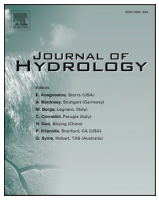
Journal of Hydrology 635 (2024) 131193

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Old-Aged groundwater contributes to mountain hillslope hydrologic dynamics

Nicholas E. Thiros [a,*], Erica R. Siirila-Woodburn [a], Matthias Sprenger [a], Kenneth H. Williams [a,b], James P. Dennedy-Frank [c], Rosemary W.H. Carroll [d], Gardner W.P. [e]

[a] Lawrence Berkeley National Laboratory, Earth and Environmental Sciences Area, Berkeley, CA, United States
[b] Rocky Mountain Biological Laboratory, Gothic, CO, United States
[c] Northeastern University, Marine and Environmental Sciences, Boston, MA, United States
[d] Desert Research Institute, Reno, NV, United States
[e] University of Montana, Department of Geosciences, Missoula, MT, United States

ARTICLE INFO

Handling Editor: Corrado Corradini

Keywords:
Groundwater age
Mountain hydrology
Bedrock groundwater
Integrated hydrologic modeling
Particle tracking
Environmental tracers

ABSTRACT

Understanding connectivity between the soil and deeper bedrock groundwater is needed to accurately predict a watershed's response to perturbation, such as drought. Yet, the bedrock groundwater dynamics in mountainous environments are typically under-constrained and excluded from watershed hydrologic models. Here, we investigate the role of groundwater characterized with decadal and longer water ages on the hydrologic and mass-transport processes within a steep snow-dominated mountain hillslope in the Central Rocky Mountains (USA). We quantify subsurface and surface water mass-balance, groundwater flowpaths, and age distributions using the ParFlow-CLM integrated hydrologic and EcoSLIM particle tracking models, which are compared to hydrometric and environmental tracer observations. An ensemble of models with varied soil and hydrogeologic parameters reproduces observed groundwater levels and century-scale mean ages inferred from environmental tracers. The numerical models suggest soil water near the toe of the hillslope contains considerable (>60 % of the mass-flux) contributions from bedrock flowpaths characterized with water ages >10 years. Flowpath connectivity between the deeper bedrock and soil systems is present throughout the year, highlighting the potentially critical role of groundwater with old ages on processes such as evapotranspiration and streamflow generation. The coupled numerical model and groundwater age observations show the bedrock groundwater system influences the hillslope hydrodynamics and should be considered in mountain watershed conceptual and numerical models.

1. Introduction

Snow-dominated mountainous watersheds provide critical water supplies for the Western United States and much of the world (Viviroli et al., 2007). With the increased occurrence of low-snow winters, earlier snowmelt, and overall hydrologic drought due to climate change (Immerzeel et al., 2020; Mote et al., 2018; Siirila-Woodburn et al., 2021), mountain watersheds will be challenged to provide sustainable water yields and balance demand for human, agricultural, and ecological uses. Better understanding the hydrologic cycle and connectivity between subsurface reservoirs and flowpaths is needed to accurately predict how mountain systems will respond to climate change and promote sustainable water resources (Meixner et al., 2016; Taylor et al., 2013).

Groundwater is recognized as a crucial, yet uncertain and understudied, mechanism of how mountain watersheds store and transmit water (Hayashi, 2020; Singha & Navarre-Sitchler, 2022; Somers & McKenzie, 2020). Bedrock groundwater in high elevation mountains can be dynamic on sub-annual timescales (Gardner et al., 2020; Manning & Caine, 2007) and essential for streamflow (Carroll et al., 2020; White et al., 2019), plant transpiration (Harmon et al., 2020; Meyers et al., 2021; Ryken et al., 2022), and solute exports (Frisbee et al., 2013). Alternatively, studies have found mountain bedrock groundwater represents a minor flux to surface waters, highlighting the complexity and site-specific dependence of surface water-groundwater connectivity (Manning et al., 2021; Tokunaga et al., 2022). Nonetheless, an

* Corresponding author.
E-mail address: nthiros@lbl.gov (N.E. Thiros).

https://doi.org/10.1016/j.jhydrol.2024.131193
Received 22 November 2023; Received in revised form 11 March 2024; Accepted 26 March 2024
Available online 13 April 2024

inconsequential bedrock groundwater system is often the *a priori* and untested assumption applied to flowpath partitioning in mountain watersheds (see Condon et al., 2020 for discussion). Research that applies observational and numerical modeling techniques sensitive to bedrock groundwater hydrologic processes are needed.

Groundwater age distributions are commonly used to understand mixing between subsurface flowpaths and water sources, such as young, shallow soil water and old, deep bedrock water (Cook & Herczeg, 2000; Sprenger et al., 2019). Recent field observations suggest bedrock groundwater in discharge zones of mountain hillslopes and catchments can have mean groundwater ages on the order of decades to hundreds of years (Manning et al., 2012; Thiros et al., 2023a; White et al., 2019). However, it generally remains uncertain whether these bedrock groundwater flowpaths characterized with considerable ages actively contribute to the shallow soil water with a meaningful flux (Hale et al., 2016; Manning et al., 2021). This uncertainty is exacerbated by the sparsity of environmental tracer observations that can constrain age distributions over broad (e.g. decades to centuries) timescales from difficult to access mountainous systems.

Integrated hydrologic models can improve our understanding of the processes controlling water fluxes, age distributions, and connectivity between subsurface reservoirs (Carroll et al., 2020; Engdahl & Maxwell, 2015; Thornton et al., 2022). Including deeper groundwater in integrated hydrologic models can influence simulated watershed streamflow and evapotranspiration (Foster & Maxwell, 2019; Kollet & Maxwell, 2008; Maina et al., 2022), leading to groundwater ages on the order of decades to hundreds of years (Carroll et al., 2020; Engdahl & Maxwell, 2015; Rapp et al., 2020). However, catchment-scale numerical models often do not include the bedrock groundwater system in adequate fidelity to capture the influence of flowpaths with potentially old water ages on the hydrodynamics over seasonal and longer timescales. For instance, while studies have shown bedrock permeability and porosity distributions have first-order controls on hillslope water age dynamics (Ameli et al., 2016; Carroll et al., 2020), it is common that numerical model domains are shallow (<10 m) and simulations are short (<10 years). Few studies have applied integrated modeling approaches to understand the transient mixing of deeper bedrock groundwater with shallow soil and stream water over broad timescales (e.g. decades to centuries).

In this work we aim to address the following questions:

(1) Do bedrock groundwater flowpaths contribute a meaningful flux to the shallow soil near the toe of a steep mountainous hillslope, which represents a critical and widespread point of convergence for groundwater and surface water?

(2) What is the age distribution of discharging groundwater and soil water along a steep mountain hillslope?

To investigate these questions, we develop a 2-D integrated hydrologic model of a mountain hillslope in the East River Watershed, Colorado (USA). With the model we evaluate the role of bedrock groundwater flowpaths on the hydrologic, mass transport, and groundwater age dynamics in the soil near the toe of the hillslope across water years with a range of annual precipitation (wet, normal, dry). Model predictions are compared to groundwater level observations (hourly over 6 years) and mean groundwater ages quantified using environmental tracers (Thiros et al., 2023a). This work provides insights on whether groundwater with old ages should be better accounted for in mountain catchment hydrologic processes and the expected response to drought.

2. Methods

2.1. Study area

The East River Watershed is a high-elevation (>2700 m above sea-level), snow-dominated watershed in the Central Rocky Mountains, Colorado (USA; Fig. 1A). Detailed descriptions of the East River

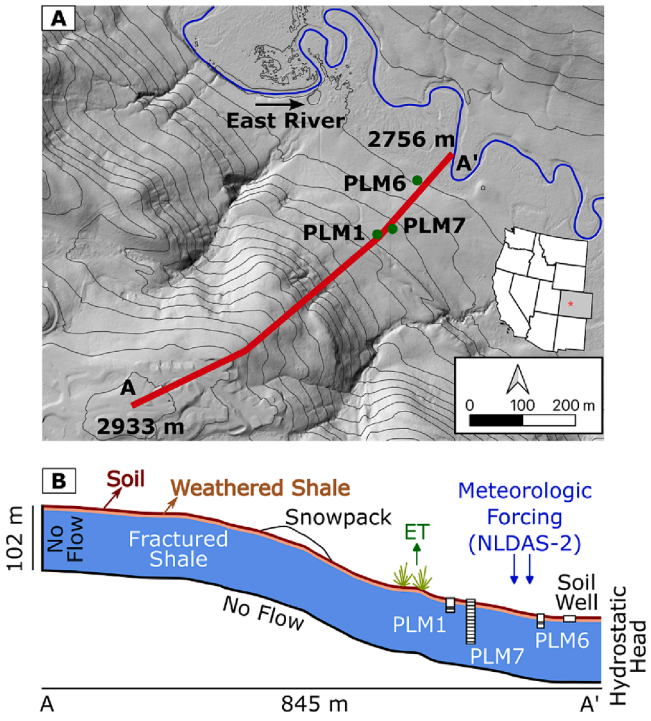


Fig. 1. (A) PLM hillslope located within the East River Watershed, Colorado, USA. Green dots correspond to groundwater wells finished in the fractured shale bedrock. (B) Conceptual and ParFlow-CLM numerical model setup for the 2-D PLM hillslope. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Watershed, summarized here, can be found in Carroll et al. (2018) and Hubbard et al. (2018). We focus on the north-east facing 'Pumphouse' lower montane (PLM) hillslope (Fig. 1). The PLM hillslope ranges in elevation from 2750 to 2930 m and is 845 m long, corresponding to an average 22° slope angle. The top of the PLM hillslope is a surface water divide and the terminus is the East River.

The East River Watershed is characterized as a continental subarctic climate with daily mean air temperatures at the Butte SNOTEL station ranging from − 8.3 °C in winter to 11 °C in summer. Annual precipitation (average ~ 700 mm) mostly occurs as winter snowfall (~80 %) and summer monsoon rains (Carroll et al., 2018). Snowpack at the Butte SNOTEL generally persists from late October into June. The East River hydrograph is tightly coupled to snowmelt timing and magnitude, with peak discharge in early June and baseflow from October to April. The lower-montane region is characterized by grasses and shrubs with annual average evapotranspiration (ET) estimates of ~500 mm (Ryken et al., 2022).

The PLM hillslope has an ~1 m deep soil underlain by weathered, fractured, and faulted Mancos Shale bedrock. Borehole drilling, geophysical logging, and geochemical depth profiles suggest bedrock fracturing decreases with depth and is broadly classified as a shallow (1–3 m below-land-surface (mbls)) weathered bedrock with higher fracture density and deeper (>3 mbls) fractured bedrock (Miltenberger et al., 2021; Tokunaga et al., 2019; Wan et al., 2021). The geophysical logging and bedrock cores from deep boreholes (~70 mbls) in the PLM hillslope show fractures at depths of at least 70 mbls (Uhlemann et al., 2022; Williams & Newman, 2020).

2.2. Observational Data

We compare numerical model predictions to groundwater level and mean age observations from two groundwater wells (PLM1 and PLM6) finished within the fractured bedrock (Fig. 1; see Tokunaga et al., 2019

for well installation details). PLM1 is farthest upslope, 10 m deep, and screened between 6.3 and 7.3 mbls within the fractured shale bedrock. PLM6 is at the toe of the hillslope, 10 m deep, and is also screened within the fractured shale bedrock along the bottom 2 m of the borehole. Both wells have continuous groundwater level measurements from 2017 to present (Faybishenko et al., 2023).

Thiros et al. (2023a) quantified mean groundwater ages at PLM1 and PLM6 using a suite of environmental tracer observations (CFC-12, SF_6, ^3H, and ^4He) and Lumped Parameter Models (Cook & Herczeg, 2000). The environmental tracers were best described using the exponential piston-flow model (EPM) and binary mixing model (BMM). The best-fit EPM was characterized with mean ages ~ 80–120 years while the BMM had older mean ages (centuries to millenia) with a young fraction mean age on the order of decades to a century and an old fraction mean age on the order of thousands of years. These tracer-based mean ages offer alternative interpretations of the field observations, yet both are generally older compared to previous studies that do not consider tracers sensitive to ages on the order of decades to centuries. The co-occurrence of young (CFC-12, SF_6, and ^3H) and old (^4He) tracers at discrete well locations in the shallow bedrock groundwater suggest variable recharge and flowpath mixing between flowpaths with young and old ages.

2.3. Numerical models

We use the ParFlow-CLM integrated hydrologic model (Kollet & Maxwell, 2008; Maxwell, 2013) to simulate subsurface and surface water fluxes along the hillslope. Parflow-CLM couples land-surface mass and energy balance processes with surface and subsurface water fluxes to simulate transient and spatially distributed water pressure heads, fluxes, and velocities. Surface runoff is simulated with the kinematic wave equation and saturated and variable-saturated subsurface flow is simulated with Richards' equation. The land-surface component solves the energy budget to calculate ET and heat fluxes, radiation partitioning, and snow processes. While ParFlow-CLM often has a larger computation expense compared to groundwater-specific models, capturing the snowpack processes with the integrated modeling framework is critical for this snow-dominated mountainous system and is feasible on today's high performance computing resources.

EcoSLIM is a Langragian particle tracking software that is coupled to ParFlow-CLM (Maxwell et al., 2019). We use EcoSLIM to simulate spatially resolved and temporally transient subsurface flowpaths, mass transport, and groundwater age distributions at all numerical grid-cells within the hillslope aquifer. Using the spatially and temporally distributed velocity fields from ParFlow-CLM, EcoSLIM adds particles during timesteps with net positive infiltration flux, moves particles within the subsurface via advection and diffusion, and removes particles as exfiltration and ET fluxes. The entering and exiting particle masses scale according to flux magnitudes allowing for calculation of flux-weighted groundwater age distributions at all subsurface spatial coordinates and times. Particles are removed from the simulation as ET when the evapotranspired (from ParFlow-CLM) to particle mass ratio is greater than a random number between 0 and 1. All simulations were initiated with 5 particles per grid-cell and assumed molecular diffusivity of 4×10^{-6} m^2/hr (broadly representing Xenon). Modeled mean ages from EcoSLIM correspond to the mass-weighted mean of all particles within a particular control volume, thus represent age distributions at a single piezometer location within the hillslope aquifer.

The model domain is a 2-D transect, 845 m long, and 102 m deep (Fig. 1B). This model depth is designed to capture the potential for deep groundwater flowpaths, which are ostensibly supported by the high hydraulic conductivity (>10^{-7} m/s) at ~ 70 mbls measured with borehole geophysical logging (Uhlemann et al., 2022). Land-surface of the model uses a terrain-following grid (Maxwell, 2013) from a 1.5 m resolution digital elevation model (DEM). The 2-D transect model has 559 columns with 1.5 m grid-cell resolution and 32 rows with variable grid-cell resolution ranging from 0.25 m near land-surface up to 10 m at the base of the model. Hydraulic and mass-transport boundary conditions include no-flow along the bottom and left (surface water flow divide) sides of the domain (Fig. 1B). The downstream boundary is at the East River and was represented as a temporally constant hydrostatic head of 0.44 m applied at land-surface, corresponding to the annual average river depth. Perturbing this hydrostatic head +/- 0.5 m had minor impacts on the simulated groundwater head distributions. Land-surface hydrologic and energy boundary conditions (precipitation, air temperature, radiation, relative humidity, and wind speed) were taken from the 1-hourly resolution North American Land Data Assimilation System (NLDAS-2) dataset from 1980 to 2022 (Xia et al., 2012).

ParFlow-CLM simulations from water year 2000 (WY2000) to WY2021 used a 1-hour timestep and were dynamically spun-up with a transient 10-year simulation using the long-term (21 water-years) hourly averaged NLDAS-2 forcing conditions. EcoSLIM was run with variable timesteps to speed-up simulations while retaining transport of particles with old ages. An initial 3000 year EcoSLIM simulation was performed with 10-day timesteps using the 10-day average fluxes from the ParFlow-CLM spin-up runs. The output from this run was used to restart a 50-year EcoSLIM simulation at 24-hour timesteps using the ParFlow-CLM spin-up fluxes. Finally, EcoSLIM was run with 24-hour timesteps from WY2000-WY2021. For all model scenarios, ~100 % of the particles initialized at the beginning of the spinup exited the domain.

We interpret model results at the PLM1 and PLM6 bedrock groundwater wells and in the soil zone at 1 m depth near the toe of the hillslope (Fig. 1B). The soil location was subjectively placed near the groundwater discharge zone, but away from the river to minimize the impact of model boundary conditions, with the explicit purpose to identify the connectivity between the deeper bedrock and shallow soil systems with the model.

2.4. Model parameterization and sensitivity analysis

The ParFlow-CLM and EcoSLIM models are simplified to three hydrogeologic layers (surficial soil, intermediate weathered shale, and deep fractured shale; Fig. 1B) that are constant thickness (Table 1) and follow land surface topography. Model layer thicknesses are based on qualitative analysis of bedrock cores (Williams & Newman, 2020) and broadly match previous models of the hillslope (Tokunaga et al., 2019). Beneath the fractured shale we assume impermeable bedrock.

Table 1
Base model hydrogeologic and soil parameters. Values in parentheses are the minimum and maximum bounds used in the Monte Carlo analysis with the constraint $K_{soil} > K_{wshale} > K_{fshale}$ (parameter is held constant if no parentheses are present).

Parameter	Soil	Weathered Shale	Fractured Shale	Source
Layer Thickness (m)	1	3	98	Tokunaga et al., 2019
Saturated K (m/s)	2.2×10^{-5} (5×10^{-6}, 5×10^{-4})	4.0×10^{-6} (10^{-7}, 10^{-5})	2.2×10^{-8} (10^{-10}, 10^{-6})	Tokunaga et al., 2019
Porosity φ (%)	40	20	10	Uhlemann et al., 2022
van-Genuchten α (m^{-1})	1.82(0.1, 5.5)	0.52(0.1, 5.5)	0.52	Rogers et al., 2021
van-Genuchten. n (–)	1.79(1.0, 5.0)	1.60(1.0, 5.0)	1.60	Rogers et al., 2021
van-Genuchten Residual Saturation (–)	0.13	0.04	0.005	Rogers et al., 2021
Specific Storage (m^{-1})	1×10^{-5}	1×10^{-5}	1×10^{-5}	Foster & Maxwell, 2019

Subsurface and surface parameters are homogeneous within the respective layer. Hydrogeologic, soil, and land-surface parameters for an initial 'base' model were estimated using field-based measurements and literature values (Table 1). In particular, the base model parameters were estimated using well recovery tests and a Guelph permeameter (Tokunaga et al., 2019), bore-hole geophysics (Uhlemann et al., 2022), and soil core (Rogers et al., 2021) field-measurements and literature values when needed. The base model represents our *a priori* best estimate of model parameters given the field characterization. For all models, a homogeneous Manning's n of 5.5 $\times$ 10^{-6} (hr $m^{-1/3}$) was applied on land-surface (Foster & Maxwell, 2019), plant-functional types were open-shrubs along the upper 781 m of the hillslope and closed-shrubs in the lower floodplain region, and transpiration was constrained to the top model 4 layers (up to 1.5 mbls). Plant parameterization (e.g. leaf area index, rooting depth, etc.) was defined by the International Geosphere-Biosphere Program (IGBP) database for each plant-functional type.

We perform an informal subsurface parameter sensitivity analysis using the ParFlow-CLM model. The adjustable parameters include (Table 1): hydraulic conductivity in the soil (K_{soil}), weathered shale (K_{wshale}), and fractured shale (K_{fshale}), and van-Genuchten α and n in the soil (α_{soil} and n_{soil}) and weathered shale (α_{wshale} and n_{soil}). A total of 128 parameter-value sets were sampled using the quasi-random Sobol sequence technique (Sobol, 1998) constrained to the parameter ranges in Table 1 (minimum and maximum values in parentheses) and $K_{soil} > K_{wshale} > K_{fshale}$. The adjustable model parameters in the sensitivity analysis were chosen due to their primary controls on the groundwater level response based on previous studies (Gardner et al., 2020; Maina et al., 2022). The parameter ranges were designed to include broad uncertainties not captured in the local scale site characterization used to parameterize the base model. While additional model parameters are uncertain (e.g. porosity, specific storage, etc.), increasing the number of adjustable parameters in the Monte Carlo analysis required computational expenses outside our resources, thus were separately investigated in a second parameter sensitivity test (described below). Recharge fluxes, which have been identified as a key uncertain boundary condition controlling water table responses (e.g. Thiros et al., 2022), were not directly varied. Recharge is not an applied boundary condition in ParFlow-CLM, but rather is simulated by the integrated hydrologic model as a function of processes such as hourly meteorologic forcing conditions, hydrogeologic property heterogeneity and structure, snowpack energy dynamics, spatially resolved evapotranspiration, and variably saturated soil flow.

The 128-run ParFlow-CLM ensemble was compared to observed groundwater levels at PLM1 and PLM6 from WY2017-WY2021. We identified the 5 models with the lowest sum of squared error (SSE) between simulated ParFlow-CLM and observed groundwater levels at PLM1 and PLM6 simultaneously. These 5 models are termed *A1* to *A5*, with *A1* having the lowest SSE (top-performing model). Due to the high computational expense (approximately 1 model run per 24 h distributed across 64 compute cores), EcoSLIM runs were performed only for the *base* and *A1-A5* models. Our goal was not to formally calibrate the model (e.g. Marçais et al., 2022), thus we do not expect to exactly fit the data. Rather, we aim to qualitatively compare the models to identify general performance and important processes controlling bedrock groundwater age distributions and contributions to the shallow soil.

2.5. Bedrock parameter perturbation analysis

Given the uncertainty in the bedrock hydrogeologic parameters when constraining the limited number of models with groundwater levels alone (McDonnell & Beven, 2014), we perform a second sensitivity test considering only the fractured shale bedrock hydraulic conductivity (K) and porosity (φ). In particular, we start with the top-performing model (*A1*) that best matches the observed groundwater levels at PLM1 and PLM6 and systematically test 5 variants of that model, including: increasing and decreasing the K_{fshale} by one-half order of magnitude (*high K* and *low K*, respectively), an exponentially decreasing K_{fshale} with depth (factor of 10 decrease per 100 m; *Exp Dec K*), and a high and low fractured shale porosity of 5 % and 15 %, respectively (*high* φ and *low* φ, respectively). For saturated groundwater flow, specific yield is directly related to, but less than the porosity in each layer due to non-zero specific storage and retention (Table 1). The range in φ was constrained using borehole geophysical data in Uhlemann et al., (2022). Both ParFlow-CLM and EcoSLIM models are performed for these scenarios that are designed to interrogate differences in groundwater age distributions as a function of bedrock properties, which are not uniquely constrained by groundwater level dynamics alone (Starn et al., 2014; Thiros et al., 2021).

3. Results

3.1. Hydrologic dynamics

Fig. 2A shows the observed and simulated groundwater levels at PLM1 and PLM6 for the base and *A1-A5* models. The *base* model poorly reproduced the observed groundwater levels, underestimating water table depths for the majority of the simulation. There was a general agreement between models *A1-A5* and the observed groundwater levels, with root mean squared errors ranging from 35 to 44 m for *A1* and *A5*, respectively. These models reasonably reproduced the groundwater level peak timing and annual fluctuation magnitudes. However, there remained systematic errors and none of the 128 simulations fully captured the details of the observed groundwater levels at both PLM1 and PLM6 simultaneously (Supplementary Information Figure S2). For instance, the simulated recession from peak groundwater level to base flow was often too fast at PLM1 and slow at PLM6. Model performance

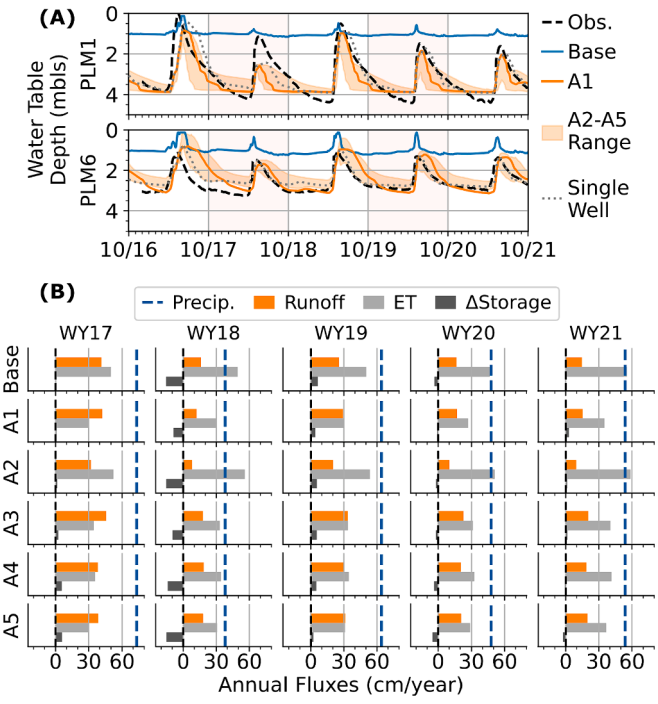


Fig. 2. (A) Observed (dashed-black line) and simulated groundwater levels at PLM1 and PLM6 for the *base* and *A1-A5* models. The shaded region is the minimum and maximum prediction from the *A2-A5* models. The dotted-grey lines are the models with the lowest SSE when considering PLM1 or PLM6 separately. (B) Simulated hydrologic mass balance components from WY2017-WY2021. Precip. (precipitation), runoff, and ET are cumulative fluxes over the water year. The change in storage is from October 1 to September 30 of the respective water year.

increased when considering either PLM1 or PLM6 alone (dotted-grey lines in Fig. 1A). However, the top performing models for each of these single well scenarios poorly predicted the groundwater level response at the other well (see Figure S2 for details).

The simulated annual hydrologic mass-balance components (precipitation, runoff, ET, and change in groundwater storage) are shown in Fig. 2B. These represent the average fluxes and storage changes over the hillslope, normalized to units of length. ET ranged from ~300–600 mm/year across all models and water years. Within a single model, ET had low variability between water years (+/- 50 mm/year). Runoff (which includes direct overland flow and groundwater exfiltration) had higher variability between water years (+/-200 mm/year) and was positively correlated to precipitation magnitude. The lowest-precipitation WY2018 (between 115 and 350 mm less than other years; Figure S1) led to a 100–250 mm decrease in groundwater storage. For all other water years, the groundwater storage change was minor compared to Precip., ET, and runoff. Mass-balance comparisons between models show the *base* and *A2* models had ET estimates of 500–600 mm/year, relative to 300–400 mm/year for the *A1*, *A3*, *A4*, and *A5* models. *A2* systematically had the lowest runoff compared to the other models, which had similar magnitudes. There are no clear patterns in the storage changes between the models. The *A1-A5* models generally have higher K_{soil} and K_{wshale} compared to the *base* model (Table 2 and Figure S3). Among the *A1-A5* models, *A2* has the lowest K_{wshale}, K_{fshale}, α_{soil}, and α_{wshale}. The soil and weathered shale α and n do not show clear clustering and span the full range used in the Monte Carlo sampling (Table 1), suggesting that the groundwater level observations were unable to reduce uncertainty in these parameters and uncertainties in K are more important.

3.2. Groundwater ages and mixing dynamics

The base and *A1-A5* models had considerable differences in flux-weighted groundwater mean ages at the soil, PLM1, and PLM6 wells (Fig. 3A). Mean ages across all models ranged from ~25–300 years for PLM1, ~40–600 years for PLM6, and ~20–800 years at the downgradient soil location. The *A5* model had the youngest mean ages (20–45 years) while the *base* and *A2* models had the oldest (275–800 years). There was low temporal mean age variability at PLM1 and PLM6 compared to the soil location. Peakflow to baseflow mean age variations at the soil location ranged between 20 and 45 years in *A5* and 390 and 720 years in *A2*. The considerable mean age variations within a single water-year highlight the influence of the young snowmelt pulse in late spring and early summer.

The EcoSLIM modeled mean ages at PLM1 and PLM6 encompassed the environmental tracer-based mean ages in Thiros et al., (2023). Fig. 3A illustrates the *A1* and *A3* models generally matched the tracer-based mean ages assuming an exponential piston-flow (EPM) model, but underpredicted the tracer mean age assuming the binary-mixing model (BMM). The *base* and *A2* models better matched the PLM1 BMM tracer mean age, but still underpredicted the PLM6 BMM tracer mean age. The *A4* and *A5* models considerably under-predicted the tracer mean ages and none of the models simulated the ~2000 year PLM6 BMM tracer mean age.

Fig. 3B shows particle ages at the soil location partitioned into discrete age-bins. For instance, the fraction younger than 100 years is the total mass of particles with ages less than 100 years and greater than 10 years (the preceding bin). In EcoSLIM, the total mass of particles is a proxy for advective water mass-flux. The total particle mass dynamics at the soil well followed the general stream hydrograph patterns with peaks in June during snowmelt and baseflow starting in September (Fig. 3B). All models suggest over 90 % of the mass-flux is characterized with particle ages >10 years during baseflow. During snowmelt, the particle mass-flux with ages >10 years dropped to values between 40 % and 60 % and the remainder generally had ages < 0.25 years. The *base* model resulted in very different water ages with minor fractions of young water (<1 year), even during snowmelt.

Differences in the age of the old fractions (100 and 1000 year bins in Fig. 3B), rather than the young fraction dynamics, dictated the disparities in simulated mean age between models. All of the models except *A5* had considerable particle mass-flux with ages >100 years. The *base* and *A2* models additionally contained a small fraction (<10 %) with ages that exceed 1000 years, leading to modeled mean ages that exceed 700 years. The mass-flux fractions characterized with old ages (>100 years) persisted throughout the year, even when the snowmelt and the groundwater levels peak (black line in Fig. 3B). Water flux with ages between 0.25 and 10 years was minimal, suggesting the age partitioning at the soil location is generally a bimodal distribution that is dominated by a young fraction (<0.25 years) during snowmelt then an old fraction (>10 years) for the remainder of the year. To facilitate comparison of the age-partitioning dynamics between the models, Fig. 4A shows the cumulative density function (CDF) of the mass-weighted particle ages, for all particles at the soil well from WY2017-WY2021. The bimodal age distribution for models *A1-A5* is represented as the flat horizontal region between approximately 0.1 and 50 years. The age CDFs show the fraction with groundwater ages >40 years represents between 60 % and 80 % of the total mass-flux over the 5 years.

The total particle mass increase at the soil location during snowmelt is not solely explained by an increase in the young (<0.25 year) fraction (Fig. 3B). The mass-flux of groundwater with older ages also increased during snowmelt, compared to base flow periods. It is also apparent the relative mass of the young fraction is lower in the low-snow WY2018 compared to the water years with larger snowpacks. However, the peak mass fluxes during WY2018 are not commensurately lower relative to the precipitation inputs, suggesting the mass flux at the soil well is buffered by groundwater with older ages during this drought year.

The models had differences in the relative particle mass that infiltrates as snow versus rain at the soil well (solid-black line in Fig. 3B and Fig. 4B). The *A5* model had a relatively stable inter-annual snow fraction that varied between 45 % and 60 %, while the *A2*, *A3*, and *A4* models had larger ranges from ~ 40 % during baseflow up to ~85 % during snowmelt. For all models except the rain-biased base model, the temporally averaged fraction of particles derived from snowmelt is between ~50 % and 60 % (Fig. 4B), which is in disagreement with studies in the East River watershed that suggest rain is primarily sourced to ET (Sprenger et al., 2022; Tokunaga et al., 2022).

3.3. Bedrock sensitivity analysis

Results from the second sensitivity analysis that only considered the bedrock hydrogeologic parameters (Section 2.5) showed the perturbed *A1* models have minor groundwater level differences compared to the *A1* model (Fig. 5A). This suggests the fractured shale bedrock K and porosity perturbations do not considerably influence groundwater level dynamics at the lower third of the hillslope where the observation wells are located. However, the *High K* scenario began to deviate from the *A1* model, suggesting an upper limit for the fractured shale bedrock K where it does impact the simulated groundwater levels.

Similar to the groundwater level response, the mass-balance components showed minor variations between the *A1* model and its variants (Fig. 5B). The annual cumulative runoff and ET inter-model range are

Table 2
Parameter values for the *base* and top 5 models that minimize the SSE considering water level observations form both PLM1 and PLM6 simultaneously.

	$\log_{10} K_{soil}$	$\log_{10} K_{wshale}$	$\log_{10} K_{fshale}$	α_{soil}	α_{wshale}	n_{soil}	n_{wshale}
base	−4.6	−5.4	−7.7	1.8	0.5	1.8	1.6
A1	−3.6	−4.3	−7.1	4.4	4.1	4.8	3.2
A2	−3.4	−5.2	−7.7	0.4	1.5	2.1	3.5
A3	−4.3	−4.3	−7.0	3.6	2.3	3.4	2.8
A4	−3.8	−4.5	−6.9	3.4	3.3	3.1	3.3
A5	−3.8	−4.3	−6.5	2.2	3.6	4.4	1.8

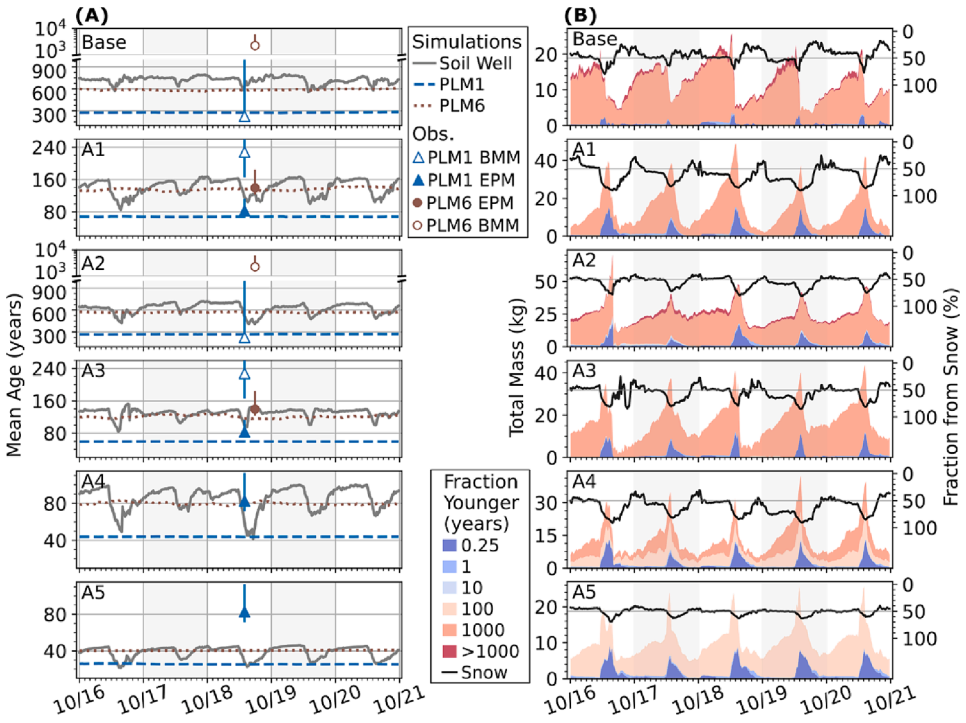


Fig. 3. (A) EcoSLIM modeled mean ages at the soil, PLM1, and PLM6 wells and mean ages (median and 95 percent uncertainty intervals) inferred from environmental tracers observations in Thiros et al. (2023a) assuming an exponential piston-flow (EPM) and binary mixing model (BMM). (B) Total particle masses at the soil location grouped into age fraction bins and the fraction that infiltrates as snowmelt (remainder is rain). Age bins represent water mass less than that bin and greater than the preceding bin.

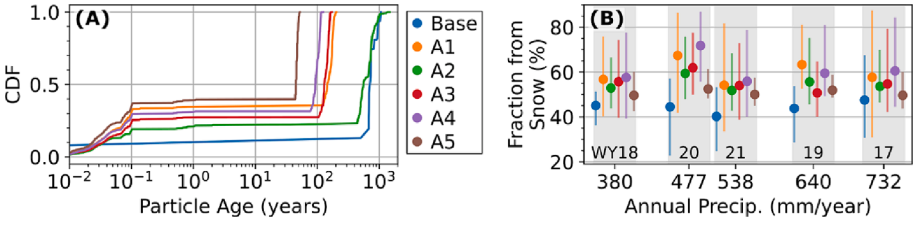


Fig. 4. (A) Cumulative density functions for the modeled age distribution considering all particles in the soil location from WY2017-WY2021. (B) Annual range (vertical bars) and median (dots) of water derived from snowmelt at the soil well for WY2017-WY2021.

~50 and <10 mm/year, respectively. Like the *A1-A5* sensitivity analysis, the groundwater storage change in the *A1* perturbation sensitivity analysis is greatest for WY2018 (~80 mm). For WY2018, the *High K* model has the largest decline in groundwater storage at ~180 mm, compared to the other models with declines of 100 +/-10 mm. This suggests the differences in the *High K* groundwater levels compared to the other models (Fig. 5A) do impact the total change in groundwater storage. In general, the annual hillslope mass-balance components are similar for all of the *A1* perturbation models, showing minor sensitivity to the bedrock *K* and porosity perturbations.

Despite the minor variations in groundwater levels and mass-balance, there were considerable differences in the simulated mean groundwater ages between the *A1* model and its variants (Fig. 6A). Fig. 6A shows the simulated mean age dynamics for the *A1* model and the mean age anomalies between the respective model scenario and *A1* model, for instance, $(High\ K)_{\text{mean age}} - (A1)_{\text{mean age}}$. The *High K* mean ages were up to 120 years younger than *A1* (negative anomaly), while the *Exp Dec K* were ~ 400 years older (positive anomaly). PLM1 had smaller mean age anomaly magnitudes compared to the downslope PLM6 groundwater well. In particular, the PLM1 anomaly ranged from ~ 40 years younger for the *High K* scenario to ~ 150 years older for the *Exp Dec K*. For these same models, the PLM6 mean age anomaly ranged from 90 years younger to ~300 years older. The mean age anomaly dynamics at the shallow soil well had even larger variations from ~80 years younger to ~500 years older across all the models. There is a seasonal structure in the mean age anomalies at the shallow soil well that is not present in the bedrock groundwater wells. The anomaly is generally lowest (closer to zero) during snowmelt conditions when the *A1* model approaches its minimum mean age for the water year. The anomaly then increases during baseflow conditions when the *A1* model is near its oldest mean ages.

4. Discussion

4.1. Model comparison with field observations

In-situ environmental tracer observations from bedrock groundwater in high-elevation mountain systems remain sparse due to the well installation challenges and high analytical costs. Nonetheless, many works have shown benefits in constraining modeled transport behavior with temporally-limited observations of groundwater mean ages inferred from environmental tracers (Carroll et al., 2020; Schilling et al., 2019; Starn et al., 2014; Thiros et al., 2021). In this work, multiple models (*A1*, *A3*, and *A4*) predict groundwater mean ages broadly

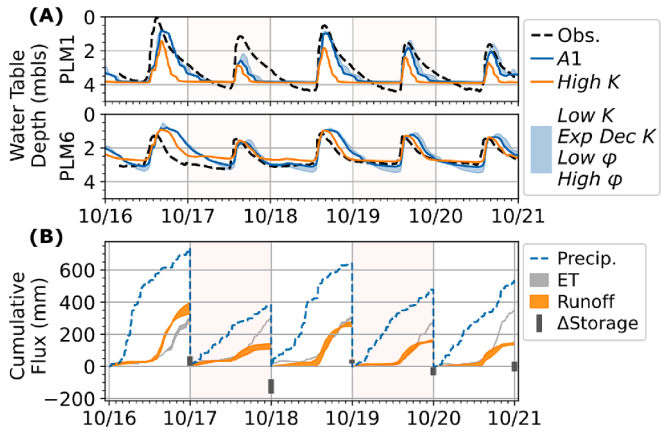


Fig. 5. (A) Simulated water level response for variants of the *A1* model that consider different hydraulic conductivity and porosity in the fractured shale bedrock. The shaded-blue region is the range for the models. (B) Simulated annual cumulative mass-balance components (hillslope average, normalized by units of length). The shaded regions are the range in ET and Runoff for the six models of different hydraulic conductivity and porosity. The red dots correspond to the mean annual storage change and the error bars are the range in predictions for all 6 models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

consistent with the environmental tracer-based mean ages assuming a single component EPM (Thiros et al., 2023a). Both the tracer-based and simulated results correspond to resident mean ages at single points (PLM1 and PLM6 piezometers) within the hillslope aquifer and do not represent the flux-weighted discharge from the entire hillslope (Berghuijs & Kirchner, 2017; McGuire & McDonnell, 2006). Alternatively, the *base* and *A2* simulated older resident mean ages (centuries), which are similar to those estimated using a BMM. Comparing these sets of models (Table 2 and Figure S3) suggests a low K_{fshale} is a plausible mechanism to generate old groundwater mean ages consistent with field observations. This is consistent with recent work that suggests low K_{fshale} is needed to satisfy the observed water mass-balance on the PLM hillslope (Tokunaga et al., 2022). Nonetheless, our results present effective model parameters due to compensation for the inherent and ubiquitous uncertainties in numerical models of complex mountainous systems.

Across all models, the average mean age increases by ~90 % from the upslope (PLM1) to downslope (PLM6) well (a distance of 135 m), which is similar to the ~60 % increase estimated with the environmental tracer observations in Thiros et al., (2023a). Given the lack of subsurface heterogeneity representation (apart from the layer differentiation, see Fig. 1B) and matrix diffusion in the numerical model setups, the mean age increase moving downslope highlights the role of topography-driven flowpaths and multi-layer macrodispersion in creating older groundwater in further downslope positions. While studies have identified groundwater mean age increases while moving down-gradient at larger catchment and watershed-scales (Engdahl & Maxwell, 2015; Frisbee et al., 2012; Gabrielli et al., 2018), few have identified this magnitude of increase over relatively short distances (~100 m) along a single hillslope transect (e.g. Kolbe et al., 2020).

PLM1 and PLM6 had moderate ^{3}H and terrigenic ^{4}He concentrations, qualitatively suggesting mixing between modern (<70 years) and pre-modern (>70 years) groundwater ages (Thiros et al., 2023a). While the EcoSLIM models broadly reproduce the environmental tracer-based mean ages (Fig. 3A), they do not capture the mixing of pre-modern and modern ages. The simulated groundwater age distributions at PLM1 and PLM6 (Figures S5) have variances on the order of decades, which cannot explain the ^{3}H and ^{4}He simultaneously. The exceptions are the *base* and *A2* models, which have age distribution variances that approach 100's of years. However, these models have no pre-modern component and suggest ^{3}H-free water. These comparisons highlight that the model is not capturing all the salient mixing processes, for instance, the impact of matrix diffusion (e.g. Rajaram, 2021), subsurface heterogeneity (e.g. Weissmann et al., 2002), and lateral flowpath connectivity (e.g. Gardner et al., 2015). This also illustrates the non-uniqueness in mean groundwater ages (Larocque et al., 2009) and the importance in comparing the

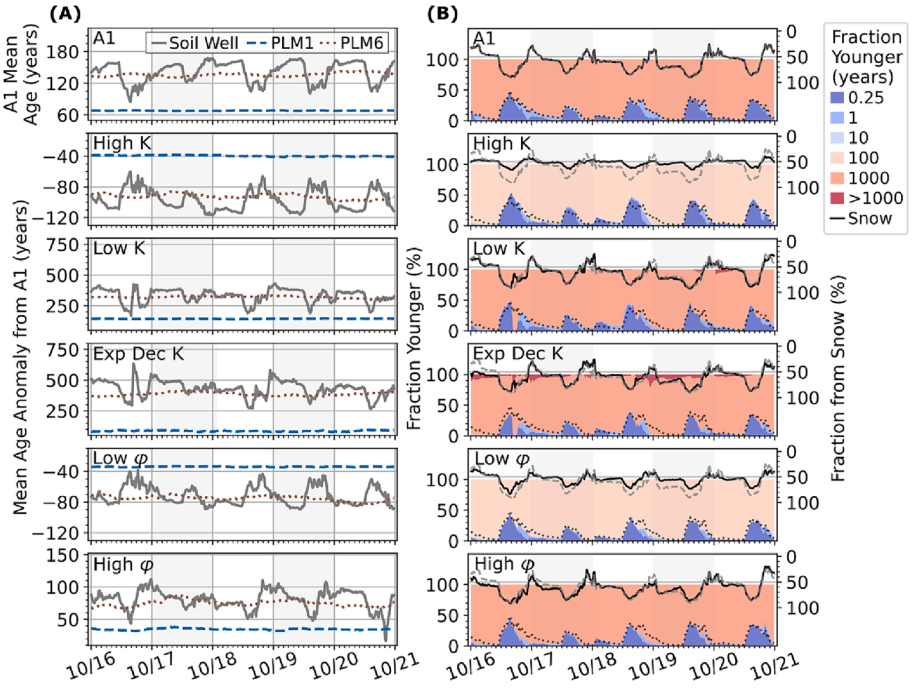


Fig. 6. (A) Simulated mean age dynamics at the shallow soil, PLM1, and PLM6 wells. A1 is plotted in absolute units and the latter 4 models are plotted as the anomaly from A1 as relative units. (B) Relative particle masses at the shallow soil well grouped into age fraction bins and the fraction that infiltrates as snowmelt (remainder is rain). Age bins represent water fractions less than that bin and greater than the preceding bin. The dotted-black and dashed-grey curves in the latter 4 subplots are the A1 Fraction Younger than 1 year and Fraction from Snow, respectively, repeated for easier comparisons.

full age distribution to multiple environmental tracers that span a broad distribution of ages (McCallum et al., 2015; McDonnell et al., 2010). Nonetheless, a limitation of this work is the model comparison to an environmental tracer observation set from a single time period. Constraining the simulated groundwater age temporal dynamics (Figs. 3 and 5) with environmental tracer observations from different hydrograph periods (e.g. baseflow and receding limb) is important future work.

4.2. Bedrock groundwater contributions to the shallow soil

Our results contrast a common assumption that bedrock groundwater fluxes do not alter the mean age within the near-surface hydrologic system. We find that the primary drivers for modeled mean age differences at the downgradient soil location are the age distributions and fluxes of the bedrock flowpaths, which can significantly vary as a function of bedrock properties (Figs. 3 and 6). Neglecting these bedrock groundwater flowpaths would bias age distributions at the shallow soil well to younger water and underestimate the deep groundwater contributions to the hillslope discharge zone. Constructing numerical models that limit the truncation of the old-age groundwater flowpaths becomes especially critical given the age distribution is a mixture of young and old water, a co-occurrence which is supported by both the numerical model and environmental tracer field observations.

Our model predicts considerable bedrock groundwater contributions into the soil throughout the entire year. This occurs even though the groundwater level rises into the conductive weathered shale and soil layers during snowmelt, which promotes lateral flow of young snowmelt water through a transmissivity feedback (Tokunaga et al., 2022). This suggests the discharge zone is continually supported by upwelling of bedrock groundwater (e.g. Gardner et al., 2020) that is characterized with a distribution of old-ages (decades to millenia) that are often not considered in hydrologic and solute transport studies. This model result is important to better understand potential sources and sustainability of groundwater that contribute to ET and runoff (Carroll et al., 2020; Foster & Maxwell, 2019; Ryken et al., 2022). Our results further suggest the hydrologic flux buffering with old-aged groundwater storage is exacerbated during the low-snow WY2018, compared to years with larger snowpacks. Ongoing work to understand this old-groundwater age buffering capacity will be important to predict the implications of a changing climate when multiple low-snow years in the Upper Colorado River Basin are projected to occur sequentially (Siirila-Woodburn et al., 2021).

4.3. Groundwater levels are insensitive to hydrogeologic parameter perturbations

The top performing models (*A1-A5*) from the 128-run ensemble all adequately reproduced the observed groundwater levels (Fig. 2A) yet resulted in different age distributions and mean ages at PLM1 and PLM6 (Fig. 3). Perturbing only the fractured shale hydrogeologic parameters (*K* and porosity) of the *A1* model led to similar results – there was little impact on the simulated water-level and mass-balance (Fig. 5), yet order of magnitude differences in groundwater mean ages at PLM1 and PLM6 (Fig. 6). These results are consistent with previous Richards' Equation sensitivity analyses that show groundwater dynamics have minor sensitivities to porosity variations relative to parameters or forcing conditions that dictate recharge fluxes (Gardner et al., 2020; Maina et al., 2022; Thiros et al., 2022). Further work to identify and constrain uncertain below and above-ground parameters and meteorologic forcing conditions that were not varied in the sensitivity analysis (e.g. plant functional types) is important to improve simulated water table dynamics (e.g. Xu et al., 2023).

The best estimate of the subsurface parameterization from field observations (*base* model) poorly reproduced the groundwater levels, suggesting further sensitivity analysis may be important to constrain hillslope water partitioning, even when in-situ observations are available. Our modeling results suggest that we need to better characterize the subsurface hydrogeologic properties, including the deeper bedrock, to improve predictions of the hydrologic and mass partitioning within the shallow soil systems near the discharge areas of steep mountain hillslopes. Our results further support studies that show constraining groundwater mass transport with groundwater levels alone leads to non-unique subsurface parameter estimates (e.g. Thiros et al., 2022) and process understanding (e.g. McDonnell & Beven, 2014).

Despite the models generally matching the groundwater levels at the PLM1 and PLM6 wells, they have large differences in groundwater table depth dynamics in the mid- and up-slope positions (Fig. 7). For instance, the *base* and *A2* models predict shallow groundwater depths over the entire hillslope that have minor annual variations compared to the *A5* and *High K* models. The differences in seasonal water table depths help explain the near-surface water content, leading to variable ET fluxes between the models (Fig. 2). Fig. 7 also illustrates the convergence of groundwater levels towards land-surface near the base of the hillslope. This is largely driven by the assumed river stage boundary condition, which influences the annual runoff similarities between models (Fig. 2). This analysis highlights limitations in constraining hillslope groundwater level dynamics with observations that are all located in the lower elevation positions. Given this non-uniqueness, our results suggest further constraining hydrologic models in mountains with solute transport field measurements, such as environmental age tracers, can be important to improve process-understanding and model predictions (e.g. Schilling et al., 2019).

4.4. Alternate models and uncertainty

Within the full 128 model ensemble (Section 2.5), there are models that match the groundwater levels at PLM1 or PLM6 with high-fidelities, but not both wells simultaneously (Fig. 2 and Figure S2). Thus, models *A1-A5* represent a compromise to match the observed groundwater level response at both wells. The inability to accurately reproduce the observations at both wells simultaneously suggests structural errors within our model, such as simplifications in the subsurface structure, lack of hydrogeologic parameter heterogeneity (Harman et al., 2009; Matonse & Kroll, 2013), forcing data and boundary condition biases (e.g. Xu et al., 2023), and broader conceptual model epistemic uncertainties that are challenging to quantify (see Wagener & Gupta, 2005). A likely important conceptual model inadequacy is our simplification of the system to a 2-D transect. A 2-D model reduces the ability for heterogeneous and lateral flowpath connectivity, which can be important for catchment hydrologic response (Bachmair & Weiler, 2012). For instance, reduced lateral connectivity in 2-D models can reduce the effective upslope accumulation areas and be one plausible mechanism causing the under-prediction of groundwater levels at PLM1 (Jencso et al., 2010), in particular during the low-snow WY2018. Furthermore, previous work has shown lower flowpath connectivity in 2-D models results in older groundwater mean ages compared to 3-D models (Gardner et al., 2015). Further work is needed to understand potential limitations and changes in groundwater age dynamics when simplifying the hillslope to a 2-D model.

The presented models do not represent the influence of bedrock fractures. Borehole characterization and geophysical surveys at the PLM hillslope have identified multiple large fractures (Miltenberger et al., 2021; Uhlemann et al., 2022), which are not discreetly included within the current model setup. Groundwater flow through fractured bedrock can influence the hydraulic response in wells (Gabrielli et al., 2012; Gardner et al., 2020), and potentially be responsible for the misrepresented groundwater level recession dynamics in the model (Figs. 2 and 4). Discrete fractures and surrounding matrix diffusion processes can also impact groundwater age distributions (Bethke & Johnson, 2008; Rajaram, 2021). The simulated groundwater ages from EcoSLIM better represent the advective age distribution, thus do not explicitly include the effects of matrix diffusion nor the full hydrodynamic

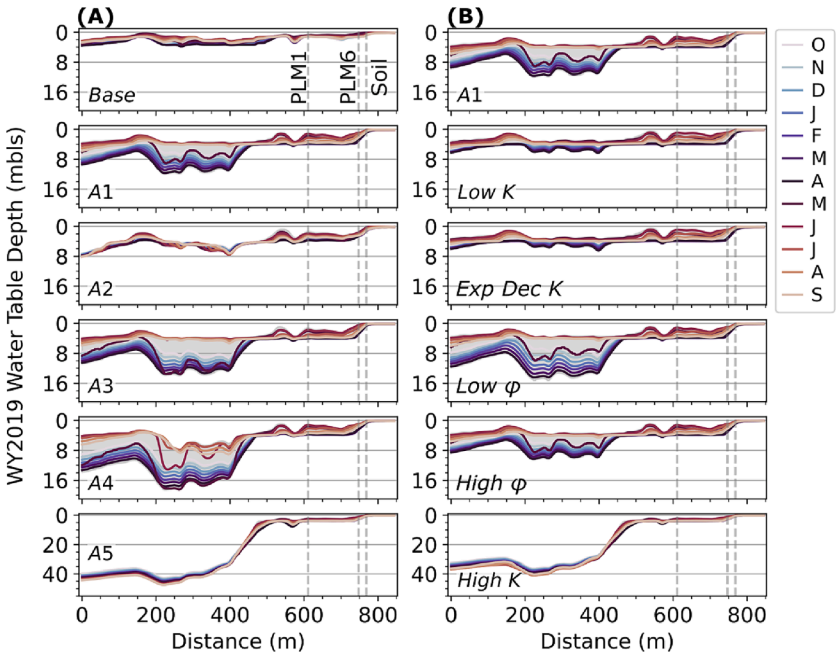


Fig. 7. Spatial and temporal plots of groundwater table depths for water year 2019. (A) *Base* model and top 5 models from the Monte Carlo sampling parameter analysis. Note the different y-axis range on *A5*. (B) Models with variations in bedrock hydrogeologic properties compared to the *A1* model. Note the different y-axis for the *A5* and High K scenarios. On both figures, the dashed vertical lines indicate the locations of PLM1 and PLM6 wells and the soil location (from left to right, respectively).

dispersion tensor. The lack of matrix diffusion is a likely cause for the simulated age distributions to be qualitatively too narrow and biased young compared to the environmental tracer field interpretations.

5. Conclusions

Water age distributions inferred from field observations remain uncertain due to the limited age-dating range of environmental tracer systems (McCallum et al., 2015) and the often sparse spatial and temporal distribution of samples. Numerical models that simulate the full groundwater age distributions can provide a valuable tool to understand spatially and temporally evolving flowpath mixing processes that are not captured in discrete field observations. We present a series of physically-based hydrologic and particle tracking simulations to investigate the contribution of bedrock groundwater to hillslope hydrodynamics, including soil water age and source (rain/snow). We find the bedrock groundwater system is important for water age partitioning at the toe of a steep mountain hillslope. For all tested models, bedrock groundwater flowpaths characterized with old ages (10′s to 1000′s of years) make-up a considerable proportion (40–90 %) of the flux within the soil near the toe of the hillslope. The contribution of old-aged groundwater is largest for the low-snow year, highlighting a key and understudied process that can buffer fluxes during drought. The sensitivity analysis highlights process equifinality when constraining hillslope mass transport with groundwater level observations alone. The result that bedrock groundwater dictates the hillslope water age partitioning builds on calls to better assimilate deeper groundwater into watershed conceptual models (Brooks et al., 2015; Condon et al., 2020; Singha & Navarre-Sitchler, 2022; Somers & McKenzie, 2020). Our analysis of a snow-dominated, soil-mantled hillslope characteristic of the Rocky Mountain region emphasizes that improved characterization of the bedrock groundwater system processes that span broad timescales can be critical to accurately forecast the implications of climate change on water resources and quality.

CRediT authorship contribution statement

Nicholas E. Thiros: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Erica R. Siirila-Woodburn:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Matthias Sprenger:** Writing – review & editing, Validation. **Kenneth H. Williams:** Supervision, Project administration, Funding acquisition. **James P. Dennedy-Frank:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis. **Rosemary W. H. Carroll:** Writing – review & editing, Validation, Conceptualization. **Gardner W.P.:** Writing – review & editing, Validation, Supervision, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The ParFlow-CLM version 3.9 (https://github.com/parflow/parflow) and EcoSLIM version 1.3 (https://github.com/reedmaxwell/EcoSLIM) softwares are open-source and publicly available. All model input files, sensitivity analysis, and post-processing Python scripts are available in the Zenodo repository (Thiros et al., 2023b). Data and model are freely available and meet FAIR principles. Groundwater level field data are in the ESS-DIVE repository (Faybishenko et al., 2023) and environmental tracer observations are presented in (Thiros et al., 2023a).

Acknowledgments

Funding from the US Department of Energy, Office of Science as part of the Lawrence Berkeley National Laboratory Watershed Function Science Focus Area (contract DE-AC02-05CH11231).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.131193.

References

Ameli, A.A., Amvrosiadi, N., Grabs, T., Laudon, H., Creed, I.F., McDonnell, J.J., Bishop, K., 2016. Hillslope permeability architecture controls on subsurface transit time distribution and flow paths. J. Hydrol. 543, 17–30. https://doi.org/10.1016/j.jhydrol.2016.04.071.

Bachmair, S., Weiler, M., 2012. Hillslope characteristics as controls of subsurface flow variability. Hydrol. Earth Syst. Sci. 16 (10), 3699–3715. https://doi.org/10.5194/hess-16-3699-2012.

Bethke, C.M., Johnson, T.M., 2008. Groundwater age and groundwater age dating. Annu. Rev. Earth Planet. Sci. 36 (1), 121–152. https://doi.org/10.1146/annurev.earth.36.031207.124210.

Brooks, P.D., Chorover, J., Fan, Y., Godsey, S.E., Maxwell, R.M., McNamara, J.P., Tague, C., 2015. Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics. Water Resour. Res. 51, 6973–6987. https://doi.org/10.1002/2015WR017039. Received.

Carroll, R.W.H., Bearup, L.A., Brown, W., Dong, W., Bill, M., Willlams, K.H., 2018. Factors controlling seasonal groundwater and solute flux from snow-dominated basins. Hydrol. Process. 32 (14), 2187–2202. https://doi.org/10.1002/hyp.13151.

Carroll, R.W.H., Manning, A.H., Niswonger, R., Marchetti, D., Williams, K.H., 2020. Baseflow age distributions and depth of active groundwater flow in a snow-dominated mountain headwater basin. Water Resour. Res. https://doi.org/10.1029/2020WR028161.

Condon, L.E., Markovich, K.H., Kelleher, C.A., McDonnell, J.J., Ferguson, G., McIntosh, J.C., 2020. Where is the bottom of a watershed? Water Resour. Res. 56 (3) https://doi.org/10.1029/2019WR026010.

Cook, P.G., Herczeg, A.L., 2000. Environmental Tracers in Subsurface Hydrology. Springer Science & Business Media.

Engdahl, N.B., Maxwell, R.M., 2015. Quantifying changes in age distributions and the hydrologic balance of a high-mountain watershed from climate induced variations in recharge. J. Hydrol. 522, 152–162. https://doi.org/10.1016/j.jhydrol.2014.12.032.

Faybishenko, B., Versteeg, R., Williams, K., Carroll, R., Dong, W., Tokunaga, T., O'Ryan, D. (2023). QA/QC-ed Groundwater Level Time Series in PLM-1 and PLM-6 Monitoring Wells, East River, Colorado (2016-2022) [dataset]. ESS-DIVE. https://data.ess-dive.lbl.gov/view/doi:10.15485/1866836.

Foster, L.M., Maxwell, R.M., 2019. Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions. Hydrol. Process. 33 (3), 332–349. https://doi.org/10.1002/hyp.13327.

Frisbee, M.D., Phillips, F.M., Weissmann, G.S., Brooks, P.D., Wilson, J.L., Campbell, A.R., Liu, F., 2012. Unraveling the mysteries of the large watershed black box: Implications for the streamflow response to climate and landscape perturbations. Geophys. Res. Lett. 39 (1), 1–6. https://doi.org/10.1029/2011GL050416.

Frisbee, M.D., Wilson, J.L., Gomez-Velez, J.D., Phillips, F.M., Campbell, A.R., 2013. Are we missing the tail (and the tale) of residence time distributions in watersheds? Geophys. Res. Lett. 40 (17), 4633–4637. https://doi.org/10.1002/grl.50895.

Gabrielli, C.P., McDonnell, J.J., Jarvis, W.T., 2012. The role of bedrock groundwater in rainfall-runoff response at hillslope and catchment scales. J. Hydrol. 450–451, 117–133. https://doi.org/10.1016/j.jhydrol.2012.05.023.

Gabrielli, C.P., Morgenstern, U., Stewart, M.K., McDonnell, J.J., 2018. Contrasting groundwater and streamflow ages at the Maimai watershed. Water Resour. Res. 54 (6), 3937–3957. https://doi.org/10.1029/2017WR021825.

Gardner, W.P., Hammond, G., Lichtner, P., 2015. High performance simulation of environmental tracers in heterogeneous domains. Groundwater 53 (S1), 71–80. https://doi.org/10.1111/gwat.12148.

Gardner, W.P., Jensco, K., Hoylman, H.Z., Livesay, R., Maneta, P.M., 2020. A numerical investigation of bedrock groundwater recharge and exfiltration on soil mantled hillslopes. Hydrol. Process. 34 (15), 3311–3330. https://doi.org/10.1002/hyp.13799.

Hale, C.V., McDonnell, J.J., Stewart, M.K., Solomon, D.K., Doolitte, J., Ice, G.G., Pack, R. T., 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. https://doi.org/10.1002/2015WR017660.

Harman, C.J., Sivapalan, M., Kumar, P., 2009. Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. Water Resour. Res. 45 (9) https://doi.org/10.1029/2008WR007392, 2008WR007392.

Harmon, R., Barnard, H.R., Singha, K., 2020. Water table depth and bedrock permeability control magnitude and timing of transpiration-induced diel fluctuations in groundwater. Water Resour. Res. 56 (5), 1–22. https://doi.org/10.1029/2019WR025967.

Hayashi, M., 2020. Alpine hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. Groundwater 58 (4), 498–510. https://doi.org/10.1111/gwat.12965.

Hubbard, S.S., Williams, K.H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., Brodie, E., Carroll, R., Dafflon, B., Dwivedi, D., Falco, N., Faybishenko, B., Maxwell, R., Nico, P., Steefel, C., Steltzer, H., Tokunaga, T., Tran, P.A., Wainwright, H., Varadharajan, C., 2018. The East River, Colorado, watershed: A mountainous community testbed for improving predictive understanding of multiscale hydrological-biogeochemical dynamics. Vadose Zone J. 17 (1), 180061 https://doi.org/10.2136/vzj2018.03.0061.

Immerzeel, W.W., Lutz, A.F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B.J., Elmore, A.C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J.S., Koppes, M., Kraaijenbrink, P.D.A., Kulkarni, A.V., Mayewski, P.A., Baillie, J.E.M., 2020. Importance and vulnerability of the world's water towers. Nature 577 (7790), 364–369. https://doi.org/10.1038/s41586-019-1822-y.

Jencso, K.G., McGlynn, B.L., Gooseff, M.N., Bencala, K.E., 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 (10) https://doi.org/10.1029/2009WR008818.

Kolbe, T., Marçais, J., De Dreuzy, J., Labasque, T., Bishop, K., 2020. Lagged rejuvenation of groundwater indicates internal flow structures and hydrological connectivity. Hydrol. Process. 34 (10), 2176–2189. https://doi.org/10.1002/hyp.13753.

Kollet, S.J., Maxwell, R.M., 2008. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. Water Resour. Res. 44 (2), 1–18. https://doi.org/10.1029/2007WR006004.

Larocque, M., Cook, P.G., Haaken, K., Simmons, C.T., 2009. Estimating flow using tracers and hydraulics in synthetic heterogeneous aquifers. Ground Water 47 (6), 786–796. https://doi.org/10.1111/j.1745-6584.2009.00595.x.

Maina, F.Z., Siirila-Woodburn, E.R., Dennedy-Frank, P.J., 2022. Assessing the impacts of hydrodynamic parameter uncertainties on simulated evapotranspiration in a mountainous watershed. J. Hydrol. 608 (February) https://doi.org/10.1016/j.jhydrol.2022.127620.

Manning, A.H., Ball, L.B., Wanty, R.B., Williams, K.H., 2021. Direct observation of the depth of active groundwater circulation in an alpine watershed. Water Resour. Res. 57 (2) https://doi.org/10.1029/2020wr028548.

Manning, A.H., Caine, J.S., 2007. Groundwater noble gas, age, and temperature signatures in an Alpine watershed: Valuable tools in conceptual model development. Water Resour. Res. 43 (4) https://doi.org/10.1029/2006WR005349.

Manning, A.H., Clark, J.F., Diaz, S.H., Rademacher, L.K., Earman, S., Niel Plummer, L., 2012. Evolution of groundwater age in a mountain watershed over a period of thirteen years. J. Hydrol. 460–461, 13–28. https://doi.org/10.1016/j.jhydrol.2012.06.030.

Matonse, A.H., Kroll, C.N., 2013. Applying hillslope-storage models to improve low flow estimates with limited streamflow data at a watershed scale. J. Hydrol. 494, 20–31. https://doi.org/10.1016/j.jhydrol.2013.04.032.

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

Maxwell, R.M., Condon, L.E., Danesh-Yazdi, M., Bearup, L.A., 2019. Exploring source water mixing and transient residence time distributions of outflow and evapotranspiration with an integrated hydrologic model and Lagrangian particle tracking approach. Ecohydrology 12 (1), 1–10. https://doi.org/10.1002/eco.2042.

McCallum, J.L., Cook, P.G., Simmons, C.T., 2015. Limitations of the use of environmental tracers to infer groundwater age. Groundwater 53 (S1), 56–70. https://doi.org/10.1111/gwat.12237.

McDonnell, J.J., 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 (6), 5342–5350. https://doi.org/10.1002/2013WR015141.

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., Lyon, S., Maloszewski, P., Newman, B., Pfister, L., Rinaldo, A., Rodhe, A., Sayama, T., Seibert, J., Solomon, K., Wrede, S., 2010. How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. Hydrol. Process. 24 (12), 1745–1754. https://doi.org/10.1002/hyp.7796.

Meixner, T., Manning, A.H., Stonestrom, D.A., Allen, D.M., Ajami, H., Blasch, K.W., Brookfield, A.E., Castro, C.L., Clark, J.F., Gochis, D.J., Flint, A.L., Neff, K.L., Niraula, R., Rodell, M., Scanlon, B.R., Singha, K., Walvoord, M.A., 2016. Implications of projected climate change for groundwater recharge in the western United States. J. Hydrol. 534, 124–138. https://doi.org/10.1016/j.jhydrol.2015.12.027.

Meyers, Z.P., Frisbee, M.D., Rademacher, L.K., Stewart-Maddox, N.S., 2021. Old groundwater buffers the effects of a major drought in groundwater-dependent ecosystems of the eastern Sierra Nevada (CA). Environ. Res. Lett. 16 (4) https://doi.org/10.1088/1748-9326/abde5f.

Miltenberger, A., Uhlemann, S., Mukerji, T., Williams, K., Dafflon, B., Wang, L., Wainwright, H., 2021. Probabilistic evaluation of geoscientific hypotheses with geophysical data: application to electrical resistivity imaging of a fractured bedrock zone. J. Geophys. Res. Solid Earth 126 (9), 1–20. https://doi.org/10.1029/2021JB021767.

Mote, P.W., Li, S., Lettenmaier, D.P., Xiao, M., Engel, R., 2018. Dramatic declines in snowpack in the western US. NPJ Clim. Atmos. Sci. 1 (1), 2. https://doi.org/10.1038/s41612-018-0012-1.

Rajaram, H., 2021. Matrix diffusion as a mechanism contributing to fractal stream chemistry and long-tailed transit time distributions. Geophys. Res. Lett. 48 (18), 1–12. https://doi.org/10.1029/2021GL094292.

Rapp, G.A., Condon, L.E., Markovich, K.H., 2020. Sensitivity of simulated mountain block hydrology to subsurface conceptualization. Water Resour. Res. 56 (10) https://doi.org/10.1029/2020WR027714.

Rogers, D.B., Newcomer, M.E., Raberg, J.H., Dwivedi, D., Steefel, C., Bouskill, N., Nico, P., Faybishenko, B., Fox, P., Conrad, M., Bill, M., Brodie, E., Arora, B., Dafflon, B., Williams, K.H., Hubbard, S.S., 2021. Modeling the impact of riparian

hollows on river corridor nitrogen exports. Frontiers in Water 3, 590314. https://doi.org/10.3389/frwa.2021.590314.
Ryken, A.C., Gochis, D., Maxwell, R.M., 2022. Unravelling groundwater contributions to evapotranspiration and constraining water fluxes in a high-elevation catchment. Hydrol. Process. 36 (1), 1–14. https://doi.org/10.1002/hyp.14449.
Schilling, O.S., Cook, P.G., Brunner, P., 2019. Beyond classical observations in hydrogeology: The advantages of including exchange flux, temperature, tracer concentration, residence time, and soil moisture observations in groundwater model calibration. Rev. Geophys. 57, 146–182. https://doi.org/10.1029/2018RG000619.
Siirila-Woodburn, E.R., Rhoades, A.M., Hatchett, B.J., Huning, L.S., Szinai, J., Tague, C., Nico, P.S., Feldman, D.R., Jones, A.D., Collins, W.D., Kaatz, L., 2021. A low-to-no snow future and its impacts on water resources in the western United States. Nat. Rev. Earth Environ. 2 (11), 800–819. https://doi.org/10.1038/s43017-021-00219-y.
Singha, K., Navarre-Sitchler, A., 2022. The importance of groundwater in critical zone science. Groundwater 60 (1), 27–34. https://doi.org/10.1111/gwat.13143.
Sobol, I.M., 1998. On quasi-Monte Carlo integrations. Math. Comput. Simul 47 (2–5), 103–112. https://doi.org/10.1016/s0378-4754(98)00096-2.
Somers, L.D., McKenzie, J.M., 2020. A review of groundwater in high mountain environments. Wiley Interdiscip. Rev. Water 7 (6). https://doi.org/10.1002/wat2.1475.
Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S.T., Benettin, P., Dubbert, M., Hartmann, A., Hrachowitz, M., Kirchner, J.W., McDonnell, J.J., Orlowski, N., Penna, D., Pfahl, S., Rinderer, M., Rodriguez, N., Schmidt, M., Werner, C., 2019. The demographics of water: A review of water ages in the critical zone. Rev. Geophys. 57 (3), 800–834. https://doi.org/10.1029/2018RG000633.
Sprenger, M., Carroll, R.W.H., Siirila-Woodburn, E.R., Newcomer, M.E., 2022. Variability of snow and rainfall partitioning into evapotranspiration and summer runoff across nine mountainous catchments. Geophys. Res. Lett. 49 (May) https://doi.org/10.1029/2022GL099324.
Starn, J.J., Green, C.T., Hinkle, S.R., Bagtzoglou, A.C., Stolp, B.J., 2014. Simulating water-quality trends in public-supply wells in transient flow systems. Ground Water 52, 53–62. https://doi.org/10.1111/gwat.12230.
Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., Macdonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Treidel, H., 2013. Ground water and climate change. Nat. Clim. Chang. 3 (4), 322–329. https://doi.org/10.1038/nclimate1744.
Thiros, N.E., Gardner, W.P., Kuhlman, K.L., 2021. Utilizing environmental tracers to reduce groundwater flow and transport model parameter uncertainties. Water Resour. Res. 57 (7) https://doi.org/10.1029/2020WR028235.
Thiros, N.E., Gardner, W.P., Maneta, M.P., Brinkerhoff, D.J., 2022. Quantifying subsurface parameter and transport uncertainty using surrogate modelling and environmental tracers. Hydrol. Process. 36 (11), e14743.
Thiros, N.E., Siirila-Woodburn, E.R., Dennedy-Frank, P.J., Williams, K.H., Gardner, W.P., 2023a. Constraining bedrock groundwater residence times in a mountain system with environmental tracer observations and Bayesian uncertainty quantification. Water Resour. Res. 59 (2) https://doi.org/10.1029/2022WR033282.
Thiros, N.E., Sprenger, M., Williams, K.H., Dennedy-Frank, J.P., Carroll, R.W.H., Gardner, W.P., Siirila-Woodburn, E.R., 2023b. Old-aged groundwater contributes to mountain hillslope hydrologic dynamics. Zenodo. https://doi.org/10.5281/zenodo.8305993.
Thornton, J.M., Therrien, R., Mariéthoz, G., Linde, N., Brunner, P., 2022. Simulating fully-integrated hydrological dynamics in complex Alpine headwaters: Potential and challenges. Water Resour. Res. https://doi.org/10.1029/2020wr029390.
Tokunaga, T.K., Wan, J., Williams, K.H., Brown, W., Henderson, A., Kim, Y., Tran, A.P., Conrad, M.E., Bill, M., Carroll, R.W.H., Dong, W., Xu, Z., Lavy, A., Gilbert, B., Romero, S., Christensen, J.N., Faybishenko, B., Arora, B., Siirila-Woodburn, E.R., Hubbard, S.S., 2019. Depth- and time-resolved distributions of snowmelt-driven hillslope subsurface flow and transport and their contributions to surface waters. Water Resour. Res. 55 (11), 9474–9499. https://doi.org/10.1029/2019WR025093.
Tokunaga, T.K., Tran, A.P., Wan, J., Dong, W., Newman, A.W., Beutler, C.A., Brown, W., Henderson, A.N., Williams, K.H., 2022. Quantifying subsurface flow and solute transport in a snowmelt-recharged hillslope with multiyear water balance. Water Resour. Res. 58 (12) https://doi.org/10.1029/2022WR032902.
Uhlemann, S., Dafflon, B., Wainwright, H.M., Williams, K.H., Minsley, B., Zamudio, K., Carr, B., Falco, N., Ulrich, C., Hubbard, S., 2022. Surface parameters and bedrock properties covary across a mountainous watershed: Insights from machine learning and geophysics. Sci. Adv. 8 (March), 1–16. https://doi.org/10.1126/sciadv.abj2479.
Viviroli, D., Dürr, H.H., Messerli, B., Meybeck, M., Weingartner, R., 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water Resour. Res. 43 (7), 1–13. https://doi.org/10.1029/2006WR005653.
Wagener, T., Gupta, H.V., 2005. Model identification for hydrological forecasting under uncertainty. Stoch. Env. Res. Risk A. 19 (6), 378–387. https://doi.org/10.1007/s00477-005-0006-5.
Wan, J., Tokunaga, T.K., Brown, W., Newman, A.W., Dong, W., Bill, M., Beutler, C.A., Henderson, A.N., Harvey-Costello, N., Conrad, M.E., Bouskill, N.J., Hubbard, S.S., Williams, K.H., 2021. Bedrock weathering contributes to subsurface reactive nitrogen and nitrous oxide emissions. Nat. Geosci. 14 (4), 217–224. https://doi.org/10.1038/s41561-021-00717-0.
White, A., Moravec, B., McIntosh, J., Olshansky, Y., Paras, B., Sanchez, R.A., Ferré, T.P. A., Meixner, T., Chorover, J., 2019. Distinct stores and the routing of water in the deep critical zone of a snow-dominated volcanic catchment. Hydrol. Earth Syst. Sci. 23 (11), 4661–4683. https://doi.org/10.5194/hess-23-4661-2019.
Williams, K.H., Newman, A. (2020). Shale drilling field notes, lithologic descriptions, and core photographs of wells PLM5, PLM8, GUM1, and GLS1 at the East River Watershed, Colorado [dataset]. Watershed Function SFA, ESS-DIVE repository. doi: 10.15485/1701756.
Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., & Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products: WATER AND ENERGY FLUX ANALYSIS. Journal of Geophysical Research: Atmospheres, 117(D3), n/a-n/a. 10.1029/2011JD016048.
Xu, Z., Siirila-Woodburn, E.R., Rhoades, A.M., Feldman, D., 2023. Sensitivities of subgrid-scale physics schemes, meteorological forcing, and topographic radiation in atmosphere-through-bedrock integrated process models: A case study in the Upper Colorado River basin. Hydrol. Earth Syst. Sci. 27 (9), 1771–1789. https://doi.org/10.5194/hess-27-1771-2023.