Water Resources Research

COMMENTARY
10.1029/2019WR026699

Key Points:
- Quantifying the watershed-scale water balance remains elusive because all components are uncertain
- Imposing water balance closure can lead to missed opportunities for identifying unknowns in the water balance
- We need more strategic quasi-replicate and nested watershed monitoring to improve water balance understanding

The Case for an Open Water Balance: Re-envisioning Network Design and Data Analysis for a Complex, Uncertain World

Stephanie K. Kampf[1], Stephen J. Burges[2], John C. Hammond[3], Aditi Bhaskar[4], Tim P. Covino[1], Abby Eurich[1], Hannah Harrison[5], Michael Lefsky[1], Caroline Martin[6], Daniel McGrath[7], Kira Puntenney-Desmond[8], and Kathryn Willi[1]

[1]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA, [2]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA, [3]Now at U.S. Geological Survey MD-DE-DC Water Science Center, Maryland, DE, USA, [4]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA, [5]Department of Civil and Environmental Engineering, University of Hawai'i at Manoa, Honolulu, HI, USA, [6]TCarta Marine, LLC, Denver, CO, USA, [7]Natural Resources Ecology Laboratory, Colorado State University, Fort Collins, CO, USA, [8]Department of Geosciences, Colorado State University, Fort Collins, CO, USA

Correspondence to:
S. K. Kampf,
stephanie.kampf@colostate.edu

Citation:
Kampf, S. K., Burges, S. J., Hammond, J. C., Bhaskar, A., Covino, T. P., Eurich, A., et al. (2020). The case for an open water balance: Re-envisioning network design and data analysis for a complex, uncertain world. *Water Resources Research*, *56*, e2019WR026699. https://doi.org/10.1029/2019WR026699

Received 5 NOV 2019
Accepted 25 MAY 2020
Accepted article online 31 MAY 2020

Abstract The discipline of hydrology has long focused on quantifying the water balance, which is frequently used to estimate unknown water fluxes or stores. While technologies for measuring water balance components continue to improve, all components of the balance have substantial uncertainty at the watershed scale. Watershed-scale evapotranspiration, storage, and groundwater import or export are particularly difficult to measure. Given these uncertainties, analyses based on assumed water balance closure are highly sensitive to uncertainty propagation and errors of omission, where unknown components are assumed negligible. This commentary examines how greater insight may be gained in some cases by keeping the water balance open rather than applying methods that impose water balance closure. An open water balance can facilitate identifying where unknowns such as groundwater import/export are affecting watershed-scale streamflow. Strategic improvements in monitoring networks can help reduce uncertainties in observable variables and improve our ability to quantify unknown parts of the water balance. Improvements may include greater spatial overlap between measurements of water balance components through coordination between entities responsible for monitoring precipitation, snow, evapotranspiration, groundwater, and streamflow. Measuring quasi-replicate watersheds can help characterize the range of variability in the water balance, and nested measurements within watersheds can reveal areas of net groundwater import or export. Well-planned monitoring networks can facilitate progress on critical hydrologic questions about how much water becomes evapotranspiration, how groundwater interacts with surface watersheds at varying spatial and temporal scales, how much humans have altered the water cycle, and how streamflow will respond to future climate change.

Plain Language Summary The water balance is a fundamental concept in hydrology that underlies many tools for predicting streamflow, soil moisture, or groundwater availability. It is often expressed as an equation that relates water inputs, outputs, and storage for a watershed. Inputs can be rainfall, snowmelt, or water imports to the watershed. Outputs include water movement into the atmosphere (evaporation, transpiration, and sublimation), streamflow, and water exports through groundwater or human diversions. Water storage can be in snow or ice, surface water bodies, or underground. Each of these water balance components is difficult to measure, and some are rarely measured. Therefore, researchers often simplify the water balance, assuming that difficult to measure quantities, like groundwater imports/exports or changes in water storage, can be neglected. Such simplifying assumptions lead to missed opportunities for discovering where these unknowns in the water balance are important controls on streamflow. This commentary advocates strategically expanding watershed monitoring networks to coordinate monitoring of different water balance components, monitor multiple similar watersheds within each geographic region, and nest monitoring of tributary streams within larger watersheds. This can accelerate progress in understanding groundwater flow, plant water availability, streamflow generation, and human impacts to the water balance.

1. Introduction

"The business of hydrology is to solve the water balance equation" proclaimed J.C.I. Dooge in his keynote lecture at the 1987 international symposium on "Water for the Future" (Klemeš, 1988). The water balance is the hydrologic version of conservation of mass, and its framework underpins many hydrologic analyses and most hydrologic models. More than three decades have passed since Dooge's proclamation, yet we still struggle to solve the water balance. Even in a <10 m long control plot with no deep percolation and carefully measured outflow, the water balance is uncertain (e.g., Kampf & Burges, 2010) because precipitation data have biases that are not correctable (Sieck et al., 2007), evapotranspiration (ET) is difficult to measure, and subsurface components are rarely measured. Extend these uncertainties to entire watersheds with spatially variable topography, vegetation, and subsurface characteristics, and the magnitude of unknown information becomes daunting even for a single research watershed, let alone the hundreds of watersheds now available in large hydrology data sets (Addor et al., 2019). A detail-oriented hydrologist could easily despair at the unlikely prospect of ever quantifying the water balance. Whereas Dooge (1988) hoped that "future availability of vast amounts of observations" could help us make progress in solving the water balance, Klemeš (1988) claimed that "for the scientist, the water balance is one of the most challenging Rubik Cubes of nature."

True to those predictions, the hydrologic research community continues to find new and better ways to convert observable information into water balance components. Radar and microwave patterns have been converted to rainfall rates (Huffman et al., 2007; Kampf et al., 2018; Krajewski & Smith, 2002), snow water equivalent (SWE; Bernier et al., 1999; Tait, 1998), and near-surface soil moisture (Entekhabi et al., 2010; Huisman et al., 2003). Cosmogenic rays (Zreda et al., 2008), electromagnetic induction, and electrical resistivity are used to estimate soil moisture (Robinson et al., 2008). Optical observations of snow cover have been assimilated into hydrologic models to reconstruct SWE (Andreadis & Lettenmaier, 2006; Molotch & Margulis, 2008), and optical and thermal imagery is incorporated into ET computations (Allen et al., 2007; Bastiaanssen et al., 2005; Mu et al., 2007). Changes in gravity detected from satellites have been converted to quantities of water stored on and in the ground (Tapley et al., 2004).

These efforts have led to advances in streamflow forecasting (Werner et al., 2009) and drought monitoring (Su et al., 2017), and they have illustrated the extent to which water extraction is changing groundwater storage (Scanlon et al., 2018). However, even with these advances, the uncertainties in the water balance remain large and should not be ignored. This commentary addresses why quantifying the water balance remains elusive and proposes strategies for collecting the data needed to advance understanding of water partitioning without assuming water balance closure.

2. Quantifying the Water Balance Remains Elusive

We begin with a basic annual water balance equation:

$$Q = P - Su - E - T \pm G \pm A \pm \frac{dS}{dt}, \tag{1}$$

where Q is the surface streamflow; P is precipitation; Su is sublimation; E is evaporation; T is transpiration; G is net input (+) of groundwater that originated outside the surface watershed boundaries or net output (−) of groundwater below the surface watershed boundaries; A is net human-constructed input or output across the watershed divides, including surface water, groundwater, and snow; and dS/dt is the change in storage with respect to time, combining all zones of storage across the watershed including SWE, glaciers, surface water, and subsurface water (Figure 1). All components are in units of volume per time or depth per time if normalized by the watershed area.

The following is a brief review of how well we can measure each water balance component. We focus on the water balance for domains the size of zero-order watersheds or larger, meaning watersheds large enough that they have a surface channel that can export water out of the domain.

2.1. Streamflow

Surface stream discharge (Q) is typically measured at a fixed location for a given watershed and integrates the water fluxes and storage dynamics of upstream contributing areas. Under careful monitoring

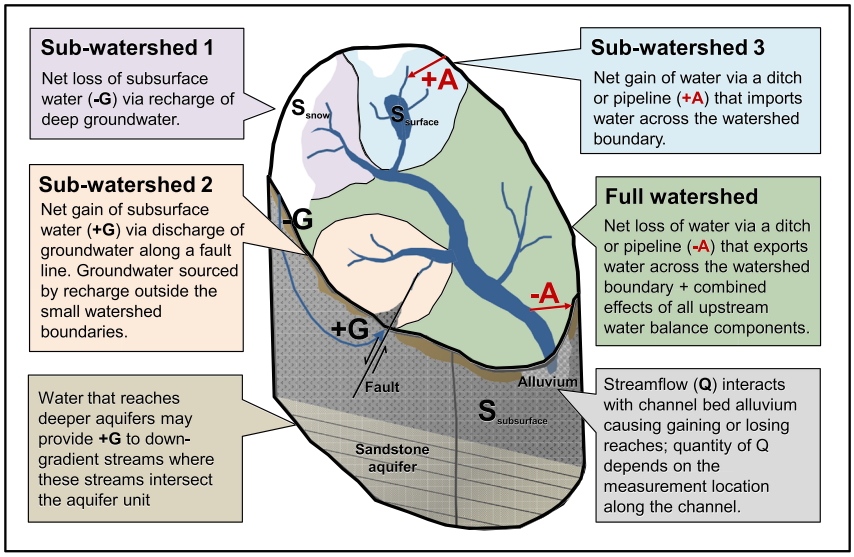

Figure 1. Conceptual diagram of a 3-D watershed and subwatersheds. Each watershed experiences spatially and temporally variable input of precipitation (P) and output via evaporation (E), transpiration (T), and sublimation (Su). Some watersheds have net import or export of subsurface water (G) or anthropogenic rerouting of water (A) across watershed divides. Storage for a given watershed can consist of snow or ice (S_{snow}), surface water ($S_{surface}$), or subsurface water ($S_{subsurface}$).

protocols with a well-developed rating curve, discharge measurements can be accurate to within 5–10% (Hirsch & Costa, 2004), but in practice most gages have >10% uncertainty (Coxon et al., 2015). High flows are often not possible to measure directly due to safety, whereas low flows may be difficult to measure with velocity meters (McMillan et al., 2012). The magnitude of uncertainty varies with location, flow level, and measurement technique (Coxon et al., 2015; Kiang et al., 2018; Rantz et al., 1982a, 1982b; Turnipseed & Sauer, 2010). Uncertainty at a given site combines the uncertainties of stage and velocity sensors (Sorensen & Butcher, 2011; Soupir et al., 2009; Tillery et al., 2001) with those of rating curves (McMillan et al., 2012). Extrapolating rating curves beyond the range of measured low- and high-flow discharge introduces additional uncertainty. For example, 30% of published flow values for reference stations in Australia are based in part on extrapolated rating curves (McMahon & Peel, 2019). Control sections such as weirs or flumes help improve discharge measurements (U.S. Department of the Interior, Bureau of Reclamation, 1997), but these can be difficult to size for watersheds with large variability in discharge (Ogden et al., 2013) and may not be a practical choice in streams where the water cannot easily be routed entirely into the control section. Weirs and flumes are cost-prohibitive for many studies, especially those that are short-term, and obtaining permission to install these structures may not be possible in protected areas.

Not all watersheds have streams that are conducive to accurate discharge measurements. For example, discharge measurements through a natural cross section may be infeasible where channel sediment transport is high, channel cross-section geometry changes, and in wide estuarine or braided reaches of rivers. Consequently, discharge monitoring tends to be concentrated in perennial streams with confined control sections and limited ranges of flow. This leaves many other types of streams under-represented by monitoring networks (Kiang et al., 2013).

2.2. Precipitation

Precipitation (P) is the most widely measured water balance component, but these measurements are usually biased low because of wind-induced undercatch. Biases typically range from 5–25% for rainfall (Groisman & Legates, 1994) and frequently exceed 50% for snow (Goodison & Yang, 1998; Rasmussen et al., 2012). A global-scale analysis estimated that measured precipitation values need to be increased by an average of 12% to remove biases (Adam & Lettenmaier, 2003). Techniques have been proposed for correcting precipitation measurement bias (e.g., Groisman et al., 1991; Sevruk, 1979), but a universally applicable correction method has not been identified because of the wide range of factors that affect precipitation measurements (Sieck et al., 2007).

Quantifying watershed-scale precipitation requires spatial precipitation information, which may be derived through statistical approaches using precipitation gage data (e.g., PRISM; Daly et al., 1994; Di Luzio et al., 2008), reanalysis products that combine climate models with observations via data assimilation (e.g., Kalnay et al., 1996), radar data (e.g., Wüest et al., 2010), and satellite-derived products (e.g., Hou et al., 2014). Most spatial precipitation products rely on ground-based precipitation measurements for either model input or bias correction, so their accuracy is tied both to gage accuracy and to the accuracy of the spatial mapping method. In mountain regions, gage networks may be sparse and biased toward lower elevations that are easily accessed, making it difficult to interpolate orographic changes in precipitation accurately. Consequently, estimates of precipitation for many mountain ranges across the globe may require large precipitation bias corrections (Beck et al., 2020; Wrzesien et al., 2018).

2.3. Sublimation, Evaporation, Transpiration, and Interception

Partitioning water into the atmosphere via sublimation (Su), evaporation (E), and transpiration (T) is an ongoing challenge, and fewer direct measurements of these fluxes are available compared to P and Q. Techniques for computing open water or reference crop ET from the Penman or Penman-Monteith equations are well established (Allen et al., 1998, 2005). Reference crop ET is for a well-watered specific vegetation type maintained at a fixed height; it is often used as a substitute for potential evapotranspiration (PET), which is the quantity of ET possible for a given set of atmospheric and land cover conditions if water is not limiting.

For the water balance we need actual ET (AET) rather than PET. AET can be measured with lysimeters that precisely track the mass balance of an enclosed container; these measurements are often considered the most accurate sources of AET rates. However, they are limited to relatively small domains, which may not be representative of the surrounding watershed's land cover or drainage characteristics (Nolz, 2016; Ruth et al., 2018). Changes in above and below-ground biomass as well as condensation of water on the lysimeter surface can cause errors in mass balance calculations (Gebler et al., 2015). Lysimeters have been applied most widely for crops (e.g., Lage et al., 2003; Sahin et al., 2007), with less data available for other types of vegetation. Some investigators have used large lysimeters or whole-tree photometers to measure tree transpiration (Knight et al., 1981; Sperling et al., 2012), but these rely on cut trees, which may not transpire in the same way as intact trees (Wullschleger et al., 1998).

For intact trees, measurements of sap flux (e.g., Granier, 1987) can be converted to transpiration using empirical equations that vary due to species type, wood properties, natural thermal gradients, sensor positioning, variability in sap flux density within the tree, inactive xylem, tree wounding, or other factors (Bush et al., 2010; Flo et al., 2019; McCulloh et al., 2007; Paudel et al., 2013; Sperling et al., 2012; Steppe et al., 2010; Sun et al., 2012; Wiedemann et al., 2016). Variability in sap flux within individual trees and between multiple trees leads to uncertainties in scaling these measurements to compute ET across larger areas (Ford et al., 2007).

For larger domains both AET and sublimation can be measured with the eddy covariance technique, which relies on measurements of water vapor and atmospheric turbulence. The method assumes turbulent air flow over uniform horizontal terrain, so in many settings corrections are required such as detrending, coordinate rotation, noise filtering, and removal of time periods when the assumptions of the method are violated (Lee et al., 2004). Eddy covariance measurements integrate over a larger domain than a lysimeter or set of sap flow measurements, but the contributing footprint of the measurement (plan view area measured) does not remain the same as wind conditions change (Schmid, 1997), making the representative spatial scale difficult to quantify (Kljun et al., 2015).

Sublimation has received less attention in the literature than ET, but it has been measured using eddy covariance (Molotch et al., 2007), aerodynamic profile (Hood et al., 1999), Bowen Ratio (Sexstone et al., 2016), snow evaporation pans (West, 1962), and artificial or natural trees on weighing devices (Lundberg & Halldin, 1994; Montesi et al., 2004).

Quantities of sublimation and evaporation are affected by interception, which modifies the susceptibility of snow (or rain) to sublimate (or evaporate). Interactions of precipitation and vegetation were widely studied in the 19th century (Friesen & Van Stan, 2019), inspiring comprehensive work by Horton (1919) on this topic. Measurements of rain interception include comparisons of rain gages or troughs above and below

the canopy (e.g., Holder, 2004; McJannet et al., 2007), mechanical displacement measurements of trees (Friesen et al., 2015), large weighing lysimeters or tree-weighing devices (Storck et al., 2002), or remote sensing techniques to infer snow storage (Lundberg & Halldin, 2001). Quantities of interception measured have varied substantially from <5% to >70% of annual precipitation (Levia & Frost, 2006) because interception varies with storm characteristics, number of and time between storms, wind speed and directions, air temperature, humidity, and land cover type (Crockford & Richardson, 2000; Horton, 1919). Consequently, it is difficult to estimate interception accurately over space and time, and uncertainties in this quantity propagate into estimates of E, T, and Su (Savenije, 2004).

2.4. Groundwater

Some watersheds are net exporters of groundwater (G), while others are net importers (Toth, 1963). Net export means that groundwater flow exports water beyond the surface watershed boundary, whereas net import means that groundwater brings in water that originated outside the watershed boundary. Estimated magnitudes of import and export from modeling studies can be substantial: 0–90% of mean annual precipitation exported and 2–50% imported, with wet, high elevation, and/or small watersheds more likely to export and lower lying basins more likely to import (Fan, 2019; McMahon et al., 2011; Philip, 1998; Schaller & Fan, 2009; Wilson & Guan, 2004). Many locations also have substantial seasonal variability in groundwater recharge (Jasechko et al., 2014), which affects the timing of G. Researchers often estimate G from some form of a water balance equation, making values subject to the uncertainties in other water balance components. Other techniques include using groundwater well observations linked with models (Ball et al., 2014); geochemical analyses that reveal information about the depth or composition of rock types along groundwater flow paths (Frisbee et al., 2017); isotopic tracers to identify water that may have originated outside the local area (Eastoe & Rodney, 2014; Scanlon et al., 2002); noble gas concentrations that allow estimating recharge temperature (Manning & Solomon, 2003); and spectral analyses of precipitation and streamflow time series (Shun & Duffy, 1999).

2.5. Storage

Water can be stored above, at, or beneath the surface of a watershed in either solid (snow/ice) or liquid phases. For the water balance equation 1. we need to quantify changes in storage over time; transient surface water storage may vary at short time scales (hours-days), whereas some glacier and groundwater storage changes may be detectable only at century or longer time scales (Jansson et al., 2003). Snow storage as SWE can be measured via in situ sensors, such as snow pillows at snow telemetry (SNOTEL) stations (Serreze et al., 1999), numerical models that assimilate surface observations (e.g., SNODAS; NOHRSC, 2004), and satellite-based approaches, such as radar backscatter observations (e.g., Lievens, 2019). Annual glacier mass balance, which corresponds to changes in storage on annual time scales, can be measured using in situ stake networks (e.g., O'Neel et al., 2019) or repeat geodetic observations from airborne or satellite platforms (e.g. Larsen et al., 2015; Menounos et al., 2019).

Storage of surface water within watersheds typically has both permanent (on decadal to century scales) and transient (on seasonal and shorter scales) components, with the relative proportion of these components varying in space and time (Pekel et al., 2016). The spatial extent of surface water can be mapped from optical satellite sensors (e.g., Landsat and Sentinel-2) (McFeeters, 1996), but determining the volume of surface storage is much more difficult, requiring an empirical or geostatistical approach to relate surface area to volume (Messager et al., 2016). NASA's GRACE and GRACE-FO time-varying satellite gravity observations provide valuable constraints on total water storage but are limited to large (>200,000 km^2) river basins (Reager et al., 2014). Hybrid approaches that assimilate multimission satellite observations (i.e., time-varying gravity, optical, microwave, and altimetry) and in situ observations with land surface models can help constrain surface water storage at finer temporal and spatial resolution (e.g., Kumar et al., 2019). In addition to surface water bodies, vegetation can also store water in tissue reserves and mature sapwood (Waring & Running, 1976), and this stored water can provide a water reserve for transpiration, accounting for up to 25% of daily water use (Köcher et al., 2013; Phillips et al., 2003).

The quantity of subsurface storage within a watershed includes both near-surface soil moisture and deeper water in regolith and bedrock. Soil moisture in the shallow subsurface can be measured with a variety of sensors (Robinson et al., 2008), which typically capture small (<1 m) length scales. Exceptions are cosmic ray

neutron (Zreda et al., 2008) and geophysical methods, which measure 10^2 m length scales (Robinson et al., 2008). Methods have been developed to retrieve soil moisture from remote sensing (e.g., review in Petropoulos et al., 2015), but these only capture the near-surface soil and vegetation water content (~0–5 cm depth), which is not necessarily a good indicator of storage in the deeper subsurface (Kim et al., 2016). At small watersheds with dense networks of soil moisture sensors and groundwater wells, watershed-scale subsurface storage can be estimated through spatial interpolation of these measurements (e.g., Western et al., 1999). Alternatively, researchers have inferred watershed-scale storage from mean transit time distributions of tracers in streams (Soulsby et al., 2009) or hydrograph recession analysis (Brutsaert & Nieber, 1977; Wittenberg & Sivapalan, 1999). Hydrograph-based approaches are best suited for small wet watersheds with perennial streamflow, where storage-discharge relationships are consistent over time. Tracer methods require that source waters have substantial differences in chemical composition, and the time-varying chemistry of source waters is well quantified (Andermann et al., 2012; Brauer et al., 2013). Stream-derived storage values usually differ from those derived from soil moisture and groundwater measurements because not all subsurface water storage is dynamically connected to the stream (Dralle et al., 2018; Lazo et al., 2019). Consequently, methods for quantifying storage can differ by up to an order of magnitude (McNamara et al., 2011; Staudinger et al., 2017).

2.6. Anthropogenic imports and exports

Although humans have altered every watershed on the planet in some direct or indirect way, common conceptualizations of the water balance exclude anthropogenic effects (Abbott et al., 2019). This focus stems from the emphasis of hydrologic research on "natural" or "reference" watersheds that have been minimally altered by human activities. However, such watersheds are scarce, and in the United States only 16% of gaged watersheds are classified as "reference" (Kiang et al., 2013). Direct water balance modifications include a wide array of water imports and exports from the watershed. Imports originate as water diversions from other watersheds and can include irrigation and associated return flows, artificial recharge, treated wastewater effluent (Townsend-Small et al., 2013), and leakage from canals, ditches, water distribution, and wastewater collection systems (Garcia-Fresca et al., 2005). Water exports can be through both intentional withdrawal of groundwater or surface water and unintentional export through leaky water infrastructure (Bhaskar & Welty, 2012). These exports can easily become a dominant term in the water balance and eclipse the outflow of water through streamflow (Sabo et al., 2010). In the United States, quantities of trans-basin imports and exports are often not available in public databases (Addor et al., 2019) and thus are commonly overlooked components of watershed analyses.

3. Recommended Strategies for Water Balance Research

The previous section demonstrated the substantial uncertainties in each component of the water balance: Q is the only "watershed-scale" measurement, and even its uncertainty is typically >10%. Other components (P, Su, E, T, G, dS/dt, and A) have substantial spatial variability within watersheds, meaning that a single measurement of each is typically not sufficient to represent the watershed-scale value. In this section we examine how uncertainties propagate through water balance analyses and present ideas for analysis and monitoring strategies that may improve water balance understanding while acknowledging these uncertainties.

3.1. Water Balance Uncertainties

3.1.1. Uncertainty Propagation in a Closed Water Balance

Imposing water balance closure means solving for one of the unknown components of Equation 1 using measurements of the other components. A potential problem with imposing closure is that it is particularly sensitive to uncertainties. P and Q are the only water balance components with widespread measurements, leaving six unknowns: E, T, Su, G, dS/dt, and A. To get around this, researchers often group E, T, and Su together into AET, assume G is negligible, assume dS/dt is negligible at annual or longer time scales, and focus on "reference" watersheds with no known A.

An example of this strategy is the Budyko framework, a water balance-based empirical method that has become an analytical tool of choice for comparing hydrologic processes between watersheds. This framework involves solving for mean annual or annual AET as P-Q, then comparing water balance components between watersheds along two axes: AET/P and PET/P. The Budyko (1974) curve is a consensus curve

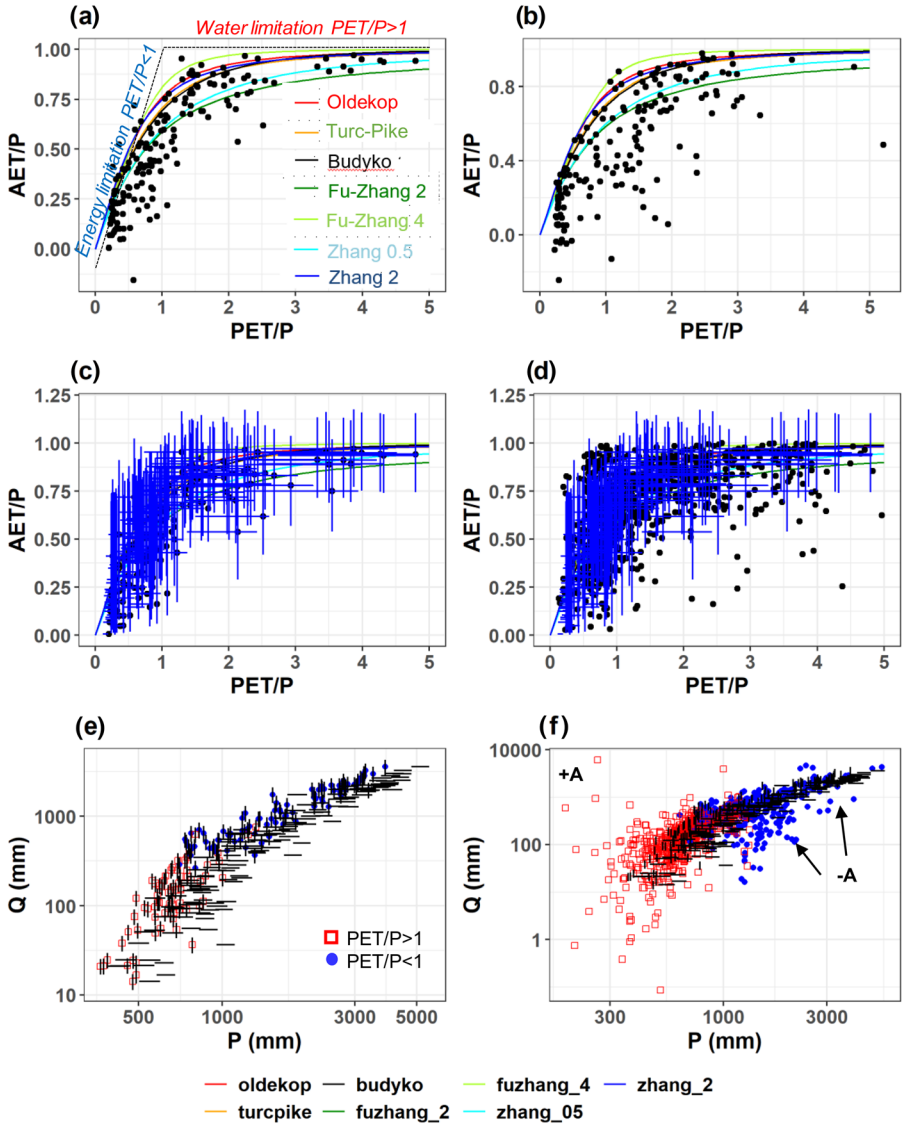

Figure 2. (a) Budyko analysis for mean annual climate variables in western U.S. reference watersheds (points; Falcone, 2011) using mean annual precipitation (*P*) from PRISM and mean annual potential evapotranspiration (PET) estimated using reference ET from gridMET; (b) same as in (a) except with *P* and PET from CAMELS (Addor et al., 2017); (c) same as in (a) but with uncertainty; blue bar ranges span from the minimum to maximum axis values from Monte Carlo analysis; (d) same as in (a) for western U.S. nonreference watersheds and uncertainty bounds overlain for reference watersheds. Lines in (a)–(d) indicate different versions of the Budyko equation using the equations compiled by McMahon et al. (2013): Oldekop (1911); Turc (1954) and Pike (1964); Budyko (1974); Fu (1981); and Zhang et al. (2004, 2001) for two different curve parameter values. (e) *Q* versus *P* for the same reference watershed data set and uncertainty analysis in (c) with black bars indicating uncertainty bounds; uncertainty bounds are offset horizontally from the data points because uncertainty analysis always assumed that *P* is higher than measured; (f) *Q* versus *P* for the same nonreference watershed data set as in (d) with uncertainty bounds for reference watersheds.

that is intermediate between other previously proposed versions (Andréassian et al., 2016; Schreiber, 1904; Oldekop, 1911; Turc, 1954). Budyko (1974) proposed that in dry conditions AET/P tends toward 1 because of limited streamflow generation (*Q*), whereas in wet conditions, AET could become limited by net radiation (Brutsaert, 1982). Thus, the bounding dashed lines outside the curve in Figure 2a represent the limits of energy (diagonal) and water (horizontal) for AET. The shape of the curve itself has no theoretical basis, and many different curve shapes have been applied in prior studies (Gerrits et al., 2009; Greve et al., 2015; Zhou et al., 2015). Curves differ from one another by up to 20% at the highest values of PET/*P* (McMahon et al., 2013; Figure 2a). Differences in curves or the anomalies from these curves have been attributed to runoff processes (Potter et al., 2005), soil variability (Wang et al., 2009), topography (Yokoo et al., 2008),

vegetation (Li et al., 2013), and amount of precipitation that is snow (Barnhart et al., 2016; Berghuijs et al., 2014).

To examine how uncertainties propagate through a closed water balance, we applied a Budyko analysis to a data set of 161 reference watersheds, which are part of the GAGES II data set (Falcone, 2011) in the western United States. Each watershed selected is <500 km^2; we selected this size limit to enable a large sample of watersheds but minimize the extent of climate variability within each watershed (Hammond & Kampf, 2020). The reference watersheds represent the least disturbed sites in the gaging network with minimal flow alteration through groundwater pumping, land use change, dams, diversions, and other human impacts. Most of the watersheds (71%) are dominated by forest land cover; 14% have mainly shrublands, and the remainder have mixed land covers. The average extent of developed land is <2%. For each watershed we compiled data from 2000–2015; we selected this time period because it corresponds with the availability of MODIS satellite data, which will be used in subsequent analyses (section 3.2). We computed the mean annual P using PRISM (Daly, 2013) and mean annual Q from USGS gaging stations. Prior Budyko applications have used free water surface or pan evaporation (Berghuijs et al., 2014; Gentine et al., 2012) to represent PET; here we used grass reference ET from gridMET (Abatzoglou, 2013) because it is available as a gridded data set. Note that open water, pan, and grass reference ET are all simplifications of PET because none account for the spatial variability in vegetation. We computed AET as P-Q and plot these data relative to a set of Budyko curves (Figure 2a). With this data set, most of the watersheds plot on the right side of the curve, possibly because the time period analyzed was relatively dry for parts of this region (Williams et al., 2020).

To examine how data source impacts this Budyko analysis, we also tested the CAMELS data set (Addor et al., 2017) for the same set of watersheds. CAMELS is a compilation of hydro-meteorological data for reference watersheds in the United States (Newman et al., 2015). It has similar watershed average P to the data set we developed, but it differs substantially in PET (−126 to 42%; mean −12%). GridMET PET values were computed using the Penman-Monteith equation for reference ET, whereas values in CAMELS were computed using the Priestley-Taylor equation. With the CAMELS data set, even more of the watershed points fall on the right side of the curves, and differences from Figure 2a are mostly due to the different PET values (Figure 2b).

In this example, the scatter of the watersheds relative the Budyko-type curves is consistent with the range of scatter for studies that have included steep and snowy watersheds (e.g., Berghuijs et al., 2014). The scatter would likely be reduced if we had used a longer time period of analysis and excluded steep watersheds that have substantial internal gradients in climate (Gentine et al., 2012). Here our focus is not on interpreting the cause of the scatter relative to the Budkyo-type curves but rather to examine how uncertainty in the input data affects the analysis. To do this, we applied a Monte Carlo approach. We assigned uncertainty ranges of ±20% to PET and Q and allowed for the possibility of G import or export at rates ranging from ±20% of P. Because P is usually biased low, we assigned increases in P from +5% to +25% and increased the probability of undercatch linearly with mean annual snow persistence (Hammond et al., 2017) to reflect the greater undercatch for snow compared to rain. For each of these hypothetical uncertainty ranges, we selected from a uniform distribution of values and computed 1,000 iterations of the Monte Carlo analysis. We excluded dS/dt because of the mean annual time scale and excluded A because the watersheds are reference watersheds. The resulting uncertainty ranges are substantial (Figure 2c) and extend beyond the full range of possible Budyko-type curves. In part this is because P is on both axes of the Budyko diagram, so uncertainties in its value affect both the x- and y-axis positions of points; having P on both axes also makes this a classic case of spurious correlation (Benson, 1965).

This example of uncertainty propagation illustrates how interpretations based on anomalies from the Budyko-type curve may vary substantially with the magnitudes of neglected G, biased values in P, and uncertainties in PET and Q. Although the assumption of $dS/dt = 0$ may be reasonable as the mean over many years, prior research demonstrates that it is an important consideration at the annual time scale (Istanbulluoglu et al., 2012; Rice & Emanuel, 2019), and it may also be non-negligible under nonstationary climate conditions. The influence of dS/dt may also vary with the definition of a water year; typically, researchers assume the same water year start and end dates for all locations (1 October to 30 September, Northern Hemisphere), which is appropriate for watersheds in which all snow has melted, and watershed

Table 1
Number of Stations for Water Balance Measurements and Their Position Relative to USGS Gaged Watersheds <500 km^2 in the Western United States

Station type	Number	% stations within gaged watersheds	% of gaged watersheds containing station
SNOTEL SWE	806	48	18
GHCN P	17,982	16	51
Ameriflux ET	128	24	2

conditions are dry in October. However, in locations with glaciers or where storage is actively changing in early October, this water year date range may not be appropriate (e.g., Huss & Hock, 2018; Jost et al., 2012; Robles et al., 2017).

In summary, assuming a closed water balance to compute AET in a Budyko analysis introduces propagating uncertainties from P and Q and potential errors of omission from neglecting G, dS/dt, and A. Similar types of uncertainty propagation or errors of omission are possible with any model that assumes a closed water balance.

3.1.2. Uncertainty Propagation in an Open Water Balance

Keeping the balance open means using observed variables directly rather than deriving variables from water balance assumptions (Andréassian & Perrin, 2012). This reduces uncertainty propagation because uncertainties are confined to one variable at a time. An example is a plot of Q versus P (Figure 2e). This type of plot can still convey information about the energy and water limitations that bracket Budyko-type curves by varying symbology for points that represent energy-limited (PET/P < 1) and water-limited (PET/P > 1) conditions.

As an example of how open water balance analysis can be applied, we compared non-reference watersheds that have known anthropogenic modifications, to the reference watersheds analyzed in Figures 2a–2c. When the non-reference watersheds are plotted in Budyko space (Figure 2d), 96% fall within the uncertainty bounds identified for the reference watersheds, meaning the anthropogenic influence may not be distinguishable from the other water balance uncertainties. In contrast, when these non-reference watersheds are plotted in Q versus P space, 41% of the watersheds plot outside the uncertainty bounds of P and Q (Figure 2f). This yields greater confidence in attributing anomalous Q values to anthropogenic influences. Watersheds plotting in the upper left quadrant in the Budyko plot (Figure 2d) and those plotting below the dominant Q versus P trend in Figure 2f are locations where Q is lower than expected. Examples watersheds highlighted with "−A" all have upstream water diversions where the water does not return to the stream above the gage. Watersheds that plot below the uncertainty range in the Budyko plot (Figure 2d) and above the dominant Q versus P trend (Figure 2f) are those with higher Q than expected. For example, the three watersheds in the upper left quadrant of Figure 2f all have higher Q due to added irrigation water (+A).

3.2. Sampling Strategies

3.2.1. Coordinate Monitoring of Water Balance Components

Reducing water balance uncertainties requires strategic monitoring plans. Current large hydrology data sets are not all well designed for water balance analysis. For example, in the United States, different agencies have been responsible for measurements of water balance components: the National Oceanic and Atmospheric Administration (NOAA) for precipitation; the Natural Resource Conservation Service (NRCS) for snow; a network of researchers monitoring ET through Ameriflux; and the U.S. Geological Survey (USGS) for Q and G. Although the USGS has been expanding into precipitation monitoring, these differences in agencies responsible for gage networks often mean that station locations are not distributed well to compute P, ET, and G across watersheds gaged for Q. For the western U.S. area analyzed in Figure 2, we examined the extent of overlap between gaged watersheds (area <500 km^2), point measurements of P from the Global Historical Climatology Network (GHCN), SWE from the Snow Telemetry (SNOTEL) network, and ET monitoring stations from the Ameriflux network (Table 1). Less than half of the P, SWE, and ET stations are placed within gaged watersheds, and many of the gaged watersheds do not have P, SWE, or ET stations within them. Of the gaged watersheds that do have stations within them, most have station densities <0.03 km^{-2}.

Next we examined whether the spatial distribution of gaging stations captures the range of conditions present in the western United States by comparing the distribution of elevation, snow persistence, and PET/P over all watersheds to those of gaged watersheds, SNOTEL, and GHCN stations (Figure 3). For the all-watershed distributions, we used the HUC-12 watersheds delineated by the USGS, which is most consistent in size with drainage areas <500 km^2. The majority of HUC-12 watersheds have intermediate elevations, with a mean at around 1,500 m. In contrast, gaged headwater watersheds are more uniformly distributed across elevations, meaning greater watershed sampling in low-elevation coastal areas and

10.1029/2019WR026699

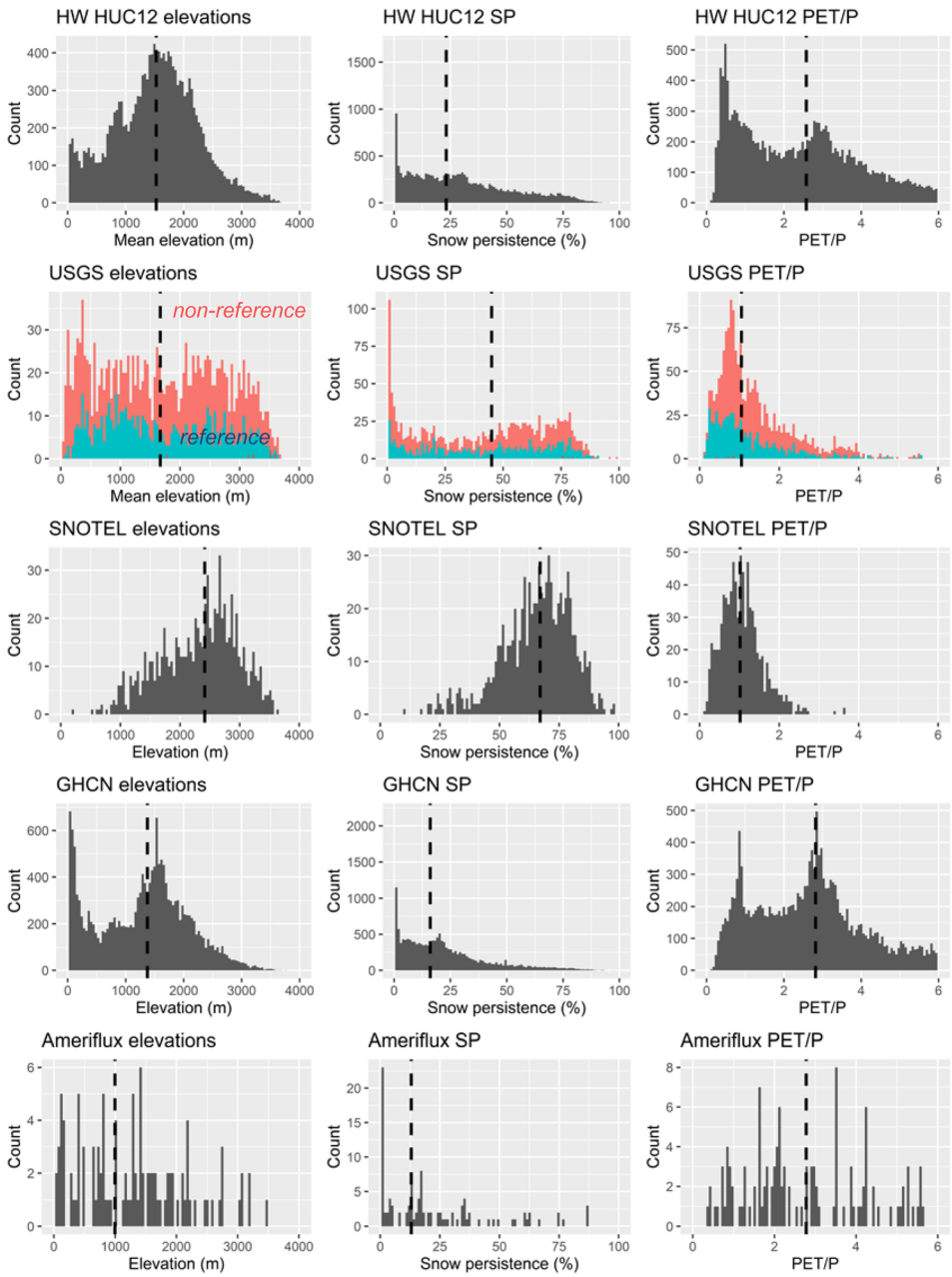

Figure 3. Histograms of mean elevation, snow persistence (Hammond et al., 2017), and PET/P (gridMET and PRISM) for USGS HUC12 headwater watersheds, USGS <500 km^2 gaged reference and non-reference headwater watersheds (Falcone, 2011), NRCS SNOTEL stations for measuring SWE, GHCN weather stations for measuring *P*, and Ameriflux stations for measuring ET.

high-elevation mountains relative to the intermediate elevations that make up the majority of the region. Gaged watersheds also have greater representation in high snow areas and in wetter areas (low PET/*P*) relative to the conditions across the region as a whole. SNOTEL stations measuring SWE are concentrated in high-elevation, high snow persistence, and low-PET/*P* locations, whereas GHCN stations measuring *P* are more common at lower elevations. Both GHCN and Ameriflux stations capture some of the more arid conditions present in the region, but Ameriflux undersamples the high snow persistence areas.

In summary, improving water balance analysis requires coordinated monitoring of water balance components. However, for the example area evaluated, *P*, SWE, and ET measurements are commonly not within watersheds gaged for *Q*, and most of the monitoring networks do not sample the full range of climate conditions present in the region.

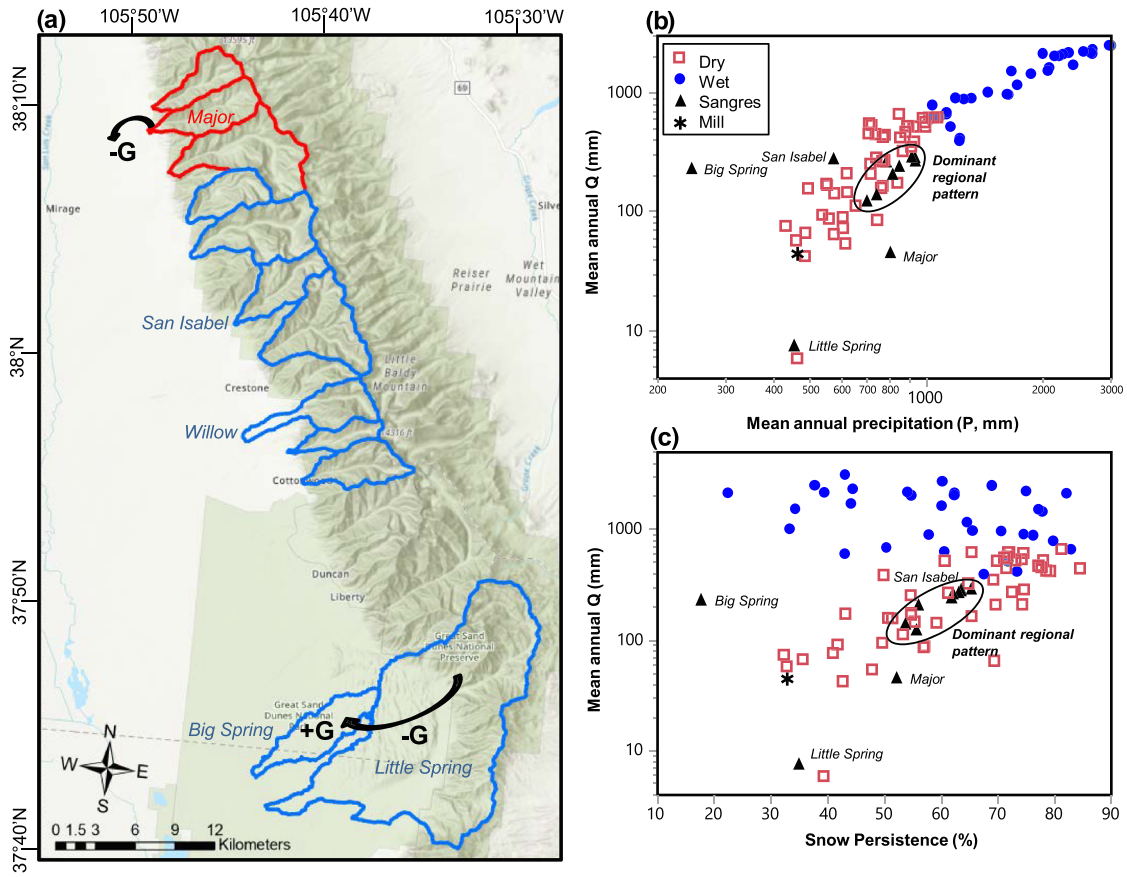

Figure 4. Watersheds draining the Sangre de Cristo Mountains in Southern Colorado (a). Streams draining the mountain front lose water as they enter the alluvial fill in the valley bottom ($-G$); some of this water emerges at valley bottom springs ($+G$). Watersheds with red outlines have small ditches diverting water from the stream, whereas those with blue lines do not have diversions. Right plots are (b) mean annual streamflow (Q) versus mean annual precipitation (P, PRISM) and (c) snow persistence for western U.S. reference watersheds (Figure 2) compared to the watersheds in the map and to Mill Creek in northern Colorado (Figure 5). Dry = PET/P > 1; wet = PET/P < 1.

3.2.2. Improve Sampling Strategies

Strategic sampling design can help identify unknown water balance components. One useful strategy for this purpose is replication. Each watershed in a data set like the one used for Figure 2 has its own unique set of characteristics, with no single watershed replicated. A lack of replicates inhibits us from knowing if the watersheds included are representative of their surroundings. Although true "replicate" watersheds likely do not exist (Dooge, 1986), in headwater areas it is usually possible to find quasi-replicate watersheds with comparable size, aspect, slope, and vegetation (van Loon et al., 2019). Comparing water balance components between quasi-replicate watersheds allows us to begin determining which patterns are characteristic of the region as a whole and which are unique to individual watersheds.

An example of quasi-replicate watershed sampling is illustrated in the series of watersheds in Figure 4, which are monitored by the Colorado Division of Water Resources. These watersheds drain the west side of the Sangre de Cristo Mountains in southern Colorado and have drainage areas ranging from 9–180 km^2. The streams enter thick alluvial fill at the base of the mountain, where they lose water to the subsurface. An exception is Big Spring Creek, which originates in the valley fill. Compared to the reference watershed data set from Figure 2, most of these watersheds produce less Q for a given amount of P, indicating that the characteristic pattern of these headwater watersheds is low streamflow generation (Figure 4b), possibly due to export of G. Several of the streams have quantities of Q that diverge from the regional pattern. For one of these watersheds, San Isabel Creek, the high Q anomaly evident in the Q versus P plot is no longer present in a Q versus snow persistence plot (Figure 4c), indicating that greater snow persistence may have caused the anomalously high streamflow. Two of the anomalous watersheds, Major Creek and Little

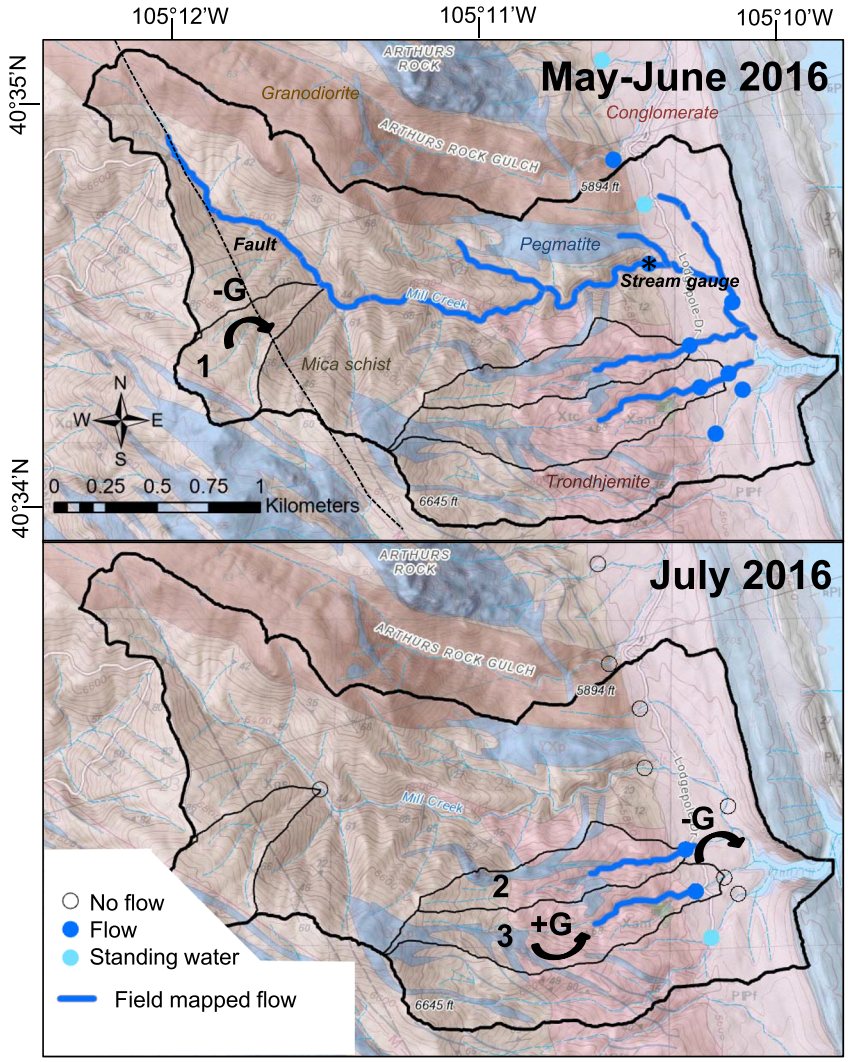

Figure 5. Mill Creek in the northern Colorado Front Range of the Rocky Mountains showing mapped locations of streamflow (Martin, 2018) and point observations of flow presence/absence. Underlying geologic map (Braddock et al., 1989) has major rock unit types labeled in italics on the upper panel. Numbers indicate subwatersheds described in text, and *G* indicates locations of possible groundwater import or export.

Spring Creek, have anomalously low *Q*, likely because they are measured further down into the valley fill than other creeks, which leads greater net export of *G*. In contrast, Big Spring Creek has anomalously high streamflow. The Great Sand Dunes separate this surface watershed from the Sangre de Cristo Mountains; the anomalously high *Q* is likely caused by imported *G* that originated higher up in the mountains and flowed through the subsurface of the sand dunes before reaching Big Spring. Quasi-replicate monitoring of this series of watersheds helps identify which of the watersheds are more representative of the region as a whole and which have unique water balance components such as import or export of *G*.

Another useful sampling strategy for water balance analysis is nesting, which helps identify how water moves between water balance components from headwaters to larger drainage areas. An example of nested sampling is shown in Figure 5, which illustrates small watersheds nested within Mill Creek, a 2.4 km^2 watershed in the northern Colorado Front Range. The watershed is semiarid, with mean annual precipitation from 450–470 mm and intermittent flow that typically lasts from winter until June. In the plots of *Q* versus *P* and *Q* versus snow persistence, Mill Creek plots within the range of other dry watersheds (Figure 4). Visual monitoring of flow presence/absence was conducted both through field mapping along the

channel (flow lines, Figure 5) and at points along trails. These visual observations indicated that flow in the main channel originated along a fault line (Martin, 2018), and flow in the southern tributaries originated near contacts between bedrock types. After the main channel dried in July, the two southern tributaries continued to flow. These tributaries drain trondhjemite bedrock, which is less fractured than the mica schist that underlies most of the rest of the watershed. The streams rapidly lost their flow when they passed from the trondhjemite into the conglomerate downstream ($-G$). These nested observations along the channel network demonstrate how the monitoring location can affect the quantity of Q. Although mean annual P and PET are similar throughout this small watershed, no Q is generated from Small Subwatershed 1 (Figure 5), which may export G that will emerge further downstream. In contrast, Small Watersheds 2 and 3, which have more sustained streamflow, may be receiving imported G that recharged outside the watershed boundaries. These types of water movement into and out of the subsurface are found in intermittent stream systems throughout the world (Beaufort et al., 2018; Godsey & Kirchner, 2014; Lovill et al., 2018; Whiting & Godsey, 2016) as well as in wetter mountain watersheds (Carroll et al., 2019; Cochand et al., 2019; Payn et al., 2009), and they are likely a relatively common feature of streamflow networks.

In summary, to facilitate water balance analyses, we can employ quasi-replicate and nested sampling to help separate dominant climate-driven water balance patterns from scale-varying storage and groundwater patterns. This approach is different from focusing on locations where the water balance is well represented by climatological frameworks such as Budyko, but it could be complementary to such analyses. The open water balance framework can facilitate identifying where surface-subsurface exchanges occur and how these scale-dependent water flow paths deviate from the climatological water balance.

4. Conclusions

No single component of the water balance can be quantified at watershed scale without substantial uncertainty. Approaches for analyzing the water balance should acknowledge these uncertainties and consider how they propagate through water balance calculations. In some cases keeping the water balance open rather than imposing closure can enable new insights about unknown components such as deep groundwater recharge. To reduce water balance uncertainties, we should invest more in nesting P, ET, and G measurements within watersheds gaged for Q. We can expand our monitoring networks strategically to enable advances in understanding how water moves through watersheds and between the surface and subsurface at multiple spatial and temporal scales. Such sampling should consider how well current gaging networks represent the range of climate, geology, vegetation, anthropogenic, and other features that affect the water balance. Quasi-replicate and nested sampling of watersheds can aid in determining individual watersheds compare to regional patterns and the climatological water balance and how water balance components change with the scale of observations. Well-designed monitoring networks will help us separate signals from noise and more reliably address key unknowns such as how large groundwater flow paths interact with surface streams, time scales of storage variability, sensitivity of ET to disturbance, and magnitudes of anthropogenic influences on the water balance.

Data Availability Statement

Data are available through Hammond et al. (2020).

Acknowledgments
Portions of this work have been supported by NSF EAR 1446870 and NASA NNX17AF96A. We are grateful for input from the Associate Editor and reviewers, whose comments helped improve the content and focus of this manuscript.

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*, 121–131. https://doi.org/10.1002/joc.3413

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., et al. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, *12*(7), 533–540. https://doi.org/10.1038/s41561-019-0374-y

Adam, J. C., & Lettenmaier, D. P. (2003). Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research*, *108*, 4257. https://doi.org/10.1029/2002JD002499

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2019). Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, *65*, 1–14. https://doi.org/10.1080/02626667.2019.1683182

Addor, N., Newman, A., Mizukami, M., & Clark, M. P. (2017). *Catchment attributes for large-sample studies*. Boulder, CO: UCAR/NCAR. https://doi.org/10.5065/D6G73C3Q

Allen, R.G., Pereira, L.S., Raes, D., & Smith, M. (1998). FAO penman-Monteith equation. Ch. 2 in crop evapotranspiration—Guidelines for computing crop water requirements. FAO irrigation and drainage paper 56. Rome.

Allen, R. G., Tasumi, M., & Trezza, R. (2007). Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model. *Journal of Irrigation and Drainage Engineering*, *133*(4), 380–394. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380)

Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., & Snyder, R.L. (2005). The ASCE standardized reference evapotranspiration equation. Technical Committee on Standardization of Reference Evapotranspiration.

Andermann, C., Longuevergne, L., Bonnet, S., Crave, A., Davy, P., & Gloaguen, R. (2012). Impact of transient groundwater storage on the discharge of Himalayan rivers. *Nature Geoscience*, *5*(2), 127–132. https://doi.org/10.1038/ngeo1356

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*(6), 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004

Andréassian, V., Mander, Ü., & Pae, T. (2016). The Budyko hypothesis before Budyko: The hydrological legacy of Evald Oldekop. *Journal of Hydrology*, *535*, 386–391.

Andréassian, V., & Perrin, C. (2012). On the ambiguous interpretation of the Turc-Budyko nondimensional graph. *Water Resources Research*, *48*, W10601. https://doi.org/10.1029/2012WR012532

Ball, L. B., Caine, J. S., & Ge, S. (2014). Controls on groundwater flow in a semiarid folded and faulted intermountain basin. *Water Resources Research*, *50*, 6788–6809. https://doi.org/10.1002/2013WR014451

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., & Schneider, D. (2016). Snowmelt rate dictates streamflow. *Geophysical Research Letters*, *43*, 8006–8016. https://doi.org/10.1002/2016GL069690

Bastiaanssen, W. G. M., Noordman, E. J. M., Pelgrum, H., Davids, G., Thoreson, B. P., & Allen, R. G. (2005). SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering*, *131*(1), 85–93. https://doi.org/10.1061/(ASCE)0733-9437(2005)131:1(85)

Beaufort, A., Lamouroux, N., Pella, H., Datry, T., & Sauquet, E. (2018). Extrapolating regional probability of drying of headwater streams using discrete observations and gauging networks. *Hydrology and Earth System Sciences*, *22*(5), 3033–3051. https://doi.org/10.5194/hess-22-3033-2018

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., et al. (2020). Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *Journal of Climate*, *33*(4), 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Benson, M. A. (1965). Spurious correlation in hydraulics and hydrology. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, *91*(HY4), 35–42.

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, *4*(7), 583–586. https://doi.org/10.1038/nclimate2246

Bernier, M., Fortin, J. P., Gauthier, Y., Gauthier, R., Roy, R., & Vincent, P. (1999). Determination of snow water equivalent using RADARSAT SAR data in eastern Canada. *Hydrological Processes*, *13*(18), 3041–3051. https://doi.org/10.1002/(SICI)1099-1085(19991230)13:18<3041::AID-HYP14>3.0.CO;2-E

Bhaskar, A. S., & Welty, C. (2012). Water balances along an urban-to-rural gradient of metropolitan Baltimore, 2001–2009. *Environmental & Engineering Geoscience*, *18*(1), 37–50. https://doi.org/10.2113/gseegeosci.18.1.37

Braddock, W. A., Calvert, R. H., O'Connor, J. T., & Swann, G. A. (1989). *Geologic map of the Horsetooth Reservoir Quadrangle*. Larimer County, CO: (No. 1625) U.S. Geological Survey.

Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F., & Uijlenhoet, R. (2013). Investigating storage-discharge relations in a lowland catchment using hydrograph fitting, recession analysis, and soil moisture data. *Water Resources Research*, *49*, 4257–4264. https://doi.org/10.1002/wrcr.20320

Brutsaert, W. (1982). *Evaporation into the atmosphere*, (pp. 241–243). Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-017-1497-6

Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, *13*(3), 637–643. https://doi.org/10.1029/WR013i003p00637

Budyko, M. I. (Ed) (1974). *Climate and life, translated from Russian by D. H. Miller*. New York: Elsevier.

Bush, S. E., Hultine, K. R., Sperry, J. S., & Ehleringer, J. R. (2010). Calibration of thermal dissipation sap flow probes for ring-and diffuse-porous trees. *Tree Physiology*, *30*(12), 1545–1554. https://doi.org/10.1093/treephys/tpq096

Carroll, R. W., Deems, J. S., Niswonger, R., Schumer, R., & Williams, K. H. (2019). The importance of interflow to groundwater recharge in a snowmelt-dominated headwater basin. *Geophysical Research Letters*, *46*, 5899–5908. https://doi.org/10.1029/2019GL082447

Cochand, M., Christe, P., Ornstein, P., & Hunkeler, D. (2019). Groundwater storage in high alpine catchments and its contribution to streamflow. *Water Resources Research*, *55*, 2613–2630. https://doi.org/10.1029/2018WR022989

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *51*, 5531–5546. https://doi.org/10.1002/2014WR016532

Crockford, R. H., & Richardson, D. P. (2000). Partitioning of rainfall into throughfall, stemflow and interception: Effect of forest type, ground cover and climate. *Hydrological Processes*, *14*(16–17), 2903–2920. https://doi.org/10.1002/1099-1085(200011/12)14:16/17<2903::AID-HYP126>3.0.CO;2-6

Daly, C. (2013). *Descriptions of PRISM spatial climate datasets for the conterminous United States (PRISM Doc., 14 p.)*. Corvallis, OR: PRISM Climate Group, Oregon State University.

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K., & Arnold, J. G. (2008). Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology*, *47*(2), 475–497. https://doi.org/10.1175/2007JAMC1356.1

Dooge, J. C. (1986). Looking for hydrologic laws. *Water Resources Research*, *22*(9S), 46S–58S. https://doi.org/10.1029/WR022i09Sp0046S

Dooge, J. C. (1988). Hydrology in perspective. *Hydrological Sciences Journal*, *33*(1), 61–85. https://doi.org/10.1080/02626668809491223

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

Eastoe, C., & Rodney, R. (2014). Isotopes as tracers of water origin in and near a regional carbonate aquifer: The southern Sacramento Mountains, New Mexico. *Water*, *6*(2), 301–323. https://doi.org/10.3390/w6020301

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Falcone, J. A. (2011). *GAGES-II: Geospatial attributes of gages for evaluating streamflow (Digit. Spat. Data set)*. Reston, VA: U.S. GeologicalSurvey.

Fan, Y. (2019). Are catchments leaky? *Wiley Interdisciplinary Reviews: Water*, *6*(6), e1386.

Flo, V., Martinez-Vilalta, J., Steppe, K., Schuldt, B., & Poyatos, R. (2019). A synthesis of bias and uncertainty in sap flow methods. *Agricultural and Forest Meteorology*, *271*, 362–374.

Ford, C. R., Hubbard, R. M., Kloeppel, B. D., & Vose, J. M. (2007). A comparison of sap flux-based evapotranspiration estimates with catchment-scale water balance. *Agricultural and Forest Meteorology*, *145*(3–4), 176–185. https://doi.org/10.1016/j.agrformet.2007.04.010

Friesen, J., Lundquist, J., & Van Stan, J. T. (2015). Evolution of forest precipitation water storage measurement methods. *Hydrological Processes*, *29*(11), 2504–2520. https://doi.org/10.1002/hyp.10376

Friesen, J., & Van Stan, J. T. (2019). Early european observations of precipitation partitioning by vegetation: A synthesis and evaluation of 19th century findings. *Geosciences*, *9*(10), 423. https://doi.org/10.3390/geosciences9100423

Frisbee, M. D., Tolley, D. G., & Wilson, J. L. (2017). Field estimates of groundwater circulation depths in two mountainous watersheds in the western US and the effect of deep circulation on solute concentrations in streamflow. *Water Resources Research*, *53*, 2693–2715. https://doi.org/10.1002/2016WR019553

Fu, B. P. (1981). On the calculation of the evaporation from land surface (in Chinese). *Scientia Atmospherica Sinica*, *5*, 23–31.

Garcia-Fresca, B., Sharp, J. M., & Jr. (2005). Hydrogeologic considerations of urban development: Urban-induced recharge. *Reviews in Engineering Geology*, *16*, 123–136.

Gebler, S., Hendricks Franssen, H. J., Pütz, T., Post, H., Schmidt, M., & Vereecken, H. (2015). Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and tipping bucket. *Hydrology and Earth System Sciences*, *19*(5), 2145–2161. https://doi.org/10.5194/hess-19-2145-2015

Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G., & Salvucci, G. (2012). Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. *Geophysical Research Letters*, *39*, L19404. https://doi.org/10.1029/2012GL053492

Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., & Pfister, L. (2009). Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model. *Water Resources Research*, *45*, W04403. https://doi.org/10.1029/2008WR007308

Godsey, S. E., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*(23), 5791–5803.

Goodison, B.E., Louie, P.Y.T, & Yang, D. (1998) WMO solid precipitation measurement intercomparison. World Meteorological Organization. Instruments and observing methods report No. 67.

Granier, A. (1987). Evaluation of transpiration in a Douglas-fir stand by means of sap flow measurements. *Tree Physiology*, *3*(4), 309–320.

Greve, P., Gudmundsson, L., Orlowsky, B., & Seneviratne, S. I. (2015). Introducing a probabilistic Budyko framework. *Geophysical Research Letters*, *42*, 2261–2269. https://doi.org/10.1002/2015GL063449

Groisman, P. Y., Koknaeva, V. V., Belokrylova, T. A., & Karl, T. R. (1991). Overcoming biases of precipitation measurement: A history of the USSR experience. *Bulletin of the American Meteorological Society*, *72*(11), 1725–1733. https://doi.org/10.1175/1520-0477(1991)072<1725:OBOPMA>2.0.CO;2

Groisman, P. Y., & Legates, D. R. (1994). The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society*, *75*(2), 215–227. https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2

Hammond, J., S. Kampf, A. Eurich (2020). Mean annual climate and watershed properties for small USGS and CDWR watersheds in the western U.S., HydroShare, https://doi.org/10.4211/hs.8c7dd37435b34b2e915c48b08504a267

Hammond, J. C., Kampf, S. K. (2020). Sub-annual streamflow responses to rainfall and snowmelt inputs in snow-dominated watersheds of the western U.S. *Water Resources Research*, *56*, e2019WR026132. https://doi.org/10.1029/2019WR026132

Hammond, J. C., Saavedra, F. A., & Kampf, S. K. (2017).MODIS MOD10A2 derived snow persistence and no data index for the western U.S. Cambridge, MA: Hydroshare. Retrieved from https://www.hydroshare.org/resource/1c62269aa802467688d25540caf2467e/

Hirsch, R. M., & Costa, J. E. (2004). US stream flow measurement and data dissemination improve. *Eos, Transactions American Geophysical Union*, *85*(20), 197–203. https://doi.org/10.1029/2004EO200002

Holder, C. D. (2004). Rainfall interception and fog precipitation in a tropical montane cloud forest of Guatemala. *Forest Ecology and Management*, *190*(2–3), 373–384.

Hood, E., Williams, M., & Cline, D. (1999). Sublimation from a seasonal snowpack at a continental, mid-latitude alpine site. *Hydrological Processes*, *13*(12–13), 1781–1797. https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1781::AID-HYP860>3.0.CO;2-C

Horton, R. E. (1919). Rainfall interception. *Monthly Weather Review*, *47*(9), 603–623. https://doi.org/10.1175/1520-0493(1919)47<603:RI>2.0.CO;2

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Huisman, J. A., Hubbard, S. S., Redman, J. D., & Annan, A. P. (2003). Measuring soil water content with ground penetrating radar: A review. *Vadose Zone Journal*, *2*(4), 476–491. https://doi.org/10.2136/vzj2003.4760

Huss, M., & Hock, R. (2018). Global-scale hydrological response to future glacier mass loss. *Nature Climate Change*, *8*(2), 135–140.

Istanbulluoglu, E., Wang, T., Wright, O. M., & Lenters, 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*, W00H16. https://doi.org/10.1029/2010WR010100

Jansson, P., Hock, R., & Schneider, T. (2003). The concept of glacier storage: A review. *Journal of Hydrology*, *282*, 116–129.

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., et al. (2014). The pronounced seasonality of global groundwater recharge. *Water Resources Research*, *50*, 8845–8867. https://doi.org/10.1002/2014WR015809

Jost, G., Moore, R. D., Menounos, B., & Wheate, R. (2012). Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada. *Hydrology and Earth System Sciences*, *16*(3), 849–860. https://doi.org/10.5194/hess-16-849-2012

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kampf, S. K., & Burges, S. J. (2010). Quantifying the water balance in a planar hillslope plot: Effects of measurement errors on flow prediction. *Journal of Hydrology*, *380*(1–2), 191–202. https://doi.org/10.1016/j.jhydrol.2009.10.036

Kampf, S. K., Faulconer, J., Shaw, J. R., Lefsky, M., Wagenbrenner, J. W., & Cooper, D. J. (2018). Rainfall thresholds for flow generation in desert ephemeral streams. *Water Resources Research*, *54*, 9935–9950. https://doi.org/10.1029/2018WR023714

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., le Coz, J., Westerberg, I. K., et al. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, *54*, 7149–7176. https://doi.org/10.1029/2018WR022708

Kiang, J. E., Stewart, D. W., Archfield, S. A., Osborne, E. B., & Eng, K. (2013). A national streamflow network gap analysis: U.S. Geological Survey Scientific Investigations Report 2013–5013, 79 p. plus one appendix as a separate file, http://pubs.usgs.gov/sir/2013/5013/.

Kim, J., Dwelle, M. C., Kampf, S. K., Fatichi, S., & Ivanov, V. Y. (2016). On the non-uniqueness of the hydro-geomorphic responses in a zero-order catchment with respect to soil moisture. *Advances in Water Resources*, *92*, 73–89. https://doi.org/10.1016/j.advwatres.2016.03.019

Klemeš, V. (1988). A hydrological perspective. *Journal of Hydrology*, *100*(1–3), 3–28. https://doi.org/10.1016/0022-1694(88)90179-5

Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2015). A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). *Geoscientific Model Development*, *8*(11), 3695–3713. https://doi.org/10.5194/gmd-8-3695-2015

Knight, D. H., Fahey, T. J., Running, S. W., Harrison, A. T., & Wallace, L. L. (1981). Transpiration from 100-yr-old Lodgepole pine forests estimated with whole-tree Potometers. *Ecology*, *62*(3), 717–726.

Köcher, P., Horna, V., & Leuschner, C. (2013). Stem water storage in five coexisting temperate broad-leaved tree species: Significance, temporal dynamics and dependence on tree functional traits. *Tree Physiology*, *33*(8), 817–832. https://doi.org/10.1093/treephys/tpt055

Krajewski, W. F., & Smith, J. A. (2002). Radar hydrology: Rainfall estimation. *Advances in Water Resources*, *25*(8–12), 1387–1394. https://doi.org/10.1016/S0309-1708(02)00062-3

Kumar, S. V., Jasinski, M., Mocko, D. M., Rodell, M., Borak, J., Li, B., et al. (2019). NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. *Journal of Hydrometeorology*, *20*(8), 1571–1593. https://doi.org/10.1175/JHM-D-17-0125.1

Lage, M., Bamouh, A., Karrou, M., & El Mourid, M. (2003). Estimation of rice evapotranspiration using a microlysimeter technique and comparison with FAO Penman-Monteith and Pan evaporation methods under Moroccan conditions. *Agronomie*, *23*(7), 625–631. https://doi.org/10.1051/agro:2003040

Larsen, C. F., Burgess, E., Arendt, A. A., O'neel, S., Johnson, A. J., & Kienholz, C. (2015). Surface melt dominates Alaska glacier mass balance. *Geophysical Research Letters*, *42*, 5902–5908. https://doi.org/10.1002/2015GL064349

Lazo, P. X., Mosquera, G. M., McDonnell, J. J., & Crespo, P. (2019). The role of vegetation, soils, and precipitation on water storage and hydrological services in Andean Páramo catchments. *Journal of Hydrology*, *572*, 805–819. https://doi.org/10.1016/j.jhydrol.2019.03.050

Lee, X., Massman, W., & Law, B. (Eds) (2004). *Handbook of micrometeorology: A guide for surface flux measurement and analysis* (Vol. 29). Dordrecht: Kluwer Academic Publishers.

Levia, D. F. Jr., & Frost, E. E. (2006). Variability of throughfall volume and solute inputs in wooded ecosystems. *Progress in Physical Geography*, *30*(5), 605–632. https://doi.org/10.1177/0309133306071145

Li, D., Pan, M., Cong, Z., Zhang, L., & Wood, E. (2013). Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, *49*, 969–976. https://doi.org/10.1002/wrcr.20107

Lievens, H. and 16 others. (2019). Snow depth variability in the Northern Hemisphere mountains observed from space. *Nature Communications*, *10*, 1, 4629, https://doi.org/10.1038/s41467-019-12566-y.

Loon, A. F. V., Rangecroft, S., Coxon, G., Breña Naranjo, J. A., Ogtrop, F. V., & Van Lanen, H. A. (2019). Using paired catchments to quantify the human influence on hydrological droughts. *Hydrology and Earth System Sciences*, *23*(3), 1725–1739. https://doi.org/10.5194/hess-23-1725-2019

Lovill, S. M., Hahm, W. J., & Dietrich, W. E. (2018). Drainage from the critical zone: Lithologic controls on the persistence and spatial extent of wetted channels during the summer dry season. *Water Resources Research*, *54*, 5702–5726. https://doi.org/10.1029/2017WR021903

Lundberg, A., & Halldin, S. (1994). Evaporation of intercepted snow: Analysis of governing factors. *Water Resources Research*, *30*(9), 2587–2598. https://doi.org/10.1029/94WR00873

Lundberg, A., & Halldin, S. (2001). Snow interception evaporation. Review of measurement techniques, processes, and models. *Theoretical and Applied Climatology*, *70*(1–4), 117–133. https://doi.org/10.1007/s007040170010

Manning, A. H., & Solomon, D. K. (2003). Using noble gases to investigate mountain-front recharge. *Journal of Hydrology*, *275*(3–4), 194–207. https://doi.org/10.1016/S0022-1694(03)00043-X

Martin, C. (2018). Spatial and temporal variability in channel surface flow across an elevation gradient on the Colorado Front Range. M.S. Thesis. Colorado State University.

McCulloh, K. A., Winter, K., Meinzer, F. C., Garcia, M., Aranda, J., & Lachenbruch, B. (2007). A comparison of daily water use estimates derived from constant-heat sap-flow probe values and gravimetric measurements in pot-grown saplings. *Tree Physiology*, *27*(9), 1355–1360. https://doi.org/10.1093/treephys/27.9.1355

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McJannet, D., Wallace, J., & Reddell, P. (2007). Precipitation interception in Australian tropical rainforests: I. Measurement of stemflow, throughfall and cloud interception. *Hydrological Processes*, *21*(13), 1692–1702. https://doi.org/10.1002/hyp.6347

McMahon, P. B., Plummer, L. N., Böhlke, J. K., Shapiro, S. D., & Hinkle, S. R. (2011). A comparison of recharge rates in aquifers of the United States based on groundwater-age data. *Hydrogeology Journal*, *19*(4), 779.

McMahon, T. A., Laaha, G., Parajka, J., Peel, M. C., Savenije, H. G., Sivapalan, M., et al. (2013). In G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, & H. Savenije (Eds.), *Prediction of annual runoff in ungauged basins. Ch. 5 in runoff prediction in ungauged basins: Synthesis across processes, places, and scales* (pp. 70–101). Cambridge, UK: Cambridge University Press.

McMahon, T. A., & Peel, M. C. (2019). Uncertainty in stage-discharge rating curves: Application to Australian Hydrologic Reference Stations data. *Hydrological Sciences Journal*, *64*(3), 255–275.

McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, *26*(26), 4078–4111. https://doi.org/10.1002/hyp.9384

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., et al. (2011). Storage as a metric of catchment comparison. *Hydrological Processes*, *25*(21), 3364–3371. https://doi.org/10.1002/hyp.8113

Menounos, B., Hugonnet, R., Shean, D., Gardner, A., Howat, I., Berthier, E., et al. (2019). Heterogeneous changes in western North American glaciers linked to decadal variability in zonal wind strength. *Geophysical Research Letters*, *46*, 200–209. https://doi.org/10.1029/2018GL080942

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 7. https://doi.org/10.1038/ncomms13603

Molotch, N. P., Blanken, P. D., Williams, M. W., Turnipseed, A. A., Monson, R. K., & Margulis, S. A. (2007). Estimating sublimation of intercepted and sub-canopy snow using eddy covariance systems. *Hydrological Processes: An International Journal*, *21*(12), 1567–1575. https://doi.org/10.1002/hyp.6719

Molotch, N. P., & Margulis, S. A. (2008). Estimating the distribution of snow water equivalent using remotely sensed snow cover data and a spatially distributed snowmelt model: A multi-resolution, multi-sensor comparison. *Advances in Water Resources*, *31*(11), 1503–1514. https://doi.org/10.1016/j.advwatres.2008.07.017

Montesi, J., Elder, K., Schmidt, R. A., & Davis, R. E. (2004). Sublimation of intercepted snow within a subalpine forest canopy at two elevations. *Journal of Hydrometeorology*, *5*(5), 763–773. https://doi.org/10.1175/1525-7541(2004)005<0763:SOISWA>2.0.CO;2

Mu, Q., Heinsch, F. A., Zhao, M., & Running, S. W. (2007). Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, *111*(4), 519–536.

National Operational Hydrologic Remote Sensing Center (2004). *Snow Data Assimilation System (SNODAS) data products at NSIDC (10/01/2003–09/30/2015)*. Boulder, CO: National Snow and Ice Data Center.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Nolz, R. (2016). A review on the quantification of soil water balance components as a basis for agricultural water management with a focus on weighing lysimeters and soil water sensors/Ein Überblick über die Ermittlung von Wasserhaushaltsgrößen als Basis für die landeskulturelle Wasserwirtschaft mit Fokus auf Lysimeter und Bodenwassersensoren. *Die Bodenkultur: Journal of Land Management, Food and Environment*, *67*(3), 133–144.

Ogden, F. L., Crouch, T. D., Stallard, R. F., & Hall, J. S. (2013). Effect of land cover and use on dry season river runoff, runoff efficiency, and peak storm runoff in the seasonal tropics of Central Panama. *Water Resources Research*, *49*, 8443–8462. https://doi.org/10.1002/2013WR013956

Oldekop, E. (1911). Evaporation from the surface of river basins (Bcgapeyie c] godepxyocnb pexysx] ,acceqyod]). Collection of the Works of Students of the Meteorological Observatory. University of Tartu-Jurjew-Dorpat, Tartu, Estonia, p. 209.

O'Neel, S., McNeil, C., Sass, L. C., Florentine, C., Baker, E. H., Peitzsch, E., et al. (2019). Reanalysis of the US geological survey benchmark glaciers: Long-term insight into climate forcing of glacier mass balance. *Journal of Glaciology*, *65*(253), 850–866. https://doi.org/10.1017/jog.2019.66

Paudel, I., Kanety, T., & Cohen, S. (2013). Inactive xylem can explain differences in calibration factors for thermal dissipation probe sap flow measurements. *Tree Physiology*, *33*(9), 986–1001. https://doi.org/10.1093/treephys/tpt070

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*, W11427. https://doi.org/10.1029/2008WR007644

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C*, *83*, 36–56.

Philip, J. R. (1998). Physics, mathematics, and the environment: The 1997 Priestley lecture. *Australian Meteorological Magazine*, *47*, 273–283.

Phillips, N. G., Ryan, M. G., Bond, B. J., McDowell, N. G., Hinckley, T. M., & Čermák, J. (2003). Reliance on stored water increases with tree size in three species in the Pacific Northwest. *Tree Physiology*, *23*(4), 237–245.

Pike, J. G. (1964). The estimation of annual runoff from meteorological data in a tropical climate. *Journal of Hydrology*, *2*(2), 116–123.

Potter, N. J., Zhang, L., Milly, P. C. D., McMahon, T. A., & Jakeman, A. J. (2005). Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. *Water Resources Research*, *41*, W06007. https://doi.org/10.1029/2004wr003697

Rantz, S. E., Barnes, H. H., Buchanan, T. J., Carter, R. W., Kilpatrick, F. A., Smoot, G. F., et al. (1982a). Measurement and computation of streamflow: Volume 1. In *Measurement of stage and discharge, geological survey water-supply paper 2175* (1–284). Washington: United States Government Printing office.

Rantz, S. E., Barnes, H. H., Buchanan, T. J., Carter, R. W., Kilpatrick, F. A., Smoot, G. F., et al. (1982b). Measurement and computation of streamflow: Volume 2. In *Computation of Discharge, Geological Survey Water-Supply Paper 2175* (285–631). Washington: United States Government Printing office.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., et al. (2012). How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society*, *93*(6), 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1

Reager, J. T., Thomas, B. F., & Famiglietti, J. S. (2014). River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nature Geoscience*, *7*(8), 588–592. https://doi.org/10.1038/NGEO2203

Rice, J. S., & Emanuel, R. E. (2019). Ecohydrology of interannual changes in watershed storage. *Water Resources Research*, *55*, 8238–8251. https://doi.org/10.1029/2019WR025164

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., et al. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, *7*(1), 358–389. https://doi.org/10.2136/vzj2007.0143

Robles, M. D., Turner, D. S., & Haney, J. A. (2017). A century of changing flows: Forest management changed flow magnitudes and warming advanced the timing of flow in a southwestern US river. *PLoS ONE*, *12*(11), e0187875. https://doi.org/10.1371/journal.pone.0187875

Ruth, C. E., Michel, D., Hirschi, M., & Seneviratne, S. I. (2018). Comparative Study of a Long-Established Large Weighing Lysimeter and a State-of-the-Art Mini-lysimeter. *Vadose Zone Journal*, *17*, (1), 1–10. https://doi.org/10.2136/vzj2017.01.0026

Sabo, J. L., Sinha, T., Bowling, L. C., Schoups, G. H. W., Wallender, W. W., Campana, M. E., et al. (2010). Reclaiming freshwater sustainability in the Cadillac Desert. *Proceedings of the National Academy of Sciences*, *107*(50), 21,263–21,269. https://doi.org/10.1073/pnas.1009734108

Sahin, U., Kiziloglu, F. M., Anapali, O., & Okuroglu, M. (2007). Determining crop and pan coefficients for sugar beet and potato crops under cool season semiarid climatic conditions. *Journal of Agronomy and Crop Science*, *193*(2), 146–152.

Savenije, H. H. (2004). The importance of interception and why we should delete the term evapotranspiration from our vocabulary. *Hydrological Processes*, *18*(8), 1507–1511. https://doi.org/10.1002/hyp.5563

Scanlon, B. R., Healy, R. W., & Cook, P. G. (2002). Choosing appropriate techniques for quantifying groundwater recharge. *Hydrogeology Journal*, *10*(1), 18–39. https://doi.org/10.1007/s10040-001-0176-2

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., van Beek, L. P., et al. (2018). *Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proceedings of the National Academy of Sciences*, *115*(6), E1080–E1089. 201,704,665

Schaller, M. F., & Fan, Y. (2009). River basins as groundwater exporters and importers: Implications for water cycle and climate modeling. *Journal of Geophysical Research*, *114*, D04103. https://doi.org/10.1029/2008JD010636

Schmid, H. P. (1997). Experimental design for flux measurements: Matching scales of observations and fluxes. *Agricultural and Forest Meteorology*, *87*(2–3), 179–200.

Schreiber, P. (1904). On the relationship between precipitation and river flow in central Europe (Über die Beziehungen zwischen dem Niederschlag und derWasserführung der Flüsse in Mitteleuropa). *Zeitschrift für Meteorologie*, *21*, 441–452.

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, *35*(7), 2145–2160. https://doi.org/10.1029/1999WR900090

Sevruk, B. (1979). Correction for point precipitation measurement. Versuchsanstalt fur Wasserbau, Hydrologie und Glaziologie anderETHZurich. *Mittellung Nr*, *41*, 267–279.

Sexstone, G. A., Clow, D. W., Stannard, D. I., & Fassnacht, S. R. (2016). Comparison of methods for quantifying surface sublimation over seasonally snow-covered terrain. *Hydrological Processes*, *30*(19), 3373–3389.

Shun, T., & Duffy, C. J. (1999). Low-frequency oscillations in precipitation, temperature, and runoff on a west facing mountain front: A hydrogeologic interpretation. *Water Resources Research*, *35*(1), 191–201. https://doi.org/10.1029/98WR02818

Sieck, L. C., Burges, S. J., & Steiner, M. (2007). Challenges in obtaining reliable measurements of point rainfall. *Water Resources Research*, *43*, W01420. https://doi.org/10.1029/2005WR004519

Sorensen, J. P., & Butcher, A. S. (2011). Water level monitoring pressure transducers—A need for industry-wide standards. *Groundwater Monitoring & Remediation*, *31*(4), 56–62.

Soulsby, C., Tetzlaff, D., & Hrachowitz, M. (2009). Tracers and transit times: Windows for viewing catchment scale storage? *Hydrological Processes*, *23*(24), 3503–3507. https://doi.org/10.1002/hyp.7501

Soupir, M. L., Mostaghimi, S., & Mitchem, C. E. Jr. (2009). A comparative study of stream-gaging techniques for low-flow measurements in two Virginia tributaries 1. *JAWRA Journal of the American Water Resources Association*, *45*(1), 110–122. https://doi.org/10.1111/j.1752-1688.2008.00264.x

Sperling, O., Shapira, O., Cohen, S., Tripler, E., Schwartz, A., & Lazarovitch, N. (2012). Estimating sap flux densities in date palm trees using the heat dissipation method and weighing lysimeters. *Tree Physiology*, *32*(9), 1171–1178.

Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M., & Stahl, K. (2017). Catchment water storage variation with elevation. *Hydrological Processes*, *31*(11), 2000–2015. https://doi.org/10.1002/hyp.11158

Steppe, K., De Pauw, D. J., Doody, T. M., & Teskey, R. O. (2010). A comparison of sap flux density using thermal dissipation, heat pulse velocity and heat field deformation methods. *Agricultural and Forest Meteorology*, *150*(7–8), 1046–1056. https://doi.org/10.1016/j.agrformet.2010.04.004

Storck, P., Lettenmaier, D. P., & Bolton, S. M. (2002). Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States. *Water Resources Research*, *38*(11), 5–1. https://doi.org/10.1029/2002WR001281

Su, Z., He, Y., Dong, X., & Wang, L. (2017). *Drought monitoring and assessment using remote sensing*, *In Remote Sensing of Hydrological Extremes* (151–172). Cham: Springer. https://doi.org/10.1007/978-3-319-43744-6_8

Sun, H., Aubrey, D. P., & Teskey, R. O. (2012). A simple calibration improved the accuracy of the thermal dissipation technique for sap flow measurements in juvenile trees of six species. *Trees*, *26*(2), 631–640. https://doi.org/10.1007/s00468-011-0631-1

Tait, A. B. (1998). Estimation of snow water equivalent using passive microwave radiation data. *Remote Sensing of Environment*, *64*(3), 286–291.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. https://doi.org/10.1126/science.1099192

Tillery, A. C., Phillips, J. V., & Capesius, J. P. (2001). Potential errors associated with stage-discharge relations for selected streamflow-gaging stations, Maricopa County, Arizona (No. 0–4224). US Dept. of the Interior, US Geological Survey.

Toth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. *Journal of Geophysical Research*, *68*(16), 4795–4812. https://doi.org/10.1029/JZ068i016p04795

Townsend-Small, A., Pataki, D. E., Liu, H., Li, Z., Wu, Q., & Thomas, B. (2013). Increasing summer river discharge in southern California, USA, linked to urbanization. *Geophysical Research Letters*, *40*, 4643–4647. https://doi.org/10.1002/grl.50921

Turc, L. (1954). Le bilan d'eau des sols: Relation entre les pre'cipitations, l'e'vaporation et l'e'coulement. *Annales Agronomiques, Se'rie A*, *5*, 491–595.

Turnipseed, D. P. & Sauer, V. B. (2010). Discharge measurements at gaging stations. U.S. Geological Survey Techniques and Methods book 3, chap. A8, 87 p.

U. S. Department of the Interior, Bureau of Reclamation. (1997). Water Measurement Manual, Third Ed, p317

Wang, T., Istanbulluoglu, E., Lenters, J., & Scott, D. (2009). On the role of groundwater and soil texture in the regional water balance: An investigation of the Nebraska Sand Hills, USA. *Water Resources Research*, *45*, W10413. https://doi.org/10.1029/2009WR007733

Waring, R. H., & Running, S. W. (1976). *Water uptake, storage and transpiration by conifers: A physiological model*, *In Water and plant life* (189–202). Berlin, Heidelberg: Springer.

Werner, M., Cranston, M., Harrison, T., Whitfield, D., & Schellekens, J. (2009). Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteorological Applications*, *16*(1), 13–22. https://doi.org/10.1002/met.124

West, A. J. (1962). Snow evaporation from a forested watershed in the Central Sierra Nevada. *Journal of Forestry*, *60*(7), 481–484.

Western, A. W., Grayson, R. B., & Green, T. R. (1999). The Tarrawarra project: High resolution spatial measurement, modelling and analysis of soil moisture and hydrological response. *Hydrological Processes*, *13*(5), 633–652. https://doi.org/10.1002/(SICI)1099-1085(19990415)13:5<633::AID-HYP770>3.0.CO;2-8

Whiting, J. A., & Godsey, S. E. (2016). Discontinuous headwater stream networks with stable flowheads, Salmon River basin, Idaho. *Hydrological Processes*, *30*(13), 2305–2316. https://doi.org/10.1002/hyp.10790

Wiedemann, A., Marañón-Jiménez, S., Rebmann, C., Herbst, M., & Cuntz, M. (2016). An empirical study of the wound effect on sap flux density measured with thermal dissipation probes. *Tree Physiology*, *36*(12), 1471–1484. https://doi.org/10.1093/treephys/tpw071

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., et al. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, *368*(6488), 314–318. https://doi.org/10.1126/science.aaz9600

Wilson, J. L., & Guan, H. (2004). Mountain-Block Hydrology and Mountain-Front Recharge. Groundwater recharge in a desert environment: The Southwestern United States, 9.

Wittenberg, H., & Sivapalan, M. (1999). Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. *Journal of Hydrology*, *219*(1–2), 20–33.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., & Shum, C. K. (2018). A new estimate of North American mountain snow accumulation from regional climate model simulations. *Geophysical Research Letters*, *45*, 1423–1432. https://doi.org/10.1002/2017GL076664

Wüest, M., Frei, C., Altenhoff, A., Hagen, M., Litschi, M., & Schär, C. (2010). A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *International Journal of Climatology*, *30*(12), 1764–1775.

Wullschleger, S. D., Meinzer, F. C., & Vertessy, R. A. (1998). A review of whole-plant water use studies in tree. *Tree Physiology*, *18*(8–9), 499–512. https://doi.org/10.1093/treephys/18.8-9.499

Yokoo, Y., Sivapalan, M., & Oki, T. (2008). Investigation of the relative roles of climate seasonality and landscape properties on mean annual and monthly water balances. *Journal of Hydrology*, *357*(3–4), 255–269. https://doi.org/10.1016/j.jhydrol.2008.05.010

Zhang, L., Dawes, W. R., & Walker, G. R. (2001). Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resources Research*, *37*(3), 701–708.

Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H., Western, A. W., & Briggs, P. R. (2004). A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, *40*, W02502. https://doi.org/10.1029/2003WR002710

Zhou, S., Yu, B., Huang, Y., & Wang, G. (2015). The complementary relationship and generation of the Budyko functions. *Geophysical Research Letters*, *42*, 1781–1790. https://doi.org/10.1002/2015GL063511

Zreda, M., Desilets, D., Ferré, T. P. A., & Scott, R. L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters*, *35*, L21402. https://doi.org/10.1029/2008GL035655