios will depend on the antecedent or postcedent interaction between multiple dynamic components (e.g., river stage, aquifer head). Because of this the simulated flow component ratios vary for specific flow percentiles (Woodward and Stenger, 2020). The changing flow paths that are activated in the rising and falling limbs of hydrographs results in scatter or variability in these contributions for any selected river flow percentile as seen in Figure 7.

Somewhat surprising, and an important motivation for writing this paper, is the indication of the continuing importance of groundwater (both slow and medium groundwater contributions) at reasonably high flow percentiles (Figure 7). Such results suggest that high river flows are as much a function of increased medium (shallow groundwater) discharges as fast (near surface) flows. For most sites groundwater components (combined slow and medium components) dominate, even to higher flow percentiles.

We focus on flow rates as our primary interest is understanding how groundwater flow paths contribute to river flows during low flow drought events and during flood events. However, we note that previous studies using BACH have focussed on flow volumes and nutrient loads. The same flow path contributions depicted in Figure 7 could also be normalised by total volume rather than flow rate, which would tend to compress the left-hand end and expand the right-hand end of the above Figure 7.

Figure 8 shows the case study distribution of the combined groundwater flow path proportions normalised by river flow, for the 25th, 75th and 95th flow percentiles of the flow exceedance

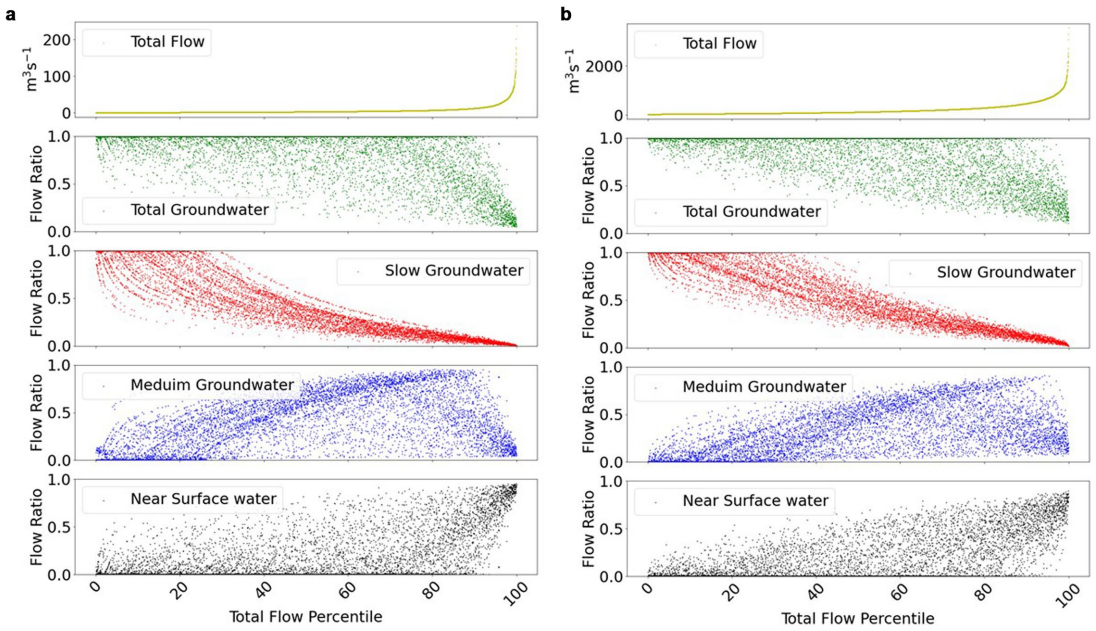


FIGURE 7
Flow and flow contributions summary at all flow percentiles at river sites: **(a)** AK1 (left), and **(b)** WA4 (right). Top panel is the river flow exceedance curve. Next is the ratio of total groundwater contribution to the measured river flow, then the ratio of slow (deep groundwater) contribution to measured river flow, the ratio of medium (shallow groundwater) contribution to measured river flow, and the bottom panel is the ratio of fast (near surface) flow path contribution to measured river flow.

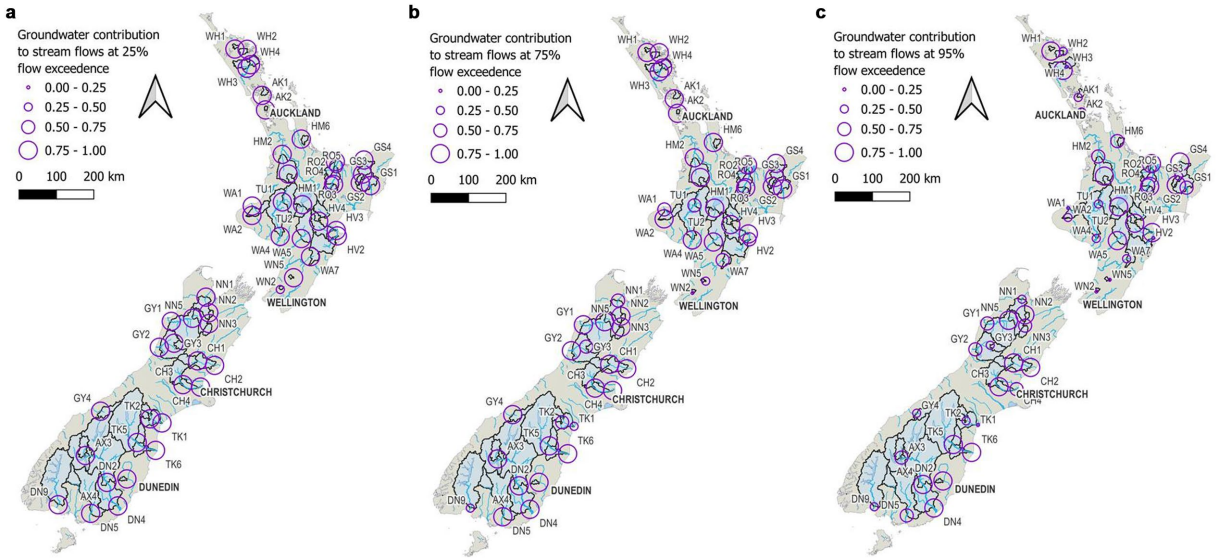


FIGURE 8
Contribution of groundwater components (slow and medium flow paths) to: **(a)** the 25%, **(b)** 75%, and **(c)** 95% river flow percentile.

distribution for all sites. These contributions are also summarised for all sites in Supplementary Tables S2, S3. The key result in Figures 8a,b is the predominance of groundwater contributions at low flows (at the 25th flow exceedance percentile), which continues at the 75th flow percentile for most sites in this study. This groundwater contribution dominance extends to many sites even at high flows (the 95th flow percentile) as depicted in Figures 8c.

These results are supported by previous work which has also indicate that groundwater plays an important contribution to river baseflows. Using a baseflow a filtering method Singh et al. (2019) found that the baseflow index indicated that on average 53% of streamflow volume in New Zealand is sustained by groundwater and other delayed sources. Similarly, Rajanayaka et al. (2020) and Ahiablame et al. (2013) found that groundwater contributed 77 and

60% of total stream volumes in the Waikato catchment (New Zealand) and sites in Indiana (USA), respectively. These studies pointed out the importance of these substantial groundwater contributions to total stream flow for adequate management of water resources and water quality.

By exploring the dynamics of the groundwater contribution to river flow, the results shown in this paper suggest that groundwater contributions to stream flow rates may be even higher, with values of groundwater contributions being commonly over 80% of river flow at the 75th flow percentile and many streams are still predominantly groundwater sourced at the 95th flow percentile.

Further, the analysis suggests that on average more than 40% of the 50th flow percentile is contributed by older water in slow flow paths. While not reflected in current water management practice, these results are consistent with those of Tetzlaff et al. (2014) who found that pre-event water (e.g., older than near surface flows) accounted for >80% of flow, even in large events. Similarly, Stewart et al. (2012), Morgenstern et al. (2010), and Stewart et al. (2010) have also identified old water contributions to river flow. The results are also consistent with Sklash and Farvolden (1979) who noted that groundwater discharge into rivers explains the temporal variations in river water chemistry which are not adequately accounted for by other theories.

Figure 9 provides a summary of the variation around the mean proportions for the 5th, 25th, 50th, 75th, and 95th percentiles for all sites. The figure shows a clear pattern of a decreasing proportion of slow flow paths and increasing proportion of medium flow paths as river flows increase at all sites. Also evident is the dominance of fast flows only at very high flows (>95% flow percentile), but with similarly high contributions also occurring from shallow groundwater flow paths. The variation in these mean flows across catchments is

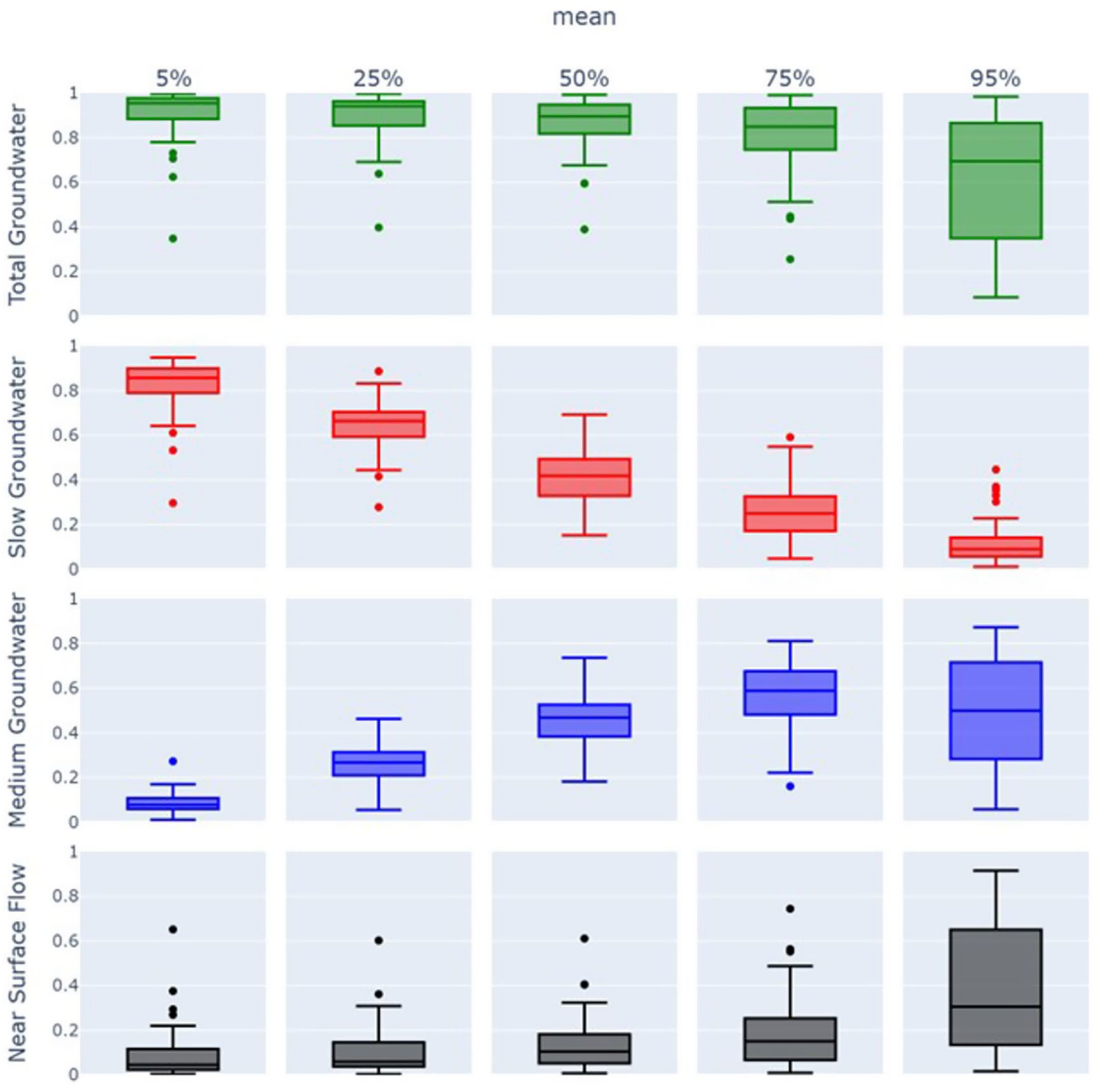


FIGURE 9
Box plot grid displaying median estimates of total, slow, medium groundwater, and near surface flow path contributions at 5, 25, 50, 75, and 95 river flow exceedance percentiles. Each row represents total, slow, medium groundwater, and near surface flow contributions respectively, with columns showing the distribution of these flow path contribution estimates at specified flow exceedance percentiles.

high and can likely be attributed to differences in surface runoff, infiltration and flashiness.

Finally, we explored the relationship between the slow flow path contributions and the FI (combining data from Table 1 and Supplementary Table S2), and found that while these are generally negatively correlated, consistent with patterns already reported in the literature (Woodward and Stenger, 2018, 2020), the correlation was not statistically significant. We note that Singh et al. (2019) and Yang et al. (2025) explored the relationship of average baseflow volumes with geological and geomorphological characteristics of catchments to extend catchment baseflow estimates and river nutrient loads to unmonitored catchments. Preliminary analyses were undertaken in this study which also indicate correlations with these characteristics, albeit with each mappable variable not correlated at a statistically significant level. Correlations of combinations of these mappable variates with the flow components would be a useful extension to this work.

4 Conclusion

Our study demonstrates two key advantages of the BACH method; (i) it can be used to provide conceptual insights on the deep and shallow groundwater contribution to stream flow very quickly and at large scales, and (ii) it can provide these insights based on only sparse datasets. This offers distinct advantages over existing chemistry assisted (e.g., EMMA) and conventional two component baseflow separation methods. These insights can then be used to guide future field programmes and modelling work.

Applying the BACH methodology to sites within the active National Water Quality Network in New Zealand provided important information on the source of river inflows. A key finding of this study, and the primary motivation for writing this paper, is the persistent and significant contribution of groundwater to river flows (both deep slow and shallow groundwater flow paths) even at high flow percentiles.

The results obtained in this study indicate that for most sites analysed, groundwater components (combined slow and medium flow components) were found to dominate flows, even at higher flow percentiles. The analyses indicate that on average more than 40% of the 50th flow percentile is contributed by slow groundwater flow paths (considered to be analogous to deeper regional groundwater flow). The combination of slow flow paths (deep groundwater) and medium flow paths (shallower groundwater) contributes more than 80% of river flows at the 75% flow percentile. For many sites this high groundwater contribution persists even at the 95% flow percentile. Meanwhile high river flows can be a function of increased medium (shallow groundwater) discharges as much as fast (near surface) flows.

Our analysis also revealed that the slow and medium baseflow components of river flows vary significantly over short time scales (days to months). This calls into question the common assumption of a smoothly varying baseflow component. Understanding these dynamics is important for a broader understanding of how the quantity and quality of baseflow may vary over time, particularly as future climate patterns change.

This study demonstrated the use of BACH, as a powerful but simple alternative to current baseflow separation methods for examining the extent and dynamics of surface- and ground-water system connectedness across the river flow regime as measured at a single location. This extends the current application of the BACH method which to date has been used to identify whether the provenance of nutrient sources in rivers originates from slow, medium or fast flow components.

A cautious approach to these results is warranted. The model is very simple, in particular the estimated magnitude of the medium flow-shallow groundwater flow pathway can be questioned, as the MCMC analysis indicates that the separation between the near surface and medium pathway is often non-unique. However, other widely used baseflow separation techniques are also simplistic, and do not have the benefit of additional information from river chemistry time series. The questioning of these previous estimates of low-moderate groundwater contributions to stream flow we consider to be the major contribution of this study. Use of water age data (Stenger et al., 2024), and groundwater chemistry concentrations to corroborate these results would be a useful extension to this study where age data becomes available.

We found that the BACH method works well wherever there is a strong relationship between flow and the TP and NNN water quality analytes, which exists in many developed catchments with intensive land use. Stenger et al. (2024) commented that other water quality analytes can be used successfully in the BACH analysis and may be more suitable in undeveloped catchments were the relationships between TP, NNN and river flows do not hold. We note that the analyses described in this paper were only applied to perennial rivers. Ephemeral rivers are estimated to be up to 51% of rivers globally (Messager et al., 2021), and it would be useful to test the performance of the BACH method in that context.

A potential extension of this study is to combine the BACH method with physically based numerical models (Ghattas and Willcox, 2021), with the goal of enhancing the efficiency and effectiveness of decision support modelling. In such an approach, the application of BACH would be to rapidly pre-process the information in hydrographs, so that only the relevant information from the time series needs to be represented in the numerical model, allowing significant run time and model complexity savings. A further benefit could be that the groundwater flow path estimation is not obfuscated by other stresses (e.g., natural stresses such as the meteorology, climate change etc.), even if they have a larger contribution to the river flow regime. In this manner the flow pathway component separation is isolated, which may provide large benefits for reducing the uncertainty in future decision-critical predictions.

Understanding these very substantial groundwater contributions to total river flow is important for any management of water resources and water quality in New Zealand and could support similar approaches in other countries. Further these results are somewhat counter to the general conceptualisation of the contributions that groundwater makes to river flows. This has important implications for management of freshwater to mitigate both drought and flood impacts, as well as for the management of river quality.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://hydrowebportal.niwa.co.nz/.

Author contributions

CM: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft. PO: Conceptualization, Formal analysis, Investigation, Software, Visualization, Writing – review & editing. BH: Formal analysis, Software, Supervision, Visualization, Writing – review & editing. SC: Supervision, Writing – review & editing. JY: Conceptualization, Data curation, Writing – review & editing. MT: Data curation, Visualization, Writing – review & editing. CR: Conceptualization, Data curation, Writing – review & editing. SW: Formal analysis, Methodology, Validation, Writing – review & editing. MM: Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research received funding from the GNS led, New Zealand Ministry for Business, Innovation, and Employment's Te Whakeheke o te Wai programme (Contract No. C05X1803).

Acknowledgments

The authors thank Dr. Roland Stenger of Lincoln Agritech for the constructive discussion of BACH application.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2025.1584947/full#supplementary-material

References

Adams, G. A., Cornish, P. S., Croke, B. F. W., Hart, M. R., Hughes, C. E., and Jakeman, A. J. (2009). A new look at uncertainty in end member mixing models for streamflow partitioning. Proceedings of 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation

Ahiablame, L., Chaubey, I., Engel, B., Cherkauer, K., and Merwade, V. (2013). Estimation of annual baseflow at ungauged sites in Indiana USA. *J. Hydrol.* 476, 13–27. doi: 10.1016/j.jhydrol.2012.10.002

Aubert, A. H., Gascuel-Odoux, C., Gruau, G., Akkal, N., Faucheux, M., Fauvel, Y., et al. (2013). Solute transport dynamics in small, shallow groundwater-dominated agricultural catchments: insights from a high-frequency, multisolute 10 yr-long monitoring study. *Hydrol. Earth Syst. Sci.* 17, 1379–1391. doi: 10.5194/hess-17-1379-2013

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W. (2004). A new flashiness index: characteristics and applications to Midwestern rivers and streams. *J. Am. Water Resour. Assoc.* 40, 503–522. doi: 10.1111/j.1752-1688.2004.tb01046.x

Ballantine, D. J., and Davies-Colley, R. J. (2014). Water quality trends in New Zealand rivers: 1989–2009. *Environ. Monit. Assess.* 186, 1939–1950. doi: 10.1007/s10661-013-3508-5

Barthold, F. K., Tyralla, C., Schneider, K., Vache, K. B., Frede, H.-G., and Breuer, L. (2011). How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis. *Water Resour. Res.* 47. doi: 10.1029/2011WR010604

Beck, H. E., Van Dijk, A. I., Miralles, D. G., De Jeu, R. A., Bruijnzeel, L., McVicar, T. R., et al. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resour. Res.* 49, 7843–7863. doi: 10.1002/2013WR013918

Bosch, D. D., Arnold, J. G., Allen, P. G., Lim, K.-J., and Park, Y. S. (2017). Temporal variations in baseflow for the Little River experimental watershed in South Georgia, USA. *J. Hydrol. Reg. Stud.* 10, 110–121. doi: 10.1016/j.ejrh.2017.02.002

Bowes, M. J., House, W. A., Hodgkinson, R. A., and Leach, D. V. (2005). Phosphorus-discharge hysteresis during storm events along a river catchment: the River Swale, UK. *Water Res.* 39, 751–762. doi: 10.1016/j.watres.2004.11.027

Broda, S., Larocque, M., Paniconi, C., and Haitjema, H. (2011). A low-dimensional hillslope-based catchment model for layered groundwater flow. *Hydrol. Process.* 26, 2814–2826. doi: 10.1002/hyp.8319

Bugaets, A., Gartsman, B., Gubareva, T., Lupakov, S., Kalugin, A., Shamov, V., et al. (2023). Comparing the runoff decompositions of small experimental catchments: end-member mixing analysis (EMMA) vs. hydrological modelling. *Water.* 15:752 doi: 10.3390/w15040752

Cartwright, I. (2022). Implications of variations in stream specific conductivity for estimating baseflow using chemical mass balance and calibrated hydrograph techniques. *Hydrol. Earth Syst. Sci.* 26, 183–195. doi: 10.5194/hess-26-183-2022

Das, K., Mukherjee, A., Malakar, P., Das, P., and Dey, U. (2021). Impact of global-scale hydroclimatic patterns on surface water-groundwater interactions in the climatically vulnerable Ganges river delta of the Sundarbans. *Sci. Total Environ.* 798:149198. doi: 10.1016/j.scitote.nv.2021.149198

de Graaf, I. E. M., Gleeson, T., van Beek, L. P. H., Sutanudjaja, E. H., and Bierkens, M. F. P. (2019). Environmental flow limits to global groundwater pumping. *Nature* 574, 90–94. doi: 10.1038/s41586-019-1594-4

Deelstra, J., and Iital, A. (2008). The use of the flashiness index as a possible indicator for nutrient loss prediction in agricultural catchments. *Boreal Environ. Res.* 13, 209–221. Available at: http://hdl.handle.net/10138/234732

Delsman, J. R., Oude Essink, G. H. P., Beven, K. J., and Stuyfzand, P. J. (2013). Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment. *Water Resour. Res.* 49, 4792–4806. doi: 10.1002/wrcr.20341

Doherty, J., and Christensen, S. (2011). Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. *Water Resour. Res.* 47:W12534. doi: 10.1029/2011WR010763

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrol. Process.* 19, 507–515. doi: 10.1002/hyp.5675

Fan, Y., Chen, Y., Liu, Y., and Li, W. (2013). Variation of baseflows in the headstreams of the Tarim River Basin during 1960–2007. *J. Hydrol.* 487, 98–108. doi: 10.1016/j.jhydrol.2013.02.037

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian data analysis: texts in statistical science series. *2nd* Edn. Boca Raton, FL: CRC Press.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136

Ghattas, O., and Willcox, K. (2021). Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica* 30, 445–554. doi: 10.1017/S0962492921000064

Gorelick, S. M., and Zheng, C. (2015). Global change and the groundwater management challenge. *Water Resour. Res.* 51, 3031–3051. doi: 10.1002/2014WR016825

Hesser, F. B., Franko, U., and Rode, M. (2010). Spatially distributed lateral nitrate transport at the catchment scale. *J. Environ. Qual.* 39, 193–203. doi: 10.2134/jeq2009.0031

Howcroft, W., Cartwright, I., and Cendón, D. I. (2019). Residence times of bank storage and return flows and the influence on river water chemistry in the upper Barwon River, Australia. *Appl. Geochem.* 101, 31–41. doi: 10.1016/j.apgeochem.2018.12.026

Kitlasten, W., Moore, C. R., and Hemmings, B. (2022). Model structure and ensemble size: implications for predictions of groundwater age. *Front. Earth Sci.* 10:972305. doi: 10.3389/feart.2022.972305

Klaus, J., and McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: review and evaluation. *J. Hydrol.* 505, 47–64. doi: 10.1016/j.jhydrol.2013.09.0006

Konikow, L. F., and Leake, S. A. (2014). Depletion and capture: revisiting "the source of water derived from wells". *Groundwater* 52, 100–111. doi: 10.1111/gwat.12204

Lott, D. A., and Stewart, M. T. (2016). Base flow separation: a comparison of analytical and mass balance methods. *J. Hydrol.* 535, 525–533. doi: 10.1016/j.jhydrol.2016.01.063

Martinez, G. F., and Gupta, H. V. (2010). Toward improved identification of hydrological models: a diagnostic evaluation of the "*abcd*" monthly water balance model for the conterminous United States. *Water Resour. Res.* 46. doi: 10.1029/2009WR008294

Messager, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., et al. (2021). Global prevalence of non-perennial rivers and streams. *Nature* 594, 391–397. doi: 10.1038/s41586-021-03565-5

Moreau, M., Herpe, M., and Santamaria Cerrutti, M. E. (2025). "2024 update of the national groundwater quality indicator" in GNS Science Consultancy Report 2024/90 (Wairakei: GNS Science), 83.

Morgenstern, U., Stewart, M. K., and Stenger, R. (2010). Dating of streamwater using tritium in a post nuclear bomb pulse world: continuous variation of mean transit time with streamflow. *Hydrol. Earth Syst. Sci.* 14, 2289–2301. doi: 10.5194/hess-14-2289-2010

NRWQN. (2024). Available online at: https://niwa.co.nz/freshwater/water-quality-monitoring-and-advice/national-river-water-quality-network-nrwqn

O'Brien, R. J., Misstear, B. D., Gill, L. W., Deakin, J. L., and Flynn, R. (2013). Developing an integrated hydrograph separation and lumped modelling approach to quantifying hydrological pathways in Irish river catchments. *J. Hydrol.* 486, 259–270. doi: 10.1016/j.jhydrol.2013.01.034

Rajanayaka, C., Weir, J., Barkle, G., Griffiths, G., and Hadfield, J. (2020). Assessing changes in nitrogen contamination in groundwater using water aging: Waikato River, New Zealand. *J. Contam. Hydrol.* 234:103686. doi: 10.1016/j.jconhyd.2020.103686

Rosenberry, D. O., LaBaugh, J. W., and Hunt, R. J. (2008). "Use of monitoring wells, portable piezometers, and seepage meters to quantify flow between surface water and ground water" in Field techniques for estimating water fluxes between surface water and ground water (Scotts Valley, CA: CreateSpace), 39–70.

Scanlon, B. R., Fakhreddine, S., Rateb, A., de Graaf, I., Famiglietti, J., Gleeson, T., et al. (2023). Global water resources and the role of groundwater in a resilient water future. *Nat. Rev. Earth Environ.* 4, 87–101. doi: 10.1038/s43017-022-00378-6

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., et al. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9320–9325. doi: 10.1073/pnas.1200311109

Seo, S. B., Mahinthakumar, G., Sankarasubramanian, A., and Kumar, M. (2018a). Conjunctive management of surface water and groundwater resources under drought conditions using a fully coupled hydrological model. *J. Water Resour. Plan. Manag.* 144:04018060. doi: 10.1061/(ASCE)WR.1943-5452.0000978

Seo, S. B., Mahinthakumar, G., Sankarasubramanian, A., and Kumar, M. (2018b). Assessing the restoration time of surface water and groundwater systems under groundwater pumping. *Stoch. Environ. Res. Risk Assess.* 32, 2741–2759. doi: 10.1007/s00477-018-1570-9

Singh, S. K., Pahlow, M., Booker, D. J., Shankar, U., and Chamorro, A. (2019). Towards baseflow index characterisation at national scale in New Zealand. *J. Hydrol.* 568, 646–657. doi: 10.1016/j.jhydrol.2018.11.025

Sklash, M. G., and Farvolden, R. N. (1979). The role of groundwater in storm runoff. *J. Hydrol.* 43, 45–65. doi: 10.1016/S0167-5648(09)70009-7

Stan Development Team (2024). RStan: the R interface to Stan T Stan Development - R package version 2.17. 3 Available online at: https://mc-stan.org/ (Accessed 2024).

Stenger, R. (2022). "Nitrogen lag review" in Report 1058-14-R1 prepared for Waikato Regional Council (Lincoln: Lincoln Agritech Ltd.).

Stenger, R., Park, J., and Clague, J. (2024). Routine stream monitoring data enables the unravelling of hydrological pathways and transfers of agricultural contaminants through catchments. *Sci. Total Environ.* 912:169370. doi: 10.1016/j.scitotenv.2023.169370

Stewart, M. K., Morgenstern, U., and McDonnell, J. J. (2010). Truncation of stream residence time: how the use of stable isotopes has skewed our concept of streamwater age and origin. *Hydrol. Process.* 24, 1646–1659. doi: 10.1002/hyp.7576

Stewart, M. K., Morgenstern, U., Mcdonnell, J. J., and Pfister, L. (2012). The 'hidden streamflow' challenge in catchment hydrology: a call to action for stream water transit time analysis. *Hydrol. Process.* 26, 2061–2066. doi: 10.1002/hyp.9262

Su, C.-H., Costelloe, J. F., Peterson, T. J., and Western, A. W. (2016). On the structural limitations of recursive digital filters for base flow estimation. *Water Resour. Res.* 52, 4745–4764. doi: 10.1002/2015WR018067

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., et al. (2013). Ground water and climate change. *Nat. Clim. Change* 3, 322–329. doi: 10.1038/nclimate1744

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., and Soulsby, C. (2014). Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. *Water Resour. Res.* 50, 969–985. doi: 10.1002/2013WR014147

Wilson, S. R., Hoyle, J., Measures, R., Di Ciacca, A., Morgan, L. K., Banks, E. W., et al. (2024). Conceptualising surface water–groundwater exchange in braided river systems. *Hydrol. Earth Syst. Sci.* 28, 2721–2743. doi: 10.5194/hess-28-2721-2024

Woodward, S. J., and Stenger, R. (2018). Bayesian chemistry-assisted hydrograph separation (BACH) and nutrient load partitioning from monthly stream phosphorus and nitrogen concentrations. *Stoch. Environ. Res. Risk Assess.* 32, 3475–3501. doi: 10.1007/s00477-018-1612-3

Woodward, S. J., and Stenger, R. (2020). Extension of Bayesian chemistry-assisted hydrograph separation to reveal water quality trends (BACH2). *Stoch. Environ. Res. Risk Assess.* 34, 2053–2069. doi: 10.1007/s00477-020-01860-7

Woodward, S. J., Wöhling, T., Rode, M., and Stenger, R. (2017). Predicting nitrate discharge dynamics in mesoscale catchments using the lumped StreamGEM model and Bayesian parameter inference. *J. Hydrol.* 552, 684–703. doi: 10.1016/j.jhydrol.2017.07.021

Yang, J., Moore, C. R., Rajanayaka, C., Shiona, H., Hemmings, B., Oluwunmi, P., et al. (2025). National nutrient contribution dynamics in New Zealand rivers. *Hydrol. Process.* 39:e70161. doi: 10.1002/hyp.70161

Yang, W., Xiao, C., Zhang, Z., and Liang, X. (2021). Can the two-parameter recursive digital filter baseflow separation method really be calibrated by the conductivity mass balance method? *Hydrol. Earth Syst. Sci.* 25, 1747–1760. doi: 10.5194/hess-25-1747-2021

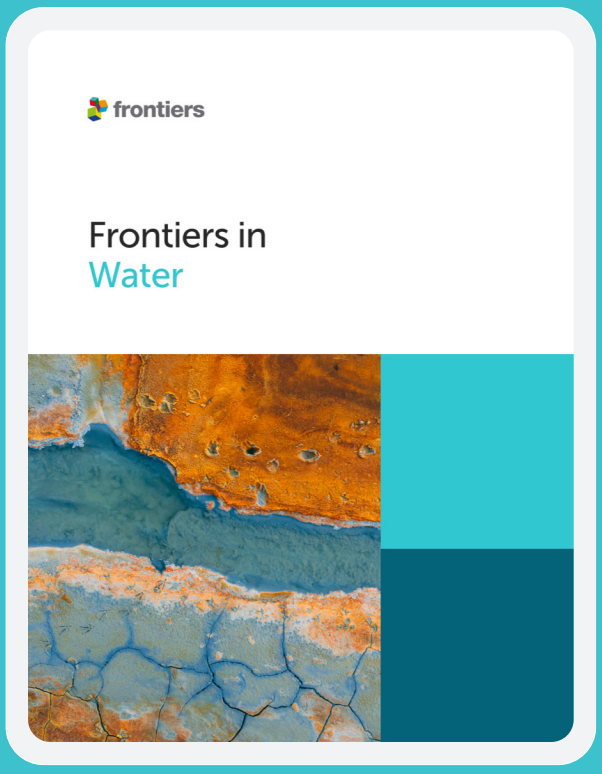
Frontiers in Water

Transforming our approach to water research and its applications

A journal dedicated to exploring challenges facing freshwater systems, including demand and supply of water resources, extreme weather events and climate change.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact