

How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds

Mohammad Safeeq^{1,2}  | Ryan R. Bart³  | Norman F. Pelak³ 
 Chandan K. Singh³  | David N. Dralle⁴  | Peter Hartsough⁵ 
 Joseph W. Wagenbrenner⁶ 

¹Division of Agriculture and Natural Resources, University of California, Davis, California

²Civil and Environmental Engineering, University of California, Merced, California

³Sierra Nevada Research Institute, University of California, Merced, California

⁴USDA Forest Service, Pacific Southwest Research Station, Davis, California

⁵Department of Land, Air and Water Resources, University of California, Davis, California

⁶USDA Forest Service, Pacific Southwest Research Station, Arcata, California

Correspondence

Mohammad Safeeq, Civil and Environmental Engineering, University of California, Merced, CA 95343.

Email: msafeeq@ucmerced.edu

Funding information

Division of Earth Sciences, National Science Foundation, Grant/Award Numbers: EAR-0619947, EAR-0725097, EAR-1239521, EAR-1331939; National Institute of Food and Agriculture, Grant/Award Number: 2018-67004-27405

Abstract

The water balance is an essential tool for hydrologic studies and quantifying water-balance components is the focus of many research catchments. A fundamental question remains regarding the appropriateness of water-balance closure assumptions when not all components are available. In this study, we leverage in-situ measurements of water fluxes and storage from the Southern Sierra Critical Zone Observatory (SSCZO) and the Kings River Experimental Watersheds (KREW) to investigate annual water-balance closure errors across large ($1016\text{--}5389\text{ km}^2$) river basins and small ($0.5\text{--}5\text{ km}^2$) headwater-catchment scales in the southern Sierra Nevada. The results showed that while long-term water balance in river basins can be closed within 10% of precipitation, in the smaller headwater catchments as much as a quarter of precipitation remained unaccounted for. A detailed diagnosis of this water-balance closure error using distributed soil moisture measurements in the top 1 m suggests an unaccounted deeper storage and a net groundwater export from the headwater catchments. This imbalance was also found to be very sensitive to the timescales over which water-balance closures were attempted. While some of the closure errors in the simple water balance can be attributed to measurement uncertainties, we argue for a broader consideration of groundwater exchange when evaluating hydrological processes at headwater scales, as the assumption of negligible net groundwater exchange may lead to an overestimation of fluxes derived from the water balance method.

1 | INTRODUCTION

Many hydrologic problems, such as understanding effects of wildfire (Roche et al., 2020), climate change (Gleick, 1987), and drought-related tree mortality (Bales et al., 2018; Goulden & Bales, 2019), involve quantifying or predicting changes in hydrologic stores and fluxes. However, since there is generally incomplete knowledge of one or more of the variables of interest at the scale of analysis, hydrologists frequently use a water mass balance to back-calculate

less well-defined variables from more well-defined ones. This process, which often imposes assumptions and biases in the application of the water balance, can increase uncertainties in answering the hydrologic problem.

A standard catchment water balance with no human withdrawal or input of water can be described as:

$$P - ET - Q - G_n - \frac{dS}{dt} = 0, \quad (1)$$

where P is precipitation, ET is evapotranspiration, Q is runoff, G_n is net subsurface or groundwater exchange relative to the catchment with positive values indicating net output and negative values suggesting input originating outside of the drainage area, and dS/dt is the change in the terrestrial water storage S over time t with positive values suggesting an increase in storage and negative values indicating loss. There are two primary challenges with the application of this water balance. First, multiple components of the water balance are often not readily measured. Second, forcing water-balance closure can be problematic even when measurements of all the water-balance components are available due to biases in the measurements.

Of the water-balance components, measurements of G_n and dS/dt are infrequent and sparse. Thus, not surprisingly, these same components are often approximated to be negligible quantities. The assumption that $dS/dt = 0$ at annual and smaller timescales is known to be problematic (Istanbulluoglu et al., 2012; Rice & Emanuel, 2019; Wang et al., 2015; Wang, Huang, et al., 2014). Similarly, net groundwater exchange is common and occurs via a number of mechanisms (Markovich et al., 2019; Meixner et al., 2016; Sophocleous, 2002; Tonina & Buffington, 2009). Headwater catchments, with relatively high topographic positions, are more likely to lose water to regional aquifers while lower elevation catchments are more likely to gain additional water (Winter et al., 1998). Catchments that are small relative to a regional aquifer are also more likely to experience net groundwater exchange (Fan, 2019). In the absence of a regional aquifer, losses may also occur due to bypass flow around or beneath a stream gauge (Boano et al., 2014; Payn et al., 2009), especially in alluvial streams.

Measurements of water-balance components are often biased due to a variety of reasons, including instrument accuracy, varying environmental conditions, and uneven distribution of gauges (Kampf et al., 2020; Wang, Huang, et al., 2014; Wang, McKenney, et al., 2014). As a consequence, achieving perfect water-balance closure, where inputs equal outputs, seems unlikely unless the biases coincidentally cancel each other out. Kampf et al. (2020) highlighted multiple issues and challenges with closing the water balance. In a closed water balance, errors in measured fluxes and assumptions in unmeasured fluxes propagate uncertainty (Kampf et al., 2020). In other words, assuming water-balance closure and estimating one of the unknown fluxes (often G_n or dS/dt) as a residual will contain uncertainties, as the biases associated with the measured variables will be reallocated to the inferred variables (Fekete et al., 2004; Wang, Huang, et al., 2014; Wang, McKenney, et al., 2014). Kampf et al. (2020) instead advocate for an open water balance, in which the known fluxes are compared to elucidate information about the characteristics of the watershed, but in which unknown water-balance fluxes are not imposed by a strict relationship with the rest of the water balance. Yet practical considerations, such as a lack of information on difficult-to-measure fluxes or an inability to quantify the uncertainty of measured fluxes, will necessitate the continued use of water balance equations to infer unknown terms, making it

absolutely critical to evaluate underlying water-balance-closure assumptions.

Independently quantifying each term in the water-balance equation along with associated uncertainty has proven to be, and will likely continue to be, extremely challenging (Flerchinger & Cooley, 2000; Mazur et al., 2011; Pan et al., 2017; Scott & Biederman, 2019). However, research catchments and observatories, where multiple components of the water balance are often measured, provide an opportunity to test the assumptions that underpin the water balance. The Southern Sierra Critical Zone Observatory (SSCZO) and the Kings River Experimental Watersheds (KREW) are co-located in the southern Sierra Nevada of California, a region that is a primary water source for agriculture, urban areas, and the environment in the San Joaquin Valley. Runoff from the southern Sierra Nevada is overallocated for downstream uses, thus making estimates of water-balance components, such as mountain-block recharge, very important (Hanak et al., 2017). In this paper we examine how realistic water balance assumptions are in the southern Sierra Nevada, with the aim of determining the scale(s) over which the water-balance closure assumptions are valid. Specifically, our objective was to assess water-balance closure at two scales, the river-basin scale ($1016\text{--}5389\text{ km}^2$) and the headwater-catchment scale ($0.5\text{--}5\text{ km}^2$). We took advantage of the multiple gauged catchments in the KREW along with evapotranspiration and detailed soil moisture measurements provided by the SSCZO to produce a quasi-replicate experiment at the headwater catchment and river-basin scales to understand closure patterns within and between watersheds.

2 | MATERIALS AND METHODS

2.1 | Study area

2.1.1 | River basins

Southern Sierra Nevada basins ($1016\text{--}5389\text{ km}^2$) of the Merced, San Joaquin, Kings, Kaweah, Tule, and Kern Rivers were selected to investigate the water balance at larger scales (Table 1). The average elevation ranges from 1638 m for the Merced to 2328 m for the Kings (Figure 1). The underlying geology is granitic with scattered metamorphic sedimentary and volcanic rocks (Jennings & Gutierrez, 2010) throughout the study area (Figure 1). Except in the Merced, Mesozoic granodiorite dominates the rock type, covering over 75% of the basin area. In the Merced, Mesozoic granodiorite occupies 45% of the basin followed by 25% Ordovician to Triassic argillite, 10% Jurassic mafic volcanic rocks, and 5% Triassic to late Jurassic slate. All of the selected basins have glacial history (Clark et al., 2003; Matthes & Fryxell, 1965; Warhaftig & Birman, 1965) and as a result rock and rock materials of glacial origin occupy as much as 8% of the basin surface area. Matthes and Fryxell (1965) documented evidence of mild glaciation in headwaters of the Tule River, but the percentage of glacial drift is negligible (Jennings & Gutierrez, 2010). Entisol (45%), Inceptisol

TABLE 1 River basin characteristics and long-term mean water balance components

River basin	Area (km ²)	Elevation (m)			\bar{P} (mm/year)	\bar{Q} (mm/year)	\bar{ET} (mm/year)	\bar{G}_n (mm/year)	ET_o (mm/year)	Aridity (-)
		Min	Max	Mean						
Merced	2760	95	3944	1647	972	428	475	69	1734	0.497
San Joaquin	4353	92	4228	2111	939	484	342	113	1702	0.439
Kings	4008	171	4300	2332	923	500	369	54	1684	0.430
Kaweah	1456	189	3818	1725	848	350	476	22	1852	0.385
Tule	1016	165	3100	1218	677	154	464	59	2028	0.290
Kern	5389	748	4405	2201	555	151	287	117	1843	0.323

Note: \bar{P} = mean annual (1981–2019) precipitation from PRISM AN81m (Daly et al., 2008, 2015). \bar{Q} = mean annual (1985–2019) full natural flow from the California Department of Water Resources (online at: <https://cdec.water.ca.gov/>). \bar{ET} = mean annual (1985–2019) evapotranspiration using Goulden and Bales (2019). \bar{G}_n = net groundwater exchange, calculated using Equation (5a). ET_o = mean annual (1970–2000) reference evapotranspiration (Trabucco & Zomer, 2018). Aridity = mean aridity-wetness index using mean annual (1970–2000) precipitation from WorldClim2 Global Climate (Fick & Hijmans, 2017) and mean annual (1970–2000) ET_o from Trabucco and Zomer (2018).

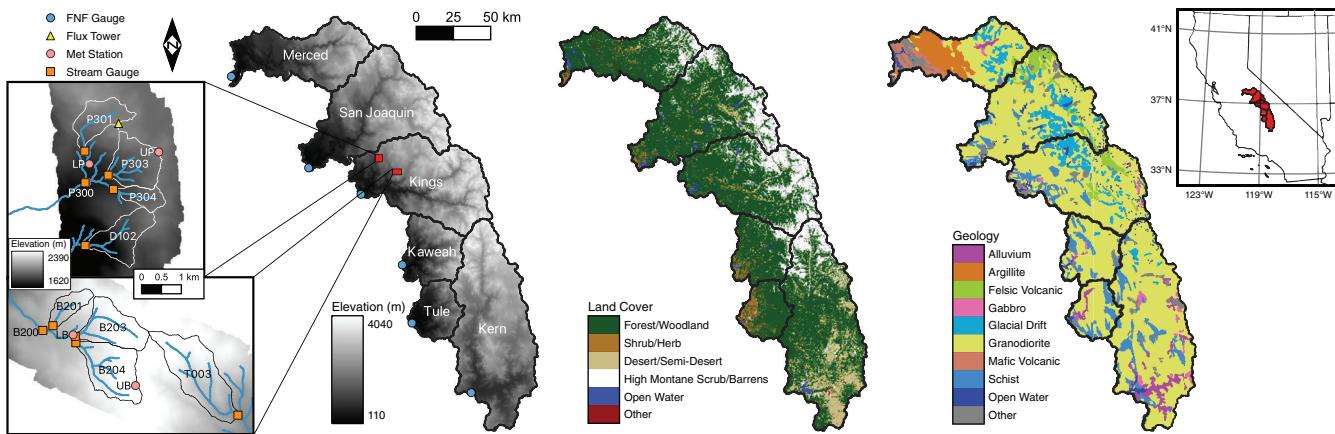


FIGURE 1 Study area showing the river basins and their elevation (left), land cover (center), and geology (right). The inset figures (far left) show the Providence (top) and Bull (bottom) headwater catchments, while the inset figure (far right) shows the location of the river basins within the state of California. FNF is full natural flow. LP, UP, LB, and UB are meteorological stations for lower Providence, upper Providence, lower Bull, and upper Bull, respectively

(25%), Alfisol (18%), and Mollisol (8%) soil orders, largely Dystric Xeropsammets (28%), Dystric Xeropchrepts (14%), Ultic Haploxeralfs (10%), and Typic Cryorthents (8%) dominate the study area (40%). The main land covers (Comer et al., 2003) are forest and woodland (68%), high montane scrub and barrens (18%), and desert or semi-desert (9%) (Figure 1). The vegetation is dominated by California dry-mesic and mesic mixed-conifer forest and woodland (18%), followed by subalpine lodgepole pine forest and woodland (9%), lower montane blue-oak-foothill pine woodland and savanna (7%), montane Jeffrey pine (7%), red fir (7%), and mixed oak woodland (6%). Nearly one-sixth (17%) of the study area is classified as alpine bedrock and scree, particularly above 3200 m of elevation (Rundel, 2011).

The climate of the river basins can be characterized as arid near the foothills of the Tule and Kern and mostly semi-arid and dry subhumid at the higher elevations. Because of the latitudinal climate gradient in California, the proportion of the basin area with a dry sub-humid climate increases between the Kern and Merced river basins. The average annual basin precipitation increases from south to north with the

Merced receiving ~43% more precipitation than the Kern. The Kings is the coldest basin with a 7°C average annual mean daily air temperature followed by the San Joaquin and the Kern (8°C), the Merced (10°C), the Kaweah (11°C), and the Tule (13°C). The percentage of the total annual precipitation falling as snow ranges from as little as 30% in the Tule to as much as 75% in the Kings (Safeeq et al., 2016).

2.1.2 | Headwater catchments

Our 10 headwater catchments (0.5–5 km²) are part of the KREW and SSCZO networks and drain into the Kings River near the tail end of the Pine Flat Reservoir (Table 2). These catchments are clustered into two groups, the lower elevation (mean elevation 1859 m) Providence and the higher elevation (mean elevation 2308 m) Bull catchments. Three catchments in each group are nested within the larger P300 and B200 catchments, respectively (Figure 1). The underlying rock is largely Mesozoic granodiorite in Providence and early Proterozoic to

TABLE 2 Headwater catchment characteristics and long-term mean water balance components

Catchment	Area (ha)	Elevation (m)			\bar{P} (mm/year)	\bar{Q} (mm/year)	\bar{ET} (mm/year)	\bar{G}_n (mm/year)	ET_o (mm)	Aridity (-)
		Min	Max	Mean						
B201	53	2151	2382	2256	1366	534	539	293	1680	0.531
B203	138	2182	2490	2372	1422	712	469	241	1672	0.522
B204	167	2194	2489	2362	1394	619	477	298	1684	0.517
T003	228	2040	2470	2285	1436	571	605	260	1688	0.521
B200 ^a	474	2120	2490	2319	1335	516	507	312	1724	0.531
P301	99	1793	2108	1974	1309	454	652	203	1772	0.551
P303	132	1722	2025	1898	1309	321	706	282	1762	0.548
P304	49	1765	1976	1895	1322	423	704	195	1769	0.543
D102	121	1476	1980	1774	1322	337	666	319	1775	0.544
P300 ^a	461	1679	2107	1881	1248	359	682	207	1724	0.532

Note: \bar{P} = mean annual (2004–2019) precipitation from KREW. \bar{Q} = mean annual (2004–2019) runoff from KREW. \bar{ET} = mean annual (2004–2019) evapotranspiration (Goulden & Bales, 2019). \bar{G}_n = net groundwater exchange, calculated using Equation (5a). ET_o and Aridity = same as Table 1.

^aNo data for 2004–2006.

Cretaceous schist in Bull. Soils are predominantly Inceptisols and Entosols in Providence and Bull, respectively. Shaver series (coarse loamy, mixed, superactive, mesic Humic Dystroxerepts) are present at lower (1750–1900 m) elevations and Gerle (coarse-loamy, mixed, superactive, frigid Humic Dystroxerepts) and Cagwin (mixed, frigid Dystric Xeropsammets) series dominate at higher (1800–2400 m) elevations (Bales et al., 2011; Hunsaker et al., 2012). Cagwin series soils have high permeability and are classified into soil hydrologic group A. Gerle and Shaver series soils have moderate permeability and are classified into soil hydrologic group B. The land cover is 40–60 m tall mature mixed conifer forest, largely white fir (*Abies concolor*), ponderosa pine (*Pinus ponderosa*), Jeffrey pine (*Pinus jeffreyi*), sugar pine (*Pinus lambertiana*), and incense cedar (*Calocedrus decurrens*). As compared to Providence, mixed conifer vegetation in Bull contains a higher percentage of red fir (*Abies magnifica*) and is sparse, with 1–4% rock outcrop or bare ground (Safeeq & Hunsaker, 2016).

The catchments have a Mediterranean climate, with wet, cold winters and dry, hot summers. Average annual precipitation varies between 1314 mm/year at Providence and 1400 mm/year at Bull. Over 90% of the annual precipitation at the KREW catchments falls between October and April. The mean average daily temperature, over water years 2002–2019, ranged between 10°C in Providence and 7.5°C in Bull. Lower temperatures in Bull relative to Providence cause a higher proportion of precipitation to fall as snow than rain. However, none of these catchments are above the rain-snow transition elevation (>2500 m), and so the type of precipitation shifts between rain and snow depending on the storm's temperature. Atmospheric rivers play a major role in the precipitation regime with as much as 30%–40% of snow accumulation (Guan et al., 2010) and 30%–45% of all precipitation (Dettinger et al., 2011), often contributed by just one or two events each year. Based on the aridity, defined as the ratio of mean annual precipitation to mean annual reference evapotranspiration (Trabucco & Zomer, 2018), all KREW catchments are classified as dry sub-humid.

2.2 | Datasets

2.2.1 | Precipitation

Daily precipitation in the KREW catchments was measured using weighing gauges at four locations (Safeeq & Hunsaker, 2016). These gauges are located at upper and lower elevations in each catchment group, Providence and Bull, to better capture the elevational patterns of precipitation (Figure 1). For the catchment average precipitation over the water year, defined as October 1–September 30, daily precipitation values from the two gauges in each catchment group were averaged, aggregated on a water year basis, and scaled using the ratio of PRISM (Daly et al., 2008, 2015) long-term (1981–2010) precipitation over the catchment to the precipitation of the grid cell underneath the gauge. These scaling factors ranged from 0.98 in B201 to 1.03 in T003, suggesting a marginal difference between catchment average precipitation and point precipitation at the gauge locations. This is not surprising considering the weak elevation control on precipitation at this elevation range in the Kings (Safeeq & Hunsaker, 2016). Water year precipitation for the six river basins was based on monthly PRISM (Daly et al., 2008, 2015) precipitation, version AN81m, available in Google Earth Engine (Gorelick et al., 2017).

2.2.2 | Discharge

Discharge at the headwater catchment outlets was measured at 15-min intervals using a combination of nested Parshall-Montana flumes, weirs, and manual stage-discharge rating curves (Safeeq & Hunsaker, 2016). Volumetric [L^3/T] 15-min catchment discharge values were aggregated to daily and water year timescales and normalized by catchment drainage area for converting into unit runoff [L/T]. Discharge for each of the river basins is regulated by hydroelectric and other multi-purpose dams. As a result, monthly full natural or

unimpaired flow data (FNF) for each basin were obtained from the California Department of Water Resources (<https://cdec.water.ca.gov/>), aggregated to a water year basis, and normalized by the drainage area for converting into unit runoff. The accuracies of these unimpaired runoff values are known to be comparable to the accuracy of the U.S. Geological Survey streamflow gauges (Gleick, 1987; U.S. Department of the Interior, 1976).

2.2.3 | Evapotranspiration

Evapotranspiration (ET) was measured using the eddy covariance method at the US-CZ3 site within catchment P301 at 30-min intervals starting in 2008 (Goulden et al., 2012). ET data was gap-filled using energy balance closure and aggregated to the daily timescale (Runge et al., 2019). Water year catchment and river basin ET values were derived in Google Earth Engine (Gorelick et al., 2017) using a linear regression ($ET \text{ (mm)} = 117.16 * \exp(2.8025 * NDVI)$, $R^2 = 0.84$) between ground-based ET from 10 California eddy covariance flux towers and Normalized Difference Vegetation Index (NDVI) from Landsat 5, 7, and 8 for the 9 nearest upwind pixels (Goulden & Bales, 2019). In an earlier study, Goulden and Bales (2014) validated this NDVI based ET model by comparing average (2003–2012) annual estimated ET against the long-term (1980–2010) mean P–Q as a proxy for measured basin-wide ET. This validation was performed in 11 major river basins on the western slope of the Sierra Nevada, including five of the six river basins from this study. More recently, Roche et al. (2020) performed a more robust independent cross-validation on NDVI-based ET estimates using the leave-one-out approach and reported a $\pm 5\%$ prediction uncertainty, mostly in high NDVI regions where saturation of surface reflectance is an issue.

2.2.4 | Soil water storage

Measurements of hourly soil moisture (2009–2016) at depths of 10, 30, 60, and 90 cm below the soil surface were obtained from the SSCZO wireless sensor network data (O'Geen et al., 2018). This network included two clusters of 10 and 17 nodes located near the lower and upper Providence precipitation gauge, respectively. These sensor nodes were strategically placed to sample soil moisture variations across aspect and tree canopy conditions. First, hourly volumetric soil moisture readings, m^3/m^3 , from each node were converted into mean daily measurements. We then multiplied each measurement by their respective zone of influence and calculated sum total volumetric water content, $S(t)$, in the top 1 m soil profile at time step t as follows:

$$S(t) = \theta_{10}(t) \times 0.2 + \theta_{30}(t) \times 0.25 + \theta_{60}(t) \times 0.3 + \theta_{90}(t) \times 0.25, \quad (2)$$

where θ is the volumetric water content and the subscript represents the depth of measurement (cm). For effective soil depth, gaps between two adjacent sensors were split in half. Second, the

volumetric water content values, $S(t)$, were multiplied by 1000 to convert the unit from volumetric (m^3/m^3) to depth of water (mm) in the one-meter profile. Finally, the depth equivalent water storage values were averaged across the 27 nodes to derive water storage for the entire P300 catchment.

2.3 | Water balance

Accounting for the measurement uncertainties in the components of the water balance, we can rewrite Equation (1) using an error term:

$$P - ET - Q - G_n - \frac{dS}{dt} = \epsilon, \quad (3)$$

where the term ϵ is the residual error in the water balance. Using Equation (3) as a baseline, we derive annual and multi-year variations of the water balance to investigate different water-balance closure assumptions at different scales.

2.3.1 | Long-term water-balance closure

Over multiple-year timescales, water-balance closure assumptions are commonly made that allow the water balance to be simplified. The mean annual water balance for a period of N years can be described using the equation:

$$\bar{P} - \bar{ET} - \bar{Q} - \bar{G}_n - \frac{\bar{dS}}{dt} = \bar{\epsilon}. \quad (4)$$

For a sufficiently large N , both mean residual error $\bar{\epsilon}$ and mean change in the terrestrial water storage \bar{dS}/dt are often approximated as zero, assuming that the basin is at a steady state with no net change in terrestrial water storage and residual error $\bar{\epsilon}$ is normally distributed. Equation (4) then becomes

$$\bar{P} - \bar{ET} - \bar{Q} - \bar{G}_n = 0, \quad (5a)$$

or

$$\bar{G}_n(N) = \frac{\sum_{i=1}^N P_i}{N} - \frac{\sum_{i=1}^N ET_i}{N} - \frac{\sum_{i=1}^N Q_i}{N}. \quad (5b)$$

These long-term water-balance equations allowed us to investigate the magnitude and spatial patterns of \bar{G}_n with varying number of N years (2, 3, 4, ...30) over which \bar{G}_n converges asymptotically to zero. For each value of N , N -years of annual P , ET , and Q were randomly drawn from the observed record for estimating $\bar{G}_n(N)$ using Equation (5b). This sampling was repeated 1000 times and the mean, standard deviation, and coefficient of variation (CV) were estimated from the 1000 values of $\bar{G}_n(N)$ for comparison between the sites.

2.3.2 | Simplified annual water-balance closure

For many water mass balance analyses at an annual time-step, measurements of G_n and dS/dt are not available and generally assumed to be zero. Similarly, since the error associated with P , ET , and Q is often unquantified, it is assumed to be negligible, which reduces the water balance Equation (3) to just three terms:

$$P - ET - Q = 0. \quad (6)$$

This simplified form of the annual water balance allows researchers to solve for one of the unknowns, often ET or Q . However, in order to use Equation (6), it is important that the assumptions used to simplify from Equation (3) to (6) be valid. In this study, detailed measurements of all three components, P , Q , and ET , allow us to test the spatial and temporal scales where the closure assumptions apply.

We first estimated the water-balance closure imbalance as:

$$\Delta = P - ET - Q, \quad (7)$$

If the assumption of a complete water-balance closure in Equation (6) holds true, then Δ will be homoscedastic, independent of P , Q , and ET , and follow a normal distribution (Xu, 2001). The homoscedasticity of the residuals was tested by plotting Δ against P , Q , and ET .

In catchment P300, soil moisture storage was available for the top 1 m (Section 2.2.4). This additional information allowed us to simplify Equation (3) without the assumption that $dS/dt \approx 0$, such that

$$P - ET - Q - \frac{dS}{dt} = 0. \quad (8a)$$

Equation (8a) was used to examine the assumption that G_n is negligible in catchment P300. It is worth noting that the simplification associated with soil moisture storage measurements limited to the top 1 m of the soil in most cases does not completely describe the S term. Past studies have reported deep water drawdown by vegetation (Fellows & Goulden, 2017; Klos et al., 2018; O'Geen et al., 2018), especially during prolonged droughts (Bales et al., 2018; Goulden & Bales, 2019). Fellows and Goulden (2017) estimated conifer forest rooting depth to be at least 3.6 m with weathered saprolite or saprock reaching up to 20 m beneath the surface (Holbrook et al., 2014; Klos et al., 2018). Moreover, our assumption of uniform soil thickness throughout the catchment is an oversimplification considering the variation in soil depth or regolith thickness within a catchment (O'Geen et al., 2018). The use of soil moisture storage measured in the top 1 m to infer overall catchment dS/dt will likely lead to an underestimation of the total terrestrial storage and an overestimation of G_n . For this reason, we further divided total terrestrial catchment water storage between shallow (top 1 m) S_s and deep (from 1 m below the surface to bedrock) S_d zone storages as follows:

$$P - ET - Q - \left(\frac{dS_s}{dt} + \frac{dS_d}{dt} \right) = 0, \quad (8b)$$

Catchment scale measurements to characterize dS_d/dt are limited. Our focus here is to investigate the $G_n \approx 0$ assumption while acknowledging the fact that changes in dS_d/dt during prolonged droughts, as shown in earlier studies, can be significant.

The precipitation-decorrelation technique of Trask et al. (2017) provides an alternative way to examine Equation (3) by resolving the ($G_n + dS/dt$) from the random residual error ε . The Trask et al. (2017) approach provides a middle ground between a fully closed water balance and an open water balance (Kampf et al., 2020). The technique allows the water balance for one of the unknown terms to be solved without making assumptions regarding the accuracy of known terms or forcing the net water-balance error ε to zero. The statistical basis of this technique is the fact that interannual variation in $G_n + dS/dt$ will closely follow the variation in P and can be approximated using a linear model with assumptions that the covariance of P and Δ is statistically significant and the covariance of P and $\varepsilon \approx 0$. This method can be applied using univariate, for example, $\Delta = f(P)$, or multivariate, for example, $\Delta = f(P, ET)$, statistical techniques. Following Trask et al. (2017), we estimated annual $G_n + dS/dt$ using the equation:

$$G_n(t) + \frac{dS}{dt}(t) = \bar{\Delta} + \beta_1[P_N(t) - 1] + \beta_2[ET_N(t) - 1], \quad (9)$$

where P_N is normalized [$P(t)/\bar{P}$] precipitation for year t , ET_N is normalized [$ET(t)/\bar{ET}$] evapotranspiration, and β_1 and β_2 are coefficients determined using the least-squares fit between $\Delta(t)$ as a dependent variable and $P_N(t) - 1$ and $ET_N(t) - 1$ as explanatory variables. The -1 terms in brackets in Equation (9) serve to subtract β_1 and β_2 and ensure that the covariance between the water balance components (ET and P) and ε is close to zero and water balance remains closed (Trask et al., 2017). In catchments B201, B203, and B204, $ET_N(t) - 1$ showed no explanatory power and was replaced with normalized discharge ($[Q_N(t) - 1]$, where $Q_N = Q(t)/\bar{Q}$). This anomaly may have been driven by the fact that ET in these high elevation catchments is energy limited with small interannual variations. We also explored 1- and 2-year lagged precipitation as additional explanatory variables in the least squares fit model to account for memory effects but neither was statistically significant.

3 | RESULTS

3.1 | Long-term water-balance closure

Long-term average net groundwater exchange \bar{G}_n (Equation (5)) from the six river basins ranged between 22 mm/year in the Kaweah to 117 mm/year in the Kern (Table 1). On average across the river basins, \bar{G}_n was equivalent to 9% of \bar{P} , 18% of \bar{ET} , and 21% of \bar{Q} . At the extreme end, in the Kern river basin \bar{G}_n was equivalent to 21% of \bar{P} , 41% of \bar{ET} , and 77% of \bar{Q} (Figure 2). At the headwater-catchment scale, \bar{G}_n ranged between 195 mm/year in P304 to as much as 319 mm/year in D102 (Table 2). In P303 and D102, \bar{G}_n was very close to being equal to \bar{Q} . On average across the 10 catchments, \bar{G}_n was

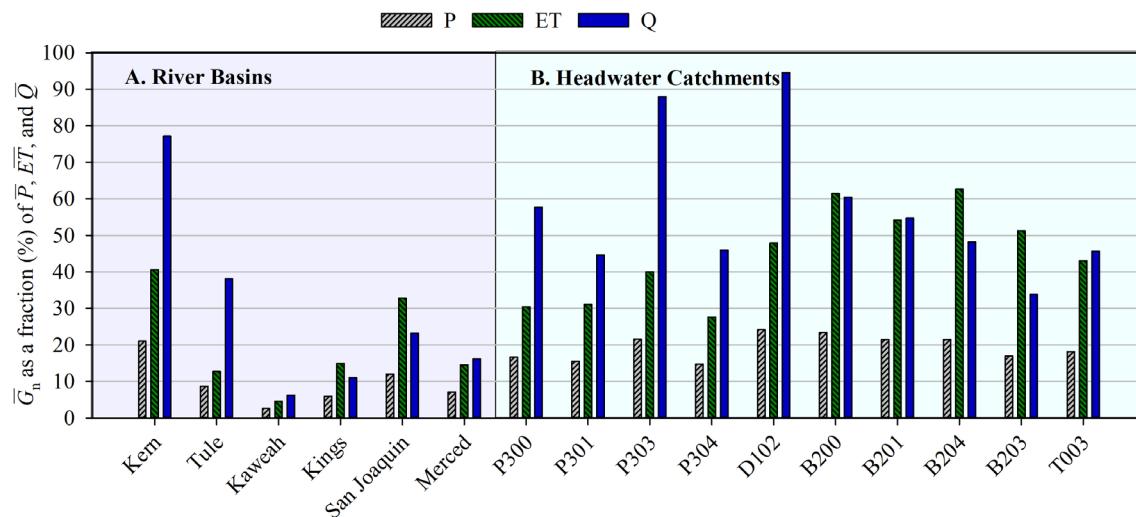


FIGURE 2 Long-term average net groundwater exchange (\bar{G}_n) across the different (a) river basins and (b) headwater catchments. Please note that the differences in the period of analysis between the river basins (1985–2019) and catchments (2004–2019). The river basin \bar{G}_n as a percentage of precipitation, evapotranspiration, and runoff during 2004–2019 is almost half of those during 1985–2019

equivalent to 20% of \bar{P} , 43% of \bar{ET} , and 54% of \bar{Q} . Despite the fact that P301, P303 and P304 are nested inside P300 and B201, B203, and B204 are nested inside B200, there was no evidence of lowering of \bar{G}_n with increasing scale or stream order (consistent with Liu et al., 2013). Overall, \bar{G}_n was highly correlated with \bar{P} ($R^2 = 0.71$, $p < 0.01$) but showed weak or no association with \bar{ET} ($R^2 = 0.3$, $p = 0.02$) or \bar{Q} ($R^2 = 0.18$, $p = 0.10$). The relationship between \bar{G}_n and elevation was also statistically insignificant ($p = 0.10$). Because of the differences in scale, relationships between \bar{G}_n , drainage area, and relief were explored separately for river basins and catchments, and only river basin drainage area showed moderate predictive power for \bar{G}_n ($R^2 = 0.47$, $p = 0.003$).

Estimated mean \bar{G}_n using different base years ($N > 1$, 1000 N-years combinations) are shown in Figure 3. The mean \bar{G}_n in the six river basins and 10 headwater catchments remained relatively constant across the different base periods (Figure 3(a, b)). However, the pattern in the CV was highly variable between and within scales (i.e., river basins vs. catchments). While the decline in CV with increasing N is statistically expected due to regression toward the mean with longer time-series, the rate at which CV declines with increasing N points to differences in the sites and hydrologic memory. Large river basins have a much higher memory effect on \bar{G}_n than the headwater catchments. It is also worth pointing out the differences within the river basins and headwater catchments. The Kaweah, and to some extent P304, show much higher CV in \bar{G}_n for the same N . These two sites have the lowest mean \bar{G}_n (Tables 1 and 2), making them more sensitive to interannual variation. These results suggest that \bar{G}_n is highly sensitive to N and the definition of long-term must be carefully evaluated. For example, in river basins, 25 years may seem like a reasonable length for assuming \bar{G}_n to be independent of N since the CV is below 0.25 for most basins. The time-scale for headwater catchments to fall below the 0.25 CV threshold is only 10 years. It is worth highlighting that while overlapping years and non-sequential years

were used in this analysis due to the limited record, a less robust analysis of non-overlapping samples showed similar results (data not shown), suggesting that the role of overlapping years may not be substantial.

3.2 | Simplified annual water-balance closure

Annual (WY) water-balance closure error (i.e., Δ) for the river basins ranged between -188 mm/year (2014) in the Kaweah and 344 mm/year (1998) in the Kern. In the headwater catchments this error was even higher, ranging between -283 mm/year (2014) in P304 and 914 mm/year (2017) in B200. Water-balance closure error was largely positive with 71% and 82% of the years having $\Delta > 0$ in river basins and catchments, respectively. In relative terms, the absolute magnitude of Δ in any given year during the study period was as much as 46% of the annual precipitation in headwater catchments and 48% of the annual precipitation in river basins. As expected, the cumulative plots indicate a strong interannual variation in Δ (Figure 4). In particular, during the droughts Δ declined as vegetation started drawing water from the storage to meet the transpiration demand (Bales et al., 2018). The 2012–2016 drought was relatively hotter and drier than the 1987–1992 drought (Goulden & Bales, 2019), causing a larger net decline in Δ . Interestingly, none of the river basins or catchments showed a return of Δ to zero over the period of record (Figure 4), implying that all of the river basins and headwater catchments are net exporters of groundwater as dS/dt over this time-scale was approximately zero (see below). This result is consistent with the occurrence of mountain-block recharge in these basins (Markovich et al., 2019). The Kaweah looks to be a potential exception, with a significant decline in Δ during 2004–2011, just before the start of 2012–2016 drought, when compared with 1985–2003 (Figure 4(a)). Other catchments also showed similar declines, but they were

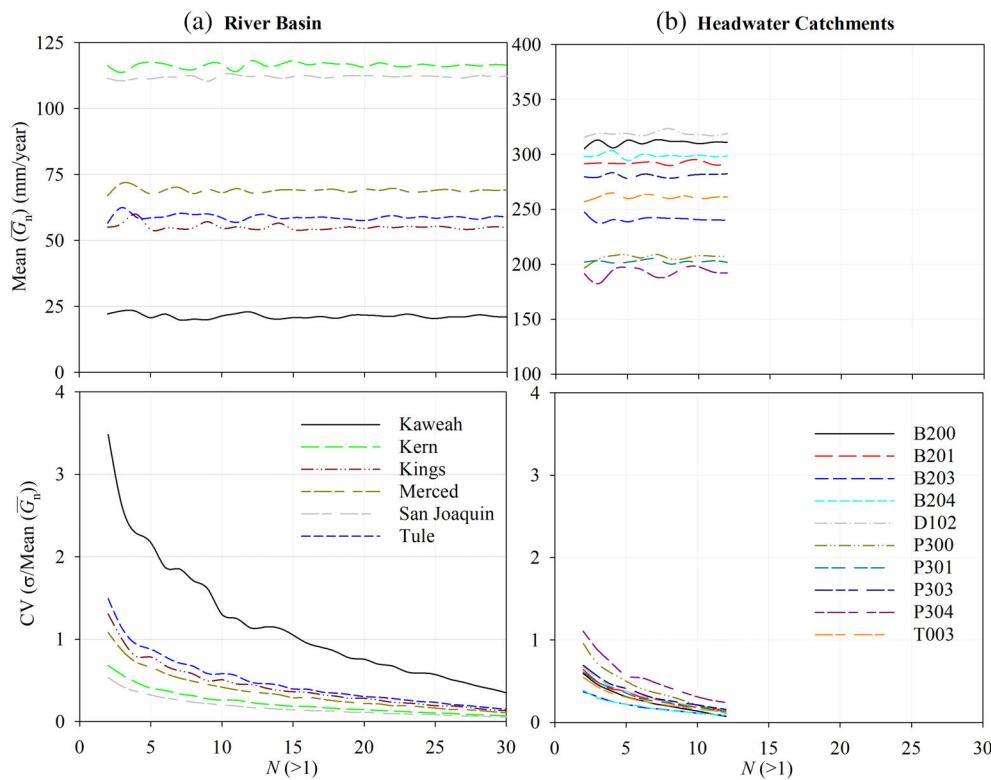


FIGURE 3 Mean of the average groundwater exchange over N -years (\bar{G}_n), and corresponding coefficient of variation (CV) estimated over multiple base periods (N) of 2 years and longer for (a) river basins and (b) headwater catchments. Note the different range of the y-axis in the top panel

statistically insignificant. This decline in Δ may have been driven by the canopy expansion, or structural overshoot of above ground biomass due to periods of favourable climatic and management conditions that facilitated abundant tree growth between 2009 and 2012 (Goulden & Bales, 2019; Jump et al., 2017). Overall, the Kaweah, along with the Tule, had the highest percentages (50% and 51%, respectively) of forest within the highly productive mid-elevation range of 1000–2500 m (Goulden & Bales, 2019), followed by the Merced (46%), the Kern (42%), the San Joaquin (39%), and the Kings (33%).

To examine the assumption of negligible G_n , we looked at the daily variations in water input (rain plus snowmelt), ET, Q, dS_s/dt , and the water-balance remainder (i.e., $dS_d/dt + G_n + \varepsilon$) between different years in P300 (Figure 5). The average dS_s/dt between 2009 and 2016 was 0.5 ± 12 mm (Table 3), suggesting that at annual and longer timescales dS_s/dt is negligible. Shallow soil water storage deficit from ET gets replenished from the rain and snowmelt within a year with no large carryover storage. At a daily timescale, maximum cumulative dS_s/dt varied marginally (standard deviation = 37 mm) among years compared to the variations in total annual P (standard deviation = 527 mm) and Q (standard deviation = 338 mm). The maximum cumulative dS_s/dt ranged between 190 mm in 2014 and 298 mm in 2011 (Table 3). Water year 2011 was one of the wettest, receiving 2152 mm precipitation compared to 555 mm in 2015 (Table 3). The cumulative water balance remainder, $dS_d/dt + G_n + \varepsilon$, remained largely negative or near zero during the first 2 months of the water year (Figure 5(e)), which are generally dry. We found $dS_d/dt + G_n + \varepsilon$ most significantly correlated with recharge or water input

(rain plus snowmelt) with a lag of zero (cross-correlation = 0.63, $p < 0.05$). The cross-correlation between water-balance remainder and water input at a lag of one-day (cross-correlation = 0.38, $p < 0.05$) suggested an additional delayed recharge of soil moisture, especially in the snow free season when the ground is largely unsaturated. Overall, the negative $dS_d/dt + G_n + \varepsilon$ values of water-balance remainder near the beginning of the water year (suggesting $dS_s/dt > P - ET - Q$) were tied to an increase in shallow storage ($r = 0.93$, $p < 0.05$) with larger negative values (Figure 5(e)) coinciding within a day following an input (Figure 5(a)). However, at the daily time-step we did notice instances where increases in soil moisture storage did not match the recharge input, suggesting measurement error. As the water year progressed, the cumulative water balance remainder peaked at the end of the snowmelt season (March to May) before declining during the summer months. The net annual total water-balance closure error between different years ranged between −23.4 mm in 2015, the driest water year in the record, and 467 mm in 2016, which had slightly below average precipitation. As water-balance remainder was positively correlated with precipitation, the unusual behaviour for water year 2016 was in part due to 2016 being a relatively warm water year (+1.5°C warmer) and followed an extremely dry year with only 555 mm precipitation in 2015 (Figure 5). The large water-balance imbalance in 2016 may reflect, in part, the refilling of the deeper storage deficit from the drought.

The cumulative increase of Δ (Figure 4) suggests the presence of a systematic component. At longer timescales, when dS/dt is zero, there are two possible end-member sources and explanations for the systematic component of Δ variation: it is entirely made of net deep

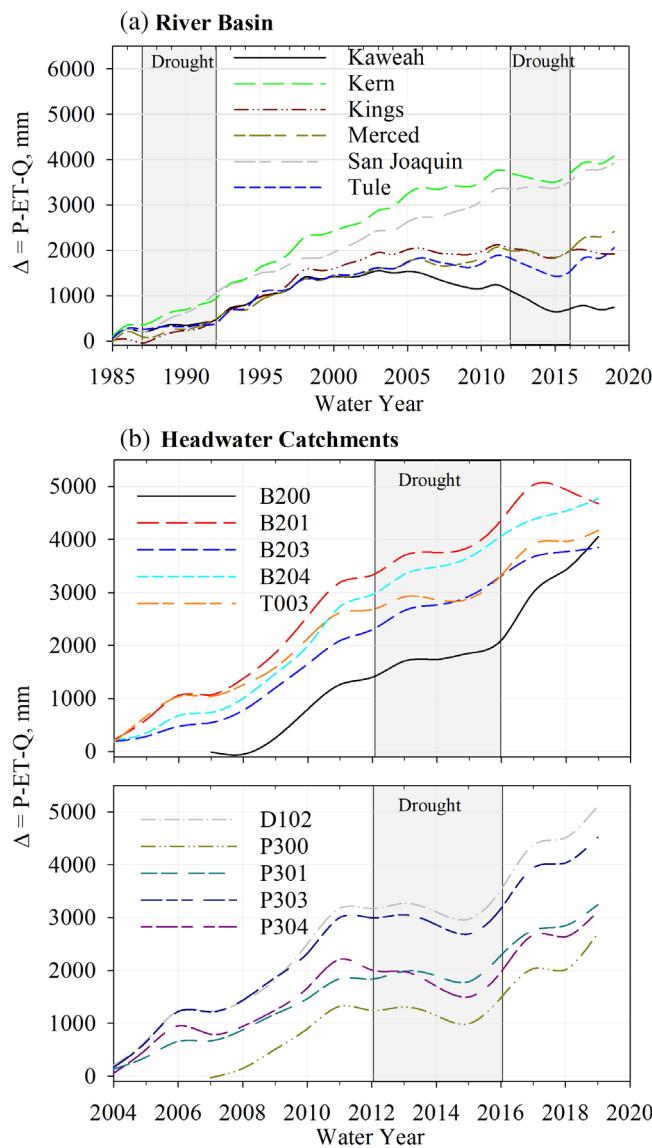


FIGURE 4 Cumulative plots of water balance closure error, $\Delta = P - ET - Q$, over time for (a) river basins and (b) headwater catchments

subsurface flows through the regolith or groundwater exchange (G_n) or it is entirely made of systematic errors associated with the measurements of P , ET and Q . In reality, it is likely neither one nor the other but rather a combination of G_n and systematic measurement errors. The relationships between annual Δ and P , ET , and Q (Figure 6) revealed that Δ is biased and largely heteroscedastic for P and Q (Kruskal Wallis test, $\chi^2 > 5.991$ for 2 degrees of freedom) with a positive slope, and unbiased and homoscedastic for ET (Kruskal-Wallis test, $\chi^2 < 4.412$ for 2 degrees of freedom). Plotting Δ against P , ET , and Q for P300 at a daily timestep revealed similar patterns, suggesting no systematic association between Δ and ET . These results are in line with the fact that temporal variation in forest ET under the Mediterranean climate of the Sierra Nevada is rather insignificant compared to variations in P and Q (Bales et al., 2018).

The components of the water balance contain both systematic and random errors. While the propagation of systematic error in the water balance is beyond the scope of this manuscript, we separated the random error from Δ using the precipitation-decorrelation method described earlier. The coefficient of determination (i.e., R^2) from the least-squares fit, developed to derive β_1 and β_2 in Equation (9) ranged between 0.56 for the Kings and 0.95 for P303. Eleven out of the 16 sites had R^2 values over 0.85, suggesting that the precipitation-decorrelation technique can be used to infer poorly resolved water balance terms, that is, $G_n + dS/dt$ and ε from the total mass balance closure error Δ .

The annual water-balance component estimates (Equation (3)) for the six river basins from 1985 to 2019 and 10 headwater catchments from 2004 to 2019 are shown in Figures 7 and 8, respectively. Overall, the resulting random residuals ε were homoscedastic (Kruskal-Wallis test, $\chi^2 < 2.469$ for 2 degrees of freedom) and showed no correlation ($p > 0.11$) with P , ET , or Q , thus satisfying the statistical criteria for being random. However, at the individual basin and catchment levels, ε was significantly correlated with P in B201 ($R^2 = 0.58$, $p < 0.001$) and B204 ($R^2 = 0.25$, $p < 0.02$), with Q in the Kern ($R^2 = 0.15$, $p = 0.01$) and the Tule ($R^2 = 0.12$, $p = 0.03$), and with ET in the Kaweah ($R^2 = 0.36$, $p < 0.001$), the Kern ($R^2 = 0.12$, $p = 0.03$), the Kings ($R^2 = 0.20$, $p < 0.001$), and the San Joaquin ($R^2 = 0.16$, $p = 0.01$). The strength of most of these correlations can be categorized as weak.

For river basins, the magnitude of $G_n + dS/dt$ ranged between -125 mm (WY 2014) in the Kaweah to as high as 344 (WY 1998) mm in the Kern (Figure 7). In contrast, the random residual ε ranged between -163 mm in the Kings (WY 2017) to as high as 171 mm in the San Joaquin (WY 1992). For the headwater catchments, the magnitude of $G_n + dS/dt$ ranged between -227 mm (WY 2014) in P304 to 946 mm (WY 2017) in D102 (Figure 8). The random residual ε ranged between -257 mm in B200 (WY 2008) to as high as 301 mm in P301 (WY 2016). These suggest that the absolute magnitudes of random residuals at annual time-scales are comparable to $G_n + dS/dt$ in river basins and smaller in the headwater catchments. At the headwater-catchment scale, the interannual variation in terms of standard deviation in annual $G_n + dS/dt$ can be explained by the mean catchment elevation ($R^2 = 0.77$) with standard deviation decreasing with increasing elevation. However, at the river-basin scale, this interannual variation in $G_n + dS/dt$ showed better agreement with mean basin runoff ($R^2 = 0.82$) and potential evapotranspiration ($R^2 = 0.69$). The standard deviation for annual $G_n + dS/dt$ declined with increasing mean river basin runoff and increased with increasing potential evapotranspiration.

Looking at the time-series, both the magnitude and interannual variations in $G_n + dS/dt$ were comparable to Q for the Tule and Kern river basins and the P303 and D102 headwater catchments (Figures 7 and 8). In terms of the magnitude, annual $G_n + dS/dt$ exceeded annual Q in 7% of the years for river basins (max. 66 mm) and 25% of the years for headwater catchments (max. 243 mm). Annual $G_n + dS/dt$ was negative in only 55 out of 210 site years in the river basins, suggesting these river basins are a net annual exporter of

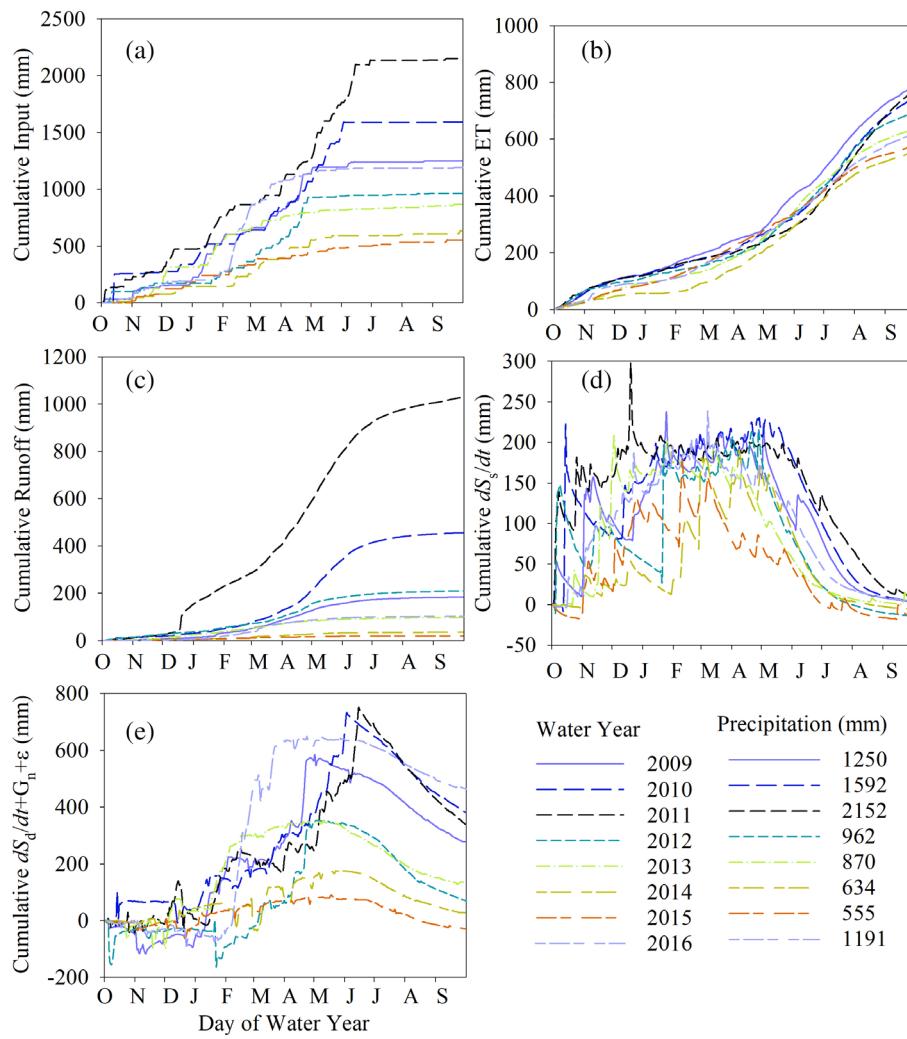


FIGURE 5 Cumulative water-balance components across water years for catchment P300: (a) input estimated as the sum of rain and snowmelt, (b) evapotranspiration (ET), (c) runoff, (d) change in shallow storage (dS_s/dt), and (e) water balance remainder ($dS_d/dt + G_n + \varepsilon = P - ET - Q - dS_s/dt$)

TABLE 3 Measured annual water-balance components for catchment P300

Water year	P (mm)	Q (mm)	ET ^a (mm)	Peak cumulative change in storage (mm)	Shallow storage change ^b (mm)	Water balance remainder ^c (mm)
2009	1250	184	785	238	2.1	279.1
2010	1592	456	750	230	5.5	380.8
2011	2152	1030	774	298	4.8	343.2
2012	962	210	698	214	-12.2	66.8
2013	870	98	634	208	4.5	133.5
2014	634	36	556	190	19.0	23.4
2015	555	20	579	178	-20.6	-23.4
2016	1191	103	620	242	1.6	466.6
Mean	1151	267	675	225	0.59	208.8
Standard deviation	527	338	90	37	12.0	182.5

^aMeasured ET at the P301 flux tower prior to scaling.

^bShallow Storage Change = dS_s/dt .

^cWater Balance Remainder = $dS_d/dt + G_n + \varepsilon$.

groundwater. This pattern was even stronger at the headwater-catchment scale with only 12 out of 154 site years having negative $G_n + dS/dt$. The majority (67%) of these negative $G_n + dS/dt$ values,

occurred in years during the 1987–1992 and 2012–2016 droughts, suggesting a water balance deficit. The cumulative water budget deficit during the 2012–2016 drought, which was more extreme than the

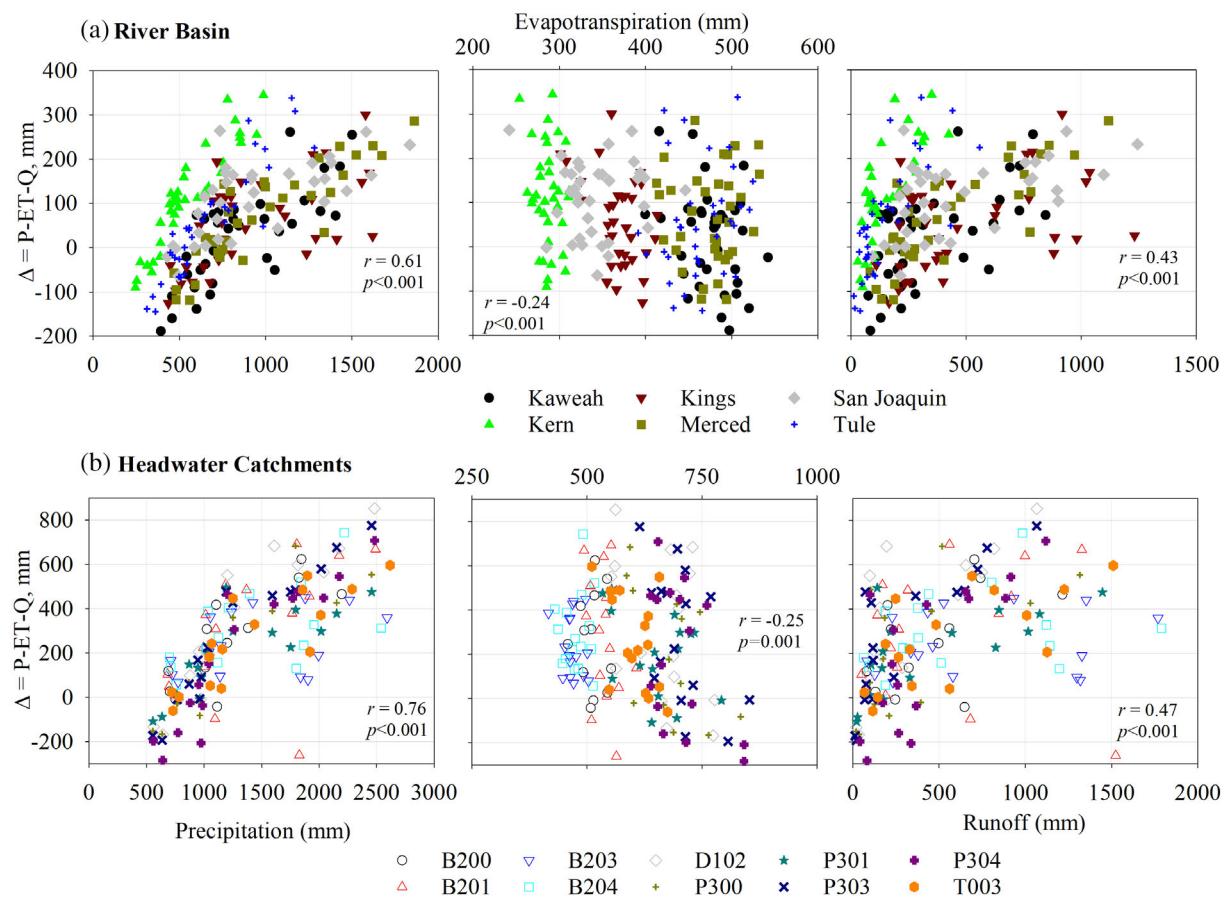


FIGURE 6 Variation in $\Delta = P - ET - Q$ with respect to precipitation (P), evapotranspiration (ET), and runoff (q) between the (a) river basins and (b) headwater catchments

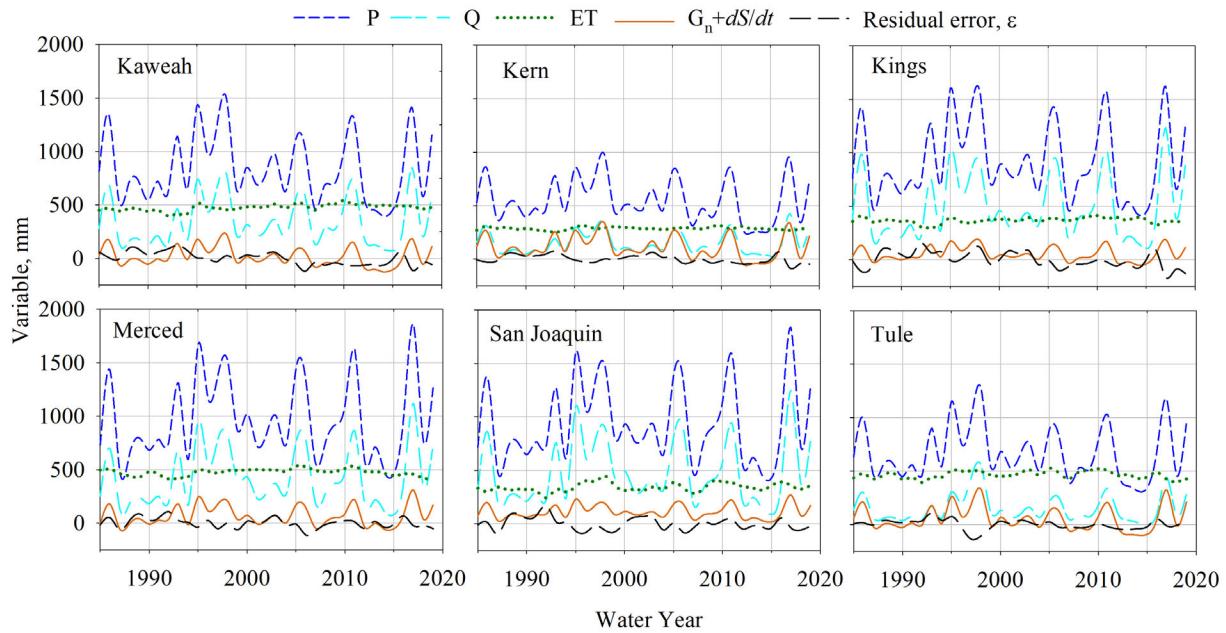


FIGURE 7 Annual water-balance components and residual error for the six studied river basins

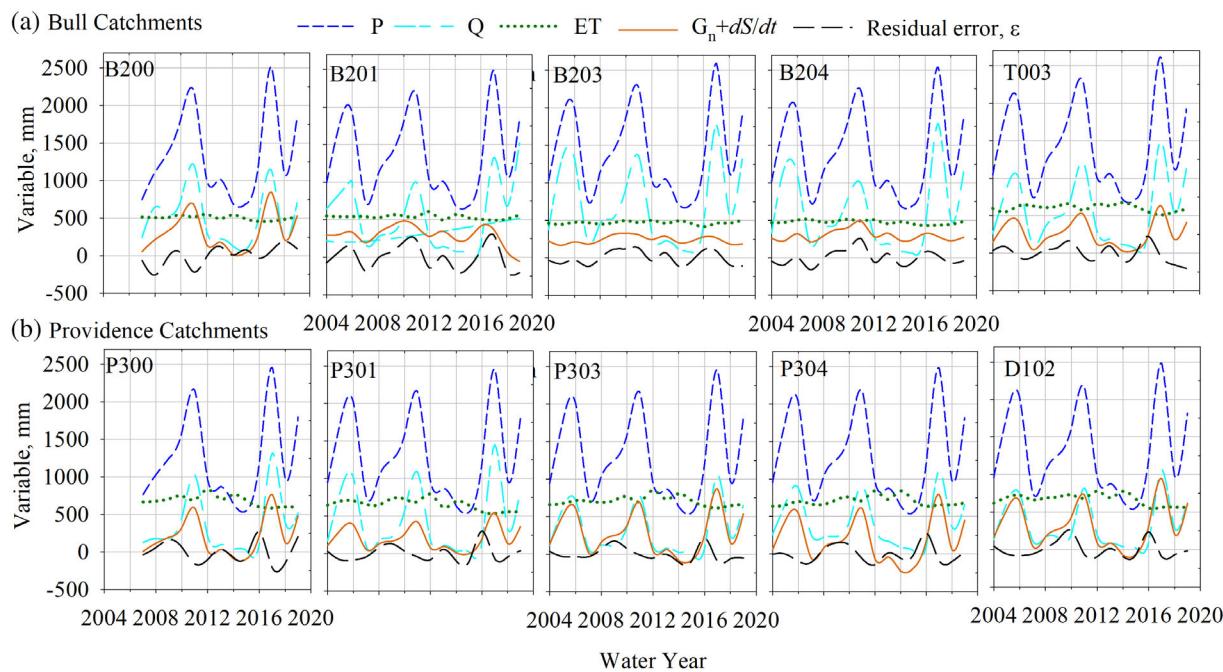


FIGURE 8 Annual water-balance components and residual error for the 10 studied headwater catchments (a) Bull and (b) Providence

1987–1992 drought, ranged from 0 mm in the San Joaquin to –405 mm in the Kaweah. These values are well within the range of multi-year deep soil drying reported earlier (Bales et al., 2018; Goulden & Bales, 2019; Roche et al., 2020). At the headwater scale, only Providence catchments P300, P303, P304, and D102 experienced overdraft (negative $G_n + dS/dt$ values) (Figure 8) with cumulative total values of –182, –166, –516, and –136 mm, respectively. These overdraft values do not concur with the shallow soil moisture measurements discussed earlier nor with deeper soil moisture measurements discussed in Section 4 below, indicating that the drying of catchments was well below the depth of our measurements, there were errors in measurements and scaling, or some combination of the two mechanisms. P304 is a groundwater driven catchment with deeper flow paths (Ackerer et al., 2020; Safeeq & Hunsaker, 2016), so a larger overdraft as reported by Goulden and Bales (2019) and Roche et al. (2020) is quite possible and may point toward limitations in soil moisture measurements.

4 | DISCUSSION AND IMPLICATIONS

We have utilized a rich dataset from the SSCZO and KREW to test water-balance closure assumptions and errors at various spatial and temporal scales. Our water-balance closure errors (i.e., \bar{G}_n and Δ) were within the range of values reported in earlier studies (Bales et al., 2018; Gao et al., 2010; Saksa et al., 2017; Wang et al., 2015; Wang, Huang, et al., 2014; Wang, McKenney, et al., 2014). Gao et al. (2010) reported a 20% water-balance closure error using remotely sensed P , ET and dS/dt for the combined San Joaquin and Sacramento River basins. Wang, Huang, et al. (2014); Wang, McKenney, et al. (2014); Wang et al. (2015) reported water-balance

closure errors between –50% and 25% of P with the majority of the watershed within a closure error of <20%. Also consistent with other studies (Abera et al., 2017; Engeland et al., 2005; Tardif et al., 2015; Wang, McKenney, et al., 2014), our results show that a high proportion of the closure error term can be attributed to the magnitude of P and Q (Figure 6).

While the reported magnitudes of water-balance closure in this study are comparable to those in the literature, differences in the scale and the various water-balance closure assumptions that were tested point to different factors and processes driving non-closure. At the river-basin scale, long-term water-balance closure in the form of \bar{G}_n was well within the inherent measurement biases that are commonly used to explain non-closure, including a reported 5%–25% negative bias in measured and gridded P (Adam & Lettenmaier, 2003; Daly et al., 2017; Groisman & Legates, 1994), 5%–10% (mostly negative) bias in naturalized and measured Q (U.S. Department of the Interior, 1976; Hirsch & Costa, 2004), and ~10% bias in ET (Wang et al., 2015). However, inherent measurement bias was not sufficient to explain the magnitude of Δ or $G_n + dS/dt$ at annual timescales and at the smaller spatial scales of the headwater catchments. Among P , Q , and ET in the water balance equation, P is the largest water flux. On average, annual P is 2.4 times Q and twice as much as ET in the river basins. This difference is even larger in the headwater catchments with P being 2.8 times Q and 2.3 times as much as ET . Hence an unbiased P will likely result in even larger magnitude of Δ or $G_n + dS/dt$.

The results presented here help quantify the spatial and temporal scales at which the water-balance closure assumption may be valid. At the annual timescales, it was not possible to resolve the components of Δ , but Figure 4(a) makes it clear that there was a systematic bias in Δ , due to some combination of G_n , dS/dt , and ε . Therefore, at the annual timescale, even for the larger river basins, the assumption of a

closed water balance should be used with caution. However, as shown in Section 3.2, the precipitation-decorrelation technique can be used to separate the error term ϵ from the combined $G_n + dS/dt$ term, and at longer timescales, dS/dt can be approximated to 0. This allows for a direct estimation of $G_n + dS/dt$ at an annual timescale and \bar{G}_n at timescales longer than 1 year, rather than assuming them to be negligible as is often done to achieve water-balance closure when P , Q , and ET are known. Annual $G_n + dS/dt$ and \bar{G}_n can then be compared to the other fluxes to evaluate the relative influence of these processes in the watershed (Figures 7 and 8). In the river basins, \bar{G}_n was a relatively low percentage of the overall water balance (only 9% of average P) and within the margin of error of the other flux measurements (Kampf et al., 2020). Therefore, over longer periods at this scale, it may be acceptable to close the water balance.

At the headwater-catchment scale, the story is similar to the river-basin scale for short periods. The positive bias in Figure 4(b) again showed that the assumptions of negligible G_n or $G_n + dS_d/dt$ are not valid at the annual timescale. While the net change in shallow storage was relatively small over the annual timescale (0.5 ± -12 mm/year), the magnitude of the water balance remainder was quite large (Table 3). O'Geen et al. (2018) measured deeper (up to 10 m) soil moisture storage changes near the flux tower in P301 (Figure 1) using a neutron probe. The estimated changes in deeper storage from this dataset ranged from a 35 mm increase to a 42 mm depletion during the peak of the drought (2013–2015). While this measurement does not extend to full regolith thickness at the SSCZO (Holbrook et al., 2014), plants are less able to access storage at increasing depths (Klos et al., 2018; O'Geen et al., 2018). Additionally, the plant-available water holding capacity for the weathered granitic bedrock is very small, less than 10%, compared to 20% in the overlying soil (Hubbert et al., 2001; Jones & Graham, 1993). While the role of measurement and scaling uncertainties in soil moisture and other water balance components on the estimated water balance cannot be ignored, the magnitude of water-balance closure error (i.e., Δ) points to additional, much larger, uncounted storage affecting the water balance. At the daily timescale there was much larger variation in cumulative $dS_d/dt + G_n + \epsilon$, which was typically negative or near zero at the beginning of the water year before peaking with snowmelt, demonstrating that even with measurements of P , ET , Q , and dS_d/dt , the water balance at a sub-annual timescale could not be considered closed (Figure 5). The higher mean G_n relative to P in the catchments was greater than the bias in P and ET reported above. However, even at longer timescales, G_n made up a larger proportion of the overall water balance for the catchments (20% of P on average) than for the river basins (Figure 2). Predictions for annual $G_n + dS/dt$ using the statistical precipitation-decorrelation technique compared to measurements of $G_n + dS/dt$ (solving Equation (3) with measurements of P , ET , Q , and dS/dt assuming $\epsilon = 0$) were biased (Table 3), suggesting that the residual error term ϵ cannot be considered negligible. Taken together, the evidence from P300 suggests that at the headwater-catchment scale, even for longer timescales and rich datasets, the assumptions that are required to close the water balance may not be valid. The general trend is that the larger the spatial or temporal scale, the more acceptable the water-balance closure assumptions become.

The error in water-balance closure associated with the assumption of $G_n = 0$ has further implications for understanding critical zone processes and mass balance. From terrestrial storage, mixing, and fluxes (Dralle et al., 2018; Sprenger et al., 2018; Włostowski et al., 2020), hyporheic flow and transport processes (Ackerer et al., 2020; Boano et al., 2014), to streamflow generation (Liu et al., 2013), inferring critical zone processes using water-balance closure without accurate knowledge of G_n can be misleading. As an example, future projections of increased impacts from extreme drought and wildfires in California (Dettinger et al., 2018) have stoked interest in understanding the effects of land management on processes such as streamflow, forest mortality, and evaporation (Bart et al., 2021; Roche et al., 2020). Setting $G_n = 0$ and neglecting changes in terrestrial storage to infer ET (as $P - Q$) or Q (as $P - ET$) could result in an overestimation of the impact of forest thinning or prescribed burns on catchment water yield by assuming that treatment-related decreases in ET directly result in corresponding increases in Q , rather than a more complex combination of changes in ET , subsurface storage, and groundwater fluxes. Furthermore, the subsurface structure of the critical zone has been identified as the biggest knowledge gap in the effort to incorporate hillslope-scale hydrological processes into Earth System Models (Fan et al., 2019), which are crucial tools for predicting the response of the critical zone to global change. Net groundwater exchange G_n is an essential component of the interaction of the subsurface with the aboveground parts of the critical zone, and so failing to accurately account for it also hampers our ability to improve models. Future studies exploring the nature and mechanisms behind the G_n flux out of the river basins and catchments can provide necessary insights on the actual magnitude of subsurface inflow of groundwater to lowland aquifers.

Fan et al. (2019) reviewed evidence for “leaky” watersheds, and found them to be widespread, and that “small catchment size, positioned at either the high or low end of a steep regional topographic and climatic gradient, underlain by deep permeable substrates that extend beyond the study catchment, and in drier climate or dry seasons and droughts” were factors making a watershed more likely to be leaky. The Providence and Bull headwater catchments examined in this study meet all of these criteria, underscoring the need to better characterize deeper flowpaths and quantify the net groundwater exchange (Frisbee et al., 2011). While flumes and weirs are often considered to have high measurement accuracy, they are often not anchored to bedrock. Morphologic changes such as lateral fluvial erosion, sediment deposition, and scour can also cause water to flow around the gauge. Indeed, some of the authors observed these effects firsthand while repairing the P300 weir in the summers of 2018 and 2019. Interestingly, the estimated long-term net groundwater exchange in P300 (207 mm) was on the low end of G_n values among the headwater catchments, suggesting that subsurface flow paths underneath the gauge may be common in the headwater catchments. Catchments may also be leaky at river-basin scales. Güler and Thyne (2006) used geochemical tracers to show that water originating in the Kern could be found in a different river basin on the eastern side of the Sierra Nevada. As a fraction of the overall water balance, this amount is likely to be smaller for river basins than for catchments.

The mechanism and magnitude of large-scale interbasin net groundwater exchange remain unclear (Markovich et al., 2019).

Due to the measurements available through the SSCZO and KREW programs, the analysis of water balance closure in headwater catchments was relatively rigorous. However, the role of biased data in water-balance closure cannot be ignored for larger river basins with complex topography. The in-situ measurement network for P is sparse for large basins, particularly in high-altitude regions. The measurement of precipitation using a sparse network can reduce the accuracy of gridded-precipitation datasets regardless of the methods used. Uncertainty in NDVI-based ET estimates may range between $\pm 10\%$ and 15% or less (Goulden et al., 2012). Regression-based approaches, like NDVI- ET , are not suitable in ecosystems where meteorological conditions and plant phenology play major roles or in areas where evaporation is dominant (Goulden & Bales, 2019).

5 | CONCLUSIONS

This study investigated water-balance closure assumptions over multiple spatial and temporal scales in the southern Sierra Nevada. The study was carried out on six river basins ($>1000 \text{ km}^2$) and 10 headwater catchments ($0.5\text{--}5 \text{ km}^2$) during the years 1985 to 2019 and 2004 to 2019, respectively. We leveraged data collected by the SSCZO and KREW to quantify the various stores and fluxes for calculating the water balance. Our results suggest that the water balance cannot be closed by current measurements without explicitly accounting for deeper storage and groundwater exchange fluxes. We found the net groundwater exchange to be much higher than the typical amount of measurement bias observed in the measurements of precipitation, streamflow, and evapotranspiration. Thus, we argue for greater consideration of groundwater exchange when evaluating and modelling hydrological processes. Long-term water balance closure at the river basin scale can be achieved but temporal scales over which the negligible net groundwater exchange assumption can be enforced must be carefully evaluated.

ACKNOWLEDGEMENTS

This work is supported by INFEWS grant (grant no. 2018-67004-27405) from the USDA National Institute of Food and Agriculture. Funding for the Kings River Experimental Watersheds (KREW) project was provided by the USDA Forest Service, the USDI and USDA Forest Service Joint Fire Science Program, from California's State Water Resources Control Board, through Proposition 50 (Water Security Clean Drinking Water, Coastal, and Beach Protection Act of 2002), the California Department of Forestry and Fire Protection, and the Sierra Resource Conservation District. SSCZO research infrastructure was supported by NSF EAR-0619947, EAR-0725097, EAR-1239521, and EAR-1331939. We thank the Sierra National Forest for their partnership and cooperation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the USFS data archive at <https://doi.org/10.2737/RDS-2018-0028>

and <https://doi.org/10.2737/RDS-2017-0037> or with the original references/web link provided in the manuscript.

ORCID

Mohammad Safeeq  <https://orcid.org/0000-0003-0529-3925>

Ryan R. Bart  <https://orcid.org/0000-0002-7235-8007>

Norman F. Pelak  <https://orcid.org/0000-0002-1069-8217>

Chandan K. Singh  <https://orcid.org/0000-0002-1802-5658>

David N. Dralle  <https://orcid.org/0000-0002-1944-2103>

Peter Hartsough  <https://orcid.org/0000-0001-7888-8028>

Joseph W. Wagenbrenner  <https://orcid.org/0000-0003-3317-5141>

REFERENCES

- Abera, W., Formetta, G., Borga, M., & Rigon, R. (2017). Estimating the water budget components and their variability in a pre-alpine basin with JGrass-NewAGE. *Advances in Water Resources*, 104, 37–54. <https://doi.org/10.1016/j.advwatres.2017.03.010>
- Ackerer, J., Steefel, C., Liu, F., Bart, R., Safeeq, M., O'Geen, A., Hunsaker, C., & Bales, R. (2020). Determining how critical zone structure constrains Hydrogeochemical behavior of watersheds: Learning from an elevation gradient in California's Sierra Nevada. *Frontiers in Water*, 2, 23. <https://doi.org/10.3389/frwa.2020.00023>
- Adam, J. C., & Lettenmaier, D. P. (2003). Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research-Atmospheres*, 108(D9), 4257. <https://doi.org/10.1029/2002JD002499>
- Bales, R. C., Goulden, M. L., Hunsaker, C. T., Conklin, M. H., Hartsough, P. C., O'Geen, A. T., Hopmans, J. W., & Safeeq, M. (2018). Mechanisms controlling the impact of multi-year drought on mountain hydrology. *Scientific Reports*, 8(1), 690. <https://doi.org/10.1038/s41598-017-19007-0>
- Bales, R. C., Hopmans, J. W., O'Geen, A. T., Meadows, M., Hartsough, P. C., Kirchner, P., Hunsaker, C. T., & Beaudette, D. (2011). Soil moisture response to snowmelt and rainfall in a Sierra Nevada mixed-conifer forest. *Vadose Zone Journal*, 10, 786–799. <https://doi.org/10.2136/vzj2011.0001>
- Bart, R. R., Safeeq, M., Wagenbrenner, J. W., & Hunsaker, C. T. (2021). Do fuel treatments decrease forest mortality or increase streamflow? A case study from the Sierra Nevada (USA). *Ecohydrology*, 14(1), e2254. <https://doi.org/10.1002/eco.2254>
- Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, 52(4), 603–679. <https://doi.org/10.1002/2012RG000417>
- Clark, D. H., Gillespie, A. R., Clark, M. M., & Burke, R. (2003). Mountain glaciations of the Sierra Nevada. In D. Easterling (Ed.), *Quaternary geology of the United States: INQUA 2003 field guide volume* (pp. 287–310). Reno: The Desert Research Institute.
- Comer, P., Faber-Langendoen, D., Evans, R., Gawler, S., Josse, C., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., & Snow, K., 2003. Ecological systems of the United States: A working classification of US terrestrial systems. NatureServe, Arlington, VA, 75.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15), 2031–2064. <https://doi.org/10.1002/joc.1688>
- Daly, C., Slater, M. E., Roberti, J. A., Laseter, S. H., & Swift, L. W. (2017). High-resolution precipitation mapping in a mountainous watershed: Ground truth for evaluating uncertainty in a national precipitation dataset. *International Journal of Climatology*, 37(S1), 124–137. <https://doi.org/10.1002/joc.4986>

- Daly, C., Smith, J. I., & Olson, K. V. (2015). Mapping atmospheric moisture climatologies across the conterminous United States. *PLoS One*, 10 (10), e0141140. <https://doi.org/10.1371/journal.pone.0141140>
- Dettinger, M. D., Alpert, H., Battles, J. J., Kusel, J., Safford, H., Fougeres, D., Knight, C., Miller, L., & Sawyer, S. (2018). Sierra Nevada summary report. California's fourth climate change assessment (no. SUM-CCCA4-2018-004). California Energy Commission/Natural Resources Agency.
- Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of California. *Water*, 3(2), 445–478. <https://doi.org/10.3390/w3020445>
- Dralle, D. N., Hahm, W. J., Rempe, D. M., Karst, N. J., Thompson, S. E., & Dietrich, W. E. (2018). Quantification of the seasonal hillslope water storage that does not drive streamflow. *Hydrological Processes*, 32(13), 1978–1992. <https://doi.org/10.1002/hyp.11627>
- Engeland, K., Xu, C.-Y., & Gottschalk, L. (2005). Assessing uncertainties in a conceptual water balance model using Bayesian methodology/estimation bayésienne des incertitudes au sein d'une modélisation conceptuelle de bilan hydrologique. *Hydrological Sciences Journal*, 50(1), 9. <https://doi.org/10.1623/hysj.50.1.45.56334>
- Fan, Y. (2019). Are catchments leaky? *WIREs Water*, 6(6), e1386. <https://doi.org/10.1002/wat2.1386>
- Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., ... Kirchner, J. W. (2019). Hillslope hydrology in global change research and earth system modeling. *Water Resources Research*, 55(2), 1737–1772.
- Fekete, B. M., Vörösmarty, C. J., Roads, J. O., & Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2), 294–304. [https://doi.org/10.1175/1520-0442\(2004\)017<0294:UICPAT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0294:UICPAT>2.0.CO;2)
- Fellows, A. W., & Goulden, M. L. (2017). Mapping and understanding dry season soil water drawdown by California montane vegetation. *Ecohydrology*, 10(1), e1772. <https://doi.org/10.1002/eco.1772>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Flerchinger, G. N., & Cooley, K. R. (2000). A ten-year water balance of a mountainous semi-arid watershed. *Journal of Hydrology*, 237(1), 86–99. [https://doi.org/10.1016/S0022-1694\(00\)00299-7](https://doi.org/10.1016/S0022-1694(00)00299-7)
- Frisbee, M. D., Phillips, F. M., Campbell, A. R., Liu, F., & 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 Resources Research*, 47(6), W06512. <https://doi.org/10.1029/2010WR009391>.
- Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., & Lettenmaier, D. P. (2010). Estimating the water budget of major US river basins via remote sensing. *International Journal of Remote Sensing*, 31(14), 3955–3978. <https://doi.org/10.1080/01431161.2010.483488>
- Gleick, P. H. (1987). The development and testing of a water balance model for climate impact assessment: Modeling the Sacramento Basin. *Water Resources Research*, 23(6), 1049–1061. <https://doi.org/10.1029/WR023i006p01049>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Goulden, M. L., Anderson, R. G., Bales, R. C., Kelly, A. E., Meadows, M., & Winston, G. C. (2012). Evapotranspiration along an elevation gradient in California's Sierra Nevada. *Journal of Geophysical Research - Biogeosciences*, 117(G3), G03028. <https://doi.org/10.1029/2012JG002027>
- Goulden, M. L., & Bales, R. C. (2014). Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion. *Proceedings of the National Academy of Sciences*, 111(39), 14071–14075. <https://doi.org/10.1073/pnas.1319316111>
- Goulden, M. L., & Bales, R. C. (2019). California forest die-off linked to multi-year deep soil drying in 2012–2015 drought. *Nature Geoscience*, 12(8), 632–637. <https://doi.org/10.1038/s41561-019-0388-5>
- Groisman, P. Y., & Legates, D. R. (1994). The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society*, 75 (2), 215–228. [https://doi.org/10.1175/1520-0477\(1994\)075<0215:TAOUP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUP>2.0.CO;2)
- Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2010). Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophysical Research Letters*, 37(20), L20401. <https://doi.org/10.1029/2010GL044696>
- Güler, C., & Thyne, G. D. (2006). Statistical clustering of major solutes: Use as a tracer for evaluating Interbasin groundwater flow into Indian Wells Valley, California. *Environmental and Engineering Geoscience*, 12 (1), 53–65. <https://doi.org/10.2113/12.1.53>
- Hanak, E., Lund, J., Arnold, B., Escriva-Bou, A., Gray, B., Green, S., Harter, T., Howitt, R., MacEwan, D., Medellín-Azuara, J., Moyle, P., & Seavy, N. (2017). Water stress and a changing San Joaquin Valley (p. 50). Public policy Institute of California. Retrieved from <http://www.ppic.org/main/publication.asp?i=1224>
- Hirsch, R. M., & Costa, J. E. (2004). U.S. stream flow measurement and data dissemination improve. *Eos, Transactions American Geophysical Union*, 85(20), 197–203. <https://doi.org/10.1029/2004EO200002>
- Holbrook, W. S., Riebe, C. S., Elwaseif, M., Hayes, J. L., Basler-Reeder, K., Harry, D. L., Malazian, A., Dosseto, A., Hartsough, P. C., & Hopmans, J. W. (2014). Geophysical constraints on deep weathering and water storage potential in the southern sierra critical zone observatory. *Earth Surface Processes and Landforms*, 39(3), 366–380. <https://doi.org/10.1002/esp.3502>
- Hubbert, K. R., Beyers, J. L., & Graham, R. C. (2001). Roles of weathered bedrock and soil in seasonal water relations of *Pinus jeffreyi* and *Arctostaphylos patula*. *Canadian Journal of Forest Research*, 31(11), 1947–1957.
- Hunsaker, C. T., Whitaker, T. W., & Bales, R. C. (2012). Snowmelt runoff and water yield along elevation and temperature gradients in California's southern Sierra Nevada. *JAWRA Journal of the American Water Resources Association*, 48(4), 667–678. <https://doi.org/10.1111/j.1752-1688.2012.00641.x>
- Istanbulluoglu, E., Wang, T., Wright, O. M., & Linters, J. D. (2012). Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. *Water Resources Research*, 48(3), W00H16. <https://doi.org/10.1029/2010WR010100>
- Jennings, C. W., & Gutierrez, C. I. (2010). *2010 geologic map of California*; California department of conservation. California Geological Survey.
- Jones, D. P., & Graham, R. C. (1993). Water-holding characteristics of weathered granitic rock in chaparral and forest ecosystems. *Soil Science Society of America Journal*, 57, 256–261.
- Jump, A. S., Ruiz-Benito, P., Greenwood, S., Allen, C. D., Kitzberger, T., Fensham, R., Martínez-Vilalta, J., & Lloret, F. (2017). Structural overshoot of tree growth with climate variability and the global spectrum of drought-induced forest dieback. *Global Change Biology*, 23(9), 3742–3757.
- Kampf, S. K., Burges, S. J., Hammond, J. C., Bhaskar, A., Covino, T. P., Eurich, A., Harrison, H., Lefsky, M., Martin, C., McGrath, D., Puntenney-Desmond, K., & Willi, K. (2020). The case for an open water balance: Re-envisioning network design and data analysis for a complex, uncertain world. *Water Resources Research*, 56(6), e2019WR026699. <https://doi.org/10.1029/2019WR026699>
- Klos, P. Z., Goulden, M. L., Riebe, C. S., Tague, C. L., O'Geen, A. T., Flinchum, B. A., Safeeq, M., Conklin, M. H., Hart, S. C., Berhe, A. A., Hartsough, P. C., Holbrook, W. S., & Bales, R. C. (2018). Subsurface plant-accessible water in mountain ecosystems with a Mediterranean climate. *WIREs Water*, 5(3), e1277. <https://doi.org/10.1002/wat2.1277>
- Liu, F., Hunsaker, C., & Bales, R. C. (2013). Controls of streamflow generation in small catchments across the snow–rain transition in the

- southern Sierra Nevada, California. *Hydrological Processes*, 27(14), 1959–1972. <https://doi.org/10.1002/hyp.9304>
- Markovich, K. H., Manning, A. H., Condon, L. E., & McIntosh, J. C. (2019). Mountain-block recharge: A review of current understanding. *Water Resources Research*, 55(11), 8278–8304.
- Matthes, F., & Fryxell, F. (1965). Glacial reconnaissance of Sequoia National Park California; characteristics and distribution of the ancient glaciers in the most southerly national park of the Sierra Nevada, U.S. Govt. Print. Off.
- Mazur, K., Schoenheinz, D., Biemelt, D., Schaaf, W., & Grunewald, U. (2011). Observation of hydrological processes and structures in the artificial Chicken Creek catchment. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(1), 74–86. <https://doi.org/10.1016/j.pce.2010.10.001>
- 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. *Journal of Hydrology*, 534, 124–138. <https://doi.org/10.1016/j.jhydrol.2015.12.027>
- O'Geen, A. (T.), Safeeq, M., Wagenbrenner, J., Stacy, E., Hartsough, P., Devine, S., Tian, Z., Ferrell, R., Goulden, M., Hopmans, J. W., & Bales, R. (2018). Southern sierra critical zone observatory and kings river experimental watersheds: A synthesis of measurements, new insights, and future directions. *Vadose Zone Journal*, 17(1), 180081. <https://doi.org/10.2136/vzj2018.04.0081>
- Pan, X., Helgason, W., Ireson, A., & Wheater, H. (2017). Field-scale water balance closure in seasonally frozen conditions. *Hydrology and Earth System Sciences*, 21(11), 5401–5413. <https://doi.org/10.5194/hess-21-5401-2017>
- Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., & Wondzell, S. M. (2009). Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. *Water Resources Research*, 45(11), W11427. <https://doi.org/10.1029/2008WR007644>
- Rice, J. S., & Emanuel, R. E. (2019). Ecohydrology of interannual changes in watershed storage. *Water Resources Research*, 55(10), 8238–8251. <https://doi.org/10.1029/2019WR025164>
- Roche, J. W., Ma, Q., Rungee, J., & Bales, R. C. (2020). Evapotranspiration mapping for Forest Management in California's Sierra Nevada. <https://doi.org/10.3389/ffgc.2020.00069>
- Rundel, P. W. (2011). The diversity and biogeography of the alpine Flora of the Sierra Nevada, California. *Madroño*, 58(3), 153–184. <https://doi.org/10.3120/0024-9637-58.3.153>
- Rungee, J., Bales, R., & Goulden, M. (2019). Evapotranspiration response to multiyear dry periods in the semiarid western United States. *Hydrological Processes*, 33(2), 182–194. <https://doi.org/10.1002/hyp.13322>
- Safeeq, M., & Hunsaker, C. T. (2016). Characterizing runoff and water yield for headwater catchments in the southern Sierra Nevada. *Journal of the American Water Resources Association (JAWRA)*, 52, 1327–1346. <https://doi.org/10.1111/1752-1688.12457>
- Safeeq, M., Shukla, S., Arismendi, I., Grant, G. E., Lewis, S. L., & Nolin, A. (2016). Influence of winter season climate variability on snow-precipitation ratio in the western United States. *International Journal of Climatology*, 36(9), 3175–3190. <https://doi.org/10.1002/joc.4545>
- Saksa, P., Safeeq, M., & Dymond, S. (2017). Recent patterns in climate, vegetation, and forest water use in California montane watersheds. *Forests*, 8(8), 278. <https://doi.org/10.3390/f8080278>
- Scott, R. L., & Biederman, J. A. (2019). Critical zone water balance over 13 years in a semiarid savanna. *Water Resources Research*, 55(1), 574–588. <https://doi.org/10.1029/2018WR023477>
- Sophocleous, M. (2002). Interactions between groundwater and surface water: The state of the science. *Hydrogeology Journal*, 10(1), 52–67. <https://doi.org/10.1007/s10040-001-0170-8>
- Sprenger, M., Tetzlaff, D., Buttle, J., Carey, S. K., McNamara, J. P., Laudon, H., Shatilla, N. J., & Soulsby, C. (2018). Storage, mixing, and fluxes of water in the critical zone across northern environments inferred by stable isotopes of soil water. *Hydrological Processes*, 32(12), 1720–1737. <https://doi.org/10.1002/hyp.13135>
- Tardif, S., St-Hilaire, A., Roy, R., Bernier, M., & Payette, S. (2015). Water budget analysis of small forested boreal watersheds: Comparison of Sphagnum bog, patterned fen and lake dominated downstream areas in the La Grande River region, Québec. *Hydrology Research*, 46(1), 106–120. <https://doi.org/10.2166/nh.2014.219>
- Tonina, D., & Buffington, J. M. (2009). Hyporheic exchange in mountain rivers I: Mechanics and environmental effects. *Geography Compass*, 3 (3), 1063–1086 1063–1086.
- Trabucco, A., & Zomer, R.J. (2018). Global Aridity Index and Potential Evapo-Transpiration (ETO) Climate Database v2. CGIAR Consortium for Spatial Information (CGIAR-CSL). Published online, Retrieved from the CGIAR-CSL GeoPortal at <https://cgciarcsli.community>
- Trask, J. C., Fogg, G. E., & Puente, C. E. (2017). Resolving hydrologic water balances through a novel error analysis approach, with application to the Tahoe basin. *Journal of Hydrology*, 546, 326–340. <https://doi.org/10.1016/j.jhydrol.2016.12.029>
- U.S. Department of the Interior (1976). Surface water supply of the United States, 1966–1970, Part II, Pacific slope basins in California, vol. 4, northern California Valley Basins, U.S. Geol. Surv. Water-Supply Pap., 2131. 747.
- Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D. W., & Sheffield, J. (2014). Assessment of water budget for sixteen large drainage basins in Canada. *Journal of Hydrology*, 512, 1–15. <https://doi.org/10.1016/j.jhydrol.2014.02.058>
- Wang, S., McKenney, D. W., Shang, J., & Li, J. (2014). A national-scale assessment of long-term water budget closures for Canada's watersheds. *Journal of Geophysical Research-Atmospheres*, 119(14), 8712–8725. <https://doi.org/10.1002/2014JD021951>
- Wang, S., Pan, M., Mu, Q., Shi, X., Mao, J., Brümmer, C., Jassal, R. S., Krishnan, P., Li, J., & Black, T. A. (2015). Comparing evapotranspiration from Eddy covariance measurements, water budgets, remote sensing, and land surface models over Canada. *Journal of Hydrometeorology*, 16 (4), 1540–1560. <https://doi.org/10.1175/JHM-D-14-0189.1>
- Warhaftig, C., & Birman, J. H. (1965). The quaternary of the Pacific mountain system in California. In H. E. Wright & D. G. Frey (Eds.), *The quaternary of the United States* (pp. 299–340). New Jersey Princeton University Press.
- Winter, T. C., Harvey, J. W., Franke, O. L., & Alley, W. M. (1998). Ground water and surface water: A single resource (report no. 1139; Circular). USGS Publications Warehouse. <https://doi.org/10.3133/cir1139>
- Wlostowski, N. A., Molotch, N., Anderson, P. S., Brantley, S., Chorover, J., Dralle, D., Kumar, P., Li, L., Lohse, K., Mallard, M. J., McIntosh, C. J., Murphy, F. S., Parrish, E., Safeeq, M., Seyfried, M., Shi, Y., & Harman, C. (2020). Signatures of hydrologic function across the critical zone observatory network. *Water Resources Research*, 57, WRCR24937. <https://doi.org/10.1029/2019WR026635>
- Xu, C. (2001). Statistical analysis of parameters and residuals of a conceptual water balance model – Methodology and case study. *Water Resources Management*, 15(2), 75–92. <https://doi.org/10.1023/A:1012559608269>

How to cite this article: Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., & Wagenbrenner, J. W. (2021). How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. *Hydrological Processes*, 35: e14199. <https://doi.org/10.1002/hyp.14199>