Combining Indirect Observations and Models to Resolve Spatiotemporal Patterns of Precipitation in Complex Terrain

Brian M. Henn

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Jessica D. Lundquist (Chair)

Dennis P. Lettenmaier

Adrian Dobra

Program Authorized to Offer Degree:

Department of Civil and Environmental Engineering

i

University of Washington

Abstract

Combining Indirect Observations and Models to Resolve Spatiotemporal Patterns of Precipitation in Complex Terrain

Brian M. Henn

Chair of the Supervisory Committee:

Associate Professor Jessica D. Lundquist

Department of Civil and Environmental Engineering

Mountain precipitation in the Western United States is critical for the water resources of the region, but resolving spatial and temporal patterns of precipitation in complex terrain is challenging due to lack of observations, measurement uncertainty and high spatial variability. We examine several gridded precipitation datasets over the Sierra Nevada mountain range of California, and find that these widely-used products exhibit substantial variation in water-year total precipitation over different areas of the range. In addition, trends in precipitation and snow computed from different datasets vary substantially. Both findings suggest that further work is needed to better resolve spatial and temporal patterns of precipitation in complex terrain.

Streamflow observations are widely made and reflect the basin's hydrologic response to precipitation input. We develop a methodology for inferring basin-mean precipitation using

lumped hydrologic models and Bayesian model calibration, which infers water-year total precipitation given daily streamflow observations. We apply this approach to several basins around Yosemite National Park in the Sierra Nevada in order to assess the sensitivity and robustness of inferred precipitation. We find that patterns of precipitation can be inferred from streamflow, both in terms of spatial and year-to-year variability. However, by using a small ensemble of hydrologic model structures to test the sensitivity of inferred precipitation, we also show that the absolute amounts of inferred precipitation are subject to significant uncertainty.

Higher-elevation basins of the Sierra Nevada are hydrologically snow-dominated, and we hypothesize that the uncertainty in inferred precipitation can be reduced by calibrating the hydrologic model to both snow and streamflow observations. We leverage the recent availability of a high-resolution distributed snow dataset from the Airborne Snow Observatory (ASO) to determine basin-mean snow water equivalent (SWE) over the upper Tuolumne River basin. We also compare point and distributed SWE measurements over the basin, to assess the reliability of using point measurements to estimate basin-mean SWE. In this case, point measurements show bias in estimating basin-mean ASO SWE, largely due to non-representative sampling with respect to elevation. When basin-mean SWE is included with streamflow in model calibration, uncertainty in inferred precipitation is reduced by up to half, and model ensemble consistency is improved.

To resolve patterns of precipitation over the Sierra Nevada, we infer precipitation from streamflow using 56 stream gauges that measure runoff from relatively unimpaired basins, over 1950-2010. We compare inferred precipitation to gauge-based gridded precipitation data, finding that significant differences exist between the mean spatial patterns of precipitation over the range. In particular, inferred precipitation suggests that gridded products underestimate

precipitation for higher-elevation basins whose aspect faces prevailing winds. Better agreement

is found in lower-elevation and leeward basins. Collectively, the findings suggest that the

development of spatially distributed precipitation datasets should not consider precipitation

gauge data in isolation, but should also consider other related hydrologic observations in order to

better resolve patterns of precipitation in complex terrain.

Dedication

To Cathy for her kind and patient love and support, and to my family for encouraging me to study the world around us.

Acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

In the Western United States, representation of precipitation patterns over complex terrain is critical for understanding hydrology and water resources. Complex, high-elevation terrain drives orographic precipitation [*Bales et al.*, 2006] and provides a large fraction of the region's runoff [*Barnett et al.*, 2005]. However, our ability to observe precipitation over mountains is limited due to factors such as lack of observations, high spatial variability and gauge errors; *Dettinger* [2014] noted our lack of awareness of hydrologic states and fluxes "in the third [*i.e.*, vertical] dimension", and *Luce et al.* [2013] showed that low-elevation observations may not properly characterize high-elevation climatic trends. Nonetheless, our ability to observe, simulate and predict hydrologic fluxes in these areas will dictate the extent to which we can efficiently manage water resources and adapt to future variability.

While there are many variables involved in characterizing the hydrology of mountain basins, precipitation is perhaps the most fundamental. Along with temperature, it is the most common type of meteorological observation, and widely-used datasets have been developed to distribute precipitation in space and time based on observations [*e.g.*, *Daly et al.*, 1994; *Hamlet et al.*, 2010]. In addition, all models of snowpack [*e.g.*, *Anderson*, 2006] and distributed soil moisture and runoff, *e.g.*, PRMS [*Leavesley et al.*, 1983] use precipitation as a primary meteorological input. Hydrologic models that require additional inputs such as the surface radiation budget, *e.g.*, VIC [*Liang et al.*, 1994] and DHSVM [*Wigmosta et al.*, 1994] utilize approaches to derive these terms from observations of precipitation and temperature, *e.g.*, MT-CLIM [*Thornton and Running*, 1999]. Thus, robust simulation and characterization of basin hydrology requires knowledge of the spatiotemporal variability of precipitation.

Despite its importance, precipitation is often poorly observed at higher elevations [*Lundquist et al.*, 2003] where climatological precipitation is generally greatest. Compounding this problem is that the spatial and temporal variability of precipitation is also typically greatest in regions of complex terrain [*Roe*, 2005; *Anders et al.*, 2007]. Spatial patterns of precipitation vary in time, sometimes confounding attempts to distribute these variables based on a fixed map [*Lundquist et al.*, 2010; *Siler et al.*, 2013]. Thus, it is challenging to resolve these fields via observations alone; errors in extrapolating point measurements of precipitation to the basin scale can be 100% or more [*Milly and Dunne*, 2002].

Due to this uncertainty, there is a need for models to supplement observations of mountain precipitation and temperature, and provide a more complete and representative description of these fields. Models that generate distributed fields of precipitation may be statistical, *e.g.*, the Parameter-elevation Regressions on Independent Slopes Model [PRISM, *Daly et al.*, 1994, 2008], or physical, *e.g.*, the Weather Research and Forecasting atmospheric model [WRF, *Skamarock and Klemp*, 2008]. While statistical models may be accurate in that they are derived empirically from observations, they suffer from errors whenever there are changes in the underlying statistical relationship. In contrast, dynamic models are based on fundamental physics, but make assumptions of unknown validity about the structure and flux of mass and energy, and often require calibration of many unknown parameters, which can lead to predictions that are not robust [*Clark et al.*, 2008; *Renard et al.*, 2010]. As a result, it is incumbent on mountain hydrologists and meteorologists to leverage the strengths of both statistical and dynamic approaches (while acknowledging and avoiding their weaknesses) to improve our understanding of mountain precipitation. This is particularly critical given that scientists are being asked to utilize spatially-distributed precipitation and temperature data in

order to assess impacts on snowpack and streamflow, either in terms of current trends [*e.g., Mote et al.*, 2005], or under future climate change scenarios [*e.g., Jeton et al.*, 1996].

Indirect measurements of precipitation are also critical to improving our knowledge of its distribution. For example, observations of streamflow and snowpack can shed light on precipitation and soil storage in the basin [*Adam et al.*, 2006; *Hood and Hayashi*, 2015; *Le Moine et al.*, 2015]. We seek to employ the additional information contained in streamflow observations from unimpaired mountain basins to "do hydrology backward" [*Kavetski et al.*, 2003; *Kirchner*, 2009] and infer precipitation amounts. A similar approach is also possible using distributed observations of snow water equivalent (SWE).

My dissertation investigates methods to improve our knowledge of the spatial and temporal patterns of precipitation in mountain basins. In particular, I seek to use indirect observations of precipitation (streamflow and SWE), along with statistical and conceptual physical models, to map precipitation across mountainous terrain. I consider streamflow and SWE observations in order to infer precipitation patterns beyond those that were directly observed via gauges. I use the Sierra Nevada mountain range of California as an example of a relatively arid region where orographic precipitation patterns are critical for water resources, and where uncertainty in high-elevation precipitation is significant.

In chapter 2, I present evidence of uncertainty in existing datasets of spatially distributed precipitation over the Sierra Nevada. In chapter 3, I develop an approach for inferring basin-mean precipitation from streamflow observations using simple hydrologic models and Bayesian parameter inference. This approach is applied to a set of basins around Yosemite National Park. In chapter 4, I extend this approach to utilize both measurements of streamflow and SWE, in particular leveraging the recent high-resolution snow depth and SWE data made available by the

Airborne Snow Observatory (ASO). In chapter 5, I extend the inference of precipitation from streamflow to a set of 56 basins with long-term, relatively unimpaired streamflow observations, spanning the Sierra Nevada over 1950-2010. I use the comparison of observed streamflow, inferred precipitation, and gridded precipitation to comment on potential biases in existing datasets, and the potential for improving representations of mountain precipitation by considering indirect observations.

Chapter 2. High Sierra precipitation uncertainty in gridded datasets

2.1. Background and datasets

The use of spatially distributed precipitation data has grown as distributed hydrologic, ecological and land surface models, which generally require distributed precipitation as an input, have made advances in resolution and process representation. Many gridded precipitation datasets have been developed that seek to represent spatial and temporal variability over the mountains of the Western United States [*Maurer et al.*, 2002; *Hamlet and Lettenmaier*, 2005; *Daly et al.*, 2008; *Hamlet et al.*, 2010; *Xia et al.*, 2012; *Livneh et al.*, 2013; *Newman et al.*, 2015]. These datasets have been used for many purposes, such as assessing climatic trends in streamflow and snowpack [*e.g., Hamlet et al.*, 2005, 2007], downscaling of global climate models [*e.g., Gutmann et al.*, 2014], estimating forest transpiration patterns and trends [*Christensen et al.*, 2008] and assessing ecological impacts of climate change [*e.g., Ogden et al.*, 2014]. Thus, the resolution of precipitation in these datasets has wide-ranging scientific implications.

In general, these datasets are produced by interpolating precipitation gauge observations over the landscape of the Western United States; for detailed comparisons of the techniques used to create these datasets, see *Lundquist et al.* [2015] and [*Newman et al.*, 2015]. The sources of precipitation gauge data vary, but most utilize the National Weather Service Cooperative Observer (COOP) gauge network, with additional gauge networks used in some cases as well. The datasets generally feature daily temporal resolution and spatial resolution between 30 arc-seconds (~800 m) and 1/8° (~12 km), which may at least partially resolve mountainous terrain and account for effects of orographic enhancement [*Roe*, 2005]. Because of orographic effects, it is necessary to include topographic corrections in the final interpolated product; most (though

5

not all) gridded datasets force long-term average grid cell precipitation values to match a precipitation climatology developed from the Parameter Regression on Independent Slopes Model [PRISM, *Daly et al.*, 2008]. However, individual datasets differ in terms of the temporal variability of grid cell precipitation, and in terms of the spatial pattern of precipitation at short to medium timescales.

Due to the challenges in estimating precipitation over complex terrain, multiple studies have demonstrated uncertainty in gridded precipitation datasets. In two cases in the Colorado Rocky Mountains where precipitation gauges had not been incorporated into the PRISM dataset, differences between predicted and observed precipitation of hundreds of mm per water year were found, as much as 50-100% of the water year total [*Gutmann et al.*, 2012; *Livneh et al.*, 2014]. *Lundquist et al.* [2015] compared daily increases in Sierra Nevada snow pillow snow water equivalent (SWE) to daily gridded dataset precipitation, and found individual water years in which the gridded products had up to 21% too little precipitation to account for snow pillow SWE increases.

These examples indicate that even though most gridded precipitation datasets are deterministic and uncertainty estimates are not provided, there exists substantial potential for substantial error in estimating precipitation in complex terrain. To address this, *Newman et al.* [2015], using the approach outlined in *Clark and Slater* [2006], generated an ensemble gridded precipitation dataset in which the uncertainty of the spatial distribution of gauge data is explicitly estimated, and showed that intra-ensemble uncertainty was on the order of 10-30% for monthly accumulations.

In this chapter, we compare four publicly-available gridded daily precipitation datasets over the Sierra Nevada, both against one another and against observations of streamflow and

SWE trends. The four datasets are *Hamlet et al.* [2010], hereafter H10; *Livneh et al.* [2013], hereafter L13; NLDAS-2 [*Xia et al.*, 2012], hereafter NLDAS; and *Newman et al.* [2015], hereafter N15. The first two datasets have 1/16° (~6 km) spatial resolution, while the latter two have 1/8° (~12 km) resolution. While the first three datasets use a PRISM precipitation climatology (either the 1961-1990 or the 1971-2000 versions) for topographic correction, they differ in the exact set of gauges used in the interpolation and in the lapse rates of the accompanying daily temperature datasets [*Lundquist et al.*, 2015]. The N15 dataset uses a regression against topographic predictors that is independent of PRISM, and explicitly samples the uncertainty of distributing precipitation using a 100-member ensemble.

We compare the spatial pattern of the water-year precipitation totals from each of the four datasets across a subdomain in the central Sierra Nevada, and for three streamflow basins (a subset of a larger 56-basin set described in chapter 5): Cherry and Eleanor Creeks (basin area 486 km^2), the San Joaquin River at Miller's Crossing (651 km^2) and Bear Creek (136 km^2). For each basin, gridded precipitation is converted to a basin total by summing the daily precipitation of the grid cells within the basin, and weighting grid cells partially within the basin by the fraction that is within. The resulting basin daily precipitation time series are summed for each water year (October 1 to September 30). Daily stream gauge observations for the San Joaquin R. and Bear Cr. (United States Geological Survey (USGS) stream gauges 11226500 and 11230500) are divided by the basin areas, and then summed over the water year. For Cherry and Eleanor Crs., we use full natural flows provided by the operator of the reservoirs at both creeks' outlets (see chapter 3 for more details). Because there is a tunnel connecting the two reservoirs, we sum both the full natural flows and the contributing areas of the two basins to create a single runoff time series; we refer to this combined basin as Cherry-Eleanor.

2.2. Comparison of water-year precipitation over the Sierra Nevada

The inter-comparison of the gridded precipitation datasets and the comparison against streamflow are shown in Figure 2.1. Figures 2.1a-2.1c show time series of water year precipitation and streamflow for the three basins. For both Cherry-Eleanor and the San Joaquin, the gridded datasets precipitation is of a similar magnitude to streamflow (San Joaquin streamflow observations were not made after 1991). Because of basin water balance, precipitation that is similar to (or less than) streamflow implies implausible near-zero evapotranspiration in these basins. These are examples of cases in which the long-term precipitation is likely being underestimated, to varying degrees, by all four of the gridded datasets (in this case, NLDAS-2 in particular). The difference between precipitation and streamflow varies substantially between years for each gridded product, suggesting that these errors may vary from year to year in an unpredictable fashion.

Bear Creek precipitation (Figure 2.1c) is greater as a fraction of streamflow for all four datasets, suggesting that underestimation is less severe in this basin. However, the agreement between the gridded datasets' water year precipitation is worse in this basin relative to the others, with differences of over 500 mm, or 40%, appearing for many years. In particular, disagreement between L13 and the other datasets in water years 2005 and 2006 is apparent. The differences between water-year total basin-mean precipitation for these four datasets, expressed as the four-dataset standard deviation divided by the four-dataset mean for each year, average 11% for Cherry-Eleanor, 18% for San Joaquin, and 21% for Bear. However, *actual* uncertainties between the datasets estimates and basin-mean precipitation may be substantially higher if the gridded datasets are biased low or high on average; low biases are almost certainly the case for some basins, given the comparisons to streamflow discussed above.

8

We plot the spatial pattern of precipitation for water year 2005 in Figures 2.1d-2.1g. In particular, differences between the datasets for the high-elevation areas around the Cherry-Eleanor and Bear Cr. basins are seen. The differences in these parts of the domain are on the order of 500-1,500 mm yr^{-1}, or 25-60% of the mean precipitation across the four datasets. While this water year in particular has large disagreement, such uncertainties presumably would be highly problematic for simulation of the hydrology of this part of the range.

2.3. Comparison of precipitation and snow trends over the Sierra Nevada

Next, we examine the trends in cool-season precipitation over the Sierra Nevada, following the analysis of *Mote et al* [2005], who argued that mountain snowpack declined across the Western United States over 1950-1997. However, in that study an exception was the southern Sierra Nevada, where snow course measurements showed increasing SWE. The model analysis in that study also failed to replicate the positive SWE trends, and we examine the gridded precipitation dataset used to drive the model [*Hamlet and Lettenmaier*, 2005], hereafter HL05, which although of coarser resolution (1/8°), was generated using a similar methodology as the H10 dataset. Importantly, both datasets explicitly control for precipitation trends, such that they match observed trends at a subset of long-term, high-quality meteorological stations. This was done to avoid spurious trends associated with changes in the station set [*Hamlet and Lettenmaier*, 2005].

We compare the 1950-1997 trends in cool-season (October 1 to April 1) precipitation totals in both the HL05 and H10 datasets. We also compute the implied snowfall in each dataset, by summing grid cell precipitation on days with mean temperature below -0.5°C; precipitation on days with mean temperature between -0.5°C and 0.5°C was reduced proportionally, and

9

precipitation on days with mean temperature above 0.5°C was excluded. We compute the 1950-1997 trends in each grid cell for snow calculated in this way. Finally, we compute the trend in April 1 SWE at 149 snow courses in the Sierra Nevada over the same time period, and compare the trends, following Figure 1 of *Mote et al.* [2005].

The results of the 1950-1997 trend comparison are shown in Figure 2.2. First, we examine trends in the HL05 dataset, which shows slightly increasing (<10%) trends in precipitation over most of the Sierra Nevada (Figure 2.2a). Trends in HL05 snow shows decreases at mid-elevations, with moderate increases at higher elevations in the southern portion of the range (Figure 2.2b). However, greater precipitation trends are seen in the more recent H10 dataset, both in precipitation and snow (Figures 2.2c-2.2d), particularly in the southern and eastern portions. Trends in April 1 snow course SWE (Figure 2.2e) show large increases in the high-elevation southern portions of the range, along with substantial decreases in the northern and mid-elevation areas. We compute streamflow trends from 23 gauges with complete or nearly complete observations over 1950-1997; the basins are part of the larger set discussed in chapter 5. The streamflow trends (Figure 2.2f) largely agree with the snow trends, showing increasing streamflow from southern Sierra Nevada basins, but a mix of slight increases and moderate decreases over the northern portion of the range.

When compared with observed SWE trends (Figures 2.2g), the H10 trends more closely approximate observations than the HL05 dataset; the observed large positive trends at almost all snow courses south of 38°N are mostly replicated in H10. However, the H10 trends still underestimate increases in SWE on the west slope in the southern Sierra Nevada, mainly in the basins of the Kings, San Joaquin and Merced rivers. The observed increases in high-elevation SWE suggest positive precipitation trends at these locations, which may have exceeded the

increases measured at the long-term precipitation gauges which control the H10 trends and are generally at lower elevations.

Using the same approach, we also compute trends in H10, N15, L13 and NLDAS over the time period when all four are available for the Sierra Nevada (1981-2006). Trends in these four gridded datasets are shown in Figure 2.3; they exhibit high spatial variability, with less coherent regional patterns. H10 precipitation trends over 1981-2006 show declines in the north and increase in the south of the domain, with magnitudes up to ±20%; snow trends are similar with the exception of more severe declines in the mid-elevation zones of the northern Sierra Nevada, presumably due to increasing temperatures (Figures 2.3a and 2.3b). However, N15 trends (when computed as the mean of the trends across all 100 ensemble members) are much more muted, showing declines in both precipitation and snow of 10% or less over most of the Sierra Nevada (Figures 2.3c and 2.3d). In contrast, the trends L13 (Figures 2.3e and 2.3f) and NLDAS (Figures 2.3g and 2.3h) are highly spatially variable. Areas of very large (>±50%) trends are present in both datasets for precipitation and snow, but the locations of those large trends are neither consistent with each other, nor with the gains and losses seen in H10. Thus, estimation of precipitation and snow trends using these datasets is likely highly uncertain.

Out of this group of datasets, H10 controls for changing station sets over time by forcing the data to match trends at long-term stations [*Hamlet and Lettenmaier*, 2005]. It is possible that the trends in L13 and NLDAS reflect the appearance and disappearance of stations as much as physically meaningful trends; temporal inhomogeneities such as these can strongly influence apparent trends in the data. N15 uses a static station set and filling of missing data (via quantile matching) to minimize the effects of inhomogeneities, but the trends in this dataset are quite different from those in H10. It is possible that the stationary relationships between stations

assumed in a quantile-matching approach to filling missing data could dampen trends in the N15 dataset. We present the mean of the trends in the N15 ensemble, which also damps out much substantial variability in the trends of the individual ensemble members (not shown). The intra-ensemble variability suggests that trends are uncertain in this dataset as well.

2.4. Discussion

The preceding examples show that gridded precipitation datasets, which rely on interpolation of gauge data across complex terrain, may contain biases in spatial patterns of precipitation at the annual timescale, and biases in precipitation trends over the long term. While the techniques to generate gridded precipitation do capture much topographic and temporal variability, certain features (*e.g.*, local precipitation maxima and high-elevation trends) may not be observed by the gauge network in areas of sparse coverage. For example, the disagreement over water year total precipitation in the Cherry-Eleanor, San Joaquin and Bear basins is likely related to the fact that they feature relatively steep terrain, large elevation gradients and limited or non-existent road access for gauge maintenance.

The high uncertainties in gridded precipitation over complex terrain have implications for their use in evaluating scientific hypothesis related to precipitation in these regions. Most gridded precipitation datasets (N15 excepted) do not provide estimates of uncertainty, and so there may be a tendency to accept these estimates as "truth" when they are applied to force a hydrologic model or for some other purpose. However, it is clear that users of these data should be aware of the potential for substantial error and consider how those errors may influence their results.

Given these findings, it will be useful to examine proxy data related to precipitation that may help to reduce the uncertainty in gridded datasets, *i.e.*, streamflow and snow observations.

Additionally, precipitation estimates based on computational weather models, *e.g.*, the Weather Research and Forecasting Model [WRF, *Skamarock and Klemp*, 2008], may provide useful information about precipitation where gauge-based estimates are uncertain. While model precipitation and precipitation inferred from streamflow and snow may have substantial uncertainties themselves, even approximate estimates may help resolve some of the disagreements between gridded datasets shown here. For example, we show in chapter 3 [*Henn et al.*, 2015], that while there is substantial uncertainty in the absolute amounts of precipitation inferred from streamflow, basin-to-basin and year-to-year variability can likely be inferred more robustly. Thus, we seek to better identify spatial and temporal patterns in precipitation using these proxy observations.

2.5. Figures
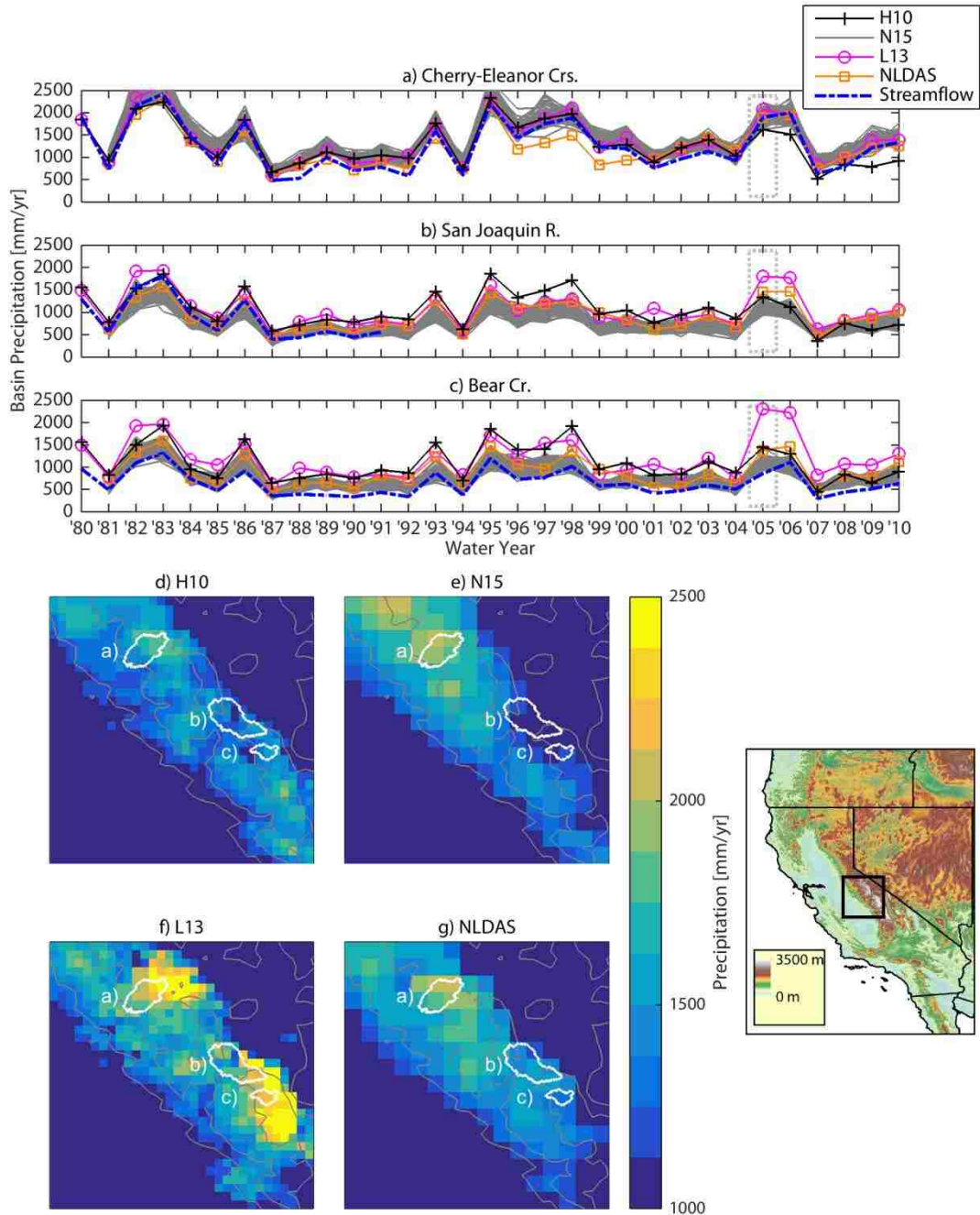
Figure 2.1. Gridded precipitation comparison for three basins with streamflow observations in the central Sierra Nevada. a) - c): Time series of water-year total, basin-mean precipitation from the four datasets for the Cherry-Eleanor, San Joaquin and Bear basins; water-year total streamflow for each basin is shown as well. Dashed box indicates example water year (2005), for which precipitation is mapped in d) - g). Basin locations are shown on each map.
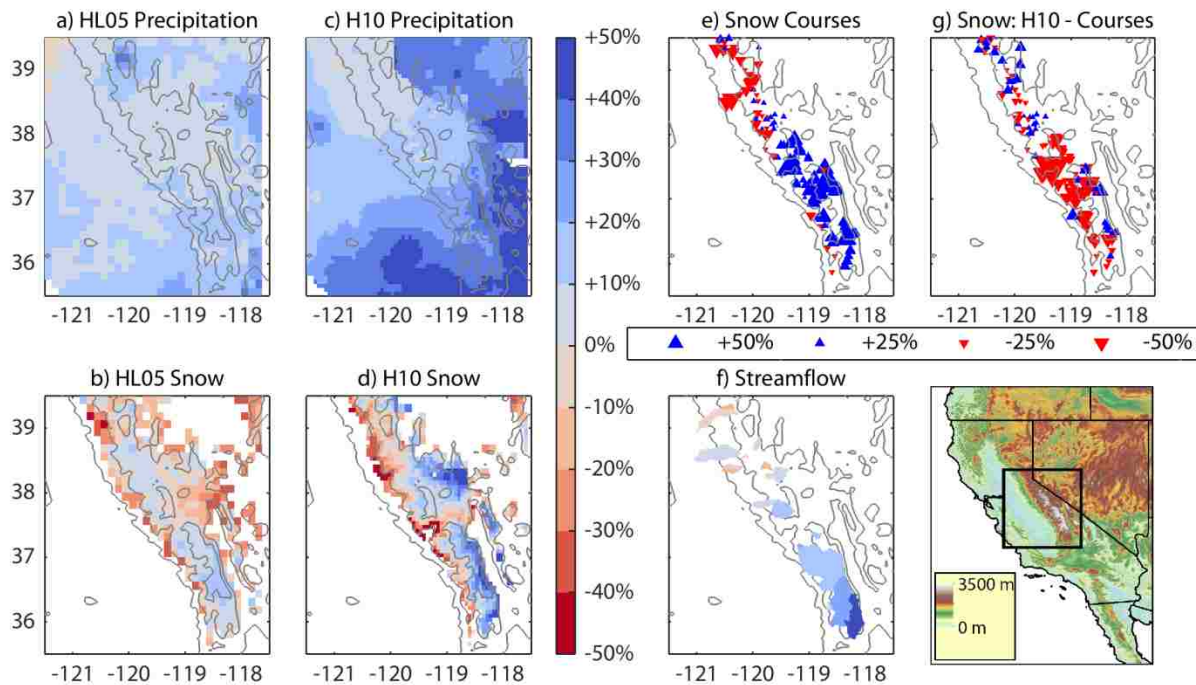
Figure 2.2. 1950-1997 cool-season precipitation, snow and streamflow trends over the Sierra Nevada. a) - d): Precipitation and snow trends for the 1/8° HL05 and 1/16° H10 datasets. e) April 1 snow course SWE trends. f) Water year total streamflow trends over 24 basins with observations spanning at least 85% of the 1950-1997 period. g) Difference between H10 and snow course trends (H10 minus courses). All trends calculated as percentage of the 1950 linear best-fit value, following *Mote et al.* [2005].
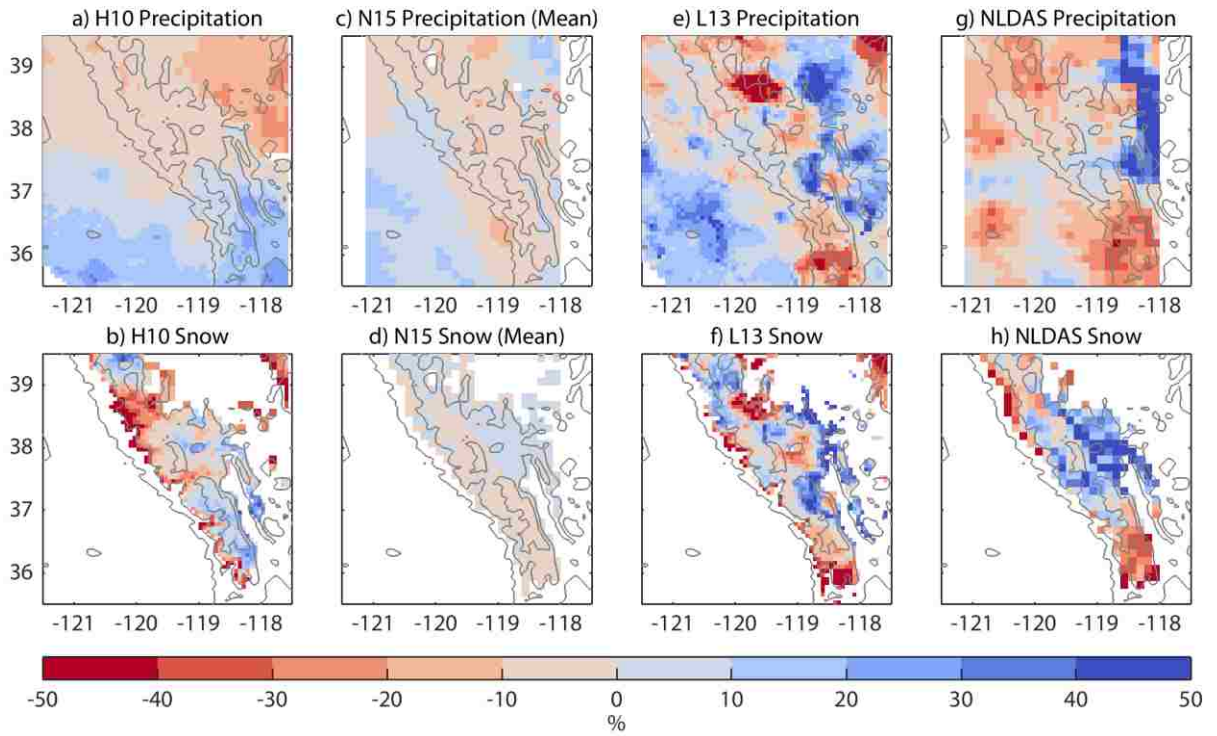
Figure 2.3. Trends in 1981-2006 Sierra Nevada cool-season precipitation (top row) and snow (bottom row) in four gridded datasets, over the same domain as Figure 2.2.

16

Chapter 3. Estimating mountain basin-mean precipitation from streamflow using Bayesian inference

Note: This chapter has been published in its current form as an article in *Water Resources Research* [*Henn et al.*, 2015]; the only differences are in section numbering and some reference information. It is used here by permission of John Wiley and Sons.

Abstract: Estimating basin-mean precipitation in complex terrain is difficult due to uncertainty in the topographical representativeness of precipitation gauges relative to the basin. To address this issue, we use Bayesian methodology coupled with a multi-model framework to infer basin-mean precipitation from streamflow observations, and we apply this approach to snow-dominated basins in the Sierra Nevada of California. Using streamflow observations, forcing data from lower-elevations stations, the Bayesian Total Error Analysis (BATEA) methodology and the Framework for Understanding Structural Errors (FUSE), we infer basin-mean precipitation, and compare it to basin-mean precipitation estimated using topographically-informed interpolation from gauges (PRISM, the Parameter-elevation Regression on Independent Slopes Model). The BATEA-inferred spatial patterns of precipitation show agreement with PRISM in terms of the rank of basins from wet to dry, but differ in absolute values. In some of the basins, these differences may reflect biases in PRISM, because some implied PRISM runoff ratios may be inconsistent with the regional climate. We also infer annual time series of basin precipitation using a two-step calibration approach. Assessment of the precision and robustness of the BATEA approach suggests that uncertainty in the BATEA-inferred precipitation is primarily related to uncertainties in hydrologic model structure. Despite these limitations, time

series of inferred annual precipitation under different model and parameter assumptions are strongly correlated with one another, suggesting that this approach is capable of resolving year-to-year variability in basin-mean precipitation.

3.1. Introduction

3.1.1. Challenges in observing precipitation in complex terrain

Estimating basin-mean precipitation in complex terrain is difficult due to the limited ability of precipitation gauges to capture spatial variability across the basin. Inadequate gauge spatial density or distribution with basin elevation can lead to errors of 100% or more in basin-mean precipitation [*Milly and Dunne*, 2002]. These problems are particularly acute in mountain basins, due to the greater spatial variability of orographic precipitation [*Roe*, 2005] and the lack of meteorological stations at higher elevations [*Lundquist et al.*, 2003; *Adam et al.*, 2006; *Daly et al.*, 2008]. Additionally, mountain precipitation gauges may contain biases of uncertain magnitude due to undercatch of snow and rain, as well as gauge icing [*Sieck et al.*, 2007; *Rasmussen et al.*, 2011b]. Manual access for wintertime gauge maintenance may be extremely difficult at mountain sites.

As a result, the observations available to estimate mountain basin-mean precipitation often have a coarse spatial resolution and high uncertainty. Uncertainty in mountain precipitation estimates hinders our ability to accurately predict rain-driven mountain basin floods, runoff volumes for water supply operations, and summertime low flows from high-elevation snowmelt [*e.g.*, *Ralph et al.*, 2005]. Problems with gauge uncertainty and representativeness have also hindered attempts to compare hydrologic model structures and forcing datasets [*Mizukami and Smith*, 2012; *Wayand et al.*, 2013]. Similarly, relatively poor knowledge of high-elevation spatial

18

precipitation patterns limits our ability to validate and improve the simulation of precipitation in research and operational weather models [*Jankov et al.*, 2009].

Streamflow is widely measured and has a well-known link to basin-mean precipitation. From mass balance, streamflow at a catchment outlet (Q) must equal basin-mean precipitation (P) minus basin-mean evapotranspiration (ET) and changes in storage (ΔS):

$$Q = P - ET - \Delta S \tag{3.1}$$

where all terms are per unit area of the basin. Thus, if Q is measured with reasonable accuracy, and ET and ΔS can also be determined, then we can infer the precipitation that the basin must have received. This idea leads to the general notion of inferring precipitation from streamflow [*Kavetski et al.*, 2003; *Vrugt et al.*, 2008; *Kirchner*, 2009], which has been termed "doing hydrology backward" [*Vrugt et al.*, 2008; *Kirchner*, 2009]. Such inference can be undertaken using Bayesian methods [*e.g.*, *Kavetski et al.*, 2003; *Vrugt et al.*, 2008; *Renard et al.*, 2011], or where possible, by directly inverting the hydrologic model [*e.g.*, *Kirchner*, 2009].

3.1.2. Objectives: Inferring spatial precipitation patterns in the Sierra Nevada

This study investigates the inference of basin-mean precipitation using streamflow observations, with the goal of reducing the uncertainty in mountain precipitation by employing the information that streamflow contains about the basin hydrologic system. We assume that errors in streamflow measurements are relatively small and unbiased at long time scales, compared to errors in estimates of basin-mean precipitation. While streamflow observations are subject to appreciable uncertainty [*Pelletier*, 1988; *Baldassarre and Montanari*, 2009], they appear more robust than spatially-interpolated estimates of precipitation in complex terrain, which have been shown to contain substantial biases [*e.g.*, *Gutmann et al.*, 2012; *Lundquist et al.*, 2015]. Thus, we hypothesize that "doing hydrology backward" will improve understanding

of basin-mean precipitation, which in turn may lead to an improvement in widely-used spatially-distributed precipitation datasets.

The approach of inferring basin-mean precipitation from streamflow observations is applied in the Sierra Nevada mountain range of California. This region has strongly topographical precipitation gradients and is critical for water supply in the state. In addition, there is substantial uncertainty regarding precipitation falling in high-altitude basins, which impacts the ability of downstream water resource managers to function efficiently. In the southern portion of the range, errors in gridded precipitation products may have interfered with hydrologic modeling efforts [*Mote et al.*, 2005] due to a lack of high-elevation precipitation gauges. In a Mediterranean (wet winter, dry summer) climate, the precipitation in the high-elevation portions of the range falls predominantly as snow. While spatial patterns of precipitation and snow have significant year-to-year consistency [*e.g.*, *Deems et al.*, 2008], changes in the spatial patterns of precipitation and snowpack between years have also been observed [*e.g.*, *Aguado*, 1990; *Lundquist et al.*, 2010], which highlights the need for precipitation estimates that do not depend on climatological spatial patterns.

The precipitation-from-streamflow approach has previously been used to address questions about regional- or range-scale hydrology. Luce et al. [2013] utilized streamflow observations and a water balance argument as evidence that high-elevation precipitation has been decreasing in the Pacific Northwest of North America. Earlier studies in the Sierra Nevada have used streamflow measurements at multiple sites to infer snowmelt timing and the location of summer thunderstorms [*Lundquist et al.*, 2003, 2009]. We seek to formalize the precipitation-from-streamflow approach so that it can be applied to resolve basin-mean precipitation patterns across multiple high-altitude basins within a region.

Inferring precipitation from streamflow in snow-dominated watersheds has unique challenges, because the basin's response to wintertime precipitation events is delayed and aggregated into the spring and summer snowmelt runoff response. In a previous study, Koskela et al. [2012] used a coupled snow and runoff model to infer precipitation from both streamflow and snowpack observations in a mixed rain-snow watershed in Finland. In this work, we use only streamflow observations, but we also explicitly address the impact of model structural differences on the inference, using the multi-model Framework for Understanding Structural Errors (FUSE) [*Clark et al.*, 2008]. We also estimate precipitation at the annual timescale, rather than for individual storm events as in *Koskela et al.*, [2012], because of the difficulty in inferring sub-annual events in snow-dominated watersheds.

We infer both long-term average precipitation amounts for a set of basins, as well as the year-by-year variations in precipitation, and then compare the inferred amounts with other spatial estimates of precipitation that are based on climatology. While data for an independent validation of basin-mean precipitation does not exist for our case study catchments, it is possible to investigate sensitivity and uncertainty associated with the precipitation-from-streamflow approach by calculating a range of basin-mean precipitation amounts from different FUSE model structures, basin soil assumptions and forcing datasets.

3.1.3. Contents

Section 3.2 provides background on the challenge of determining spatially-distributed precipitation in complex terrain, and on the inference of precipitation from streamflow. Section 3.3 describes the multi-model hydrological modeling framework (FUSE) and the temperature index snow model; the Bayesian inference methodology (BATEA) used to estimate basin-mean precipitation; the hydrological, topographical and meteorological data used in this study; and the

21

methodology of the case studies in the Sierra Nevada basins. Section 3.4 presents the inferred, long-term, basin-mean precipitation and the results of inferring year-by-year variations in the precipitation rates. This section also evaluates the sensitivity of the precipitation-from-streamflow approach to model structural choices, forcing data sources and parameter uncertainty. Section 3.5 discusses the applicability of the precipitation-from-streamflow approach in improving understanding of mountain precipitation, and section 3.6 summarizes the major findings from these experiments.

3.2. Background

3.2.1. Current approaches for distributing precipitation in complex terrain

The most common strategy for estimating spatially-distributed precipitation in mountain basins is to interpolate observations between available precipitation gauges. Because topography plays a dominant role in precipitation patterns, spatial weights derived from precipitation regression against topography are applied to distribute precipitation from gauges [*e.g.*, *Clark and Slater*, 2006; the Parameter-elevation Regression on Independent Slopes Model (PRISM), *Daly et al.*, 2008]. The accuracy of these approaches is governed by the quality and spatial density of the precipitation data in the regressions. In particular, a challenge for these methods is to identify time-varying patterns of precipitation which may deviate from better-established climatological patterns [*e.g.*, *Lundquist et al.*, 2010, 2015].

While precipitation deviations will ultimately average out to produce stable spatial weights, this timescale may be too long for many hydrological applications. Spatial distributions of mountain precipitation have been found to vary with many factors in the Western United States, *e.g.*, indices of the El Nino Southern Oscillation (ENSO) [*Dettinger et al.*, 2004; *Siler et*

al., 2013]; extent of wind blocking by terrain [*Hughes et al.*, 2009]; and microphysical differences in precipitation type and formation [*Pavelsky et al.*, 2012]. All of these features may vary during a particular storm or study period, leading to non-stationary spatial patterns of precipitation. As a result, we are interested in methods of estimating spatial patterns beyond what can be gleaned from the often sparse network of gauges.

3.2.2. Precipitation inferred from streamflow

The water balance of the basin indicates the link between precipitation and streamflow. For a hydrologically-simple basin that behaves approximately as a single-reservoir system, it is possible to invert the reservoir model equation and express precipitation as a function of the time series of streamflow values [*Kirchner*, 2009]. When such direct estimation is not applicable, it may still be possible to estimate precipitation from streamflow using Bayesian approaches [*Kavetski et al.*, 2006a; *Vrugt et al.*, 2008]; a useful review of Bayesian methods in hydrologic modeling is given by *Liu and Gupta* [2007]. In these approaches, the precipitation forcing data used to drive the hydrologic model are assumed to be corrupted by errors, such as those due to spatial unrepresentativeness and undercatch. Precipitation multipliers (or similar correction factors) are established which correct the gauge observations to basin-mean precipitation. Prior knowledge about the basin's hydrologic and hydrometeorologic characteristics, in combination with the goodness-of-fit of modeled and observed streamflow, are used to infer the precipitation multipliers along with other model parameters [*Renard et al.*, 2010, 2011].

Several frameworks have been proposed to calibrate hydrologic models under the assumption of precipitation input uncertainty. In this study, we use the Bayesian Total Error Analysis (BATEA) methodology [*Kavetski et al.*, 2006a, 2006b; *Kuczera et al.*, 2006; *Renard et al.*, 2011] in order to explicitly specify the level of confidence in the precipitation forcing data,

as well as in the other model parameters such as soil properties. Importantly, BATEA allows exploring the interplay between multiple sources of uncertainty, thus providing insights into their contribution to total predictive uncertainty [*Kavetski et al.*, 2006b; *Thyer et al.*, 2009; *Renard et al.*, 2010, 2011].

3.3. Methods and data

3.3.1. Models

3.3.1.1. FUSE conceptual rainfall-runoff models

We use the Framework for Understanding Structural Errors (FUSE) [*Clark et al.*, 2008] to provide a set of conceptual rainfall-runoff models and evaluate the effects of model structural uncertainty on inferred precipitation [*Clark et al.*, 2011a; *McMillan et al.*, 2011b]. FUSE provides multiple options for representing soil moisture storage and fluxes, such as ET, surface runoff and baseflow, using a spatially-lumped approach with upper and lower soil zones. Streamflow, soil moisture and ET time series are simulated at a daily timestep. The use of lumped models results in relatively few model parameters, which tends to reduce the computational cost of calibration and improve parameter identifiability.

In this study, FUSE is used to explicitly test the uncertainty associated with model structural choices. This is accomplished by calibrating multiple FUSE structures to the same basin forcing data, and then evaluating the variance in the inferred basin-mean precipitation. We use an ensemble of six model structures which were previously used to investigate the impact of structure on model performance [*Clark et al.*, 2011a; *McMillan et al.*, 2011b]. More information about the specific model structures is found in the appendix and Figure 3.10.

3.3.1.2. Snow Model

FUSE is coupled to a temperature index snow model based on Snow-17 [*Anderson*, 2006], which tracks snow water equivalent (SWE) based on precipitation and melt. Precipitation is added to SWE if the air temperature is below a threshold; otherwise, it is treated as rain. Snowmelt is initiated if the air temperature is above a threshold, and is proportional to the temperature above that point:

$$SM = \max(\min[f\{T - T_0\}, SWE], 0) \tag{3.2}$$

where SM is snowmelt per timestep, f is the snowmelt factor, T is the air temperature, and T_0 is the melt threshold temperature. The melt factor varies in a sinusoidal manner over the year, with minimum and maximum values on the winter and summer solstices, respectively. The rain-snow partition temperature, melt initiation temperature, and winter and summer melt factors are all inferred parameters. Snowmelt is combined with rainfall, and the two are routed to the soil model, which otherwise simulates fluxes and storages independently of the snow model.

In order to simulate the strong elevation dependence found within the basins, the snow model is run at 100 m elevation bands (unlike the soil model, which is lumped over the entire catchment). Elevation bands are a method of simulating the spatial snowpack variability that is necessary to reproduce the streamflow recession [*Luce and Tarboton*, 1998; *Clark et al.*, 2011b].

For each band, forcing temperature is lapsed to the ith band midpoint elevation:

$$T_i = T_f + \Gamma(z_i - z_f) \tag{3.3}$$

where T_i is the ith band temperature, T_f is the forcing temperature, z_i is the band midpoint elevation, z_f is the forcing data elevation, and Γ is the temporally-invariant temperature lapse rate (<0, °C km^{-1}), which is also inferred.

Similarly, precipitation is distributed to each band via a multiplicative scaling of the forcing precipitation:

$$P_i = MP_f(1 + \alpha[z_i - z_f])$$
<div align="right">(3.4)</div>

where P_i is the ith band precipitation, P_f is the forcing precipitation, M is the gauge-to-basin precipitation multiplier, and α is the orographic precipitation gradient (OPG) in km^{-1} [*Lundquist et al.*, 2010]. The negative lapse rate and the non-negative OPG ensure that the higher bands are colder and receive at least as much precipitation as the forcing elevation, allowing the model to simulate high-elevation snowpack that can last into the summer.

Finally, water fluxes to the lumped soil model are calculated by a weighted average of rain plus snowmelt from each elevation band:

$$F_{soil} = \sum_{i=1}^{n_{bands}} AF_i(R_i + SM_i)$$
<div align="right">(3.5)</div>

where F_{soil} is the flux of water to the soil model, and AF_i is the fraction of the basin area, R_i is the rain, and SM_i is the snowmelt, all from band i.

For simplicity, the snow model does not simulate sublimation from the snowpack, and in that regard it may underestimate basin evaporative losses. The soil model also permits evapotranspiration regardless of the degree of snow cover in the basin, so long as there is sufficient energy and soil moisture. These two errors likely compensate each other to some degree.

3.3.2. Precipitation inference using BATEA

3.3.2.1. Conceptual approach

Figure 3.1 shows a schematic of information flow in the precipitation-from-streamflow approach using BATEA and FUSE. For a given set of parameter values, including the precipitation multiplier, OPG and hydrologic model parameters, a corrected estimate of basin precipitation can be routed through the snow and FUSE models to produce a simulated runoff time series. The simulated runoff is compared within BATEA's residual error model (in the

likelihood function) to the observed streamflow record. The combination of the priors on the multiplier, OPG and other model parameters with the likelihood function yields the Bayesian posterior distributions of inferred parameters. For a given value of the precipitation multiplier and OPG, basin-mean precipitation can be calculated using (3.4); the posterior distribution of these parameters can then be used to obtain the posterior distribution of the basin-mean precipitation.

We infer basin-mean precipitation over timescales of a year or longer, given observed streamflow records for a basin. The Bayesian inference is applied as follows. Following *Kavetski et al.* [2006a], the posterior probability density functions (PDFs) of the multiplier M, the OPG α, and other model parameters θ, given observed precipitation $\tilde{P}$ and streamflow $\tilde{Q}$, are described by:

$$\boldsymbol{P}\big(M, \alpha, \theta \big| \tilde{Q}, \tilde{P}\big) \propto \boldsymbol{P}\big(\tilde{Q}\big|\mathrm{M}, \alpha, \theta, \tilde{P}\big) \cdot \boldsymbol{P}(M, \alpha, \theta) \qquad (3.6).$$

The first product term on the right hand side of (6) is the likelihood function, which describes the probability distribution of the observed streamflow, given particular values of M, α, θ, and the observed gauge precipitation $\tilde{P}$. The likelihood function is constructed by combining the hydrologic model with an error model. The second product term in (3.6) is composed of the prior distributions of the precipitation multiplier, OPG, and other model parameters.

The posterior distribution of the precipitation multiplier, OPG and the other model parameters, $\boldsymbol{P}\big(M, \alpha, \theta \big| \tilde{Q}, \tilde{P}\big)$, represents an update of the prior distributions given streamflow and precipitation observations.

The combination of the soil, snow and multiplier parameters results in a parameter space with at least 18 dimensions, and so to sample the posterior probability distributions, multiple-start Quasi-Newton optimization and a Monte Carlo Markov Chain (MCMC) routine are used

within BATEA. The MCMC routine generates 4,000 samples from a chain, with convergence determined by the stabilization of the posterior distributions [*Kavetski et al.*, 2006a]. The result of the MCMC sampling routine is an ensemble of parameter sets, with more samples clustered around parameter values of greater posterior probability.

3.3.2.2. Parameter prior distributions

The ability to infer precipitation from streamflow is limited by the information contained in the streamflow time series and in the model parameter priors. Lack of identifiability may limit the precision of the parameter inference, but can be reduced if prior information is available [*Renard et al.*, 2010]. In the case of mountain basins, this requires some knowledge of the soil characteristics and snow model parameters, as well as the temperature and potential evapotranspiration (PET) forcing data. Table 3.1 shows a summary of the parameter prior distributions for model structure FUSE-070 and the sources of the prior information. In general, uniform distributions are used, and when no information is available about a parameter in the study domain, the distribution is set to the default limits for the FUSE model [*Clark et al.*, 2008]. For simplicity and due to a lack of information about parameter covariance, the prior distributions for each parameter are independent of one another.

We set the prior distribution of the precipitation multipliers to be uniform over [0.1 2.0]. This prior distribution represents high uncertainty in the precipitation gauge observations as they relate to the actual basin water input. The uniform distribution and wide bounds are set because of the paucity of high-elevation precipitation gauges in the study watersheds, which requires the use of forcing gauges that are at substantially lower elevation than the basin midpoints (3.3.3.1.). Although gridded daily precipitation datasets are available which cover the study basins and could be used to generate priors of the multipliers [*e.g.*, *Hamlet and Lettenmaier*, 2005], the lack

of gauges also may result in potentially large uncertainties in these products. Thus, we elect to use vague priors in order to avoid biasing the posteriors. However, we note that uniform priors may introduce problems, such as making it easier for model structural errors to affect posterior distributions of the precipitation multipliers [*e.g.*, *Renard et al.*, 2010].

3.3.2.3. Weighted least squares error model

In the BATEA inference we use a weighted least squares (WLS) residual error model. The residual error model translates differences between modeled and observed streamflow into the likelihood of the corresponding parameter sets; as such residual errors are used to represent all sources of simulation error that are not explicitly accounted for (*e.g.*, errors in streamflow observations and in temperature and PET forcing data). In the WLS approach, it is assumed that higher streamflow is subject to greater uncertainty, so that errors at different timesteps are treated differently, unlike simple least squares in which all timesteps are treated equally. Following *Thyer et al.* [2009], we assume that residual streamflow errors follow a Gaussian distribution with zero mean and a standard deviation that is a linear function of streamflow, *i.e.*,

$$\sigma_t = a\hat{Q}_t + b \tag{3.7}$$

where a and b are residual error parameters inferred in the calibration, and $\hat{Q}_t$ is the modeled streamflow at a given timestep, t.

WLS error models have been shown to improve the robustness of the model calibration by avoiding overfitting to large, potentially anomalous flood events [*Thyer et al.*, 2009]. It is further assumed that there is no temporal autocorrelation between model residuals [*e.g.*, see *Koskela et al.*, 2012]. This simplification is used to avoid the problematic interaction between model parameters and the mass balance parameters, as reported by *Evin et al.* [2014]

3.3.3. Data

In this section we describe the data required for the precipitation-from-streamflow approach: basin elevation distributions, daily streamflow observations, daily precipitation, temperature and PET forcing data, and information about the probability distributions of soil and snow model parameters. All meteorological forcing data, precipitation and streamflow observations, and model-simulated values are at daily resolution. We also describe the PRISM precipitation dataset used for comparison with our results.

3.3.3.1. Yosemite-area basins

Our study includes basins within the Tuolumne, Merced, Walker and San Joaquin Rivers' watersheds in the Sierra Nevada of California (Figure 3.2): the Tuolumne above Hetch Hetchy Reservoir, the combined basins above Cherry Lake and Lake Eleanor (which drain to the Tuolumne River), the Merced above both Happy Isles and Pohono, the upper West Walker basin, and the Bear Creek and Pitman Creek basins (both small tributaries of the San Joaquin River). The first four basins are wholly or partially within the boundaries of Yosemite National Park; the Merced and Tuolumne basins compose the majority of the park's alpine wilderness areas. The West Walker basin is located just outside the park on the eastern slope of the range, while the Bear and Pitman basins are located further south. The basins feature mixes of coniferous forests, meadows and high alpine open areas.

Topographical boundaries of the basins are delineated based on 30 m United States Geological Survey (USGS) topographical data, which are also used to calculate basin area distributions within 100 m elevation bands. The basins' elevations range from approximately 1000-4000 m above sea level. The basin areas, elevation distributions and mean streamflow over the 1981-2006 study period are shown in Table 3.2.

30

3.3.3.2. Streamflow observations

We use daily streamflow observations from USGS stream gauges at Happy Isles and Pohono on the Merced River in Yosemite Valley (USGS Gauges 11264500 and 11266500). These gauges provide a long-term record of unimpaired runoff from the upper Merced basin, with the Happy Isles basin nested within the Pohono basin. We also obtained USGS gauge data for the West Walker, Bear and Pitman basins (USGS Gauge Nos. 10296000, 11230500 and 11237500). All gauges are listed in the Hydro-Climate Data Network of basins without significant dams or diversions [*Slack and Landwehr*, 1992]. Uncertainties in the streamflow measurements from the rating curve have been estimated at 10% or less [*Rockwell et al.*, 1996a], though studies of other rating curve-based streamflow techniques have suggested uncertainties of up to 20-40% at very high flows [Baldassarre and Montanari, 2009].

Because the Tuolumne, Cherry, and Eleanor basins drain to reservoirs that are operated for water supply and hydroelectric generation, we use reconstructed full natural flows from the reservoir operators (provided by Bruce McGurk, personal communication, 2013). The full natural flows estimate the discharge in the absence of the dams, based on recorded reservoir releases and water levels. The uncertainties associated with reconstructed flows are less well known, but are assumed to be generally similar to those of standard streamflow observations. Because the adjacent Cherry and Eleanor Lakes are connected via a pumping tunnel with uncertain flow, we summed their watershed areas and discharge time series to create a "combined" basin, herein referred to as "Cherry-Eleanor."

3.3.3.3. Precipitation and temperature observations

We use precipitation observations from three mid-elevation gauges (Figure 3.2): two Cooperative Observer Network (COOP) stations operated by the San Francisco Public Utilities

Commission (Hetch Hetchy and Cherry Valley Dam), and one United States Forest Service

Remote Automatic Weather Station (RAWS) site (Buck Meadows). The gauges' mean annual

precipitation over 1981-2006 were 960, 1303 and 935 mm, respectively, and their elevations are

1180, 1453 and 976 m above sea level. The gauges are manually maintained during the winter

and are at elevations where snow is less likely to cause significant measurement biases; few

reliable, long-term, high-elevation records are available in the study area. We use an average of

the three gauges' daily precipitation as the forcing series for the hydrologic model. Averaging

different gauges' records may result in temporal smoothing, but given our inference timescales

of a year or longer, we assume this did not impact our results.

Daily high and low temperatures were also available from the three precipitation sites.

We average the high and low temperatures at each site to create a daily mean temperature, and

then average the three stations' mean temperatures, to create one temperature forcing time series.

Temperatures in the models are lapsed to the elevation bands from the average elevation of the

three sites.

3.3.3.4. PET forcing data

The FUSE models require a time series of basin-mean PET. PET is estimated using the

Makkink equation, which uses temperature and shortwave radiation as predictors [*e.g.*, *Cristea et

al.*, 2013]. Daily high and low temperatures from the forcing sites are used, and the Bristow-

Campbell parameterization is used to estimate shortwave radiation from temperature [*Bristow

and Campbell*, 1984]. Daily PET is estimated at each elevation band assuming a constant lapse

rate (-6.5°C km^{-1}), and then basin-mean PET is aggregated from the bands.

We calibrate the Bristow-Campbell and Makkink methods using data from the Dana

Meadows meteorological station (2987 m above sea level, Figure 3.2) for water years 2003

through 2009. Shortwave radiation and temperature are used to calibrate the Bristow-Campbell coefficients A (0.7) and C (2.49). Shortwave radiation, wind speed, temperature and relative humidity are applied to estimate PET using the Penman-Monteith equation, following the methods of Allen et al. [1998]; this PET estimate is then used to calibrate the Makkink coefficient (0.668). In this way, we estimate PET over the long term using daily temperature observations alone. PET averaged 995 mm yr^{-1} over the Tuolumne basin during 1981-2006.

3.3.3.5. Soils

We use information about typical basin soil properties to provide prior information for the BATEA inference of the model parameters. The basins are near the crest of the Sierra Nevada and have steep topography that is underlain by granitic bedrock. Typical soils are shallow and sandy, with a depth of no more than 1 m and a porosity of about 0.4 [*NRCS*, 2007]. Field studies have indicated that the soil has a high conductivity typical of sand and gravel, but that the percolation rate into the underlying bedrock is much lower [*Flint et al.*, 2008]. A modeling study of the upper Merced basin indicated that the soils' conductivity is likely at least 1 m d^{-1} and that field capacity is likely no higher than 0.2-0.25, in order to match observed streamflow patterns [*Lundquist and Loheide*, 2011].

3.3.3.6. PRISM precipitation data

In order to obtain an alternative estimate of basin-mean precipitation in the region, we use the PRISM climatological precipitation product [*Daly et al.*, 2008], specifically the 1981-2010 normals at 800 m resolution. For each basin, all PRISM cells with centers within the watershed boundaries are used to find the basin-mean annual precipitation. The PRISM cells containing each of the three precipitation gauges are also averaged, and the PRISM ratio of basin-to-gauge precipitation is calculated. For each basin, the observed precipitation from the

gauges is then scaled by this ratio, thus providing a PRISM-based estimate of the basin-mean precipitation. This approach allows us to use the 1981-2010 PRISM normal over our slightly different study period (1981-2006) without bias, while still retaining the spatial information from the normals.

In addition to precipitation gauges, PRISM normals also make use of long-term snow courses' April 1 SWE observations in this portion of the Sierra (Chris Daly, personal communication, 2015). April 1 SWE is translated into precipitation by distributing it among the winter months, with some correction for melt and sublimation. Thus the PRISM precipitation values are derived from both precipitation gauges and snow observations in this region.

3.3.4. Case study methodology

3.3.4.1. Model validation

First, we conduct a validation of the parameter inference by using distributions inferred from the first half of the study period (water years 1982-1995) to simulate the second-half validation period (water years 1996-2006). We calculate the Nash-Sutcliffe model efficiency coefficient and the percentage of timesteps within the 90% model confidence interval during the latter period, as means of validating the model's predictions, confidence limits and error model assumptions. The validation serves as a check that the model can replicate observed streamflow, before examining inferred precipitation.

3.3.4.2. Inferring long-term mean annual precipitation

Separate from the calibration-validation experiment, we infer long-term basin-mean precipitation rates over the study period of water years 1982-2006 for each basin. The time-invariant soil, snow and precipitation parameters are inferred. The long-term precipitation multiplier and OPG are inferred for each basin in this way, resulting in a posterior PDF of basin-

mean precipitation. The relatively long 25-year inference period is used to avoid errors associated with anomalous water years. Water year 1981 is used for model spin-up and is not included in the calibration period. Only one model structure (FUSE-070) is considered at this stage.

3.3.4.3. Inferring time series of annual precipitation

Following the inference of the long-term precipitation, we infer time series of basin-mean annual precipitation for each basin over the same 1982-2006 period. We apply a two-step calibration approach: First, the soil and snow model parameters are inferred from the long-term inference. Then, these values (including the OPG) are used as fixed quantities when inferring annual variability in the basins' precipitation multipliers. Only the multipliers are allowed to vary with each water year, but they are held constant within the year. While in practice the OPG and lapse rate may vary from year to year, in this case they are fixed to limit the dimensionality and computational burden of the inference, and to improve the identifiability of the precipitation multipliers [*e.g.*, see *Renard et al.*, 2010].

We also examine the behavior of the approach under different hydroclimatic conditions. We do this by comparing inferred precipitation, ET and runoff ratios in certain wet (2005 and 2006) and dry (1987) years of the study period.

3.3.4.4. Sensitivity of inferred precipitation

We also assess the sensitivity of the inferred precipitation to assumptions made during the inference process, including model structure, soil depth, PET forcing dataset, and year order of forcing data.

To examine the impact of the choice of PET forcing data estimation method, we consider several approaches. Above we describe the Bristow-Campbell/Makkink approach for estimating

35

PET. One alternative approach utilizes pan evaporation measurements made at the Hetch Hetchy

Reservoir during summers from 1955-1978. We regress pan evaporation on daily high and low

temperatures and precipitation at the site. This model is used to extend the PET estimate to the

entire study period. In order to reflect the likely decrease in PET with elevation, the regression is

run at each 100 m elevation band, using a -6.5°C km^{-1} lapse rate to adjust temperatures from the

observation sites' elevation. The basin-averaged PET is then aggregated from the bands. Under

these assumptions, the Tuolumne basin averages 738 mm yr^{-1} of PET, with near-zero wintertime

values and summertime maxima of 5-6 mm d^{-1}. Finally, another approach would simply be to

use the pan evaporation regression time series without introducing the uncertainty of scaling for

change with elevation. The average PET of this approach is 1629 mm yr^{-1} for Tuolumne; while

this is likely an overestimate of basin-average PET, it allows us to test the method's sensitivity to

PET uncertainty.

To examine sensitivity of the inference to year order, we conduct alternate simulations in

which the forcing and streamflow series are altered. In one, the order of the water years is

reversed (2006 to 1982), and in another, the years are arranged in a random order. Because the

total precipitation and streamflow remains constant, and yearly carryover of SWE and soil

moisture is very small in these basins, the inferred long-term precipitation should not

significantly change between these calibrations.

3.4. Results

3.4.1. Model calibration and validation

We check the validation of the calibrated model parameters to see if the models are able

to robustly simulate streamflow in the basins. Figure 3.3 shows a subset of the calibrated and

validated simulation of the Tuolumne basin using the FUSE-070 model structure, with

parameters inferred from 1982 to 1995 used to simulate 1996 to 2006 (only water years 1995 and

1996 are shown for clarity). The simulation has Nash-Sutcliffe values of 0.90 for the calibration

period and 0.83 for the validation period. In both the calibration and validation periods, 96% of

the timesteps fall within the 90% model confidence interval (shaded area in Figure 3.3). While

this indicates that the confidence limits are slightly overestimated, the diagnostics do not suggest

any major problems in the calibration; similar levels of model performance are obtained for the

other basins. Importantly, the inferred multipliers all fall well within the prior range of [0.1 2.0],

indicating that the choice of these bounds did not impact the results.

3.4.2. Long-term basin-mean inferred precipitation

The inferred mean annual precipitation over the study period (water years 1982-2006)

from the FUSE-070 structure is shown in Figure 3.4. Figure 3.4a shows histograms of the

inferred basin-mean precipitation from the BATEA inference. The Tuolumne and the Merced at

Pohono and at Happy Isles have relatively similar inferred precipitation (mean of the BATEA

histogram): 1105, 967 and 1030 mm yr^{-1}, respectively. Pitman has an inferred precipitation rate

of 884 mm yr^{-1}. However, Cherry-Eleanor is inferred to be significantly wetter, with a mean

precipitation rate of 1648 mm yr^{-1}. The West Walker and Bear basins are drier, with mean

precipitation rates of 792 and 856 mm yr^{-1}, respectively. The uncertainty of the precipitation

estimates, as indicated by the mean of standard deviation of the sample histograms of the seven

basins, is ±10 mm yr^{-1}.

Figure 3.4b is a map of inferred precipitation for each basin. Within-basin precipitation

variability is not resolved in this experiment; inferred OPG within each basin is very close to

zero. This was likely due to the forcing stations being near the precipitation maximum on the

west slope of the Sierra Nevada and the linear OPG formulation, which would have produced very large precipitation rates for the highest elevation bands if $\alpha \neq 0$. As a result, each basin is shown with spatially-uniform precipitation rates. Using only basin-mean values, however, it is possible to see two precipitation gradients: a north-to-south, wet-to-dry gradient along the west slope of the Sierra Nevada, and an east-west gradient, in which basins that are rain-shadowed or in the lee of the crest receive less precipitation than those on the west slope.

To illustrate the inference of precipitation from streamflow, we compare the inferred precipitation for each basin against its mean annual streamflow (Figure 3.5a). The basins' inferred precipitation fell above the 1:1 line and increased in a linear fashion with streamflow, as would be expected given that precipitation is inferred from streamflow. The vertical difference between the 1:1 line and the precipitation estimate also suggests the inferred ET rate for each basin; the points imply a runoff ratio of 0.7-0.8 and basin-average ET of 200-400 mm yr^{-1}. To compare the inferred precipitation against another estimate of basin-mean precipitation, in Figure 3.5b we plot them against PRISM estimates for each basin. The positive slope of the points indicates that both PRISM and the inferred precipitation agree on the order of the basins' precipitation. However, the inferred precipitation is less than precipitation derived from PRISM in the Tuolumne (16% less than PRISM), Merced at Pohono (26% less), Merced at Happy Isles (21% less), West Walker (27% less), Bear (31% less) and Pitman (30% less) basins, while in Cherry-Eleanor inferred precipitation is 14% higher than PRISM. The variability between basins is higher than in PRISM, with 855 mm yr^{-1} separating the wettest and driest basins as inferred from streamflow, compared to 346 mm yr^{-1} in PRISM.

Finally, we plot PRISM against observed streamflow (Figure 3.5c). While there is also rank-order agreement between the two, the implied PRISM actual ET may conflict with the land

cover characteristics of some basins, in particular in Cherry-Eleanor. Implied PRISM actual ET

is 157 mm yr^{-1} in this basin, of which evergreen forests cover 25%, and shrubland, grassland,

wetlands and open water cover another 45%, according to the National Land Cover Database

2011 [*Jin et al.*, 2013]. While studies of actual ET are limited, measurements from flux towers

[*Goulden et al.*, 2012] and of conifer sap flow [*Kurpius et al.*, 2003] suggest that Sierra

evergreen forests have actual ET of 350-850 mm yr^{-1}. Estimates based on meteorological data

suggest Southern Sierra forests have actual ET of 275-350 mm yr^{-1} [*Stephenson*, 1998, their

Figure 3.6]. Thus, the land cover mix for Cherry-Eleanor suggests that PRISM precipitation is

too low in this basin, at least assuming that the runoff data are not strongly biased. Additionally,

in the West Walker and Bear basins, the comparisons of PRISM precipitation and streamflow

imply >500 mm yr^{-1} actual ET. These basins have similar levels of forest cover but are at higher

mean elevations than Cherry-Eleanor; it is not obvious why actual ET would be much greater in

these basins.

3.4.3. Inference of yearly precipitation

Figure 3.6 shows inferred yearly precipitation using the FUSE-070 model structure. The

basins' yearly precipitation is strongly correlated. The basins' inferred precipitation also roughly

tracks observed gauge precipitation (Figure 3.6, dashed line). The coefficient of variability of the

inferred annual precipitation series is about 0.4 for each basin, consistent with findings of

significant inter-annual variability in California. The order of the basins from wet to dry is

relatively consistent from year to year, but not completely so, which suggests that the relative

spatial pattern of precipitation does vary at the yearly timescale.

3.4.4. Case studies in annual variability: 1987, 2005 and 2006

To further examine the inferred precipitation, ET and streamflow, we focus on three years: 1987, the driest year in the study period, and 2005 and 2006, two-above average precipitation years with certain unique aspects (dashed boxes, Figure 3.6). First, we compare the inferred ratios of ET and streamflow in 1987 vs. 2005 and 2006, to assess how the model partitions precipitation under dry and wet conditions. In 1987, inferred precipitation averages 497 mm across the seven basins, while inferred ET averages 237 mm, with a mean modeled runoff ratio of 52%. In 2005 and 2006, the mean inferred precipitation values are 1268 and 1531 mm, the mean ET values are 352 and 364 mm, and the runoff ratios are 73 and 77%, respectively. Inferred ET increases linearly in years with greater observed streamflow, though not as rapidly as inferred precipitation, leading to inferred higher runoff ratios in wetter years.

Next, we examine differential spatial patterns between 2005 and 2006. In 2005, the Gin Flat snow pillow had relatively greater peak SWE than the pillow at Tuolumne Meadows, while in 2006, the order was reversed (Figure 3.7, upper row). The relative patterns in 2005 and 2006 were also reflected in the observed streamflow of the Tuolumne and the Merced at Pohono: the two basins had similar streamflow volumes in 2005, but in 2006 the Tuolumne had approximately 15% greater streamflow volume (Figure 3.7, middle row). The Gin Flat pillow is at lower elevation, and temperature differences between the two years could have led it to receive more of its precipitation as rain in 2006. However, an examination of the pillows' SWE and temperature records shows that SWE diverged in January, and temperatures at that time do not suggest that Gin Flat received rain instead of snow. Thus, the spatial patterns of precipitation appear to have differed between the two years.

We then normalized the inferred time series of precipitation, to produce annual estimates of basin precipitation as a percentage of the 1982-2006 average. In 2005, it is inferred that precipitation in the basin of the Merced at Pohono was 122% of average, while in the Tuolumne basin precipitation was 108% of average (Figure 3.7, bottom left). In 2006, both basins received more precipitation; it is inferred that the Tuolumne received 138% of normal, while the Merced received 139% of normal (Figure 3.7, bottom middle). Thus, the increase in precipitation between 2005 and 2006 was more pronounced in the Tuolumne basin (Figure 3.7, bottom right). The Merced received 103 mm more precipitation than the Tuolumne in 2005, but the Tuolumne received 68 mm more in 2006.

The reversal of the order of the precipitation between the two basins over these two years is consistent with the observations of snowpack. Additionally, the Bear basin has the largest increase in inferred precipitation relative to normal from 2005 to 2006; the spatial shift may reflect a greater increase in precipitation in 2006 for higher-elevation basins closer to the crest of the Sierra Nevada. Notably, this shift is reflected neither in the precipitation gauges used for forcing the models (Figure 3.6), nor in the PRISM monthly 4 km gridded precipitation maps for the region, both of which estimate that 2005 had more precipitation than 2006.

3.4.5. Method sensitivity and uncertainty

3.4.5.1. Sensitivity to model structure

To investigate the effects of structural uncertainty in the FUSE models, we repeat the long-term calibration for each basin using five additional FUSE model structures, and then examine the range of the inferred precipitation results. The inferred precipitation from the six structures is shown in Figure 3.8a. The vertical range of points for each basin suggests that the inferred precipitation uncertainty associated with the choice of model structure has a standard

41

deviation of ± 275 mm yr^{-1} – much larger than that inferred from the individual calibrations. Some model structures (-014, -016, -170) produce consistently higher inferred precipitation estimates than others (-070, -072, -160). Across all model structures, higher ET is associated with higher precipitation. Neglecting storage terms in (1), it follows that precipitation should be approximately equal to observed streamflow plus modeled ET. Therefore, the sensitivity of the inferred precipitation to model structure is likely due to the variability in modeled ET between the different structures.

3.4.5.2. Sensitivity to soil capacity

To examine the impact of uncertainty in soil depths on inferred precipitation, we repeat the calibration with a range of upper bounds on the uniform prior distribution on the maximum upper-level soil water capacity. We test three different soil water capacities (200, 800 and 1600 mm) in addition to the best-estimate value (400 mm) used in all other experiments herein. The two higher values are considered unlikely given the regional soil types (primarily exposed bedrock interspersed with alpine meadows), but are tested to establish sensitivity. The inferred precipitation variability from a range of soil capacities is shown in Figure 3.8b. Increasing the soil capacity is shown to increase inferred precipitation across all basins. Presumably, greater soil capacity leads to increased soil water storage and ET, thus requiring greater precipitation to generate the observed streamflow rates. Uncertainty in basin soil depths could be a limiting factor in the ability to infer precipitation from streamflow, but in this case, the largest soil capacity (1600 mm) is implausible given the basins' soil types and geology; excluding this run results in uncertainty in inferred precipitation associated with soil capacity of ± 110 mm yr^{-1}.

3.4.5.3. Sensitivity to PET forcing

The precipitation values inferred using the three different PET forcing series and the FUSE-070 model structure are shown in Figure 3.8c. The primary Bristow-Campbell/Makkink approach is labeled as PET_0, the scaled pan evaporation is PET_{lo}, and the unscaled pan evaporation is PET_{hi}. Variations in inferred precipitation between PET forcing runs have an average standard deviation of ± 76 mm yr^{-1}. The uncertainty associated with PET is somewhat less than that associated with soil capacity, and much smaller than that associated with model structure. Given the relatively thin soils in the basins and the prolonged dry summer season, limitations on water availability likely prevent actual ET from increasing significantly with PET.

3.4.5.4. Sensitivity to year order

The results of the year-order sensitivity experiment are shown in Figure 3.8d. The differences in the inferred precipitation between simulations for the same basin in which the water years are in actual, backwards and random order had standard deviation of ± 82 mm yr^{-1}. The exact mechanism for the calibration's dependence on year order is currently not known. The calibration uncertainty is smaller than that associated with model structure, but it may suggest the limit of precision that can be obtained with this approach.

3.4.6. Precision of annual precipitation estimates

The preceding analyses suggest substantial uncertainties associated with inferring climatological (1982-2006) basin-average precipitation rates. However, the magnitudes of both spatial (between basins) and temporal (between years) variance in precipitation are large in this region. We test whether the signals of year-to-year variability in basin precipitation inferred from streamflow are robust, by repeating the inference for each of the six FUSE structures. Year-by-year inferred multipliers for the Tuolumne basin are shown in Figure 3.9a. The multipliers have been normalized by the 1982-2006 mean for each structure in order to highlight yearly

variations, rather than biases between structures. The high temporal correlation of the multipliers across model structures (average correlation coefficient of +0.87) suggests that these variations are driven by robust annual variability in streamflow and precipitation. For comparison, the average correlation coefficient among the seven basins' annual streamflow volumes over this period was +0.97.

The inferred precipitation rates for each year from 1982-2006 in the Tuolumne basin are shown in Figure 3.9b, plotted against the observed streamflow for the year. The six model structures are shown, along with a best-fit line relating inferred precipitation to observed streamflow in each model. The case study years of 1987, 2005 and 2006 are shown in dashed boxes. The varying slopes of the best-fit lines reflect the implied runoff ratios; each structure tends to behave linearly in inferring more precipitation in years with more streamflow. We can also compare the uncertainty in the inferred precipitation between wet and dry years. For 1987, the range of the six model structures' inferred precipitation is 424 mm or 59% of the mean, while for 2005 and 2006, it is 793 and 838 mm, or 54 and 45% of the mean, respectively. This suggests that in drier years, relative uncertainty in inferred precipitation is greater, though absolute uncertainty is less than in wet years.

3.5. Discussion

3.5.1. Precision and applicability of the precipitation-from-streamflow approach

The sensitivities discussed above suggest that there is substantial uncertainty in precipitation inferred from streamflow. In particular, uncertainties associated with model structure appear to be the limiting factor of this approach. However, some improvement may be made by favoring model structures which appear to better simulate the basins than others. For

example, Nash-Sutcliffe coefficients were not equal across the model structures: the high-ET and high-precipitation models averaged 0.72, while the low-ET and low-precipitation models averaged 0.83. The calibrated soil moisture time series for structures -014 and -170 (not shown) show persistently high soil moisture relative to other model structures even in summer, as well as higher inferred field capacity parameters. This suggests that in order to correctly simulate the surface runoff dynamics, these model structures require higher soil moisture, and thus produce actual ET of greater than 600 mm yr^{-1}.

Previous modeling has suggested that basin-mean actual ET is approximately 250-400 mm yr^{-1} for the Merced at Pohono [*e.g.*, *Lundquist and Loheide*, 2011]. Point-scale measurements of actual ET from flux towers located south of the study basins showed actual ET over forested canopies of 350-850 mm yr^{-1}, decreasing at higher elevations [*Goulden et al.*, 2012]. However, no direct observations of basin-mean actual ET exist; thus, it is difficult to exclude model structures on this basis alone.

Reducing model structural uncertainty may be aided by the inclusion of other data. Qualitative indicators, such as the match between modeled and observed seasonal patterns of soil moisture, or comparisons between inferred ET and minimum ET needed for known land cover types [*Stephenson*, 1998; *Kurpius et al.*, 2003; *Goulden et al.*, 2012], may be useful in this regard. Additionally, over the Sierra, observations of SWE are made at many snow courses and pillows [*Rice and Bales*, 2010; *Meromy et al.*, 2013], snow depth and snow covered area are observed via remote sensing (*e.g.*, MODSCAG [*Painter et al.*, 2009] and the Airborne Snow Observatory [*Painter et al.*, 2015]), and SWE may be reconstructed from remotely-sensed snow disappearance date [*Rittger et al.*, 2011; *Raleigh and Lundquist*, 2012]. Numerical weather models now produce spatially-distributed precipitation fields at sufficient resolution to resolve

the study basins [*Wayand et al.*, 2013]. Thus, approaches exist to reduce uncertainty in basin-mean precipitation.

A comparison of the uncertainty from the precipitation-from-streamflow approach with that from commonly-used gridded precipitation datasets reveals that errors of a similar magnitude may sometimes be present in the gridded products. For example, disagreement between gridded products such as PRISM and precipitation inferred from snowpack measurements may be on the order of hundreds of mm per year in certain mountain ranges and under certain wind conditions [*Gutmann et al.*, 2012; *Lundquist et al.*, 2015]. In the context of these issues, the precipitation-from-streamflow approach may provide independent information about precipitation patterns that does not depend on the sparse precipitation gauge network.

Additionally, it appears possible to robustly infer annual variations in precipitation from streamflow in these basins, despite the significant uncertainty and biases between the different model structures. While each model structure had a different bias regarding ET and thus inferred precipitation, the biases were consistent from year to year, allowing for robust inference of years with more or less precipitation from streamflow observations (Figure 3.9). Overall, the precipitation-from-streamflow approach appeared capable of detecting year-to-year shifts in spatial precipitation patterns, which may not be as well resolved in other distributed precipitation products.

The approach defined here depends on the availability of unimpaired or full natural flow data from basins which have a relatively straightforward runoff response to water input. Both of these restrictions may limit its use to certain areas. Basins with gentler slopes, deeper soils and more significant groundwater dynamics would require more complex models in order to infer precipitation from streamflow, which would likely result in greater uncertainty in the results.

While it is reasonable to apply this approach only to certain basins, the types of basins where this approach may work best are also water supply regions across the Western United States (*e.g.*, the Sierra Nevada and Colorado headwaters). The method also provides further motivation for USGS to maintain mountain basin stream gauges.

3.5.2. Comparison with findings of Koskela et al. [2012]

Koskela et al. [2012] inferred precipitation multipliers in a snow-affected basin in Finland. They defined multipliers for individual storm events, and used a simultaneous inference of snow, soil and time-varying multiplier parameters, using both streamflow and snow observations as calibration targets. In their study, multipliers for each storm event could not be robustly identified. Our results contrast, in that precipitation multipliers are robustly identified, with the posterior distributions clustered around unique values for each water year, despite our vague prior distribution.

Several factors may have led to this outcome. First, we use a longer timescale for each multiplier (annual vs. several-day storm event), which may provide a more distinct runoff signal for identification of the multipliers. Second, we choose study basins with steep slopes and thin soils, such that there would be a direct runoff response from rain and snowmelt. Another difference is that we use an error model with fewer degrees of freedom and no temporal autocorrelation, which was shown to be more robust in *Evin et al.* [2014].

In contrast to the join calibration approach used by *Koskela et al.* [2012], we use a two-step calibration approach to infer time series of precipitation multipliers for each water year for each basin, in which the hydrologic model parameters were inferred first, and then held fixed as the time-varying multipliers were inferred in the second step. The two-step calibration was not the cause of the difference between our study and *Koskela et al.* [2012], however. We performed

a test of a simultaneous inference of the snow and soil parameters and the series of annual multipliers, and the results were qualitatively similar to the two-step calibration.

3.6. Conclusions

Estimating basin-mean precipitation in complex terrain is difficult due to uncertainty in the representativeness of precipitation gauges relative to the basin. Inadequate distribution with elevation and spatial density can lead to large estimation errors in basin-mean precipitation. Streamflow offers additional information about the water balance of the basin with which to estimate precipitation. We infer basin-mean precipitation rates using a Bayesian methodology which calibrates a hydrologic model with precipitation multipliers. We apply this approach to a cluster of basins in the snow-dominated Sierra Nevada of California, inferring both the long-term average precipitation for each basin, as well as time series of annual precipitation using a two-step calibration approach.

The application of the precipitation-from-streamflow approach yields the following major conclusions:

1) Precipitation inferred from streamflow using multiple nearby basins appears to appropriately reflect the climatic gradients of the region.

2) Precipitation inferred from streamflow has greater variability from basin to basin than does PRISM-scaled precipitation. In fact, comparing PRISM precipitation to streamflow implies PRISM runoff ratios that, in some study basins, may be climatically implausible. Incorporating streamflow in the development of spatially-distributed precipitation datasets provides an independent source of hydrologic information for use in these products.

3)   Long-term average precipitation have uncertainties with standard deviations of ±110 and ±275 mm yr^{-1}, associated with soil depths and the model structure, respectively. Inferred precipitation is less sensitive to the method used to estimate the PET forcing.

4)   Despite these uncertainties, year-by-year inferences of precipitation show robust signals across different model structures. This suggests that inferring precipitation from streamflow may be useful in discerning changes in spatial patterns of precipitation at annual and basin scales.

5)   Times series of precipitation multipliers inferred for each water year in the study can be robustly identified, even when using a relatively vague prior distribution for the multipliers. This suggests that in basins with thin soils and a pronounced dry season, the precipitation-from-streamflow approach is likely to be robust at the annual timescale.

3.7. Appendix: Description of FUSE model structures

We use six FUSE model structures: FUSE-014, -016, -070, -072, -160 and -170. FUSE-070 (Figure 3.10, upper left) has the fewest number of states and parameters; it has single upper and lower zone storage, evapotranspiration only from the upper layer, percolation based on the water content above the wilting point (which is considered zero water storage) in the upper zone, and a single linear baseflow reservoir. FUSE-072 (Figure 3.10, bottom left) is identical to FUSE-070, except that the baseflow reservoir discharges with a power-law relationship with storage, rather than a linear one.

FUSE-016 (bottom middle) is very similar to FUSE-072, except that the percolation between upper and lower zones is based on storage above the amount of soil water held in tension, thus potentially limiting percolation. The fraction of soil water held in tension, as compared to free storage, is a tunable parameter in FUSE, somewhat similar to the field capacity of the soil.

 FUSE-014 and FUSE-170 are similar to FUSE-016, but differ in their choice of the upper layer architecture and the evapotranspiration parameterization. FUSE-014 (upper middle) has an upper zone with two cascading buckets of tension storage, and a separate store of free water (storage above field capacity). Evapotranspiration in FUSE-014 is sequential, in which the upper zone first is first depleted before the lower zone. FUSE-170 (bottom left panel) also features the cascading buckets in the upper tension zone, but evapotranspiration is weighted between the two zones based on fraction of roots between the zones, a tunable parameter. This differentiates FUSE-014 and FUSE-170 from FUSE-016, which permits evapotranspiration only from the upper storage. Finally, FUSE-160 (upper right) differs from FUSE-016 based on the lower zone architecture; FUSE-160 has two parallel linear baseflow reservoirs, allowing for more complex baseflow behavior, and parameterizes ET using the same root-weighting scheme as FUSE-170.

All models calculate saturation excess surface runoff via a Pareto distribution related to storage in the upper zone. None of the model structures permit interflow, the transfer of water from the upper storage zone directly to the stream; thus, all modeled streamflow is generated either by surface runoff or baseflow.

3.8. Tables

Table 3.1. Parameters information for the FUSE-070 model structure.

Parameter	Category	Description	Units	Lower bound	Upper bound	Source
FRACTEN		Fraction soil water stored in tension	-	0.05	0.75	[*Lundquist and Loheide*, 2011]
MAXWATR1		Maximum storage in the upper soil zone	mm	1	400	[*NRCS,* 2007]
MAXWATR2		Maximum storage in the lower soil zone	mm	1	400	[*NRCS,* 2007]
QB_PRMS	Soil	PRMS baseflow exponent	-	0.001	0.25	[*Clark et al.,* 2008]
PERCRTE		Percolation rate	mm d^{-1}	100	10000	[*Lundquist and Loheide*, 2011]
PERCEXP		Percolation exponent	-	1	20	[*Clark et al.,* 2008]
AXV_BEXP		ARNO-Xiang-VIC B exponent	-	0.001	3	[*Clark et al.,* 2008]
TIMEDELAY	Routing	Streamflow routing delay time	d	0.01	5	[*Clark et al.,* 2008]
RFERR_MLT		Precipitation multiplier	-	0.1	2	[*Milly and Dunne,* 2002]
OPG	Meteorology	Orographic precipitation gradient	km^{-1}	0.01	2.5	[*Dettinger et al.,* 2004; *Lundquist et al.,* 2010]
LAPSE		Temperature lapse rate	°C km^{-1}	-7	-4	[*Lundquist and Cayan,* 2007]
MBASE		Base snowmelt temperature	°C	0	3	[*Anderson,* 2006]
MFMAX		Maximum snowmelt factor	mm d^{-1} °C^{-1}	2	8	[*Anderson,* 2006]
MFMIN	Snow	Minimum snowmelt factor	mm d^{-1} °C^{-1}	0	6	[*Anderson,* 2006]
PXTEMP		Rain-snow partition temperature	`°C	-1	3	[*Anderson,* 2006]

Table 3.2. Characteristics of Yosemite-area streamflow basins used to infer precipitation.

Basin Name	Area [km^2]	Area Elevation Distribution Percentiles [m]			# 100 m bands	1981-2006 Mean Streamflow [mm yr^{-1}]
		10th	50th	90th		
Tuolumne R. at Hetch Hetchy	1181	1998	2789	3238	30	879
Merced R. at Pohono	834	2097	2548	3142	29	762
Merced R. at Happy Isles	470	2184	2779	3265	28	723
Cherry L.-L. Eleanor	492	1681	2310	2801	20	1287
West Walker R.	469	2171	2674	3067	16	570
Bear Cr.	136	2683	3234	3614	20	681
Pitman Cr.	60	2181	2335	2655	10	738

3.9. Figures
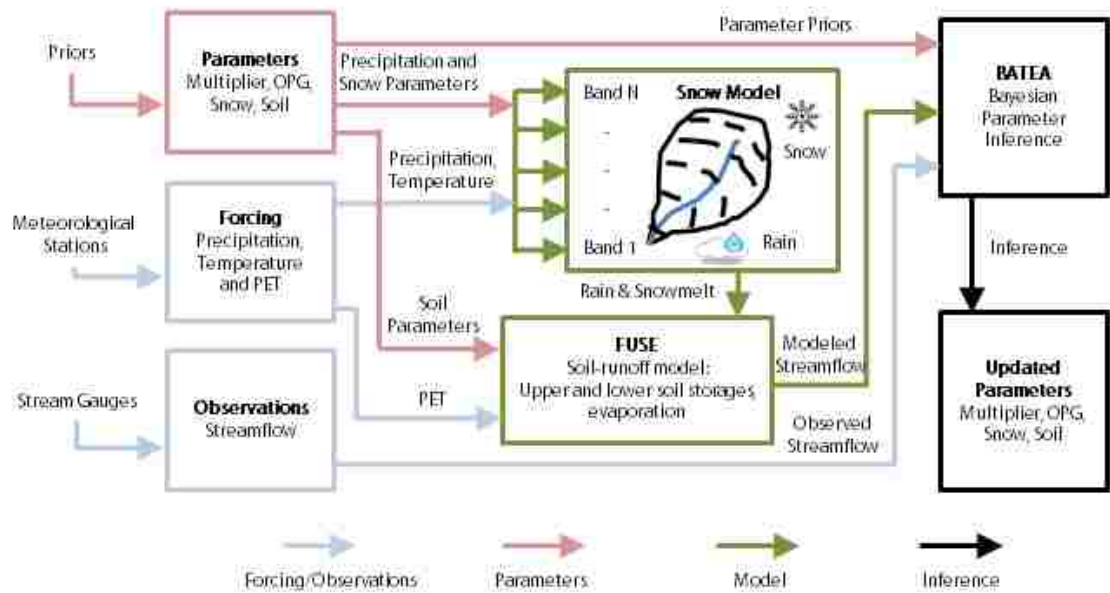
Figure 3.1. Schematic of information flow in the calibration routine using the snow model, FUSE and BATEA. Forcing data from meteorological stations are combined with parameters such as the precipitation multiplier, the OPG, and soil and snow parameters, to simulate basin snowpack and runoff. The simulated runoff is then compared with observed values, using a Bayesian approach that incorporates prior information. The result is a posterior distribution of parameters, including the multiplier and OPG, which describe basin-mean precipitation inferred from streamflow.
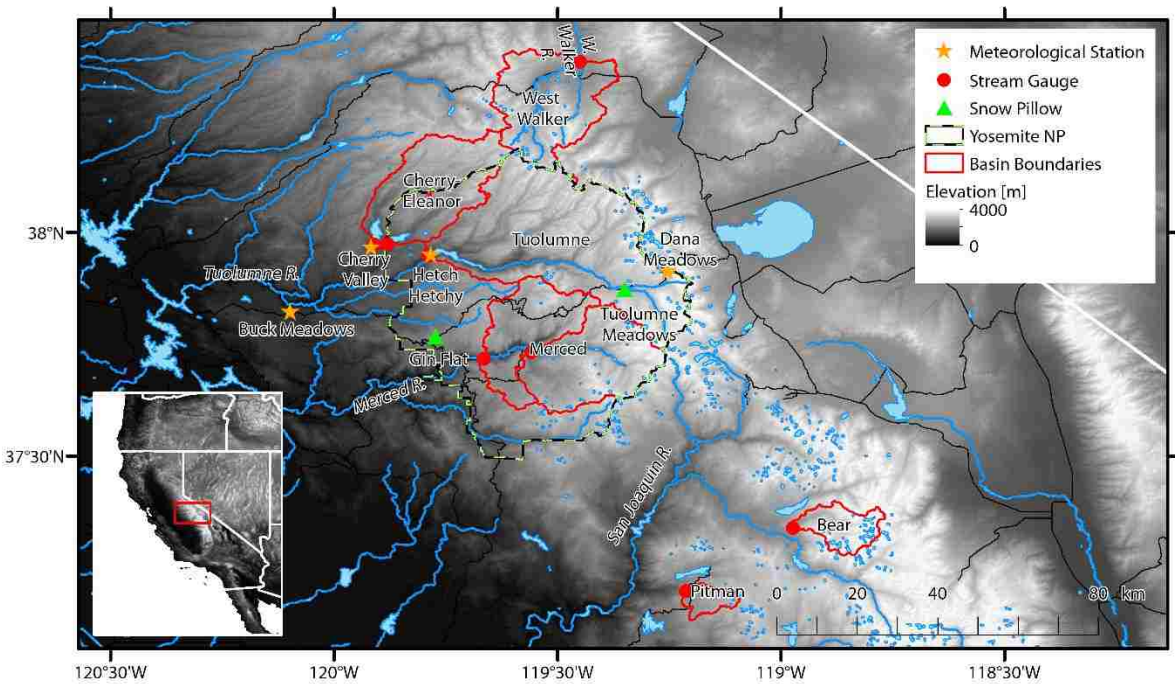
Figure 3.2. Topographical map of the Yosemite area, showing the Cherry-Eleanor, Tuolumne, Merced, West Walker, Pitman and Bear basins. The larger of the two nested Merced basins is referred to as the Merced at Pohono; the smaller is the Merced at Happy Isles. The stream gauges used in the inference and the three mid-elevation precipitation gauges used to force the model are shown. A high-elevation site measuring temperature, wind, and shortwave radiation used for PET calibration (Dana Meadows), and two snow pillows (Gin Flat and Tuolumne Meadows) are shown as well.
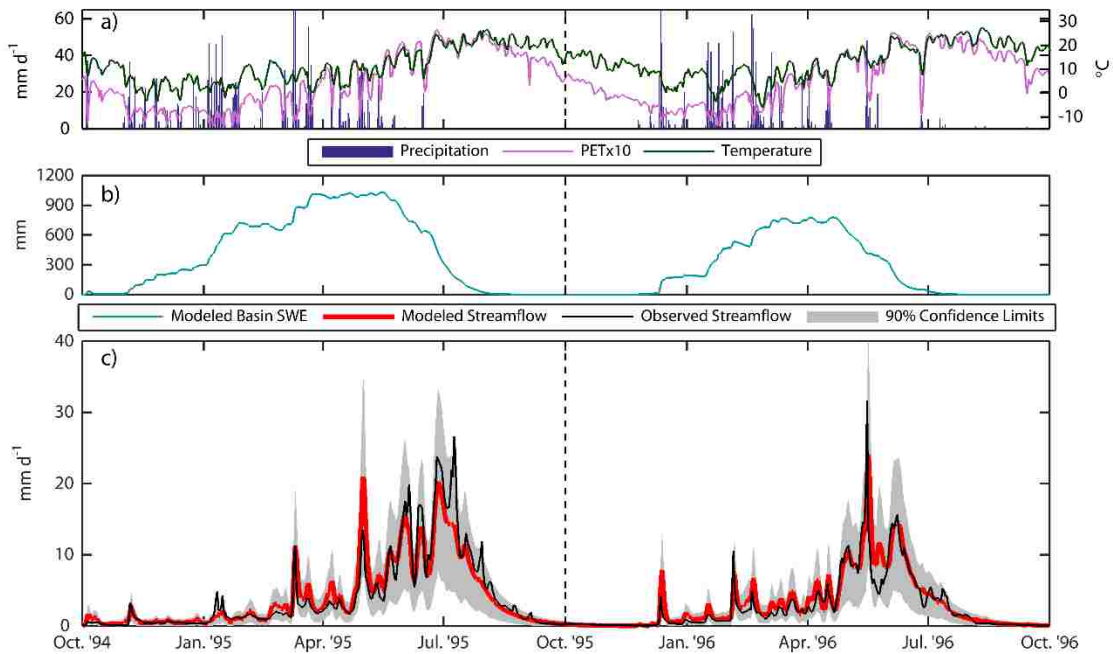
Figure 3.3. Example of the calibration and validation of the models used in the inference for the Tuolumne basin. For clarity, only water years 1995 (last year of calibration) and 1996 (first year of validation) are shown. a) Forcing data: precipitation (left axis), PET (left axis; shown tenfold for visibility) and temperature (right axis). b) Modeled basin-mean SWE using the mean BATEA parameter set. c) Modeled streamflow using the mean BATEA parameter set, compared against observed streamflow. The 90% confidence limits of the BATEA parameter set are also shown for comparison.
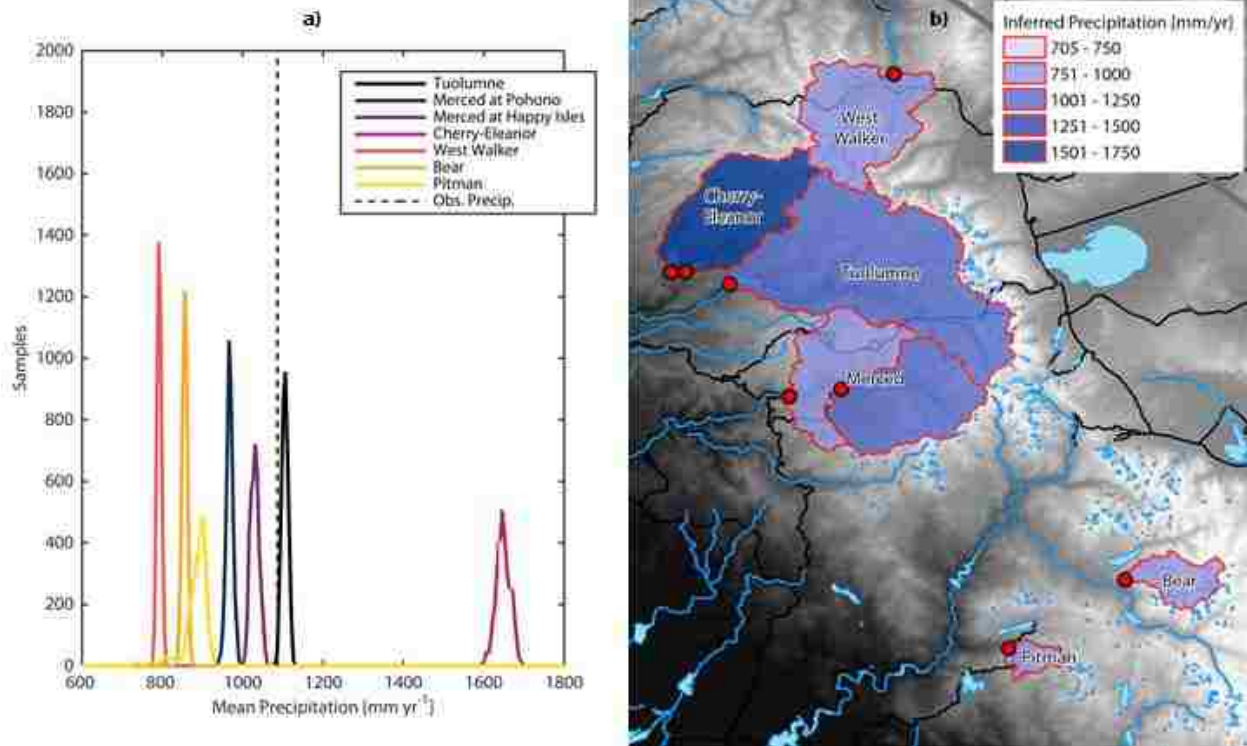
Figure 3.4. Inference of long-term basin-mean precipitation. a) 1982-2006 average inferred precipitation for each basin, shown as PDFs of annual precipitation, as well as the mean from the three precipitation gauges (dashed line). b) Map of inferred 1982-2006 precipitation.
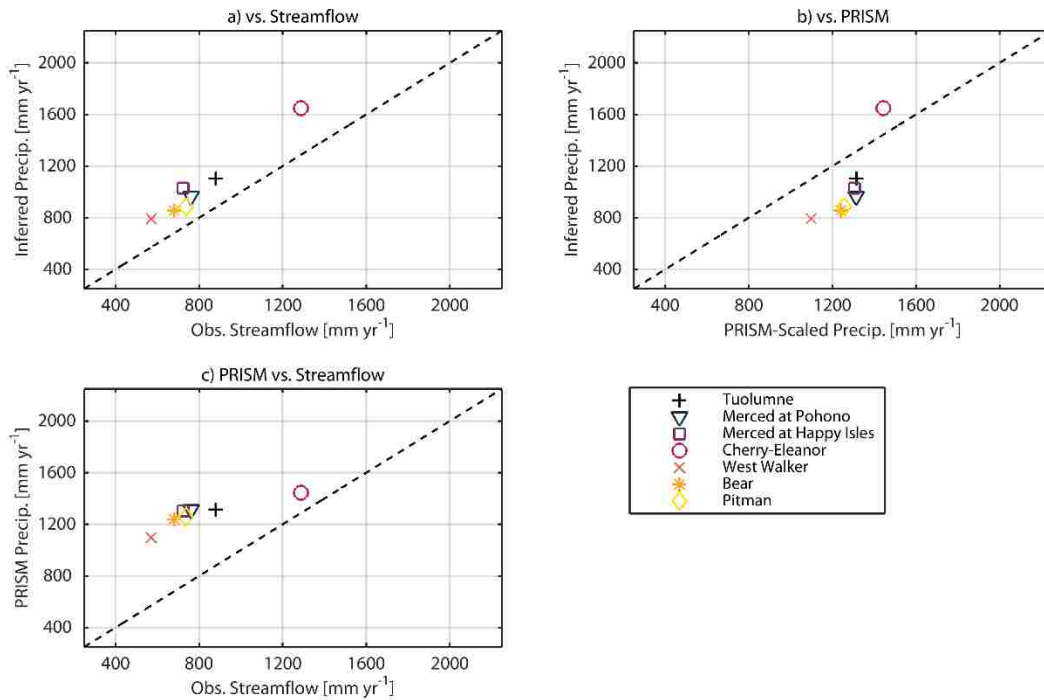
Figure 3.5. Comparison of inferred climatological precipitation to other basin hydrologic indicators. a) Observed streamflow. b) Precipitation derived from PRISM 1980-2010 climatological weights. c) PRISM-derived precipitation plotted against observed streamflow.
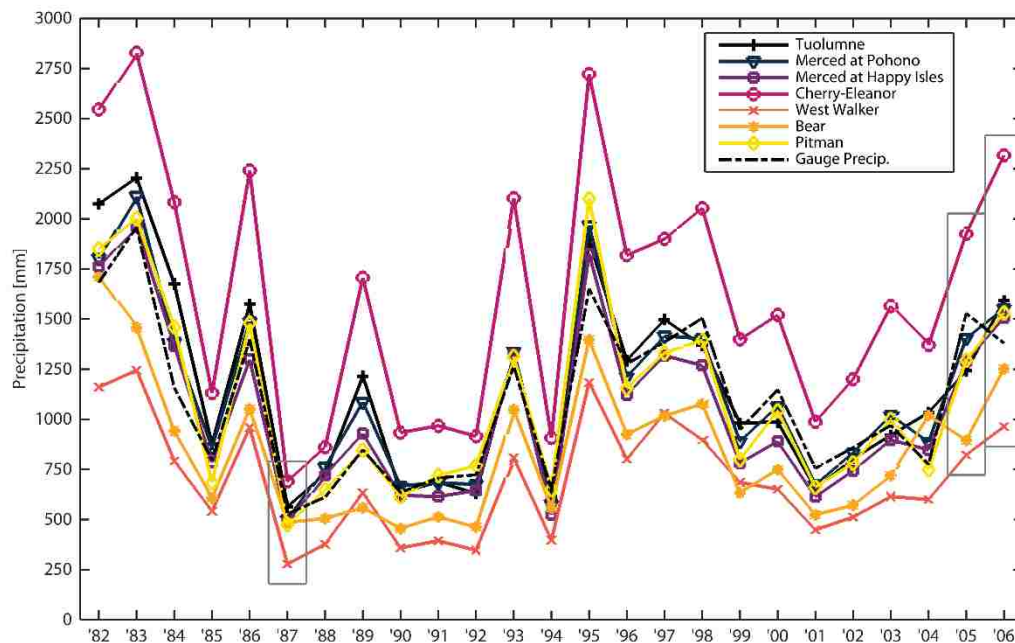
Figure 3.6. Annual precipitation inferred using the FUSE-070 model structure for the seven basins. Observed annual gauge precipitation is shown as the dashed line.
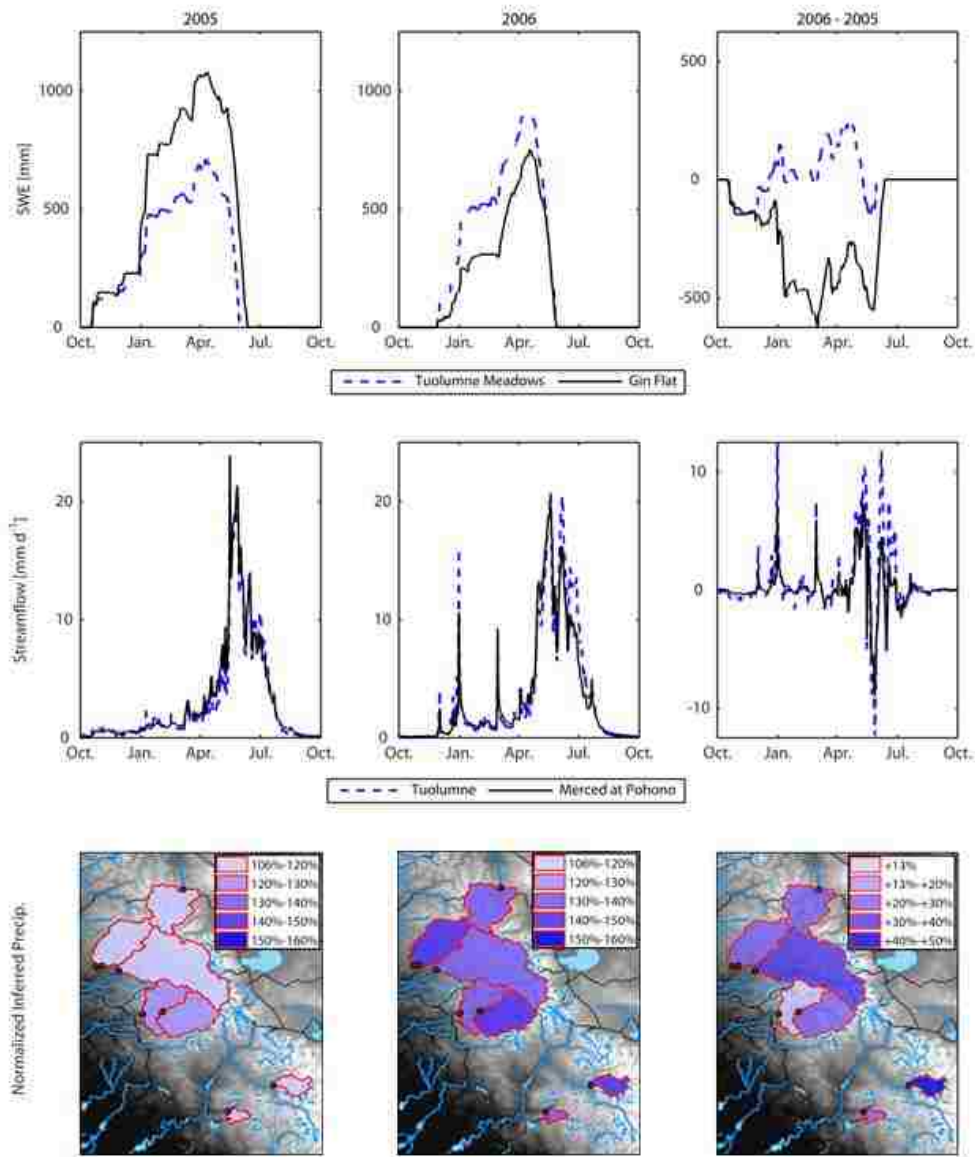
Figure 3.7. Comparison of snowpack, streamflow and inferred precipitation between 2005 and 2006. Left column shows the results for 2005; center column shows results for 2006, and right column shows 2005 subtracted from 2006. The top row shows observed snowpack at Gin Flat and Tuolumne Meadows. The middle row shows observed streamflow in the Tuolumne and the Merced at Pohono. The bottom row shows inferred precipitation from the FUSE-070 model structure, normalized by the 1982-2006 mean for each basin. The order of the Tuolumne and Merced basins in terms of snowpack, streamflow and inferred precipitation is reversed between 2005 and 2006.
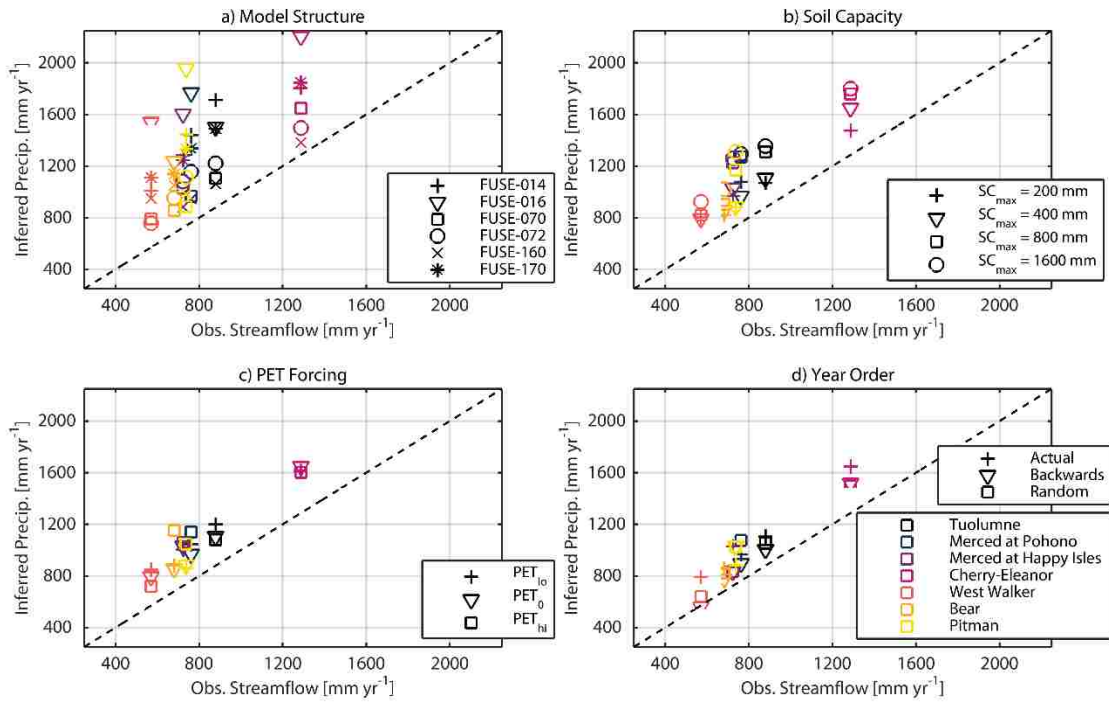
Figure 3.8. Sensitivities of inferred precipitation to various sources of uncertainty. a) For each basin, inferred precipitation corresponding to the six FUSE model structures. b) Inferred precipitation from each basin with four different upper-zone soil capacities. c) Potential evapotranspiration forcing: three forcing series with different magnitudes of average PET. d) Sensitivity of inferred precipitation to the order of years in forcing data. "Actual" refers to the true observations, while "Backwards" and "Random" show alternative scenarios in which the order of the water years 1982-2006 has been changed.
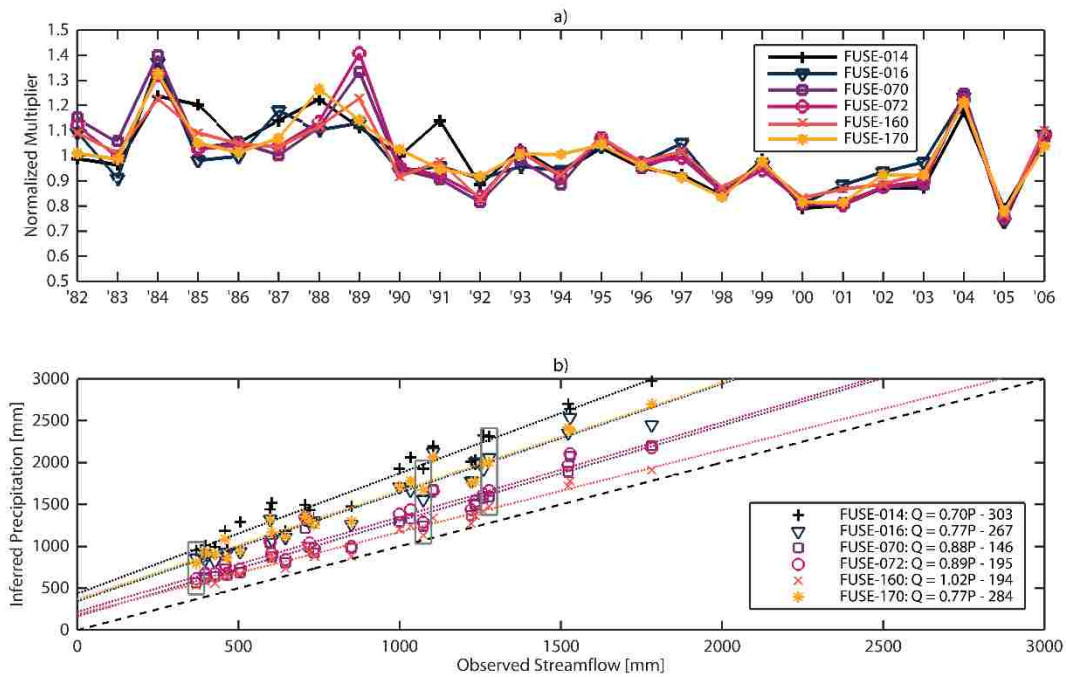
Figure 3.9. Annual inference in the Tuolumne basin. a) Precipitation multipliers inferred for each year from 1982-2006. Multipliers have been normalized by the mean 1982-2006 multiplier for each model structure. b) Inferred yearly precipitation plotted against observed annual streamflow for each model structure. Best-fit lines and accompanying equations show the response of each model structure to streamflow and precipitation.

60
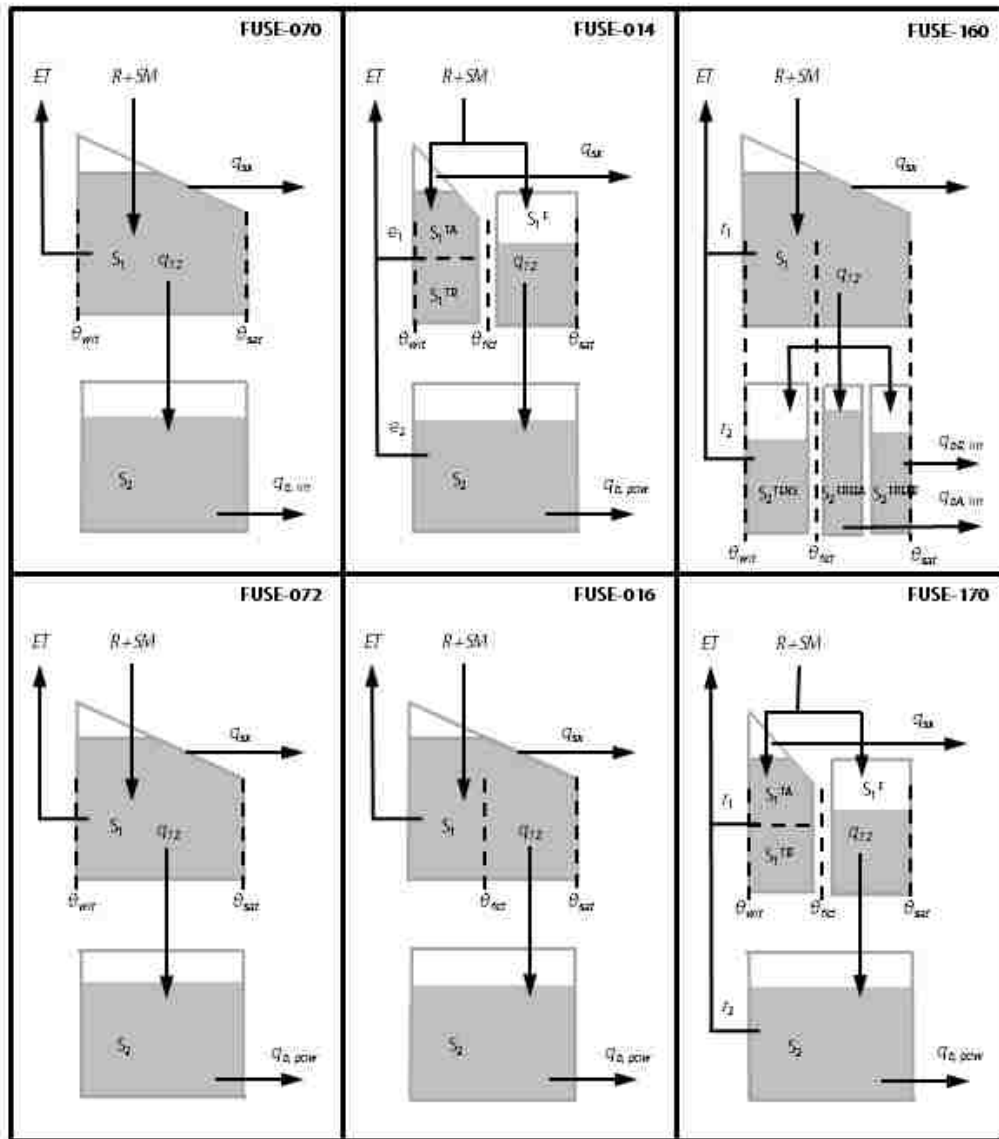
Figure 3.10. (Appendix Figure 1) Schematics of the FUSE conceptual rainfall-runoff models used in this study. Rain plus melt from the snow model (R+SM) enters the soil storages. Upper and lower storages (S_1 and S_2) are shown using boxes. Fluxes are shown with arrows; q_{12} is percolation from the upper to lower layer, q_b is baseflow, q_{sx} is surface runoff, and ET is evapotranspiration. Dashed lines indicate soil storage thresholds; θ_{wlt} is wilting point, θ_{fld} is field capacity, and θ_{sat} is saturation.

Chapter 4. Combining snow, streamflow and precipitation gauges observations to infer basin-mean precipitation

Abstract: Precipitation data in mountain basins is typically sparse and subject to substantial uncertainty due to difficulties in measurement and due to significant spatial variability. Streamflow provides indirect information about basin-mean precipitation, but the inference of precipitation from streamflow requires assumptions about hydrologic model structure that influence the precipitation amounts. In this study, we test the extent to which the use of both snow and streamflow observations reduces uncertainty in inferred precipitation compared to precipitation inference from streamflow alone. The case study area is the upper Tuolumne River basin in the Sierra Nevada mountain range of California, where distributed and basin-mean snow water equivalent (SWE) observations have recently been made available via the Airborne Snow Observatory (ASO). To extend our estimate of basin-mean SWE to periods prior to the ASO campaign, we compare recent SWE estimates from ASO with point observations from snow courses and snow pillows. This comparison shows that point estimates of SWE in the Tuolumne region tend to overestimate SWE at a given elevation, but undersample high-elevation areas. We then compare the inferences obtained with multiple hydrologic model structures. When included in precipitation inference, snow observations reduce differences in the resulting precipitation amounts by up to half, and improve the consistency between hydrologic model structures in terms of the yearly variability of precipitation. Some hydrologic model structures appear better suited to simulate both streamflow and SWE correctly. We reiterate previous findings that multiple data types be used in calibration of simplified hydrologic models to improve the robustness of model predictions.

62

4.1. Introduction

Precipitation observations across high-elevation and complex terrain are often highly uncertain due to a lack of gauges, high spatial variability and problems with undercatch and wintertime gauge maintenance [*Milly and Dunne*, 2002; *Sieck et al.*, 2007; *Daly et al.*, 2008; *Rasmussen et al.*, 2011b]. Gridded precipitation datasets over complex terrain necessarily reflect this uncertainty, as evidenced by differences between and within gridded datasets (chapter 2 and *Newman et al.* [2015]), between gridded precipitation and precipitation estimated from snow pillows [*Lundquist et al.*, 2015], and between gridded precipitation and precipitation from gauges not included in the grid interpolation [*Gutmann et al.*, 2012; *Livneh et al.*, 2014]. As a result, additional types of observations may be necessary to supplement precipitation gauge observations and reduce uncertainty in spatially distributed precipitation in complex terrain.

Streamflow observations have been used to infer precipitation patterns and amounts over mountain basins [*Adam et al.*, 2006; *Weingartner et al.*, 2007]. These earlier studies attempted to infer multi-year average precipitation rates from runoff, but other studies have attempted to infer precipitation from streamflow at shorter timescales with varying degrees of success. *Koskela et al.* [2012] found that precipitation from individual storm events was difficult to robustly infer in a snow-affected basin in Finland, while *Le Moine et al.* [2015] were able to infer daily precipitation amounts from streamflow and snow observations using the assumption of consistent spatial patterns of precipitation associated with various synoptic conditions. However, most approaches have not explicitly computed the uncertainty of the inferred precipitation.

In chapter 3 [*Henn et al.*, 2015], we tested the robustness of inferring annual and long-term precipitation rates from streamflow in snow-dominated basins in the Sierra Nevada mountain range of California. We used lumped hydrologic models and multipliers on the annual

precipitation forcing data to infer basin-mean precipitation by calibration against daily streamflow observations. This approach was generally able to identify year-to-year variability in precipitation in such basins, and identified potential disagreements between gridded precipitation climatology and precipitation inferred from streamflow. However, *Henn et al.* [2015] also showed that the inferred precipitation amounts strongly depend on model structural assumptions about evapotranspiration (ET) and soil moisture fluxes.

Calibrating only to streamflow in snow-dominated basins potentially allows for large model variability in the simulation of snow water equivalent (SWE). In these basins, both streamflow and SWE are major components of the water balance, and so it is reasonable to expect that a joint calibration to both types of observations would constrain models better than calibration to a one type alone. For example, *Hood and Hayashi* [2015] found that using both snowpack and runoff observations in an alpine basin allowed for inference of other quantities of the water balance, such as soil storage.

Using snow observations in addition to streamflow observations is especially applicable in the snow-dominated basins of the Western United States, where a network of pillows and snow courses make regular observations. Measurement of snowpack at individual points in the mountains Western United States has a long and successful history, beginning with early snow surveys in the 1930s [*Church*, 1933], and continuing through the current set of snow courses and the Natural Resources Conservation Service Snow Telemetry (SNOTEL) network of over 700 automated snow pillows [*Serreze et al.*, 1999]. Point measurements have allowed near real-time estimation of snowpack and improved skill in seasonal streamflow prediction.

Nonetheless, the spatially variable nature of mountain snowpack challenges our ability to quantify it at the basin scale, which is necessary for the calibration of lumped hydrologic models.

Point measurements of snowpack only begin to sample the spatial variability at a wide range of scales of SWE across the mountain landscape. Due to variability from topography, wind, precipitation, radiative forcing, vegetation and other factors, snow courses and pillows generally do not reflect the amount of SWE in areas around them, even at relatively small spatial scales [*Molotch and Bales*, 2005; *Clark et al.*, 2011b]. These biases are not always positive or negative, nor do they remain constant between the accumulation and ablation seasons or across different water years [*Molotch and Bales*, 2005; *Rice and Bales*, 2010; *Meromy et al.*, 2013]. Further, practical constraints prevent installation and maintenance of snow sensors in the highest-elevation areas, which have some of the deepest and most persistent snowpacks. Thus, in situ networks tend to oversample mid-elevation SWE and report melt earlier in the year as compared to higher-elevation areas [*Rice et al.*, 2011]. All of these factors make it challenging to directly estimate basin-mean SWE from point observations alone.

Given the unpredictable biases in using point measurements of SWE to estimate basin-mean SWE, much effort has been devoted to remote sensing of snow. For example, LiDAR provides an airborne or ground-based means of measuring snow depth with high spatial resolution (~1 m) across relatively large domains (a few to thousands of km^2) [*Deems et al.*, 2013]. Remote sensing approaches such as LiDAR allow for fully sampling the spatial distributions of snowpack across the landscape. Since 2013, the Airborne Snow Observatory (ASO) has employed LiDAR to regularly observe the complete distribution of snow depth over several basins in the Sierra Nevada of California during the snow ablation season [*Painter et al.*, 2015]. When combined with modeled snow density, a spatially-distributed snapshot of SWE is produced. Importantly, ASO allows for a comparison of traditional point estimates of snow

(pillows and courses) to spatially-distributed observations, potentially allowing us to retrospectively assess the representativeness of point-derived estimates of basin-mean SWE.

This study has two primary aims. First, using the Tuolumne River basin above Hetch Hetchy Reservoir as a testbed, we investigate spatially-distributed snow observations. We compare basin-mean SWE from ASO data to basin-mean SWE estimated from snow pillows and courses. We hypothesize that a robust relationship may exist between point-based SWE indicators and spatially-distributed ones, and so we compare these two types of observations over water years 2013-2015. We then use this approach to develop a long-term time series of basin-mean SWE using point observations alone.

Our second goal is to assess the value of multiple types of observational data in hydrologic model calibration. To do this, we use snow observations as an additional calibration target for a set of hydrologic models of the Tuolumne basin. We use also different sources of SWE observations, and test the hypothesis that inferring precipitation based on both SWE and runoff will yield more robust results when compared to conducting the inference based on streamflow alone. We use a multi-model framework to assess the degree to which hydrologic model consistency is improved via calibration to additional types of observations.

Section 4.2 reviews our previous study inferring precipitation from streamflow in the Sierra Nevada, which this work builds upon. Section 4.3 describes the hydrologic models and the inference routine used to estimate hydrologic model parameters and basin-mean precipitation. Section 4.4 describes multiple types of data used in this study: meteorological forcing data, streamflow data, prior information about model parameters, and ground- and LiDAR-based observations of SWE. Section 4.5 describes the model calibration experimental approach. Section 4.6 compares the two types of SWE observations, and then presents the results of the

precipitation inference over both a single water year (2014), and over a long-term run (water years 1982-2014). Section 4.7 discusses the implications of the comparison of point and distributed SWE on estimating basin-mean SWE, and of the different precipitation inference scenarios, and section 4.8 concludes with the main findings.

4.2. Background

Chapter 3 [*Henn et al.*, 2015] showed that when the precipitation input is allowed to vary, different lumped model structures can produce nearly equally satisfactory reproductions of observed streamflow, while having substantially different internal model states and fluxes. Figure 4.1 shows example data from that study, in which water year 1983 precipitation is inferred from runoff data in the basin of the Tuolumne River above Hetch Hetchy Reservoir. Using the Framework for Understanding Structural Errors (FUSE, [*Clark et al.*, 2008]), time series of SWE and streamflow were generated by calibrating six different model structures to streamflow observations (Figure 4.1a). There are large differences between the simulated SWE amounts in each of the model structures. This is the result of substantial differences in the simulation of soil moisture and ET between the model structures, and as such, their inferred amounts of water-year precipitation disagree by more than 1,000 mm for this water year (Figure 4.1b). However, the six structures all provide at least an adequate match to streamflow in terms of Nash-Sutcliffe coefficient (captions, Figure 4.1b). The differences in inferred SWE and precipitation are due to different structural assumptions in the lumped hydrologic models (Figure 4.1c).

The preceding appears to be an example of model equifinality, in which disparate model assumptions (structure, parameters, etc.) match observations more or less equally well after

67

calibration [*Beven and Binley*, 1992]. Equifinality can hinder process understanding and model predictions. In addition, the example suggests that inferring precipitation from both snow and streamflow observations may reduce the uncertainty in inferred precipitation: the divergence of the modeled SWE in Figure 4.1a presumably reflects biases in the model structures that can be reduced by calibration of the models against SWE observations. Calibration of models to both streamflow and snowpack has been shown to improve process representation and robustness [*e.g.*, *Finger et al.*, 2015].

4.3. Hydrologic models and calibration approach

Figure 4.2 shows the information flow in this study's approach. To infer precipitation from streamflow and SWE observations, we assume that the precipitation forcing data used to drive the hydrologic model are corrupted by error, due to spatial unrepresentativeness or other factors. Following the approach of *Kavetski et al.* [2003] and *Renard et al.* [2010, 2011], we use precipitation multipliers that attempt to correct the precipitation gauge observations to better represent basin-mean precipitation. For a given set of parameter values, which includes the precipitation multiplier and hydrologic model parameters, a corrected estimate of basin precipitation is routed through snow and hydrologic models to produce simulated runoff and SWE time series. Using the Bayesian Total Error Analysis [BATEA, *Kavetski et al.*, 2006a, 2006b] calibration framework, we compare the simulated and observed streamflow and SWE (section 4.3.4), and the goodness of fit to both targets is used to establish the likelihood of the parameter set. The inferred basin-mean precipitation amount is then the product of the inferred multiplier and the original precipitation gauge record. A more detailed description of the precipitation inference routine is found in *Henn et al.* [2015].

4.3.1. FUSE conceptual rainfall-runoff models

We use FUSE model structures (shown schematically in Figure 4.1c; [*Clark et al.*, 2008]) to provide a set of conceptual hydrologic models and to help evaluate the effects of model structural uncertainty on inferred precipitation [*Clark et al.*, 2011a; *McMillan et al.*, 2011b]. This is accomplished by calibrating multiple FUSE structures to the same basin forcing data, and then evaluating differences in the inferred basin-mean precipitation.

FUSE provides multiple options for representing soil moisture storage and fluxes, such as ET, surface runoff and baseflow, using a spatially lumped approach with upper and lower soil zones. In this study, we use the six model structures previously used to investigate the impact of structure on model performance in the Mahurangi basin in New Zealand [*Clark et al.*, 2011a; *McMillan et al.*, 2011b] and to infer precipitation from streamflow in chapter 3 [*Henn et al.*, 2015]. The six structures have different upper- and lower-zone storage architectures and fluxes (Figure 4.1c):

- FUSE-016, -070 and -072 are the simplest, with a single storage in each zone, but they differ in their baseflow and percolation parameterizations;

- FUSE-014 and -170 use a two-storage upper zone to represent free and tension water;

- FUSE-160 subdivides lower zone storage into tension storage and two parallel reservoirs;

- ET parameterization varies: FUSE-014, -160 and -170 allow ET from both zones in different fashions, while FUSE-016, -070 and -072 allow ET only from the upper zone.

4.3.2. Snow model

FUSE is coupled to a temperature index snow model based on Snow-17 [*Anderson*, 2006], which tracks SWE based on precipitation and melt. Precipitation is added to SWE if the

air temperature is below a threshold; otherwise, it is treated as rain. Snowmelt is initiated if the air temperature is above a threshold, and is proportional to the temperature above that point:

$$SM = \max(\min[f\{T - T_0\}, SWE], 0) \qquad (4.1)$$

where SM is snowmelt per timestep, f is the snowmelt factor, T is the air temperature, and T_0 is the melt threshold temperature. The melt factor f varies in a sinusoidal manner over the year, with minimum and maximum values on the winter and summer solstices, respectively. The rain-snow partition temperature, melt initiation temperature, and winter and summer melt factors are all inferred parameters. Snowmelt is combined with rainfall and provided as input to the hydrologic model.

In order to simulate the strong elevation dependence within the basin, the snow model is discretized over 100 m elevation bands (unlike the hydrologic model, which is lumped over the entire basin). The use of elevation bands is generally necessary to reproduce basin-mean SWE and summertime streamflow recession in mountain basins [*Luce and Tarboton*, 1998; *Clark et al.*, 2011b].

For each elevation band, forcing temperature is lapsed to the *i*th band midpoint elevation:

$$T_i = T_f + \Gamma(z_i - z_f) \qquad (4,2)$$

where T_i is the *i*th band temperature, T_f is the forcing temperature, z_i is the band midpoint elevation, z_f is the forcing data elevation, and Γ is the temporally-invariant temperature lapse rate (<0, °C km^{-1}), which is also inferred.

Similarly, precipitation is distributed to each band via a multiplicative scaling of the forcing precipitation:

$$P_i = MP_f(1 + \alpha[z_i - z_f]) \qquad (4.3)$$

where P_i is the ith band precipitation, P_f is the forcing precipitation, M is the gauge-to-basin precipitation multiplier, and α is the orographic precipitation gradient (OPG) in km[-1], following *Lundquist et al.* [2010]. The negative lapse rate and the non-negative OPG ensure that the higher bands are colder and receive at least as much precipitation as the forcing elevation, allowing the model to simulate high-elevation snowpack that can last into the summer.

Finally, water fluxes to the lumped soil model are calculated by a weighted average of rain plus snowmelt from each elevation band:

$$F_{soil} = \sum_{i=1}^{n_{bands}} AF_i(R_i + SM_i) \qquad (4.4)$$

where F_{soil} is the flux of water to the soil model, and AF_i, R_i and SM_i, respectively, are the fraction of the basin area, the rain and the snowmelt, all from band i.

Full snow and hydrologic model details can be found in *Henn et al.* [2015].

4.3.3. Precipitation inference using BATEA

4.3.3.1. Parameter priors

To constrain inference within the Bayesian framework, prior distributions are placed on the multiplier and each hydrologic and snow model parameter. We use the same set of parameter priors as in *Henn et al.* [2015]: in general, uniform distributions are used, and when no information is available about a parameter, the distribution is set to the default limits for the FUSE model [*Clark et al.*, 2008]. We set the prior distribution of the precipitation multiplier to be uniform over 0.1 to 2.0. This represents high uncertainty in the precipitation gauge observations as they relate to basin-mean precipitation, which is because of the paucity of high-elevation gauges in the study watershed and the use of forcing gauges that are at substantially lower elevation than the basin midpoint. Thus, we elect to use vague priors in order to avoid biasing the results with our choice of prior bounds.

4.3.3.2 Multi-response error model

Within BATEA, error model are used to translate differences between observed and simulated model responses (streamflow and SWE) into the likelihood of the corresponding parameter sets. For streamflow, we use a weighted least squares error model that assumes that higher streamflow values are subject to greater uncertainty. Following *Thyer et al.* [2009], we assume that residual streamflow errors follow a Gaussian distribution with zero mean and a standard deviation that depends linearly on simulated streamflow:

$$\sigma_{Q,t} = a_Q \hat{Q}_t + b_Q \tag{4.5}$$

where a_Q and b_Q are error model parameters and $\hat{Q}_t$ is the modeled streamflow at timestep t. Similarly, the BATEA calibration of the model to SWE assumes that errors are normally distributed, and with a standard deviation, σ_{SWE}. To obtain the overall goodness-of-fit, BATEA sums the squares of the errors from both streamflow and SWE, each normalized by their relative standard deviations (σ), and compares this sum against that of other parameter sets.

Thus, the error model parameters a_Q, b_Q and σ_{SWE} dictate the relative weight of each type of observation in the calibration. Increasing $\sigma_{Q,t}$ would reduce the weight of streamflow errors, effectively worsening the model fit to streamflow and improving fit to SWE; the effect of increasing σ_{SWE} would be the reverse. We set the three error model parameters based on exploratory calibrations in which we sought to balance the fit to both observed time series; we use $a_Q = 0.4$, $b_Q = 0.2$ mm d^{-1}, and $\sigma_{SWE} = 25$ mm. Given that we use daily streamflow observations, but yearly (April 1) SWE observations, the standard deviation for SWE errors should be less than that for streamflow errors, in order to produce an appropriate balance of model fit. For typical streamflow (~10 mm d^{-1}) and SWE (~500 mm) magnitudes, the parameters above produce standard deviations of 42% and 5%, respectively, which reflects this balance.

4.3.3.3. Computational methods for BATEA inference

We infer basin-mean precipitation over timescales of a year or longer, given observed streamflow and snow for a basin. The combination of the hydrologic, snow and multiplier parameters results in a parameter space with at least 18 dimensions, and so to posterior sample the joint probability distribution, multiple-start Quasi-Newton optimization and a Monte Carlo Markov Chain (MCMC) routine are used within the BATEA algorithm. The result of the MCMC sampling routine is a posterior ensemble of parameter sets, with more samples clustered around parameter values of greater posterior probability.

4.4. Data

4.4.1. Tuolumne basin topographic and soils data

The basin of the Tuolumne River above Hetch Hetchy Reservoir is located within the boundaries of Yosemite National Park. The basin features a mix of coniferous forests, meadows and high alpine open areas. Topographical boundaries of the basin are delineated based on 30 m United States Geological Survey topographical data, which are also used to calculate basin area distributions over 100 m elevation bands. The basin's elevation ranges from 1,162 m above sea level at Hetch Hetchy Reservoir to 3,997 m above sea level at Mount Lyell. Figure 4.3 shows the basin's boundaries and topography.

We use information about typical basin soil properties to provide prior information for the BATEA inference of the model parameters. The basin is near the crest of the Sierra Nevada and has steep topography underlain by granitic bedrock. Soils are shallow and sandy, with an average depth of no more than 1 m and a porosity of about 0.4 [*NRCS*, 2007], a high

conductivity in the surface layer (>1 m d^{-1}) but very low conductivity in the bedrock beneath, and fairly rapid drainage [*Flint et al.*, 2008; *Lundquist and Loheide*, 2011].

4.4.2. Tuolumne precipitation, temperature and potential evapotranspiration data

We use precipitation observations from two Cooperative Observer Network (COOP) gauges operated by the San Francisco Public Utilities Commission (Hetch Hetchy and Cherry Valley, Figure 4.3). The gauges' mean precipitation over water years 1981-2014 was 923 and 1284 mm yr^{-1}, respectively, and their elevations are 1,180 and 1,453 m above sea level. We use an average of the two gauges' daily precipitation as the forcing series for the hydrologic model. Daily high and low temperatures were also available from the two sites. We averaged the high and low temperatures at each site to create a daily mean temperature series, and then averaged the stations' mean temperatures to create one temperature forcing time series. Temperatures in the models are lapsed to the elevation bands from the average elevation of the two sites.

The FUSE models require a time series of basin-mean potential evapotranspiration (PET). PET is estimated using the Makkink equation, which uses temperature and shortwave radiation as predictors [*e.g.*, *Cristea et al.*, 2013]. Daily high and low temperatures from the forcing sites are used, and the Bristow-Campbell parameterization is used to estimate shortwave radiation from temperature [*Bristow and Campbell*, 1984]. Daily PET is estimated at each elevation band assuming a constant lapse rate (-6.5 °C km^{-1}), and then basin-mean PET is aggregated from the bands. We calibrated the Bristow-Campbell and Makkink methods using data from the Dana Meadows meteorological station from water years 2003 through 2009. In this way, we estimate PET over the long term using daily temperature observations alone. PET averaged 984 mm yr^{-1} in the Tuolumne basin over water years 1981-2014. In chapter 3, we showed that sensitivity of inferred precipitation to PET is relatively small [*Henn et al.*, 2015].

4.4.3. Tuolumne runoff data

Because the Tuolumne basin drains to a reservoir operated for water supply and hydroelectric generation, we use reconstructed full natural flows from the reservoir operators [see supplemental material of *Henn et al.*, 2015]. The full natural flows estimate the discharge in the absence of the dams, based on recorded reservoir releases and water levels. We use a daily streamflow series which begins in October 1980 and extends to early August 2015.

4.4.4. SWE data

4.4.4.1. Point SWE observations

For point-scale observations of SWE, we use a collection of observations from Sierra Nevada snow pillows and snow courses in the vicinity of Yosemite National Park. We use 23 pillows and 40 courses both within and outside of the Tuolumne basin (Figure 4.3). In addition to the Tuolumne basin, the sites are located in the Merced, San Joaquin, Walker, Stanislaus, Mono and Owens drainages. The sites were chosen based on their relatively complete data records for water year 2014, which facilitates comparison with the ASO observations: the courses have at least three monthly observations and pillows have at least 60% of days with observations in this year. Monthly course SWE and daily pillow SWE were obtained the California Data Exchange Center (CDEC, http://cdec.water.ca.gov/). We take the daily mean (for pillows) or monthly mean (for courses) of all sites, to generate a robust indicator of the 1981-2015 temporal pattern of SWE over this region.

We use sites outside the Tuolumne basin because of the small number of sites within the basin: there are four pillows and seven courses in the basin, with fewer generally available during any one year due to missing data. While measurements made outside the basin do not

sample from its SWE distribution, they provide useful information because of the high degree of correlation between sites.

We examined the correlation of April 1 SWE between Sierra Nevada snow courses and pillows within the Tuolumne basin, outside the basin but within the set used here, and outside the set, calculating the mean pairwise correlation coefficient between station pairs with at least 5 years of concurrent data. The mean correlation of snow courses and pillows within the basin was +0.93 and +0.85, respectively. The mean correlation between courses within the basin to those outside the basin but in the set used here was +0.91, compared with +0.84 for courses not used here. For pillows, the equivalent values were +0.85 and +0.79. These values suggest that the sites used here may potential serve as proxies for Tuolumne basin SWE. In addition, high-elevation courses from south of the basin (Figure 4.3) may be useful predictors of un-sampled high-elevation terrain within the Tuolumne basin. For example, snow courses separated by 1° of latitude have a similar correlation (+0.8) to those separated by 500-1,000 m of elevation difference. This supports the use of sites in the upper San Joaquin basin as indicators of high-elevation Tuolumne basin SWE.

4.4.4.2. ASO SWE observations

We use LiDAR observations of snow collected on 27 ASO flights over the Tuolumne basin in water years 2013-2015. The flights begin in February or March of each year and end in June, and occur at approximately weekly intervals. The LiDAR observations cover the entire Tuolumne basin, providing a direct observation of the spatial distribution of snow depth and water equivalent over the landscape. However, the timing of the ASO campaign has largely coincided with the ongoing severe drought in California that began in water year 2012, and so it

is important to remember that all types of SWE observations over this period have values below the long-term average.

For each flight, we use a 50 m resolution SWE product that has been developed by the ASO program. The high-resolution (3 m) LiDAR snow depth data are first aggregated to 50 m resolution. Then a snow model is run over the basin in order to simulate snow density, with the model forced to match the LiDAR depth data and constrained by reliable in situ snow course and snow pillow densities [*Painter et al.*, 2015]. The density and depth data are integrated to give a spatial map of SWE. For each flight, we average the SWE of the 50 m pixels within the basin to give a basin-mean value. We also average the SWE of pixels within each of the 100 m elevation bands in the Tuolumne basin to provide a vertical profile of the SWE distribution.

4.5. Comparison and inference methodologies

4.5.1. Comparison of point and distributed SWE, 2013-2015

We use the sequence of basin-mean and elevation band-mean ASO SWE observations over 2013-2015 to compare against the time series of SWE from the pillows and courses. For example, we consider whether ASO or point observations show greater SWE at different times during the year, as well as the elevation distributions of April 1 SWE. In addition, we note the ratio of April 1 mean course and pillow SWE to the ASO estimates of basin-mean SWE, from the flight closest in time to April 1 each year. We hypothesize that the ASO basin-mean SWE is a much less biased indicator of basin-mean SWE than the mean point measurements, which may not be representative of spatial SWE patterns. Using the April 1 ratio retrospectively, we can develop a time series of basin-mean SWE that is derived from the mean April 1 course SWE, but

77

is multiplied by a correction factor that is based on the ASO comparison from 2013-2015. The corrected time series spans water years 1981-2014.

4.5.2. Inference of water year 2014 precipitation under different calibrations

Using water year 2014 as a case study, we infer precipitation in the Tuolumne basin under the following calibration scenarios: a) streamflow only, b) streamflow and mean April 1 course SWE, c) streamflow and mean April 1 course SWE scaled to match ASO, and d) streamflow and ASO SWE. Water years 2012 and 2013 are simulated but used for model spin-up and data from those years are discarded during the inference. We compare the inferred basin-mean precipitation values across the calibrated model structures; the extent to which the inferred precipitation agrees (or disagrees) indicates the robustness of the model calibrations, and the certainty to which we can infer precipitation using this approach. We hypothesize that the addition of snow information to the calibration will improve the inference of the basin-mean precipitation, which is not well constrained when inferred from streamflow alone (Figure 4.1).

4.5.3. Inference of annual precipitation from 1982-2014

Using similar approaches as in the water year 2014 study, we infer precipitation in the Tuolumne basin for each water year from 1982 to 2014. We conduct the 1982-2014 inference under three calibration scenarios: a) streamflow only, b) streamflow and mean April 1 course SWE, and c) mean April 1 course SWE scaled by the ASO ratio. Water year 1981 is used for model spin-up and is discarded. To infer basin-mean precipitation in individual water years, we use time-varying multipliers, which are assumed to constant within a single water year, but statistically independent across multiple water years. All other model parameters are inferred and assumed to be constant over the simulation.

4.6. Results

4.6.1. Comparison of point and ASO SWE

We first compare SWE observations from point estimates against spatially-distributed ASO SWE. Fig. 4.4a shows the 2013-2015 time series of the snow course and pillow SWE measurements, as well as the basin-mean ASO SWE for each flight. The mean course and pillow measurements are also shown in Figure 4.4a using bold lines. Water year 2013 had the greatest SWE accumulation at the pillows and courses, with water year 2014 showing less SWE, and water year 2015 less still. All three water years had significantly below-average snowpack: the average April 1 SWE over 1981-2014 using the mean of the courses analyzed here was 780 mm.

There is a great deal of SWE variability among the pillows and courses at any given time, but most show peak SWE accumulation around late March or early April in 2013 and 2014, but with an earlier (and smaller) peak in 2015. Cumulative streamflow shows the importance of snowmelt in runoff generation, as streamflow occurs primarily during the period of snowmelt, decreasing sharply once most snow has melted each year; the total streamflow volume is largely determined by the peak SWE for that year.

The ASO SWE series (Figure 4.4a) mainly captures the ablation season, starting around peak SWE and generally showing declines in SWE with each flight, to nearly zero SWE by the time of the last flight in June. In comparison with the pillow and course SWE, ASO shows somewhat less SWE around the time of peak accumulation. On April 1 of each year (dashed vertical line, Figure 4.4a), ASO SWE is 66, 80 and 95% of mean course SWE for the three years, respectively, and averaging the three years produces a mean April 1 ratio of 80%. Despite showing less April 1 SWE, ASO has a slower rate of melt than the pillows and courses during spring. By approximately May 1 of each year, ASO shows more SWE than the point

79

observations, and by the time of the last flight, ASO continues to show non-zero SWE when all pillows and courses have melted out. The differences between the two types of observations likely reflect the different spatial sampling, with ASO covering the entire elevation range while the point observations are limited spatially.

To examine the different profiles of SWE with elevation, we plot them on April 1 of each year (Figures 4.4b-4.4d). Point observations show SWE generally increasing with elevation, though few measurements exist above 3,000 m, and none are made below 2,000 m. However, we note that the pillows and courses show great variability in SWE for a given elevation. The extreme lack of snow in water year 2015 (Figure 4.4d) is also apparent, as nearly half of all pillows and courses had no snow on April 1.

In each year, the mean ASO SWE in the elevation bands increases uniformly with elevation up to a point, though the elevation where SWE begins is much higher in 2015 than in 2013 and 2014. In all three years, the greatest SWE is observed in the 3100 to 3200 m band, with SWE gradually decreasing above this elevation. When comparing ASO and course and pillow SWE, it is apparent that ASO sees less SWE on average than the point observations at a given elevation. For example, 86% of the non-zero point measurements had greater SWE than the mean ASO SWE at that elevation. Assuming that neither ASO SWE nor the point measurements suffer from strong measurement bias, this supports the idea that most courses and pillows in this region are situated at locations that are have anomalously high April 1 SWE for their elevation [*e.g.*, *Rice et al.*, 2011], and that extrapolating point measurements to represent the spatial domain likely leads to overestimation of basin-mean SWE.

Finally, we note the distribution of SWE and measurements relative to the elevation distribution of the Tuolumne basin (dashed line, Figures 4.4b-4.4d). Almost all point

observations occur between the 10th and 85th percentile of the basin elevation distribution; when

only pillows and courses actually within the Tuolumne basin boundaries are considered, only

areas between the 10th and 69th percentile are sampled. The ASO elevation distribution of SWE

shows that the 27% of the watershed area between 3,000 and 3,500 m both receives heavy

snowfall and is not well observed by point measurements.

Based on the findings here, we use a time series of mean April 1 course SWE as one

calibration target, and the same series multiplied by the 80% ASO-course factor as a second

calibration target.

4.6.2. Comparison of model calibrations using streamflow and snow observations

4.6.2.1. Water wear 2014 case study

We now describe the results of using snow observations, in addition to streamflow

observations, to infer water year 2014 precipitation in the Tuolumne basin. As a baseline, we

first calibrate the set of six FUSE model structures using only streamflow observations and

compare simulated SWE and streamflow of the six model structures. Figure 4.5 shows time

series of modeled snowpack for all structures; SWE from the calibration to streamflow only is

shown in dark blue lines. Large variance exists between the different models' SWE, and all of

the models simulate more April 1 SWE than was observed by either ASO or the mean of the

courses.

Next, we consider the calibration of the models to both streamflow and April 1 mean

snow course SWE (Figure 4.5, light blue lines). SWE among the six model structures now tracks

somewhat more closely relative to the streamflow-only case, and is much closer to the SWE

calibration target of 282 mm. This result is not surprising given that the model SWE series are

calibrated to match April 1 observations. When the model is calibrated to match streamflow and

81

the scaled mean of the April 1 course, simulated SWE is reduced further (green lines). A similar result is found when calibrating to streamflow and the ASO April 1 SWE (yellow lines). Note that the calibration targets for these last two cases are very close to one another.

In terms of streamflow, the time series of cumulative runoff for each model structure and calibration scenario are shown in dashed lines in Figure 4.5, along with observed streamflow. Qualitatively, all of the simulated traces match the timing of runoff, with biases in water year total flow of up to 15%. Taken together with the different SWE time series, this suggests that calibration to both types of observations can improve the simulation of snowpack without a substantial degradation of the streamflow simulation.

We can quantify the performance of each calibration scenario by considering the consistency of water year 2014 inferred precipitation and the goodness of fit to both streamflow and SWE. Figure 4.6a shows bar charts of inferred precipitation across the four calibration scenarios, with the six FUSE structures in each scenario. The streamflow-only calibration has a range between model structures of 394 mm in inferred precipitation, relative to a mean of 953 mm. When the models are calibrated to both streamflow and mean April 1 course SWE, the range of inferred precipitation decreases, to 323 mm, as does the mean, to 768 mm. When scaled mean April 1 course SWE or ASO SWE are used for calibration in addition to streamflow, the mean inferred precipitation across the model structures decreases further, to 678 and 698 mm, respectively. The range between structures for these two scenarios is 248 and 336 mm, respectively. Thus, including the SWE calibration target, especially the lower ASO-based targets, reduces the inferred precipitation, and somewhat reduces the variability of inferred precipitation across the multiple model structures.

We can further consider the reliability of the inferred precipitation by considering the different model structures' ability to match streamflow and SWE calibration targets. In terms of streamflow bias (Figure 4.6b), the streamflow-only calibration appears to produce unbiased simulations in terms of streamflow. When SWE is included, negative biases of up to 15% appear, suggesting that the calibration is trading off between matching a lower SWE target and retaining the streamflow skill. The bias in April 1 SWE (Figure 4.6c), however, shows that some model runs substantially overestimate SWE. The positive bias is most severe in the streamflow-only calibration, but also appears in the dual calibrations, mostly with the FUSE-014, -016 and -170 model structures. Simulations that highly overestimate SWE (>30% bias) likely also overestimate precipitation, and if these scenarios are discarded, a much more consistent set of inferred precipitation values results.

Finally, a comparison of streamflow Nash-Sutcliffe coefficients (Figure 4.6d), shows that all scenarios reproduce observed streamflow well (coefficients of 0.85 to 0.95). This suggests that a) streamflow Nash-Sutcliffe coefficient is not a good indicator of whether the models are correctly simulating SWE, and b) adding SWE to streamflow in the calibration does not reduce the ability to simulate streamflow (in fact, there is slight improvement in the streamflow Nash-Sutcliffe coefficients for the three scenarios that include snow information as a calibration target).

4.6.2.2. Inferring annual time series of water year 1982-2014 precipitation

We infer the multiplier and basin-mean precipitation for each water year in the 33-year period under the first three scenarios from the previous section. ASO data are not available prior to 2013, so we cannot use this scenario historically, but instead we use the scaled snow course mean SWE. The yearly multipliers and the inferred precipitation from each model structure are

plotted in Figure 4.7, with the multipliers shown in the left column. In comparing the streamflow-only calibration (Figure 4.7a) to the April 1 course SWE scenarios (Figures 4.7c, 4.7e), a greater degree of correlation between the model structures is observed when SWE is included in the calibration. Second, the latter two scenarios have less variation across model structures for each year's multiplier: the standard deviation of the multipliers across model structures averaged over all years is 0.22 for the streamflow-only calibration, while it is 0.12 and 0.13, respectively, for the two streamflow and SWE calibrations. Thus, calibrating to both streamflow and SWE results in less uncertainty in the inferred precipitation multiplier for each year.

The impact of the calibration scenarios on inferred precipitation (as opposed to the multiplier) is shown in the right column of Figure 4.7. Here, we see that the year-to-year variability in precipitation is greater than variability in the multiplier, due to the high inter-annual variability in the climate of the Sierra Nevada. However, the greater consistency between precipitation multipliers in the streamflow and SWE calibration scenarios, as compared to the streamflow-only scenario, is also apparent in the inferred precipitation. The difference between model structures' inferred precipitation is smaller and more consistent from year to year (*cf.* Figure 4.7b to Figures 4.7d and 4.7f). The average standard deviation of annual precipitation across model structures is 258 mm yr^{-1} in the streamflow-only calibration, 171 mm yr^{-1} in the streamflow and course mean SWE calibration and 177 mm yr^{-1} in the streamflow and scaled course mean SWE calibration. Also note that the model structures that exhibited high SWE bias in Figure 4.6c (FUSE-014, -016 and -170) tend to have the highest inferred precipitation amounts. This suggests that uncertainty could be reduced further by excluding calibrations that fail to adequately match SWE from the estimate of basin-mean precipitation

4.7. Discussion

4.7.1. Relationship between ASO SWE and course and pillow SWE

In using a constant 80% ratio to scale mean April 1 course SWE to the ASO basin-mean SWE over the 1982-2014 inference period, we are assuming that the differences between ASO and course SWE are robust and repeat each year. While some studies of yearly SWE variability suggest this is not the case [*e.g.*, *Meromy et al.*, 2013], in other cases strong elevation dependence of SWE have been observed [*e.g.*, *Rice et al.*, 2011]. If the difference between ASO and the courses is driven largely by their different (but temporally consistent) sampling from the elevation distribution of SWE (*e.g.*, Figures 4.4b-4.4d), then the ratios between them should repeat at least approximately each year. In that case, we can bias-correct the long-term snow course record to generate a more accurate historical record of basin-mean SWE.

An examination of Figures 4b-4d suggests that it may be possible to determine the relationship between basin-mean SWE and a regional index of mean snow course SWE, based on the elevation distribution of the basin. For example, 29% of the Tuolumne basin area is above 2,950 m, which is the 90th percentile elevation of the snow courses used in this study, while 10% of the basin area is below 1,975 m, the 10th percentile elevation of the snow courses. The Tuolumne basin had April 1 ASO basin-mean SWE of 66-95% of mean April 1 course SWE. The courses tend to overestimate spatially-distributed SWE at their particular elevation, but they underestimate higher-elevation SWE because this area is under-sampled. Thus, a basin with an area-elevation distribution similar to that of the courses would be expected to have less basin-mean SWE than the courses. However, one with much of its area higher in elevation than most of the courses may have more basin-mean SWE.

It is also important to recognize that this approach is based on three years of data. Water years 2013-2015 were much warmer and drier than the observed long-term climate of the Sierra Nevada. Out of water years 1981-2015 at the two sites used for model forcing (Cherry Valley and Hetch Hetchy), water year 2014 was both the second-warmest (1.9°C above average), and fourth-driest (58% of average precipitation). Water year 2015 was even more extreme: it was both the warmest (2.2°C above average) and third-driest (57% of average precipitation). As a result, snowpack from these two years was far below normal, and so relationships between point and ASO SWE drawn from this period may be different from those in the long term.

At this time, ASO snow observations are ongoing; when they are made during a non-drought year, the climatic sensitivity of the ratio between ASO and point measurements should become much clearer. Given the very high spatial resolution and large domain of the ASO observations, it is now possible to conduct analyses of the spatial pattern of SWE across the high-elevation Sierra Nevada landscape that were not possible from point measurements alone. In particular, ASO offers the potential to observe spatial distributions of SWE accumulation from individual storm cycles, which would shed light on atmospheric dynamics and orographic enhancement occurring over the Sierra Nevada during these events.

4.7.2. Impacts of including SWE in precipitation inference

The results presented in Figures 4.5-4.7 suggest that when we use SWE along with streamflow to calibrate the hydrologic models, the consistency of the model states and their inferred precipitation is greater. Specifically, we find that the standard deviation of the inferred precipitation across model structures (*i.e.*, its uncertainty) is reduced by nearly half with the inclusion of SWE in the calibration, particularly in the long-term run (Figure 4.7). The improvement in model robustness is similar to earlier findings [*Kuczera and Mroczkowski*, 1998;

Finger et al., 2015], and suggests a general rule that the inclusion of additional types of data in calibrations will improve model consistency.

There appears to be a tendency towards overestimation of SWE in several model structures, FUSE-014, -016 and -170. Structures FUSE-014 and -170 (as well as -160) allow for ET from both the upper and lower storages. FUSE-016 restricts percolation from the upper to lower zone. The effect of both of these choices would be to increase ET due to greater soil moisture availability, which in turn requires greater precipitation and SWE to match observed streamflow. The high bias in SWE in these structures is only partially eliminated by the inclusion of snow information in the calibration (Figure 4.6c). While the FUSE model structures are lumped and highly-conceptualized representations of the basin, the model structural assumptions do appear to affect how realistically the models behave. Structures with high SWE biases may be less well suited for simulating the steep, shallow-soiled Tuolumne basin.

In addition to variability, the SWE calibration also has an influence on magnitude of the inferred precipitation. In the streamflow-only calibrations, the set of FUSE models tends to overestimate basin-mean SWE (Figure 4.5), and thus likely overestimates basin-mean precipitation. When both streamflow and SWE are used in calibration, a better match to both targets is achieved, which likely reduces bias in inferred precipitation. However, the match to streamflow and SWE depends to some extent on their relative error weights. For some model structures, our exploratory weighting of streamflow and SWE errors showed that it is not possible to produce unbiased simulations of both: increasing the weight of one type of error to reduce bias resulted in greater bias in the other.

Additionally, inferred precipitation is sensitive to the choice of SWE calibration target. The April 1 course mean SWE is greater than the ASO SWE for the Tuolumne basin, and using

the course SWE for calibration results in greater inferred precipitation than using ASO SWE (Figure 4.6a). Thus, correctly resolving the magnitude of basin-mean SWE is important for inferring basin-mean precipitation under this approach.

For simplicity, the snow model does not simulate sublimation from the snowpack, and in that regard it may underestimate basin evaporative losses. However, the climate of the Sierra Nevada generally is warmer and more humid than basins where sublimation is most significant [*e.g., Gustafson et al.*, 2010]. The soil model also permits evapotranspiration regardless of the degree of snow cover in the basin, so long as there is sufficient energy and soil moisture. This has been observed in some situations in the Sierra Nevada [*Goulden et al.*, 2012], but may result in overestimates of ET. Altogether, there remains considerable uncertainty in the absolute ET over these basins as simulated by the FUSE models.

4.8. Conclusions

This study has two major aims. We compare different means of estimating the basin-mean SWE amounts that are needed for model calibration, distinguishing between point-based and distributed data for estimating basin-mean SWE. We also investigate the value of snow data to constrain the inference of precipitation from streamflow, using lumped hydrologic models and an elevation-band snow model.

Using the Tuolumne basin as a test case, we compare indices of snow based on point observations (courses and pillows), with spatially-distributed snow observations from the ASO LiDAR platform. The comparison of the two types of data is consistent with the idea that the current point observation network tends to oversample snowy, mid-elevation locations: the mean of the snow courses has greater April 1 SWE than ASO, but melts out more rapidly.

Inclusion of basin-mean SWE in the inference of basin-mean precipitation is shown to reduce biases in the inferred precipitation, and to somewhat reduce uncertainty in the inferred precipitation due to model structural uncertainty. The temporal correlation of different model structures' inferred precipitation is higher in the calibrations with both streamflow and SWE as targets than in the streamflow-only calibration.

From our findings, we suggest that the inclusion of multiple types of hydrologic observations is necessary to constrain the water balance of high-elevation basins such as that of the Tuolumne above Hetch Hetchy Reservoir. Because precipitation gauges are few, difficult to maintain during the winter, and suffer from undercatch of snow, it is difficult to directly estimate basin-mean precipitation in this basin. Incorporating multiple other types of observations, such as streamflow and SWE, is likely necessary to correctly estimate this component of the basin's water balance.

4.9. Figures
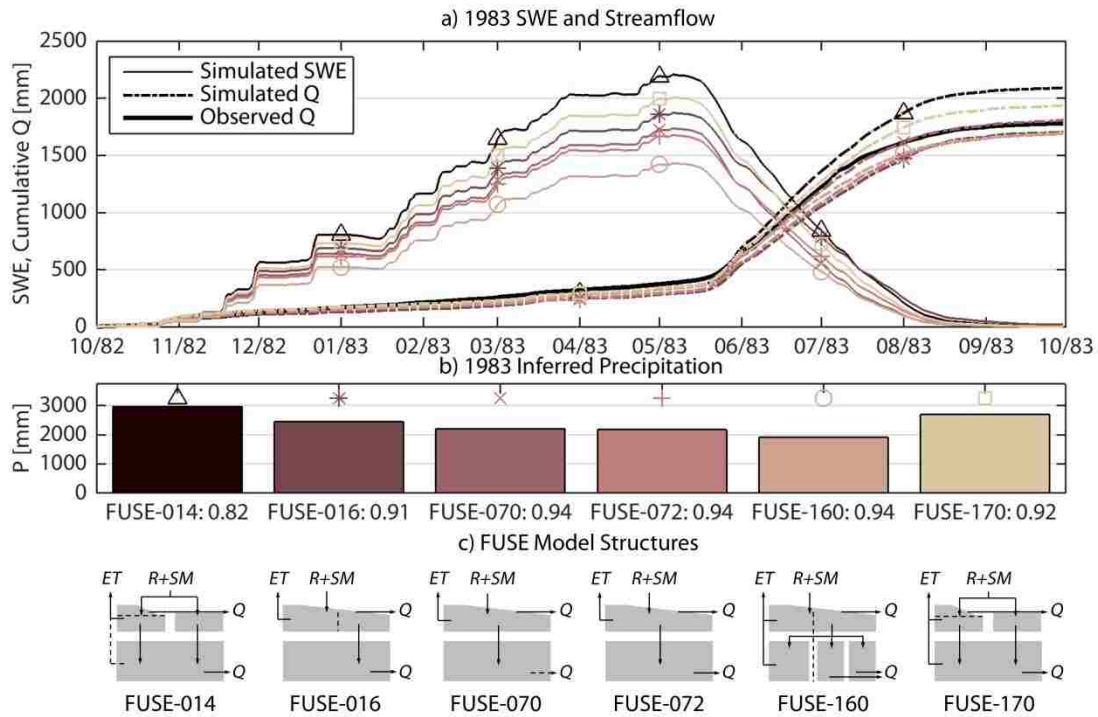
Figure 4.1. Comparison of water year 1983 simulations of the Tuolumne River basin above Hetch Hetchy Reservoir using six FUSE model structures (data from *Henn et al.* [2015]). a) Simulated SWE and streamflow, along with observed streamflow. b) Inferred precipitation for the six structures for this water year, with streamflow Nash-Sutcliffe coefficients of each structure shown after the structure name; symbols and colors indicate FUSE structures in a). c) Schematic diagram of the upper and lower zone storages and fluxes in each FUSE structure. Boxes represent storages and arrows represent fluxes (*ET* = evapotranspiration, *R + SM* = rain plus snowmelt, and *Q* = streamflow); dashed arrows indicate alternate parameterizations, and dashed lines indicate storage partitions.
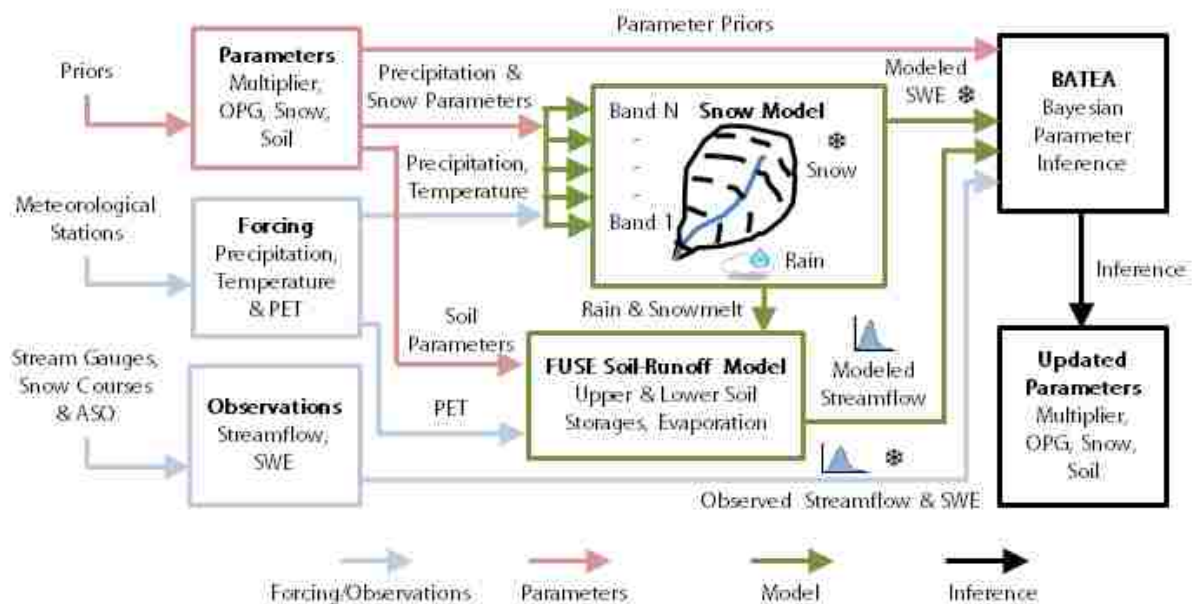
Figure 4.2. Information flow in inference of precipitation from streamflow and SWE. Forcing data from meteorological stations are combined with parameters such as the precipitation multiplier, orographic precipitation gradient (OPG), and soil and snow parameters, to simulate basin snowpack and runoff. The simulated runoff and SWE are then compared with observed values, using a Bayesian approach that incorporates prior distributions on the parameters, to estimate the posterior probability of each parameter set. Posterior distributions of parameters, including the multiplier and OPG, describe basin-mean precipitation inferred from streamflow and SWE.
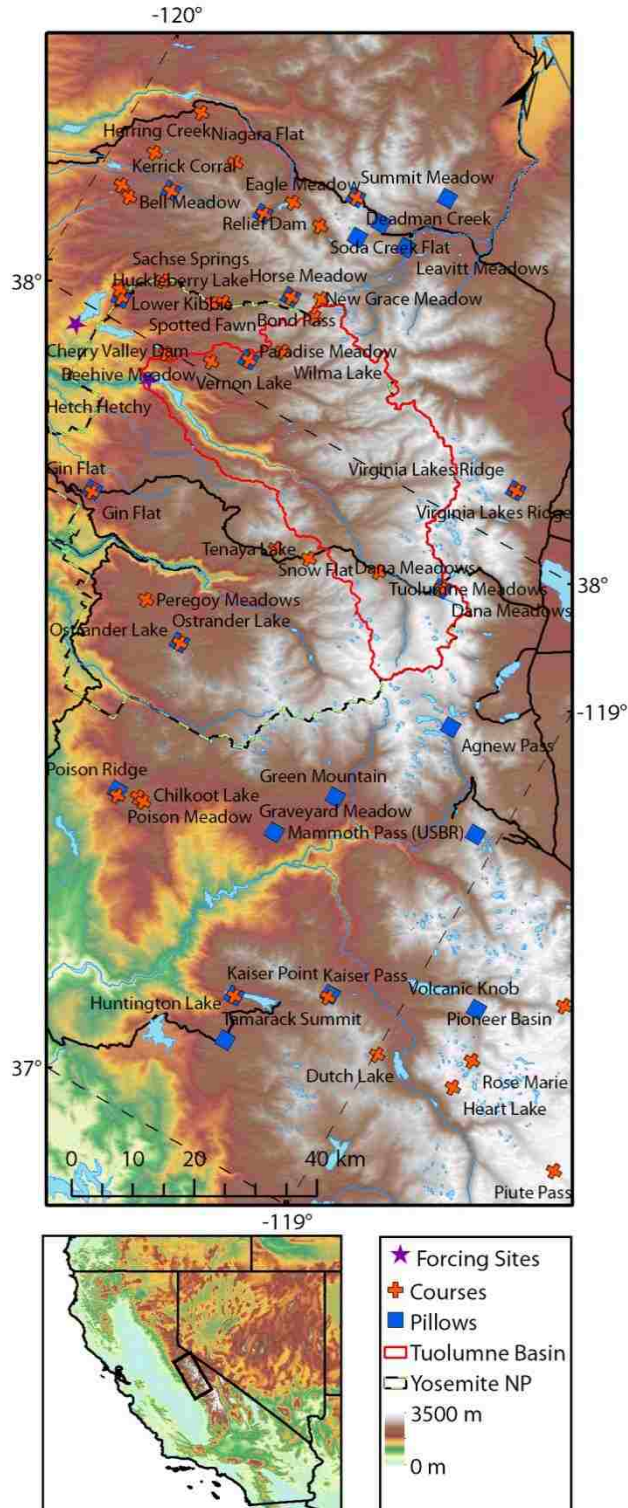
Figure 4.3. Topographical map of the Sierra Nevada domain, showing the Tuolumne basin boundaries. The snow courses and pillows used in the study are shown, as are the meteorological sites used to force the snow and hydrologic models.

92
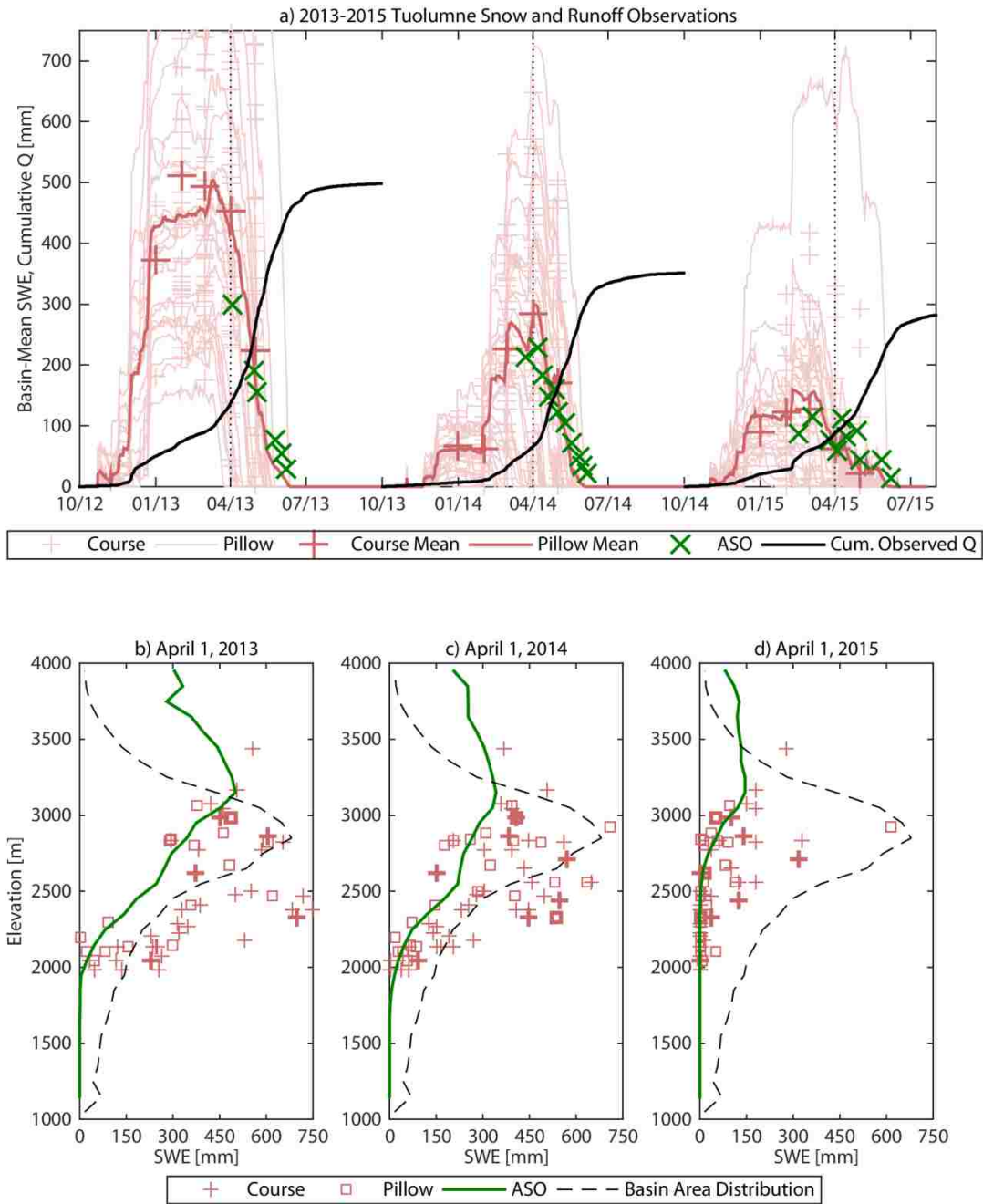
Figure 4.4. Comparison between point and distributed measurements of SWE in water years 2013-2015. a) Time series of course, pillow and ASO basin-mean SWE; cumulative runoff from the Tuolumne basin is also shown for each water year. b) - d) Elevation profiles of SWE from ASO and course and pillow measurements. The area distribution of the Tuolumne basin is also shown (in arbitrary units that integrate to unity).

93
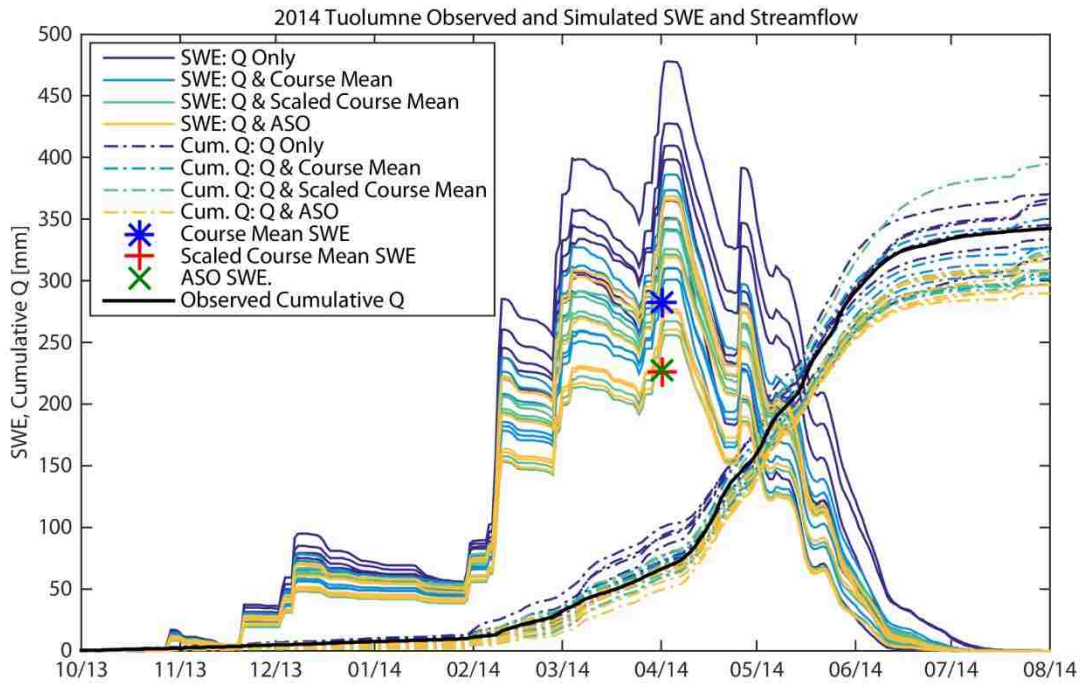
Figure 4.5. Water year 2014 simulated and observed SWE in the Tuolumne basin. For each calibration scenario (colors), SWE (solid lines) and streamflow (dashed lines) from the six FUSE model structures are shown. April 1 SWE calibration targets are shown, along with cumulative observed streamflow.
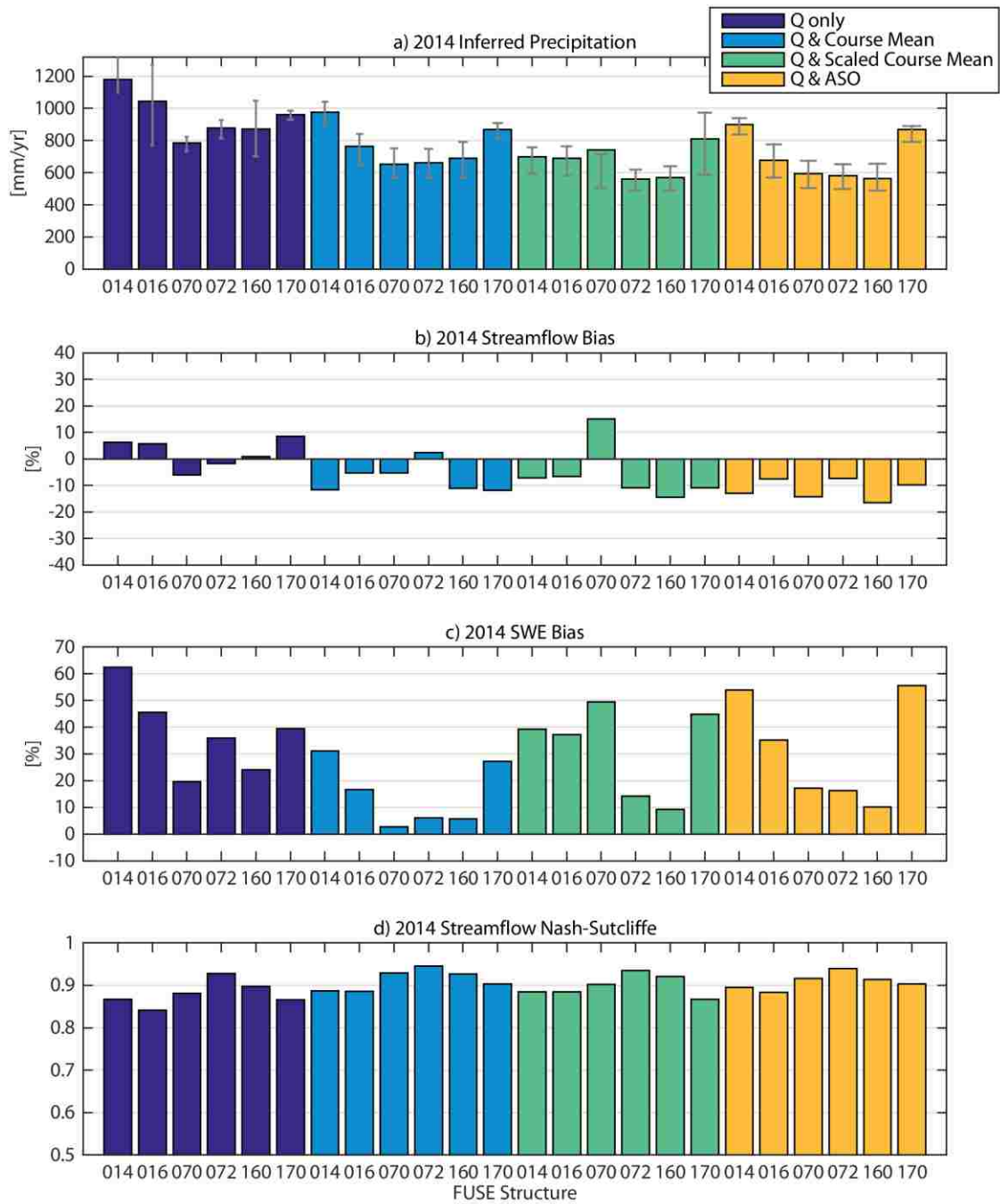
Figure 4.6. Calibration metrics for the water year 2014 precipitation inference; colors indicate the calibration scenario and the individual FUSE model structures are indicated on the x-axis. a) Inferred precipitation. b) Streamflow bias. c) SWE bias. d) Streamflow Nash-Sutcliffe coefficient.
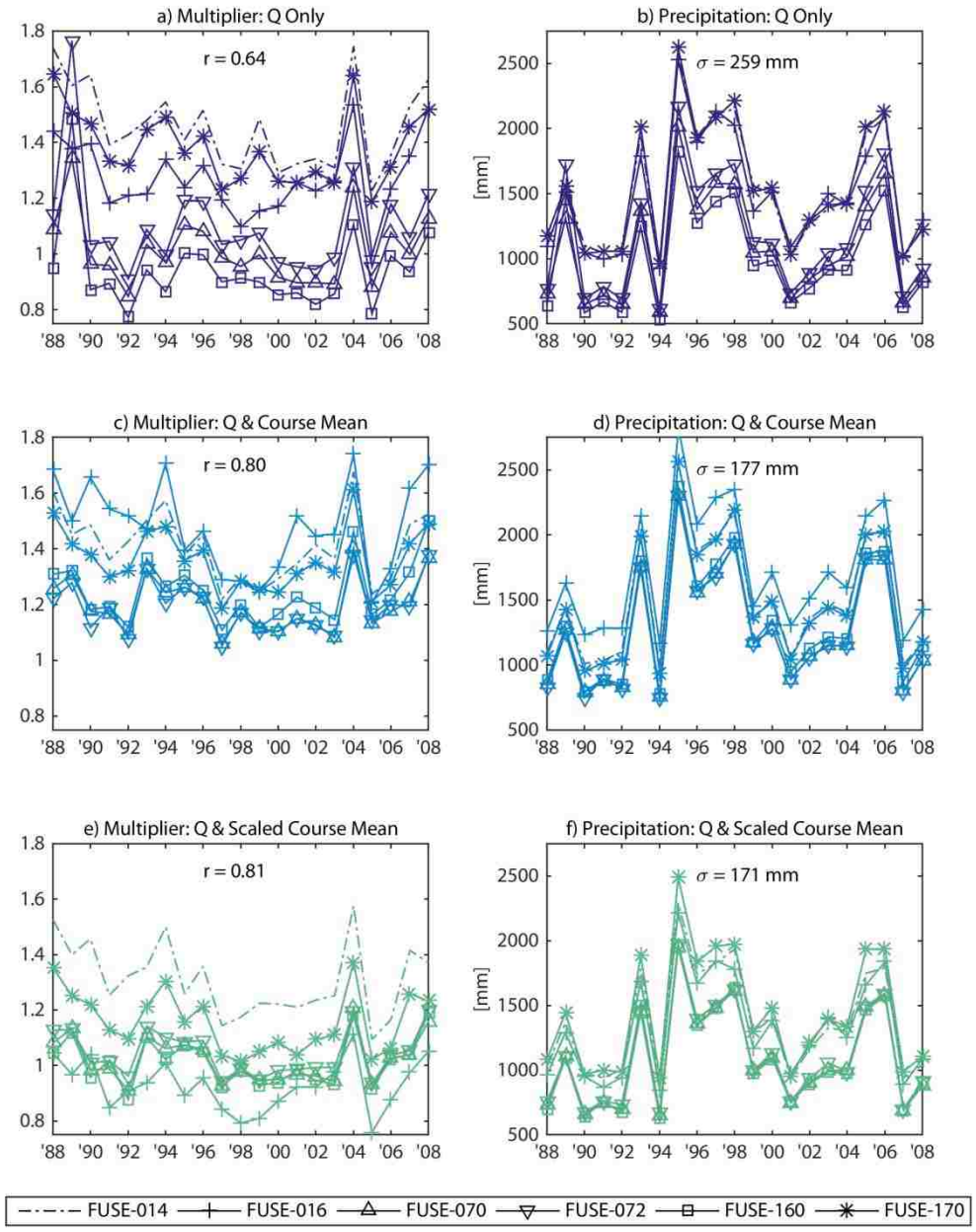
Figure 4.7. Results of precipitation inference over water years 1982-2014 in the Tuolumne basin, under different calibration scenarios. For clarity, a subset of water years is shown (1988-2008). a), c) and e): Annual time series of inferred precipitation multipliers; mean correlation between multiplier times series (r) is indicated. b), d) and f): Annual time series of inferred precipitation; mean standard deviation of water year precipitation (σ) is indicated. a) and b): Streamflow-only calibration. c) and d): Streamflow and April 1 course mean SWE calibration. c) Streamflow and scaled April 1 course mean SWE calibration.

Chapter 5. Spatiotemporal precipitation patterns across the Sierra Nevada of California inferred from streamflow observations

Abstract: Given uncertainty in precipitation gauge-based gridded datasets over complex terrain, we use multiple streamflow observations as an additional source of information about precipitation to identify spatial and temporal differences between a gridded precipitation dataset and precipitation inferred from streamflow. In particular, we test whether gridded datasets capture across-crest and regional spatial patterns of variability, as well as year-to-year variability and trends in precipitation, in comparison to precipitation inferred from streamflow. We use a Bayesian model calibration routine with multiple simple hydrologic model structures to infer the most likely basin-mean water year total precipitation for 56 basins with long-term (>30 year) streamflow records in the Sierra Nevada mountain range of California. We compare basin-mean precipitation derived from this approach with basin-mean precipitation from a 1/16° gridded precipitation dataset that has been used to simulate and evaluate trends in Western United States streamflow and snowpack over the 20th century. The long-term average spatial patterns differ. In particular, there is less precipitation in the gridded dataset in higher-elevation basins whose aspect faces prevailing cool-season winds, as compared to precipitation inferred from streamflow. In some years and basins, there is less gridded precipitation than there is observed streamflow. Lower-elevation, southern, and east-of-crest basins show better agreement between gridded and inferred precipitation. Implied actual evapotranspiration (calculated from precipitation vs. streamflow) also varies between the streamflow-based estimates and the gridded dataset. Uncertainty in inferred precipitation is substantial, but the basin-to-basin and year-to-year differences in precipitation inferred from streamflow are likely robust. The findings suggest

that considering streamflow when spatially distributing precipitation in complex terrain may improve the representation of basins whose orientations (*e.g.*, windward-facing) are favored for orographic precipitation enhancement.

5.1. Introduction

5.1.1. Reducing uncertainty in spatially distributed precipitation estimates over complex terrain

As discussed in chapters 1, 2 and 3, large errors can occur during estimation of spatially distributed precipitation in complex terrain, due to both sparse and uncertain gauge data and uncertainty in the interpolation of gauge data to a grid. As a result, it may be useful to apply other types of observations and estimates to infer precipitation in these areas and reduce potential errors in gridded datasets. In chapters 3 and 4, we discussed using streamflow and snow observations to infer basin-mean precipitation over basins in the area of Yosemite National Park in the Sierra Nevada of California. In this chapter, we describe a larger streamflow dataset from the Sierra Nevada, the method for inferring precipitation from streamflow, and a set of experiments to compare inferred precipitation against the gridded precipitation dataset. We do this to identify patterns of where and when precipitation estimates diverge the most in the Sierra Nevada.

Given the limits of precipitation gauge observations in high-elevation areas, streamflow observations offer an additional method with which to infer precipitation patterns and evaluate gridded precipitation datasets. Streamflow has been used to correct gridded precipitation products for high-elevation underestimation at the global scale [*Adam et al.*, 2006], to evaluate sources of precipitation data [*Wayand et al.*, 2013] and precipitation distribution schemes [*Livneh et al.*, 2014] at the basin scale, and to identify errors in precipitation forcing data

[*Renard et al.*, 2010; *McMillan et al.*, 2011a]. In general, these studies have shown that substantial errors in spatially distributed precipitation can be identified through a comparison with observed streamflow.

In chapter 3 [*Henn et al.*, 2015], we developed an approach for inferring basin-mean precipitation from streamflow in the Sierra Nevada, and tested the sensitivity of the inferred precipitation from this approach to various assumptions. This approach uses Bayesian calibration of simple snow and hydrologic models to estimate the most likely water-year total precipitation in a basin given its daily streamflow observations. In effect, this approach calibrates the model to match streamflow under uncertain precipitation inputs, while estimating the evapotranspiration (ET) needed to close the basin's water balance. We use a small ensemble of six hydrologic model structures to estimate the uncertainty of the inferred precipitation associated with model structural assumptions.

Based on the seven Yosemite National Park-area basins tested in *Henn et al.* [2015], we concluded that streamflow allows for robustly inferring the climatological patterns of precipitation across the mountain range, *i.e.*, identifying wetter and dryer areas within the range. Streamflow also highlighted annual variability between wet and dry years. We also found that long-term inferred basin-mean precipitation disagreed by up to 25% with Parameter Regression on Independent Slopes Model [PRISM, *Daly et al.*, 1994, 2008] precipitation climatology. However, the absolute amounts of inferred precipitation are sensitive to the model structures used to simulate the basins, suggesting that this approach is only suited for identifying errors greater than 100 mm yr^{-1} in spatially distributed precipitation estimates.

We apply the approach of inferring precipitation from streamflow in 56 basins in the Sierra Nevada that have largely unimpaired, long-term streamflow observations. We do this to compare inferred precipitation against the *Hamlet et al.* [2010] dataset (hereafter H10) over water years 1950-2010. We choose this dataset as it has been used to resolve spatial patterns and trends in Western United States snowpack and streamflow [*Hamlet et al.*, 2005, 2007], and because it explicitly attempts to reduce temporal inhomogeneity by using long-term station records to control trends in the precipitation data [*Hamlet and Lettenmaier*, 2005].

We test for potential mismatches between the spatial patterns of the H10 precipitation dataset and the streamflow records. We hypothesize that areas of high elevation and complex terrain in the Sierra Nevada, where precipitation gauge density is low, may have substantial biases in gridded precipitation datasets, and these biases can be identified by precipitation inferred from streamflow. We further hypothesize that these biases may be greatest in areas that are topographically favorable or unfavorable for precipitation, and that they may vary in time depending on characteristics of a particular water year. We use precipitation inferred from streamflow to evaluate whether gridded precipitation datasets capture areas of orographic enhancement, the ratios of windward to leeward precipitation (*i.e.*, rain shadows), year-to-year variability in the spatial patterns of precipitation, and long-term trends in basin-mean precipitation.

In chapter 4, we considered the use of snow observations in addition to streamflow to infer precipitation. While this yielded a reduction in the uncertainty of inferred precipitation, it also showed sensitivity to the estimate of basin-mean snow water equivalent (SWE). Since high-resolution SWE data is available for the Tuolumne basin, we used both streamflow and SWE in

that study. However, because we have lower confidence in basin-mean SWE estimates for other Sierra Nevada basins in which only point measurements of SWE are available, in this chapter we use only streamflow to infer precipitation.

Also, unlike in previous chapters, here we develop quantitative estimates of prior uncertainty in basin-mean precipitation by considering the agreement (or lack thereof) for each basin between independent precipitation data sources. For example, weather model estimates of precipitation over complex terrain have increased in fidelity as models have been run at higher resolution and with more process-based precipitation parameterizations [*e.g.*, *Rasmussen et al.*, 2011a]. The Weather Research and Forecasting model [WRF, *Skamarock and Klemp*, 2008] has been shown to generate precipitation estimates that are equal or better in terms of simulated hydrologic responses, as compared to gauge-based gridded precipitation, in some mountain basins [*Gutmann et al.*, 2012; *Wayand et al.*, 2013]. Thus, we consider modeled precipitation as a source of information about the patterns of precipitation over complex terrain. We also consider a gridded precipitation dataset [*Newman et al.*, 2015] that does not depend on PRISM climatology, as well as a PRISM-based gridded dataset [*Hamlet et al.*, 2010], as part of our prior estimate of basin-mean precipitation (see sections 5.2 and 5.3).

Sections 5.2 and 5.3 describe the data and models, respectively, used in the inference of precipitation from streamflow. Section 5.4 presents the results of the precipitation inference and the comparison against the gridded dataset. Section 5.5 discusses the implications of the differences between gridded and inferred precipitation, and the potential value of streamflow observations in improving gridded precipitation datasets. Section 5.6 offers conclusions from the study.

5.2. Data

In this section we describe the data used to compare H10 gridded precipitation with precipitation inferred from streamflow: unimpaired daily streamflow observations, basin topography and soils information, gridded precipitation data from several sources, and gridded temperature and evapotranspiration data.

5.2.1. Streamflow and basins

For comparison against precipitation, we identify 56 basins in the Sierra Nevada (Figure 5.1; Table 5.1) in which daily streamflow observations have been made for extended periods of time. Forty-one of the gauges are west of the crest of the Sierra Nevada, and 15 are on the east slope. Streamflow observations included in this study have at least 30 water years of complete or nearly complete records between 1950 and 2010; in a few cases we include stream gauges with 20-30 years of data because they were in areas with few gauges (the southeastern Sierra Nevada). Thirty-one of the 56 continue to operate, and 34 are listed in the Hydro-Climate Data Network (HCDN) of basins without significant regulation from dams or diversions to agricultural or domestic use [*Slack and Landwehr*, 1992]. We also include 22 basins not in the HCDN but with relatively few or "small" impairments, based on descriptions provided by the United States Geological Survey (USGS) water data reports for each site [*Anderson et al.*, 1996; *Rockwell et al.*, 1996a, 1996b]. In a few cases, we include basins with upstream dam regulation, so long as the storage was small compared to annual runoff volumes, or the regulation effect could be removed from the streamflow record because complete storage records were available. We exclude most stream gauges with observations that were described as having poor quality by the USGS; 3 gauges in the southeastern Sierra Nevada rated "poor" (Big Pine Cr., Independence Cr. and Cottonwood Cr.) were included due to a lack of gauges in this area. All streamflow data

were downloaded from the USGS National Water Information System Web Interface
(http://waterdata.usgs.gov/nwis/sw), with the exception of two basins (Cherry-Eleanor and the
Tuolumne River at Hetch Hetchy), in which full natural flows were provided by the operator of
the reservoirs at the basin outlets, Hetch Hetchy Water and Power (see supplemental material of
Henn et al. [2015]). More detailed information on the individual stream gauges and their
associated basins is included in Table 5.1.

Uncertainties in the daily streamflow measurements from the rating curve have been
estimated at 10% for observations of "good" quality, and 15% for "fair" records [*Rockwell et al.*,
1996a], in that 95% of the daily measurements should fall within the stated margin of error.
While other studies of other rating curve-based streamflow techniques have suggested
uncertainties of up to 20-40% at very high flows [*Baldassarre and Montanari*, 2009], we are
primarily interested in precipitation and runoff over timescales of a water year or longer, and so
we presume the USGS estimates to be a reasonable representation of uncertainty.

Topographical boundaries of the basins are delineated based on 30 m USGS elevation
data, which are also used to calculate basin area distributions within 100 m elevation bands. The
basins' areas (Table 5.1) range from 10.1 km^2 (Daggett Cr.) to 3,472 km^2 (Kings R.); the median
basin area is 165 km^2. The basins' mean elevations range from 744 m (White R.) to 3,457 m
(Middle Fork of Bishop Cr.); most basins span at least 1,500 m in elevation.

We use typical basin soil properties to provide information for the lumped hydrologic
model in the inference of precipitation from streamflow (section 5.3). Most of the basins are near
the crest of the Sierra Nevada and have steep topography that is underlain by granitic bedrock.
Typical Sierra Nevada soils are shallow and sandy, with a depth of no more than 1 m and a
porosity of about 0.4 [*NRCS*, 2007]. Field studies have indicated that soils in higher elevation

areas of the Sierra Nevada have a high conductivity typical of sand and gravel, but that the percolation rate into the underlying bedrock is much lower [*Flint et al.*, 2008]. A modeling study of the upper Merced River basin indicated that the soils' conductivity is likely at least 1 m d^{-1} and that field capacity is likely no higher than 0.2-0.25, in order to match observed streamflow patterns [*Lundquist and Loheide*, 2011].

5.2.2. Gridded precipitation

5.2.2.1. H10 precipitation

We use the H10 daily precipitation dataset as a basis for comparison with streamflow and precipitation inferred from streamflow. The dataset uses National Weather Service Cooperative Observer (COOP) daily precipitation gauge observations, and the Parameter Regression on Independent Slopes Model [PRISM, *Daly et al.*, 1994, 2008] 1971-2000 monthly precipitation climatology as a topographic corrector in the interpolation of gauge precipitation over complex terrain. We choose this dataset because of its relatively high spatial resolution (1/16°, ~6 km) and long-term span (1915-2010 over California), and because H10 and its predecessors have been used to simulate snow and streamflow in Western United States [*e.g.*, *Mote et al.*, 2005]. However, we note that other long-term gridded datasets [*e.g., Livneh et al.*, 2013] are also available.

Here we use H10 data covering water years 1950-2010 (1949 is used for model spin-up), over a spatial domain which covers the streamflow basins: from -121.5°W to -117.5°W and from 35.5°N to 40°N. For each basin we calculate the fraction of each H10 grid cell that is within the basin, and calculate the H10 basin-mean time series as the weighted sum of the grid cells' precipitation. We use this daily time series of precipitation, combined with a multiplicative

precipitation scaling factor inferred in calibration, to force the hydrologic and snow models that are the basis for inferring precipitation from streamflow for each basin (see section 5.3).

5.2.2.2. N15, WRF and other precipitation datasets

In order to ascertain the uncertainty of the H10 precipitation estimates and how that uncertainty varies of the Sierra Nevada domain, we examine independently-generated gridded precipitation data from additional sources. A quantitative estimate of precipitation, and its associated uncertainty, is necessary for the Bayesian inference of precipitation from streamflow (see section 5.3). For one additional source of precipitation information, we use the N15 dataset [*Newman et al.*, 2015]; while it interpolates gauge data in a qualitatively similar manner to H10, it notably does not use PRISM precipitation climatology as a topographic corrector, and thus may be a more independent estimate from H10 than other available gridded datasets (*e.g.*, *Livneh et al.* [2013]). N15 is a 100-member ensemble dataset, such that uncertainty in the data can be explicitly quantified. N15 uses a 1/8° (~12 km) spatial grid, and so there may be greater topographic smoothing compared to the 1/16° datasets.

Second, we use WRF model [*Skamarock and Klemp*, 2008] daily precipitation totals over California. The regional WRF simulations were generated using WRF version 3.6.0 [*Skamarock et al.*, 2008] to dynamically downscale the North American Regional Reanalysis (NARR), with a methodology very similar to that used in *Hughes et al.* [2012]. The WRF downscaling has two nested domains, with an outer 18 km grid extending across the northeastern Pacific and much of the U.S. intermountain west, and a 6 km grid domain that covers all of California (similar to Figure 1 of *Hughes et al.* [2012]). Both domains have 82 vertical levels, with 50-75 m vertical resolution within the lowest 2 km of atmosphere, with model top at 100 hPa. Both domains use Morrison double-moment microphysics [*Morrison et al.*, 2009], Dudhia shortwave radiation

[*Dudhia*, 1989] and RRTM longwave radiation [*Mlawer et al.*, 1997] called every 6 minutes, Kain-Fritsch convective parameterization [*Kain*, 2004], the Yonsei University planetary boundary layer scheme [*Hong et al.*, 2006], and the Noah land surface model [*Tewari et al.*, 2004]. WRF data were generated for 11 calendar years, 2004-2014, and over this time period, the model was re-initialized approximately every 5 days using NARR data, with the first three hours of each 5-day period discarded as model spin-up. The model is run with a 30-second (10-second) timestep in the 18 km (6 km) domain; output data is aggregated to daily totals for precipitation. For both the N15 and WRF precipitation, gridded precipitation is aggregated to single time series for each basin using the fractional grid cells approach described above for H10.

For comparison, we also briefly examine two other gridded precipitation datasets: *Livneh et al.* [2013], hereafter L13, and North American Land Data Assimilation System, phase 2 [*Xia et al.*, 2012], hereafter NLDAS. We include these datasets to see whether patterns of differences between gridded and inferred precipitation are consistent across different gridding methodologies. While each gridded dataset considered here varies in the set of precipitation gauges used, all of them (with the exception of N15) use PRISM climatology as the topographic corrector. Thus, we hypothesize that patterns of inferred biases may be fairly consistent across the datasets.

5.2.3. Temperature and potential evapotranspiration

To infer precipitation from streamflow, time series of daily mean temperature and potential evapotranspiration (PET) are needed to force the hydrologic model that is run for each basin. For temperature, we calculated basin-mean time series of daily high and low temperature from the H10 data in the same fashion as for precipitation, and averaged the high and low temperatures to create a daily mean temperature series.

106

PET is estimated using the Makkink equation, which uses temperature and shortwave radiation as predictors [*e.g., Cristea et al.*, 2013]. The Bristow-Campbell parameterization is used to estimate shortwave radiation from temperature [*Bristow and Campbell*, 1984]. We calibrated the Bristow-Campbell and Makkink methods using data from the Dana Meadows meteorological station (2,987 m above sea level) for water years 2003 through 2009. Shortwave radiation and temperature are used to calibrate the Bristow-Campbell coefficients *A* (0.7) and *C* (2.49). Shortwave radiation, wind speed, temperature and relative humidity are applied to estimate PET using the Penman-Monteith equation, following the methods of *Allen et al.* [1998]; this PET estimate is then used to calibrate the Makkink coefficient (0.668).

H10 daily high and low temperatures for each grid cell are used with the approach above to estimate daily PET at each grid cell in the spatial domain. The grid cells' PET are then aggregated to basin-mean PET series for each basin. Due to reduced diurnal temperature range at higher elevations in the H10 data, PET calculated in this way decreases somewhat with elevation, such that 1,000 m elevation has roughly 20% more PET than 3,000 m elevation (about 1,200 mm yr^{-1} vs 1,000 mm yr^{-1}), though there is substantial grid cell variation for a given elevation. In chapter 3 [*Henn et al.*, 2015], we found that inferred precipitation and inferred actual evapotranspiration are only mildly sensitive to PET, presumably due to the largely water-limited climate of the Sierra Nevada, so we do not investigate PET further here.

5.3. Methods

Precipitation is inferred from streamflow in this study using methods similar to those shown in Figure 3.1. H10-based precipitation, temperature and PET forcing data are used to drive snow and hydrologic models for each basin and produce simulated streamflow time series.

We simulate all water years for which streamflow observations are available for each basin from 1950-2010; at least one water year before the start of the analysis (*i.e.*, 1949 for basins with streamflow in 1950) is used for model spin-up and is discarded. The parameters of the models include a multiplicative error correction [*Kavetski et al.*, 2003] on the precipitation time series, along with snow and soil parameters. The simulated runoff is compared to the observed streamflow record, and based on the goodness of fit of the streamflow and the prior likelihood of the model parameters, the Bayesian joint posterior distribution of the parameters is generated. From the posterior distribution of the precipitation multiplier, the corrected basin-mean precipitation can then be calculated. This section describes the snow and hydrologic models and the calibration routine used to infer precipitation from streamflow; for a more detailed description of this approach, see chapter 3 [*Henn et al.*, 2015].

5.3.1. Models

5.3.1.1. FUSE conceptual hydrologic model structures

We use the Framework for Understanding Structural Errors (FUSE) [*Clark et al.*, 2008] to provide a set of conceptual hydrologic models and to evaluate the effects of model structural uncertainty on inferred precipitation [*Clark et al.*, 2011a; *McMillan et al.*, 2011b]. FUSE provides multiple options for representing soil moisture storage and fluxes, such as ET, surface runoff and baseflow, using a spatially-lumped approach with upper and lower soil zones. Streamflow, soil moisture and ET time series are simulated at a daily timestep. In this study, multiple FUSE structures are included to estimate the uncertainty of the precipitation inferred from streamflow. We use six model structures previously used to investigate the impact of structure on model performance in the Mahurangi basin in New Zealand [*Clark et al.*, 2011a;

McMillan et al., 2011b] and to infer precipitation from streamflow in chapters 3 and 4. For greater details on the FUSE model structures, see *Henn et al.* [2015] and [*Clark et al.*, 2008].

5.3.1.2. Snow model

Each FUSE structure is coupled to a temperature index snow model based on Snow-17 [*Anderson*, 2006], which tracks SWE based on precipitation and melt. Precipitation is added to SWE if the air temperature is below a threshold; otherwise, it is treated as rain. Snowmelt is initiated if the air temperature is above a threshold, and is proportional to the temperature above that point:

$$SM = \max(\min[f\{T - T_0\}, SWE], 0) \qquad (5.1)$$

where SM is snowmelt per timestep, f is the snowmelt factor, T is the air temperature, and T_0 is the melt threshold temperature. The melt factor f varies in a sinusoidal manner over the year, with minimum and maximum values on the winter and summer solstices, respectively. The rain-snow partition temperature, melt initiation temperature, and winter and summer melt factors are all inferred parameters. Snowmelt is combined with rainfall, and the two are routed to the soil model, which otherwise simulates fluxes and storages independently of the snow model.

H10 basin-mean precipitation is corrected for errors using a precipitation multiplier M_i, where the subscript i indicates the daily timestep:

$$P_i = M_i P_{f,i} \qquad (5.2).$$

Here, $P_{f,i}$ is the H10 basin-mean precipitation, M_i is the multiplier and P_i is the model forcing precipitation, all at time i. The multipliers are defined such that they are different for each water year, but constant within the water year, and each yearly multiplier is inferred in calibration. In this way, we infer water year precipitation from streamflow.

In order to simulate the strong elevation dependence of snow in the Sierra Nevada, the snow model is run at 100 m elevation bands for each basin (unlike the soil model, which is lumped over the entire catchment). For each band, forcing temperature is lapsed to the ith band midpoint elevation:

$$T_i = T_f + \Gamma(z_i - z_f) \tag{5.3}$$

where T_i is the ith band temperature, T_f is the forcing temperature, z_i is the band midpoint elevation, z_f is the basin's mean elevation (because basin-mean temperature and precipitation forcing is used), and Γ is the temporally-invariant temperature lapse rate (<0, °C km^{-1}), which is also inferred in calibration.

Finally, water fluxes to the lumped soil model are calculated by a weighted average of rain plus snowmelt from each elevation band:

$$F_{soil} = \sum_{i=1}^{n_{bands}} AF_i(R_i + SM_i) \tag{5.4}$$

where F_{soil} is the flux of water to the soil model, and AF_i is the fraction of the basin area, R_i is the rain, and SM_i is the snowmelt, all from band i.

Unlike in chapters 3 and 4, we do not use an orographic precipitation gradient (OPG) to distribute precipitation to the basins' elevation bands and infer intra-basin precipitation distributions. Instead, the multiplier-corrected precipitation is distributed evenly to each elevation band. This is done because the results in chapter 3 and 4 indicated that the inference was likely not able to discern intra-basin precipitation patterns; almost all inferred OPGs were at or near zero. Thus, precipitation inferred from streamflow represents the basin-mean amount for each water year.

5.3.2. Bayesian model calibration

5.3.2.1. Bayesian inference of precipitation

We infer basin-mean precipitation from streamflow using Bayes' Theorem using the Bayesian Total Error Analysis package [BATEA, *Kavetski et al.*, 2006a]. The posterior probability density functions (PDFs) of the multipliers M and other model parameters θ (the snow and soil parameters and the lapse rate), given observed precipitation $\tilde{P}$ and streamflow $\tilde{Q}$, are described by:

$$\boldsymbol{P}(M,\theta|\tilde{Q},\tilde{P}) \propto \boldsymbol{P}(\tilde{Q}|M,\theta,\tilde{P}) \cdot \boldsymbol{P}(M,\theta) \qquad (5.5).$$

The first product term on the right hand side of (5.5) is the likelihood function, which describes the probability distribution of the observed streamflow, given particular values of M, θ and the observed gauge precipitation $\tilde{P}$. The likelihood function interprets the difference between observed and simulated streamflow, using an error model. The second product term in (5.5) is composed of the prior distributions of the precipitation multiplier and other model parameters. The posterior distribution of the precipitation multiplier and the model parameters, $\boldsymbol{P}(M,\theta|\tilde{Q},\tilde{P})$, represents an update of the prior distributions given streamflow and precipitation observations.

We use a weighted least squares (WLS) residual error model to translate differences between modeled and observed streamflow into the likelihood of the corresponding parameter sets. In the WLS approach, it is assumed that higher streamflow is subject to greater uncertainty, unlike simple least squares in which all timesteps are treated equally. Following *Thyer et al.* [2009], we assume that residual streamflow errors follow a Gaussian distribution with zero mean and a standard deviation that is a linear function of streamflow, *i.e.*,

$$\sigma_t = a\hat{Q}_t + b \qquad (5.6)$$

where a and b are residual error parameters inferred in the calibration, and $\hat{Q}_t$ is the modeled streamflow at a given timestep, t. The error model calculates likelihood as a function of the squared differences between observed and modeled streamflow, which are first normalized at each timestep by σ_t.

The combination of the soil, snow and yearly multiplier parameters results in a parameter space with 50-80 dimensions, and so to sample the posterior probability distributions, multiple-start Quasi-Newton optimization and a Monte Carlo Markov Chain (MCMC) routine are used within BATEA [*Kavetski et al.*, 2006b]. The result of the MCMC sampling routine is an ensemble of parameter sets, with more samples clustered around parameter values of greater posterior probability.

5.3.2.2. Parameter prior distributions

To constrain the inference within the Bayesian framework, prior distributions are placed on the multiplier and each hydrologic and snow model parameter. We use similar snow and soil parameter distributions as in *Henn et al.* [2015]: uniform distributions are generally used, and when no information is available about a parameter in the study domain, the distribution is set to the default limits for the FUSE model [*Clark et al.*, 2008]. For more details on the soil and snow model parameters, see Table 3.1.

The prior distributions of the precipitation multipliers are normal distributions and each yearly multiplier is independent of the others. We use estimates of each basin's mean precipitation, and the differences in precipitation estimates between H10, N15 and WRF, to set the mean and standard deviation of the multipliers' prior distribution. An example of this process is shown in Figure 5.2, for the Cherry-Eleanor basin in water year 2008.

For the mean of the prior distribution we use:

$$\mu_M = (P_{H10} + \overline{P}_{N15} + P_{WRF})/(3P_{H10}) \qquad (5.7)$$

where P_{H10}, $\overline{P}_{N15}$ and P_{WRF} are the H10, N15 and WRF basin-mean precipitation for the water

year; N15 precipitation is the mean of the ensembles. For each basin, this procedure is applied

over each water year from 2005 to 2010, and then averaged. In other words, the mean of the

prior distribution for each basin is the average of the three datasets' precipitation, divided by the

H10 precipitation; this is an *a priori* guess of how H10 precipitation should be corrected, before

considering streamflow (see H10, N15 and WRF precipitation in Figure 5.2a and 5.2b).

For the standard deviation of the distribution we use:

$$\sigma_M = \frac{1}{\overline{P}} \sqrt{\frac{1}{3} \left(\left[P_{H10} - \overline{P} \right]^2 + \left[P_{N15} - \overline{P} \right]^2 + \left[P_{WRF} - \overline{P} \right]^2 \right)} \quad (5.8)$$

where $\overline{P}$ is the mean precipitation of the three datasets, or:

$$\overline{P} = \mu_M P_{H10} \qquad (5.9).$$

Thus, the standard deviation of the multiplier prior distribution is the standard deviation of basin

precipitation in H10, N15 and WRF, normalized by the mean of the three. As with the mean,

(5.8) is applied to estimate the standard deviation for each year from 2005 to 2010 and then

averaged; this process is repeated for each of the basins. Thus, we use the agreement between the

three datasets (or lack thereof) to estimate the prior distribution of uncertainty in precipitation for

a given basin. To illustrate the result of this process, the prior distribution for Cherry-Eleanor in

water 2008 is plotted in Figure 5.2b.

5.3.3. Experimental approach

First, we examine the Bayesian inference, comparing the prior distribution of the

multiplier with the posterior distribution for each of the six FUSE model structures. This

comparison allows us to evaluate whether the choice of prior distribution and precipitation

uncertainty appears appropriate given the precipitation inferred from streamflow, as well as the uncertainty in inferred precipitation associated with the choice of model structure.

Next, we compare H10 precipitation and precipitation inferred from streamflow across the 56 study basins in the Sierra Nevada, at several timescales. We present results for several water years, focusing on those for which many streamflow gauges were operational (1983 and 1977), which allows for greater spatial comparison of precipitation. We also compare inferred precipitation and streamflow to other gridded precipitation datasets (to test whether the findings are consistent across different methodologies). Then we present long-term average differences between gridded and inferred precipitation. When testing the differences between gridded and inferred precipitation, we consider their spatial patterns in terms of distribution with elevation, and distribution topographically across the crest of the Sierra Nevada. Finally, we consider the patterns of implied ET from the gridded and inferred datasets across the range.

5.4. Results

5.4.1. Precipitation multiplier prior and posterior distributions

Figure 5.2 shows an example of the prior and posterior distribution of the precipitation multiplier for the Cherry-Eleanor basin. We first plot the different precipitation estimates over time for this basin, showing the 2005-2010 comparison on which the multiplier prior distribution is based. The H10 precipitation is generally less than N15 and WRF, and so the mean prior estimate of the multiplier is 1.48 for this basin. The disagreement between the different datasets is also relatively high compared to other basins, with the three sources of precipitation spanning 600-800 mm yr^{-1}. Therefore, the standard deviation of the prior distribution is set to reflect this at 0.26; together with the mean, this indicates that prior to considering streamflow, we expect

basin-mean precipitation to be between 0.97 and 1.99 times H10 precipitation with 95% confidence. Using water year 2008 (boxed, Figure 5.2a) as an example, the prior distribution is plotted as a precipitation distribution for Cherry-Eleanor in Figure 5.2b, *i.e.*, as the product of the multiplier distribution and the 840 mm of precipitation in H10. We also plot water year 2008 streamflow (790 mm) and N15 (1,052 mm) and WRF precipitation (1,499 mm), showing that the prior distribution largely spans the three precipitation data sources.

The precipitation inferred from streamflow by the six FUSE model structures for Cherry-Eleanor is also shown in Figure 5.2. The time series of the inferred precipitation ensemble shows substantial differences between FUSE structures, but also robustly reflects yearly variability in streamflow. In general, the ensemble of precipitation inferred from streamflow is greater than H10, similar or greater than the N15 ensemble, and contains the WRF precipitation. In the water year 2008 distribution (Figure 5.2b), the 95% confidence intervals of the inferred precipitation from each structure are shown, with the mean indicated by the marker. While the individual structures' intervals are about 100 mm wide, the range of the ensemble is almost 800 mm, showing that structural uncertainty in inferred precipitation is more substantial than the individual structures imply; this is consistent with the findings of chapter 3 and 4. The mean of the FUSE ensemble is also shown (1,332 mm); it is less than WRF but greater than H10 and N15.

A qualitative examination of this basin and others indicates that the posterior precipitation amounts generally fall within reasonable (95%) limits of the prior distributions. This suggests that the prior distributions are not grossly incompatible with the posterior distributions inferred from streamflow using the FUSE model structures. When this is the case, the specific choice of the prior distribution likely does not exert undue influence on the inferred

precipitation. Instead, the posterior distributions of inferred precipitation largely reflect the best fit of the different model structures to streamflow.

Precipitation uncertainty (as measured by the agreement between the H10, N15 and WRF datasets) appears greatest in the central Sierra Nevada. When the difference between the datasets for each basin are normalized by the basin mean value and plotted spatially (not shown), greater relative uncertainty is seen in the Tuolumne, Merced and San Joaquin R. basins. Differences for these basins frequently exceed 20% of the mean, while in the southern Sierra Nevada differences are generally 10-15% or less, and the northern Sierra Nevada has intermediate values. However, the spatial pattern of uncertainty may be at least partially dependent on the choice of the particular precipitation datasets used in the comparison.

5.4.2. Comparing H10 precipitation and inferred precipitation for wet and dry years

For each of the 56 basins, we compile the inferred precipitation from the FUSE structures and calculate the mean of the FUSE ensemble for each basin-year. We exclude the basin-years in which the model calibrations produced very poor fits to streamflow (streamflow Nash-Sutcliffe coefficient of -2 or less), or about 8% of all basin-years simulated. We then compare the inferred precipitation mean to H10 basin-mean precipitation for individual water years from 1950 to 2010.

Figures 5.3 and 5.4 show this comparison for two water years (1983 and 1977). Water year 1983 was one of the wettest on record in the Sierra Nevada, with many gauges reporting over 2,000 mm of precipitation. H10 precipitation (Figure 5.3a) reflects the heavy precipitation that year, with most basins on the west slope of the Sierra Nevada north of 37°N averaging at least that much. Maximum accumulations are seen in the North Fork of the American R. and the South Fork of the Yuba R., as well in small adjacent basins. H10 precipitation generally

decreases to the south along the west slope of the Sierra Nevada. In comparison, inferred precipitation (Figure 5.3b) shows a somewhat more spatially variable pattern: while precipitation of 2,500 mm or greater is seen in the Yuba and American basins, similarly large accumulations are seen in the basins in the Stanislaus, Tuolumne, San Joaquin, Kings and Kaweah drainages.

Water year 1983 streamflow (Figure 5.3c) shows a similarly spatially variable pattern. Of particular note are the strong gradients in streamflow seen over the central Sierra Nevada, with a local maxima over Cherry-Eleanor and Falls Crs., and much drier (>1,000 mm less runoff) basins to the east in the Walker R. basin, and to the southwest in the lower Tuolumne R. basin. The spatial gradients are much less sharp in the H10 dataset than they are in streamflow, but are (not surprisingly) reflected in the precipitation inferred from streamflow.

To check H10 and inferred precipitation, we subtract observed streamflow from both datasets' basin-mean precipitation (Figures 5.3d and 5.3e). Notably, H10 precipitation minus streamflow is negative in several basins in water year 1983: Cherry-Eleanor (-175 mm), the North Fork of the Stanislaus R. (-16 mm) and Duncan Cr. (-75 mm). Several other basins have streamflow within about 100 mm of H10 precipitation, including the San Joaquin R. and its tributary Granite Cr. On the other hand, H10 precipitation minus streamflow is much greater for basins further from the crest of the Sierra Nevada. For example, the low-elevation Cosumnes and Mokelumne basins have H10 precipitation more than 800 mm greater than runoff. Inferred precipitation is greater than streamflow by more consistent amounts across the Sierra Nevada (Figure 5.3e), typically 400-800 mm, though some local variability exists.

Finally, we plot the difference between H10 and inferred precipitation (Figure 5.3f), which highlights the different spatial patterns of precipitation from the two sources of information. Inferred precipitation is greater than H10 over most high-elevation basins,

particularly in the central Sierra Nevada (Stanislaus, Tuolumne, San Joaquin, Kings and Kaweah basins). Precipitation inferred from streamflow is less than H10 for many of the lower-elevation basins, and for many of the basins east of the crest of the range.

The equivalent water year 1977 comparison (Figure 5.4) shows spatial patterns during an anomalously dry year. Comparing Figures 5.3 and 5.4 shows that streamflow is on the order of 90% less in 1977 as compared to 1983. As a result, the absolute differences in streamflow and precipitation between basins in 1977 are relatively small, but as a fraction of water year totals and considering the value of water in California in a drought year, they remain significant. For example, streamflow over 200 mm was observed in some central and northern basins (*e.g.*, San Joaquin, Cherry-Eleanor, Falls Cr., the Stanislaus R. basins, and the Yuba R.), as compared to the southern portion of the range, where every basin west of the crest except the upper Kaweah R. had less than 200 mm of streamflow (Figure 5.4c). Streamflow was also strongly elevation-dependent, with lower-elevation basins producing less than 100 mm.

The spatial pattern of H10 precipitation in 1977 is relatively smooth, and the greatest basin-mean precipitation is in the south in the Kaweah R. basin (Figure 5.4a). In contrast, inferred precipitation reflects the northerly and high-elevation patterns of streamflow (Figure 5.4b). Thus, H10 precipitation minus streamflow tends to be greater in the south than in the north, while inferred precipitation minus streamflow is more spatially uniform (Figures 5.4d and 5.4e). The north-south difference is also highlighted in Figure 5.4f.

To estimate the long-term differences between H10 and inferred precipitation, we calculate the mean difference between the two datasets, using the years in which streamflow is available for each basin (generally at least 30 for each basin). While the resulting map (Figure 5.5) does not reflect a consistent time period, it does give a sense of the overall spatial

118

differences between the two sources of precipitation information. Requiring at least 30 years of

data means that each basin samples close to the full range of yearly variability, as opposed to

only a relatively wet or dry period. As discussed for the example years above, the H10

precipitation pattern appears spatially smoother than the pattern of precipitation inferred from

streamflow (Figures 5.5a-5.5b). The spatial pattern of average streamflow shows the same strong

gradients between windward and leeward basins seen in single water years (Figure 5.5c). While

no basin has long-term H10 precipitation less than long-term streamflow (Figure 5.5d),

streamflow does suggest very high runoff ratios from H10 precipitation for many windward,

high-elevation basins (Figure 5.5f).

5.4.3. Comparison to other precipitation-gauge based gridded datasets

For comparison, we also plot the water year 1983 precipitation pattern of three other

gridded datasets (N15, L13 and NLDAS) against streamflow (Figure 5.6). Qualitatively, the

patterns of precipitation across the three datasets are similar, and also similar to H10, though

NLDAS appears to have slightly less precipitation across the Sierra Nevada than the other

datasets, and L13 appears to have somewhat more. In Figures 5.6d-5.6f, a set of basins shows

low values of precipitation minus streamflow: the upper Kaweah R. basins, the San Joaquin R.

and Granite Cr., Cherry-Eleanor and Falls Cr., the North Fork of the Stanislaus R. and Highland

Cr., several of the small basins draining to Lake Tahoe, Duncan Cr., and the South Fork of the

Yuba R. This set of basins is quite similar to those with low H10 precipitation minus streamflow

(*cf.* Figure 5.3d), suggesting that the spatial patterns of precipitation across different gridded

datasets are relatively similar.

Given that the H10, L13 and NLDAS datasets use PRISM precipitation climatology as a

topographic corrector in the interpolation of gauge observations, we compare the PRISM spatial

pattern against streamflow observations. In Figure 5.7, we show the 1971-2000 PRISM precipitation normal for each basin (Figure 5.7a); the 1971-2000 normal is used by both the H10 and L13 datasets. When compared against streamflow (Figures 5.7b and 5.7c), it is apparent that the spatial pattern of precipitation in PRISM is similar to those seen in the other precipitation datasets, because largely the same set of basins shows low values of precipitation minus streamflow (*e.g.*, San Joaquin, Cherry-Eleanor, and the Stanislaus basins). As in the previous cases, the spatial pattern of PRISM precipitation is much smoother than the spatial pattern of streamflow.

5.4.4. Cross-range transects of precipitation and streamflow

To resolve the spatial patterns of precipitation and streamflow with respect to the topography of the Sierra Nevada, we plot them along two cross-range transects (Figure 5.8). The first transect begins in the Tuolumne R. basin on the west slope and travels northeast through the Walker R. basin, while the second begins in the Mokelumne R. basin, passes through the North Fork of the Stanislaus R. basin before crossing the range into the Carson R. basin. Given predominant cool-season winds out of the southwest [*e.g.*, *Lundquist et al.*, 2015], these transects pass from windward to leeward basins.

In the first transect (Figure 5.8b), the greatest streamflow is seen near the top of the windward slope in the Cherry-Eleanor basin, while steep declines to both the southwest (Big Cr.), and in particular, the northeast (Walker R.) are seen. H10 precipitation generally places the greatest precipitation near the crest, but the maximum is likely too little given observed streamflow (Figure 5.8c). However, H10 precipitation appears compatible with streamflow and nearly matches inferred precipitation for Big Cr. and the Walker R. basins.

In the second transect (Figure 5.8d), streamflow also reaches its peak on the high windward face of the range, in the North Fork of the Stanislaus R. basin. Similar to the first transect, H10 precipitation appears insufficient in this basin; H10 precipitation is slightly less than streamflow in water year 1983 (Figure 5.8e). Also similar to the first transect, the leeward basin (East Fork of the Carson R.) has much better agreement between H10 and inferred precipitation, but the lower-elevation windward basin (South Fork of the Mokelumne R.) has much greater H10 precipitation than inferred precipitation. In both transects, the higher-elevation windward basins have generally southwest-facing aspects; winds from this direction encounter a consistent upslope aspect and little upstream blocking, conditions favorable for orographic enhancement [*Roe*, 2005].

5.4.5. Interannual variability in spatial patterns of precipitation and streamflow

By considering multiple stream gauges in the Sierra Nevada, we can examine changes in the spatial pattern of streamflow and inferred precipitation from year to year. For this analysis, we select a subset of years and basins with complete records, so differences between years are due to changes in streamflow, and not due to changes in the set of basins. We select 1974-1987, a period in which 43 basins have complete streamflow records (Figure 5.9a). Within that set, we identify 15 basins on the west slope of the Sierra Nevada that are potentially underestimated by the gridded datasets based on the findings in the previous sections, and average these basins' runoff and precipitation to represent the windward portions of the range. We identify 9 basins on the eastern side of the crest and average them to create a leeward group. Finally, we divide the basins into northern and southern groups (Figure 5.9a); the windward and leeward basins are also included in the appropriate northern and southern groups. With these indices of streamflow and precipitation, we can identify windward vs. leeward and north vs. south shifts between years.

In Figures 5.9b and 5.9c, we show windward-leeward and north-south ratios, respectively, of precipitation and streamflow. Streamflow is on average more than twice as great from the windward basins than from the leeward basins; however, the ratio varies from year to year, with 1978 in particular showing greater windward streamflow relative to the lee of the range (Figure 5.9b). This shift is reflected in the inferred precipitation, which has a lower mean windward-leeward ratio (~1.7), but generally follows the variability in streamflow. H10 windward-leeward precipitation ratios, however, are between 1.25 and 1.5, suggesting a weaker rain shadow, and show very little interannual variability. North-south ratios (Figure 5.9c) show substantial interannual variability, with precipitation and streamflow twice as large in the northern portion of the range compared to the south in some years, but near equal in others. Streamflow inferred precipitation and H10 precipitation track one another over this time period, though agreement is worse during the 1976-1977 drought years.

The annual difference between precipitation and streamflow in the Sierra Nevada implies the actual evapotranspiration (ET), as drainage to deep aquifers and changes in soil moisture storage are generally small compared to precipitation and streamflow. In some studies, ET has been estimated from the differences between basin-mean gridded precipitation and streamflow [*e.g.*, *Christensen et al.*, 2008; *Lutz et al.*, 2010], and so we plot implied ET from the different basin groups under both H10 and inferred precipitation in Figure 5.10. When considering windward-leeward differences, H10 precipitation implies that ET is greater in the leeward basins than the windward basins (Figure 5.10a). The difference from H10 precipitation between leeward and windward implied ET is variable, and appears larger in wetter water years (1978, 1982 and 1983). However, inferred precipitation implies that windward ET is greater than leeward ET, by a fairly constant margin. There is greater agreement between H10 and inferred precipitation

122

regarding north-south differences in implied ET (Figure 5.10b): both datasets generally show greater ET in the northern basins, though in some years this is not the case for the H10 dataset. We discuss patterns of ET in more detail in section 5.5.

5.4.6. Trends in precipitation and streamflow

Using a subset of basins with at least 90% complete streamflow records over the study period (1950-2010), we calculate trends in the water year totals of observed streamflow, H10 basin-mean precipitation and inferred basin-mean precipitation. While there are fewer basins with sufficiently long records to conduct the trend analysis, the spatial patterns of precipitation and streamflow trends can still be discerned (Figure 5.11). Trends in H10 precipitation show a strong north-south divide; the Kern and Kaweah basins in the south have positive precipitation trends, while all basins to the north have declines of 5-15%. In contrast, inferred precipitation (Figure 5.11b) shows increases in precipitation over the central Sierra Nevada in the Merced and San Joaquin basins, weak trends in the southern portion of the domain, and decreases in the northern basins. In general, this follows from the observed streamflow trends, which are most positive over the central Sierra Nevada and are negative for all basins from the Mokelumne R. northward (Figure 5.11c). The differences between the H10 and inferred precipitation trends are thus greatest over the central Sierra Nevada (Figure 5.11d).

The trends in Sierra Nevada precipitation and streamflow are generally consistent with the trends in precipitation, April 1 SWE and streamflow shown in chapter 2. Since 1950, precipitation, streamflow and SWE appear to have increased over the Sierra Nevada from approximately Yosemite National Park southward, while the northern Sierra Nevada has seen declines in precipitation, streamflow and SWE. The H10 dataset appears to capture some, though not all, of these shifts in spatial patterns. While it does show the positive trends in precipitation

123

in the southern Sierra Nevada that would be expected from the observed positive trends in streamflow and SWE, it does not capture the correct sign or magnitude of the increasing precipitation inferred over the central basins, in particular the San Joaquin, Merced and Walker basins.

5.4.7. Comparing gridded, modeled and inferred precipitation patterns

In Figure 5.12, we plot water year 2008 basin-mean precipitation for H10, WRF and inferred precipitation. These three datasets have independent sources of information: the spatial pattern of H10 is driven by PRISM-based precipitation gauge interpolation, while WRF simulates atmospheric dynamics, and inferred precipitation is derived from streamflow. While in comparison to earlier years there are fewer operating stream gauges in 2008 (when WRF precipitation is available), precipitation patterns can still be compared across the three datasets. WRF has the greatest precipitation amounts, particularly for basins in the northern portion of the range. By contrast, H10 has significantly less, while inferred precipitation is generally in between the other two. When precipitation minus streamflow is plotted for each (Figure 5.12, bottom row), different spatial patterns are seen: H10 again shows relatively low values over higher-elevation windward basins, while both WRF and inferred precipitation show patterns of precipitation minus streamflow that are greatest in the northern and high-elevation windward basins. Notably, WRF precipitation minus streamflow does not show low values in the basins previously identified with H10 precipitation and the other gridded datasets, suggesting that it is capturing the dynamics of local orographic enhancement and may better reproduce local runoff maxima in streamflow.

5.5. Discussion

5.5.1. Precipitation minus streamflow and ET estimates

Precipitation inferred from streamflow using the approach here essentially uses basin water balance closure and assumes that precipitation is the sum of observed streamflow and modeled actual ET, using the flexible precipitation input outlined in previous sections. Therefore, inferred precipitation represents a physically-based estimate of basin-mean precipitation given streamflow observations. Figure 5.2 and the results in chapters 3 and 4 make it clear that this formulation does produce substantial uncertainties in the absolute amounts of inferred precipitation, specifically due to uncertainties in hydrologic model structure and the parameterization of fluxes such as percolation and ET. Therefore, the absolute inferred precipitation quantities presented here should be viewed with some care due to the uncertainty in inferred ET. However, the results of chapter 3 also suggest that the year-to-year and basin-to-basin variability of inferred precipitation is more robust than the absolute amount of inferred precipitation [*Henn et al.*, 2015]. While each FUSE model structure has a bias relative to the others, this bias is quite consistent across water years and basins (Figures 3.5 and 3.9). So we interpret the spatial pattern of inferred precipitation, as well as the yearly wet vs. dry differences in inferred precipitation, as being physically-based. In contrast, gridded precipitation products such as H10 leverage available precipitation observations to produce interpolated patterns, but interpolating gauge observations to a grid is a statistical process, not a physical one, and as such is subject to statistical uncertainty.

Given these considerations, it is reasonable to check precipitation minus streamflow from both approaches against available observations and estimates of actual ET over the Sierra

Nevada. However, measurements of ET over the Sierra Nevada are few, and uncertainty is high. Observational studies have estimated actual ET to be between 400 and 800 mm yr^{-1} at the point scale [*Stephenson*, 1998; *Kurpius et al.*, 2003; *Ichii et al.*, 2009; *Goulden et al.*, 2012]. Studies of forest cover and soils have shown that deep rooting allow trees and vegetation to access water stored in weathered bedrock [*Rose et al.*, 2003; *Bales et al.*, 2011], allowing for greater ET and maintenance of forests despite shallow (<1 m) soils and nearly precipitation-free summers in the Sierra Nevada's Mediterranean climate.

However, ET observations are exclusively made at the point scale in or over forest cover, and the extrapolation to spatially distributed, basin-mean ET has greater uncertainty. *Goulden et al.* [2012] and *Goulden and Bales* [2014] developed a relationship between the normalized difference vegetation index (NDVI) observed by the Moderate-Resolution Imaging Spectroradiometer (MODIS) and observations of ET at multiple flux towers in California, and used it to estimate basin-mean ET and elevational gradients of ET. Basin-scale studies can also infer ET through consideration of precipitation and streamflow, but both must be known well since ET is often the remainder of these two larger terms. (Our findings suggest that gridded precipitation datasets may have too much uncertainty to apply this approach to reliably infer ET in many basins.) A two-year study in the forested Southern Sierra Critical Zone Observatory (SSCZO) in the Kings basin [*Bales et al.*, 2011] estimated ET to be 760 mm yr^{-1}, respectively. This suggests that basin-mean ET is generally less than point estimates made in forests, presumably due to the areas of limited transpiration (*i.e.*, rock, snow and grasses); a two-year water balance study of an above-treeline Kaweah basin found ET of 400 mm yr^{-1} [*Kattelmann and Elder*, 1991]. The relationship between forest and basin-mean ET may depend on the fraction of forest cover in the basin.

In Figure 5.13 we consider implied ET from the different precipitation data sources against observation-based estimates of ET over the Sierra Nevada. We plot precipitation, streamflow and precipitation minus streamflow against each basin's mean elevation in Figure 5.13a. The distribution of precipitation with elevation for H10 shows an increase at lower elevations, then a maximum at perhaps 2,000 m, followed by a slight decline above that elevation. The pattern for streamflow, and inferred precipitation, is more sensitive to elevation: streamflow increases rapidly from the lower elevation basins up to 2,000-2,500 m; there are few basins above this elevation. Inferred precipitation follows a similar pattern, but with a positive offset from streamflow. Precipitation minus streamflow (Figure 5.13b) for H10 shows a general decline with elevation, with the negative values noted above apparent at 1,800 m to 2,400 m in elevation. However, inferred precipitation minus streamflow is much less sensitive to elevation, with most values at all elevations between 400 and 800 mm yr^{-1}, presumably reflecting a relatively constant pattern of evapotranspiration in the FUSE models across the basins.

Point observations of ET from the Sierra Nevada are plotted against elevation in Figure 5.13c, including estimates from Ameriflux tower sites [*Kurpius et al.*, 2003; *Ichii et al.*, 2009] and SSCZO flux towers and basin water balance [*Bales et al.*, 2011; *Goulden et al.*, 2012]. We also apply the regression of ET against MODIS NDVI from *Goulden et al.* [2012] to estimate basin-mean ET for each of the streamflow basin (gray circles). Collectively, these observations show that ET likely peaks in the middle elevations (1,000-2,000 m), with declines at higher and lower elevations. We can interpret this difference given arguments about ET over the elevation range of the Sierra Nevada. While higher, colder basin have shorter growing seasons and less PET, they also have greater water availability compared to lower-elevation basins due to greater precipitation. Given the Sierra Nevada's very arid summers and shallow soils, actual ET is

127

generally driven more by water availability than PET, *i.e.*, the climate is generally water-limited at all but the highest elevations. The resulting tradeoffs appear to favor the greatest actual ET in the middle elevations, particularly on windward slopes [*Christensen et al.*, 2008; *Goulden et al.*, 2012], which is also where the greatest forest cover is located [*Stephenson*, 1998]. The NDVI-based approach in particular suggests that there is low (<400 mm yr^{-1}) ET in high-elevation and leeward basins. However, because *Goulden et al.* [2012] based their regression on flux towers that are located primarily at sites at lower elevations and in more arid regions than the high Sierra Nevada, they note that it may underestimate ET in energy-limited alpine areas where NDVI is not a major indicator of water availability. For example, studies of evaporation and sublimation from the snowpack in the Sierra Nevada have estimated it to be from near zero to more than 300 mm yr^{-1}, the substantial uncertainty is due to a lack of direct measurements [*Kattelmann and Elder*, 1991; *Leydecker and Melack*, 1999]. Thus, the NDVI-based estimates are probably best considered as a lower bound of ET in the higher-elevation basins.

In the lower row of Figure 5.13, we plot the spatial pattern of implied ET from H10 (d), inferred precipitation (e) and from the NDVI-based regression (f). The spatial patterns exhibit significant differences, with H10 and inferred precipitation showing greater implied ET than the NDVI-based approach over many of the higher-elevation basins in the central and southern Sierra Nevada as well as east of the crest. The different ET estimates also suggest that the precipitation inferred from streamflow using the FUSE models may overestimate high-elevation ET in basins with little vegetative cover. However, the NDVI-based approach does suggest that basins of similar elevation and forest cover (*e.g.*, the Tuolumne and Merced basins in the central portion of the range) have similar amounts of ET. While the NDVI-based ET amounts are less than those from inferred precipitation, these spatial patterns of variability (Figures 5.13e-f) are

more similar than the H10 ET pattern (Figure 5.13d). The NDVI-based approach does not support aspects of the H10 dataset in which certain basins (*e.g.*, Cherry-Eleanor) have much less ET than basins of similar elevation, or that ET is greater on the leeward slope of the Sierra Nevada as compared to the windward slope. We interpret those features of the H10 dataset to be more likely due to underestimation of precipitation in grid interpolation.

5.5.2. Gridded underestimation of precipitation in favored windward basins

The maps (Figures 5.3-5.7) and transect plots (Figures 5.8) of the difference between gridded and inferred precipitation yield several consistent spatial patterns. First, the likely underestimation of precipitation by H10 occurs over higher-elevation windward basins with generally southwest aspect (Duncan Cr., Cole Cr., the North Fork of the Stanislaus R. and Highland Cr., Cherry-Eleanor and Falls Cr., Granite Cr. and the San Joaquin R., and the Marble Fork and main stem of the Kaweah R.). Lower-elevation basins on the west slope, basins in the far southern Sierra Nevada, and basins east of the crest generally show better agreement, or even potential overestimation by H10. This pattern is illustrated in Figures 5.8, which shows the greatest disagreement occurring in relatively high-elevation windward basins. The apparent enhancement and mainly southwestern aspect of these basins agrees with findings that the predominant moisture transport into the Sierra Nevada is from low-level southwesterly winds during cool-season storms [*e.g.*, *Lundquist et al.*, 2015].

Given that these findings are essentially the result of basins with higher streamflow than others at similar elevation and latitude, the reliability of the streamflow observations deserves consideration. Streamflow measurements are subject to uncertainty in the rating curve and depth measurements, which produce instantaneous errors in flow measurement, and may lead to biases if not validated with reliable instream flow measurements. However, we consider the likelihood

of large, long-term biases in streamflow data to be relatively small, assuming that the gauges are well-maintained and that the USGS considers the records to be of reasonable quality. It is possible that any individual gauge has a substantial bias; however, the patterns of streamflow seen in Figures 5.3-5.5 appear to be spatially coherent. They exhibit as a generally north-to-south, wet-to-dry pattern, and appears to favor southwest-facing basins with no upstream blocking. The apparent enhancement in these topographically favored basins is supported by theoretical and observational studies of relatively small-scale precipitation patterns in complex terrain [*Frei and Schar*, 1998; *Roe*, 2005; *Anders et al.*, 2007; *Minder et al.*, 2008]: mountainous areas receiving the greatest precipitation generally are those in which prevailing moisture-bearing winds are forced upslope due to favorable topographic aspect. These factors suggests that the long-term streamflow measurements are relatively robust.

The patterns appears also fairly consistent in time across the water years surveyed and in the long-term average pattern. Thus, it suggests that there may be low bias in the spatial mean pattern of precipitation in the gridded datasets, which in H10 (and most similar datasets) is set to match a long-term precipitation climatology from PRISM [*Daly et al.*, 1994, 2008]. Other PRISM-based gridded datasets (L13, NLDAS) show relatively similar patterns of potential underestimation (Figure 5.6), suggesting that this climatology may be a source of the underestimation. PRISM uses a detailed topographic interpolation scheme that reflects the spatially-variable relationship between precipitation and elevation on slopes of different aspect. However, it ultimately relies on gauge observations (as well as April 1 snow course observations in some cases) to obtain the elevation-precipitation relationship. If there are few or no gauges on certain preferential slopes where air parcels are forced upward due to aspect and topographic

convergence [*Roe*, 2005], it will be difficult for the regression to capture these local maxima. As a result, it may underestimate precipitation for basins in these areas.

Because many datasets use PRISM climatology to establish the mean spatial pattern of precipitation over complex terrain, the potential biases seen in Figures 5.3-5.6 do not just apply to the H10 dataset, but to a broader array of gridded precipitation products. In fact, even the gridded precipitation dataset that does not rely on PRISM (N15) shows a relatively similar spatial pattern of precipitation. Given these findings, it appears that underestimation of local precipitation maxima in complex terrain could be a widespread feature of gridded precipitation datasets. However, it is notable that WRF precipitation does appear to better represent precipitation in some of the basins with very high streamflow (Figure 5.12). Some have argued that WRF precipitation over mountains is generally too high [*e.g., Caldwell et al.*, 2009], and the WRF data shown here does produce high values of precipitation minus streamflow overall. But given that gridded precipitation datasets are usually used to evaluate WRF, it appears possible that the apparent wet biases in WRF in some areas actually may be dry biases in the gridded products.

We also find evidence of year-to-year variability in the potential biases in H10 precipitation. While the differences between inferred and H10 precipitation appear to be largely temporally consistent, in that the sign of the differences tends to be the same for a given basin from year to year, there are also differences between the wet and dry years shown in Figures 5.3 and 5.4. Water year 1977 inferred precipitation and streamflow, while very low, is mostly in the northern Sierra Nevada. H10 precipitation is more uniform from north to south, and so the comparison with inferred precipitation (Figure 5.4f) suggests H10 underestimation primarily in the northern Sierra Nevada. In contrast, for the very wet water year 1983, the greatest differences

131

between H10 and inferred precipitation occur throughout the Sierra, but primarily at higher-elevation, windward basins (Figure 5.3f). In addition, the gridded dataset appears to underestimate changes in year-to-year ratios of windward and leeward precipitation (Figure 5.9). These findings suggests that the biases in gridded precipitation may vary depending on the particular storm tracks and spatial patterns of precipitation in a water year [*e.g., Lundquist et al.*, 2010].

5.6. Conclusions

From the comparison between streamflow from 56 long-term, largely unimpaired Sierra Nevada stream gauges, gridded precipitation, and precipitation inferred from streamflow, over water years 1950 to 2010, we draw the following conclusions:

1) There are differences between the mean spatial pattern of gridded precipitation in H10 (and other gridded datasets, most of which are set by PRISM climatology) and the spatial pattern of precipitation inferred from streamflow. In particular, streamflow implies greater precipitation in higher-elevation, southwest-facing basins than is seen in H10 and other gridded datasets.

2) The spatial and temporal variability of precipitation inferred from streamflow (*e.g.*, precipitation gradients over the crest of the Sierra Nevada) is higher than in H10 gridded precipitation. While absolute uncertainties in inferred precipitation are relatively high, the spatial patterns are relatively robust, such that the gradients in streamflow also likely represent gradients in the precipitation pattern.

3) The relationship between implied actual ET (precipitation minus streamflow) across the crest of the Sierra Nevada is different between the gridded precipitation dataset and

inferred precipitation: H10 implies actual ET is similar or greater in a set of basins to the lee of the crest relative to a set on the windward side, while implied actual ET is greater on the windward side with inferred precipitation.

4) 1950-2010 trends in inferred precipitation show increases over the central and southern Sierra Nevada; in contrast, H10 precipitation trends show declines over the central Sierra Nevada.

5.7. Tables

Table 5.1. Sierra Nevada unimpaired streamflow gauge and basin information.

USGS Gauge Number	USGS Gauge Name	Watershed	Latitude	Longitude	Gauge Elevation [m]	Basin Mean Elevation [m]	Basin Peak Elevation [m]	Basin Area [km2]	Starts	Ends	Years	HCDN	Quality	Notes
10270872	MF BISHOP C BL LK SABRINA NR BISHOP CA	Owens	37°12''50''	118°36''34''	2,762	3,457	4,212	43.6	1990	-	26	Y	-	Regulation by Lake Sabrina (USGS station 10270870); mass balance correction applied to flow record
10276001	COMBINED Q BIG PINE C & GIROUX DITCH CA		37°08''42''	118°18''52''	1,372	2,997	4,336	100.5	1931	1978	48	N	Poor	Includes both river and diversion
10281800	INDEPENDEN CE C BL PINYON C NR INDEPENDEN CE CA		36°46''43''	118°15''49''	1,616	2,996	4,130	31.3	1924	1978	55	N	Poor	No diversions or regulations
10286001	COTTONWOO D C NR OLANCHA CA.+ PH CA		36°26''20''	118°04''48''	1,421	3,058	4,263	103.8	1907	1978	72	N	Poor	Includes both river and diversion
10291500	BUCKEYE CK NR BRIDGEPORT, CA	Walker	38°14''20''	119°19''30''	2,104	2,804	3,612	114.1	1954	-	62	Y	Good	No regulation or diversion above station; WYs 1980-1995 missing
10295500	L WALKER RV NR BRIDGEPORT, CA		38°21''39''	119°26''38''	2,070	2,647	3,527	163.0	1945	2008	64	Y	Good	Small diversions above stations; WYs 1987-1995 missing

10296000	W WALKER R BL L WALKER R NR CO		38°22''47"	119°26''57"	2,010	2,696	3,577	469.1	1939	-	77	Y	Fair	Small diversions upstream; slight regulation by Poore Lake
10296500	W WALKER R NR COLEVILLE, CA		38°30''48"	119°26''56"	1,683	2,624	3,577	632.0	1958	-	58	Y	Fair	Small diversions upstream; slight regulation by Poore Lake
10308200	E FK CARSON RV BLW MARKLEEVILLE CK NR MARKLEEVILLE		38°42''53"	119°45''50"	1,646	2,412	3,489	717.2	1961	-	52	Y	Fair	Small diversions upstream; slight regulation by small reservoirs
10310000	W F CARSON R AT WOODFORDS, CA	Carson	38°46''11"	119°49''58"	1,754	2,458	3,309	170.0	1939	-	74	Y	Fair	Small diversions upstream; small regulation upstream
10310400	DAGGETT CK NR GENOA, NV		38°57''52"	119°50''57"	1,555	2,220	2,915	10.1	1972	-	39	Y	Good	No diversions above station; WYs 1984-1988 missing
10336660	BLACKWOOD CREEK NR TAHOE CITY		39°06''27"	120°09''40"	1,901	2,217	2,704	29.6	1961	-	55	Y	Fair	No known diversion or regulation; ice error in winter
10336676	WARD C AT HWY 89 NR TAHOE PINES CA		39°07''56"	120°09''24"	1,899	2,222	2,704	24.9	1973	-	43	Y	Good	Minor diversions upstream
10336780	TROUT CREEK NR TAHOE VALLEY CA	Tahoe	38°55''12"	119°58''17"	1,903	2,426	3,318	93.9	1961	-	55	Y	Good	Minor diversions upstream; ice error in winter
10343500	SAGEHEN CREEK NR TRUCKEE CALIF		39°25''54"	120°14''13"	1,927	2,164	2,656	27.6	1954	-	62	Y	Good	No storage or diversion upstream

Station ID	Name	Basin	Latitude	Longitude										Comments
11186001	COMBINED FLOW OF KERN R AND KERN R NO 3 CA		35°56''43"	118°28''36"	1,104	2,632	4,403	2,190.2	1911	-	104	Y	-	Includes both river and SCE diversion
11187000	KERN RIVER AT KERNVILLE CALIF	Kern	35°45''16"	118°25''21"	799	2,495	4,403	2,610.5	1954	1993	40	N	Good	A few small diversions for irrigation upstream from station
11189500	SF KERN R NR ONYX CALIF		35°44''15"	118°10''22"	884	2,357	3,691	1,371.0	1946	-	66	Y	-	<15 cfs diversions upstream
11199500	WHITE R NR DUCOR CA		35°48''36"	118°55''03"	218	744	2,526	234.9	1972	2005	33	N	Good	No regulation or diversion
11202001	NF OF MF TULE R NR SPRINGVILLE CALIF (TOTAL FLOW)	Tulare	36°10''29"	118°41''41"	890	2,057	3,116	101.4	1940	-	72	Y	-	Includes both river and PG&E diversion
11204500	SF TULE R NR SUCCESS CA		36°02''33"	118°51''24"	235	1,238	2,831	285.2	1931	1990	60	N	-	
11206501	MF KAWEAH R NR POTWISHA CAMP CALIF (TOTAL FLOW)		36°30''47"	118°47''27"	640	2,287	3,826	266.1	1950	2002	53	N	-	Includes both river and SCE diversion
11208001	MARBLE FK KAWEAH AT POTWISHA CP CALIF TOTAL FLOW	Kaweah	36°31''19"	118°47''54"	674	2,403	3,494	132.7	1950	2002	53	N	-	Includes both river and SCE diversion
11208731	COMBINED FLOW EF KAWEAH R NR THREE RIVERS CA		36°27''05"	118°47''15"	762	2,315	3,791	222.5	1957	2002	29	N	-	Includes both river and SCE diversion; WYs 1978-1993 missing

11209900	KAWEAH RIVER AT THREE RIVERS CALIF		36°26''38"	118°54''09"	247	1,894	3,826	1,085.3	1959	1990	42	N	Good	Diversions of up to 200 acres upstream
11210100	SOUTH FORK KAWEAH RIVER AT THREE RIVERS, CALIF.		36°25''00"	118°54''48"	246	1,578	3,472	230.3	1959	1990	42	N	Good	Several small diversions upstream from station for irrigation
11213500	KINGS RIVER AB NF NR TRIMMER CALIF		36°51''48"	119°07''24"	305	2,621	4,338	2,463.4	1932	1982	51	N	Good	No diversions or regulations
11218501	COMBINED FLOW KINGS R BL N F & KINGS R PP CA	Kings	36°52''29"	119°08''27"	287	2,551	4,338	3,471.5	1953	1993	41	n	Good	NF Kings portion regulated post-1957; includes both river and powerplant flow
11221700	MILL CREEK NEAR PIEDRA CALIF		36°49''07"	119°20''27"	168	782	2,197	328.3	1958	1994	37	N	Good	Some small diversions upstream from station for irrigation
11226500	SAN JOAQUIN R AT MILLER CROSSING CALIF		37°30''38"	119°11''47"	1,393	2,781	4,009	650.5	1922	1991	47	N	-	WYs 1929-1951 missing
11228500	GRANITE C NR CATTLE MOUNTAIN CA		37°31''36"	119°15''28"	2,073	2,709	3,539	123.9	1966	1986	21	N	-	No diversion from station
11230500	BEAR CR NR LAKE T.A.EDISON CALIF	San Joaquin	37°20''22"	118°58''21"	2,246	3,244	4,177	136.3	1922	-	94	Y	-	No storage or diversion upstream of station
11237500	PITMAN C BL TAMARACK CREEK CALIF		37°11''55"	119°12''46"	2,140	2,436	3,002	59.9	1928	-	88	Y	-	No diversion upstream from station

11264500	MERCED R AT HAPPY ISLES BRIDGE NR YOSEMITE CALIF	Merced	37°43''53''	119°33''29''	1,225	2,747	3,980	469.8	1916	-	100	Y	Good	Up to 5 cfs can be diverted upstream from station
11266500	MERCED RIVER AT POHONO BRIDGE NEAR YOSEMITE, CAL		37°43''01''	119°39''55''	1,177	2,577	3,980	834.5	1916	-	100	Y	Good	No diversions between Happy Isles and Pohono
11275000	FALLS C NR HETCH HETCHY CA	Tuolumne	37°58''15''	119°45''48''	1,631	2,599	3,545	119.4	1916	1983	68	N	Good	No regulation or diversion
-	TUOLUMNE R A HETCH HETCHY		37°58''15''	119°47''18''	1,067	2,723	3,984	1,172.8	1970	-	46	Y	Fair	HHWP full natural flows
-	CHERRY CR A CHERRY LAKE + ELEANOR CR A LAKE ELEANOR		37°58''29''	119°54''36''	1,357	2,344	3,275	485.7	1974	2013	40	Y	Fair	HHWP full natural flows; includes flow of Eleanor basin, connected by tunnel
11281000	SF TUOLUMNE RIVER NR OAKLAND RECREATION CAMP CAL		37°49''18''	120°00''43''	854	1,692	2,813	225.6	1924	-	85	N	Good	No diversion, small regulation upstream; USGS record ends in 2002; reappears on CDEC in 1/2009 as STO
11282000	MIDDLE TUOLUMNE R AT OAKLAND RECREATION CAMP CAL		37°49''42''	120°00''38''	854	1,869	2,930	165.1	1917	-	92	N	Good	No regulation, small diversions; USGS record ends in 2002; reappears on CDEC in 1/2009 as MTO
11283500	CLAVEY RIVER NEAR BUCK MEADOWS, CALIF.		37°54''02''	120°04''15''	724	1,802	2,825	378.2	1960	1994	35	N	Good	No storage or diversion upstream of station; WYs 1984-1986

138

													missing
11284400	BIG C AB WHITES GULCH NR GROVELAND CA		37°50"31"	120°11"02"	781	961	1,210	41.7	1970	-	46	Y	Good	No storage or diversion
11292000	MF STANISLAUS R AT KENNEDY MDWS NR DARDANELLE CA		38°17"51"	119°44"25"	1,929	2,725	3,520	123.6	1938	-	78	Y	-	No diversion, but higher summer low flows due to releases from Relief Reservoir
11292500	CLARK FORK STANISLAUS RIVER NEAR DARDANELLE, CAL	Stanislaus	38°21"50"	119°52"13"	1,679	2,497	3,428	174.8	1951	1994	44	N	Good	No regulation or diversion upstream
11293500	NF STANISLAUS R BL SILVER CREEK CALIF		38°26"22"	120°00"53"	2,036	2,303	2,868	74.4	1953	1987	35	N	-	Low/medium flows regulated by Union/Utica/Alp ine reservoirs; no diversions
11294000	HIGHLAND C BL SPICER MEADOWS RES CALIF		38°23"35"	119°59"53"	1,933	2,343	3,024	119.8	1953	1987	60	Y	-	
11315000	COLE C NR SALT SPRINGS DAM CALIF		38°31"09"	120°12"42"	1,805	2,260	2,842	54.1	1928	-	85	Y	-	Occasional pumping upstream from station for summer homes
11316800	FOREST CREEK NEAR WILSEYVILLE, CALIF.	Mokelumne	38°24"12"	120°26"45"	899	1,419	2,108	54.6	1960	-	53	Y	-	No regulation, minor diversions upstream for irrigation and domestic use

11317000	MF MOKELUMNE R A WEST POINT CA		38°23''23"	120°31''32"	747	1,336	2,248	178.0	1911	-	105	Y	-	Up to 10 cfs diversion from this basin to SF Mokelumne; slight regulation by Schaads Res. since 1940; other small diversions
11318500	SF MOKELUMNE R NR WEST POINT CA		38°22''06"	120°32''40"	595	1,250	2,113	194.6	1933	-	82	Y	-	Up to 10 cfs diversion from MF Mokelumne to this basin; up to 9 cfs diversion out of basin to Jeff Davis Res.
11335000	COSUMNES R A MICHIGAN BAR CA		38°30''01"	121°02''39"	51	934	2,357	1,387.3	1907	-	109	Y	Good	Regulation and diversion by Jenkinson Lake, some transfers out of basin
11414000	SOUTH YUBA RIVER NEAR CISCO, CALIF.	Yuba	39°19''17"	120°33''48"	1,683	2,106	2,754	140.3	1943	1994	52	N	Fair	Low flow regulated by several small PG&E reservoirs
11427000	NF AMERICAN R AT NORTH FORK DAM CALIF	American	38°56''10"	121°01''22"	218	1,329	2,736	884.0	1942	-	74	Y	Good	Minor regulation by Lake Clementine, Big Reservoir, Lake Valley Reservoir; diversion into Bear River basin; minor effect on natural flows
11427700	DUNCAN CREEK NR FRENCH MEADOWS CALIF		39°08''09"	120°28''39"	1,607	1,950	2,263	25.6	1961	-	55	Y	-	No regulation or diversion upstream from station

11431800	PILOT CREEK ABOVE STUMPY MEADOWS RESERVOIR, CALI		38°53'41"	120°34'02"	1,305	1,568	1,883	30.2	1960	2008	48	N	Good	No regulation or diversion upstream from station
11440000	ALDER C NR WHITEHALL TOTAL FLOW CA		38°45'19"	120°22'17"	1,171	1,801	2,274	57.2	1923	1981	59	N	-	Includes flow diverted into pipeline
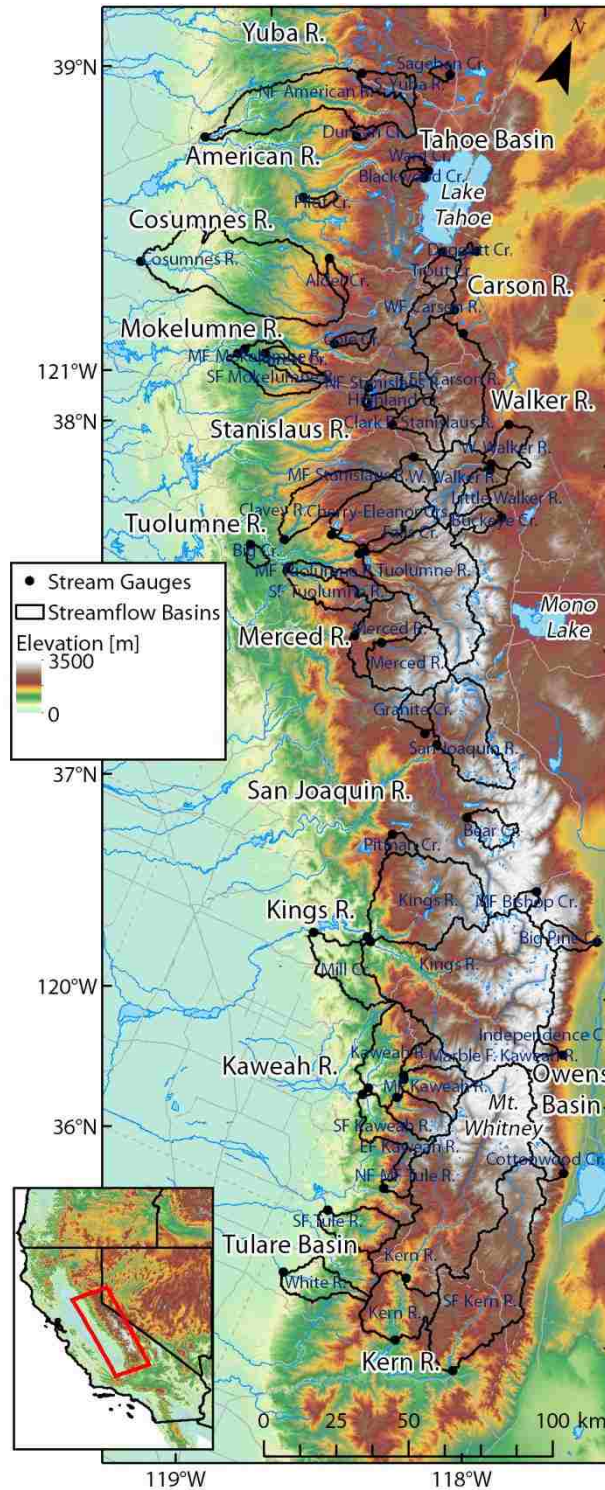
Figure 5.1. Topographic map of the Sierra Nevada showing streamflow basins used in this study; major drainages are labeled.
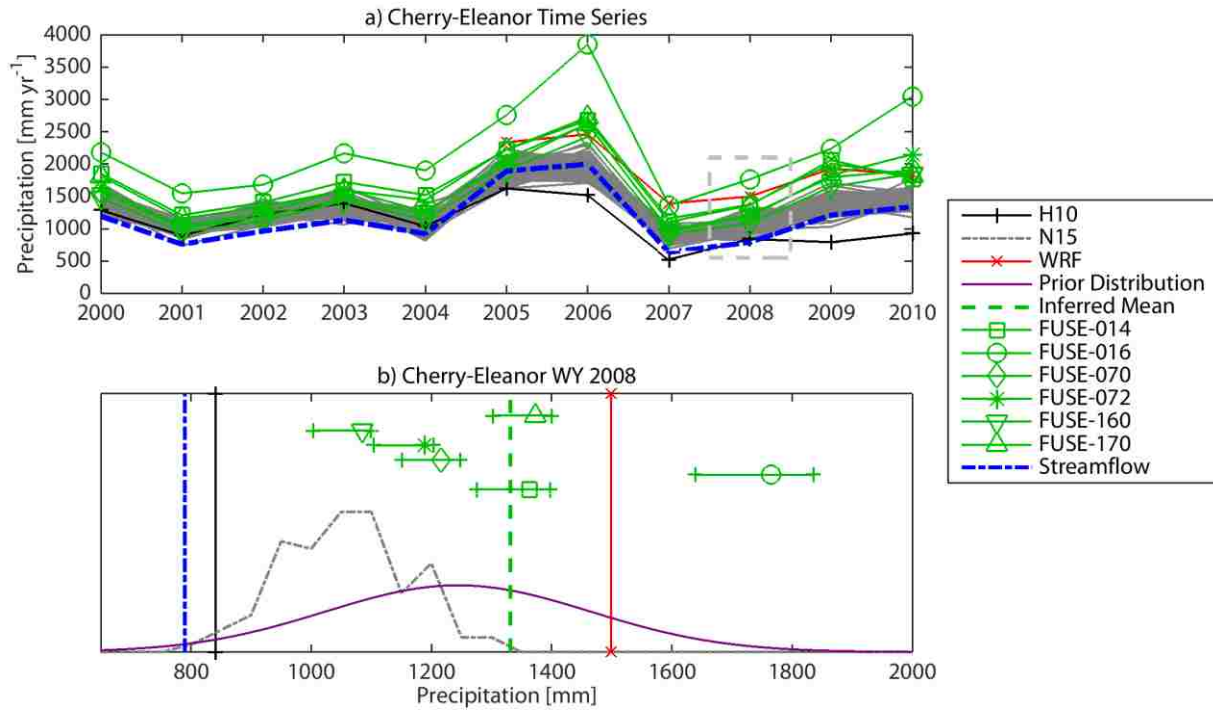
Figure 5.2. Comparison of prior and posterior precipitation distributions for the Cherry-Eleanor basin. a) Time series of water year total precipitation from H10, N15 and WRF, along with streamflow, and the precipitation inferred from streamflow (FUSE structures). b) Probability distributions of precipitation for Cherry-Eleanor in water year 2008, showing fixed precipitation dataset values, the prior distribution that is based on the datasets, and the posterior distribution of precipitation from the six FUSE structures. Vertical axis is normalized so that distributions sum to 1; inferred precipitation is shown as credible intervals with the symbol indicating the mean value. FUSE intervals are plotted at arbitrary vertical positions to clearly show their horizontal distributions.
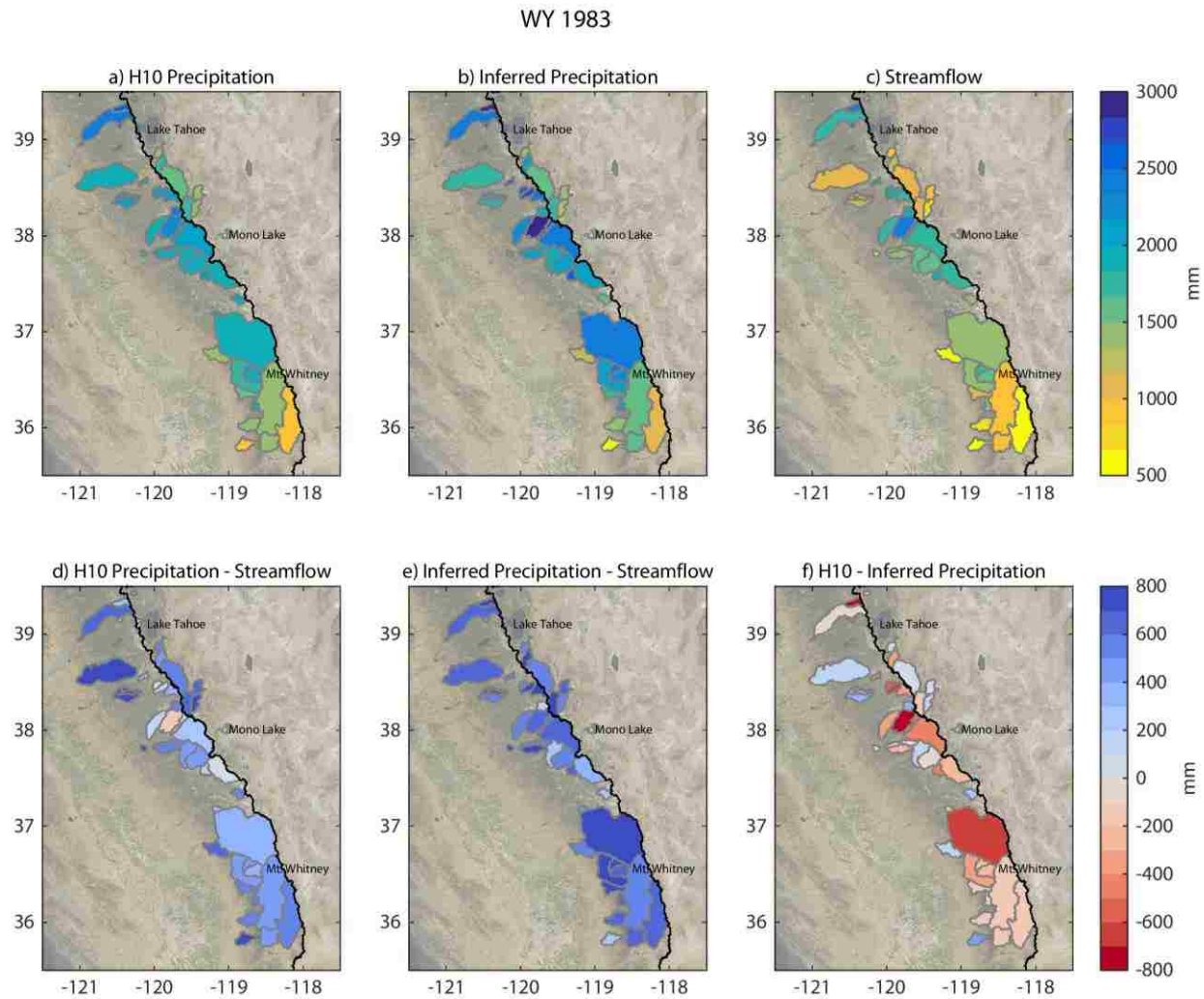
Figure 5.3. Basin-mean precipitation for water year 1983 from H10 (a) and from inferred precipitation (b). c) Water year streamflow totals. d) and e): Basin-mean differences between H10 and streamflow, and inferred precipitation and streamflow, respectively. f) Basin-mean differences between H10 and inferred precipitation. The crest of the Sierra Nevada is shown in black; background image from LANCE Rapid Response MODIS images (http://lance-modis.eosdis.nasa.gov/).
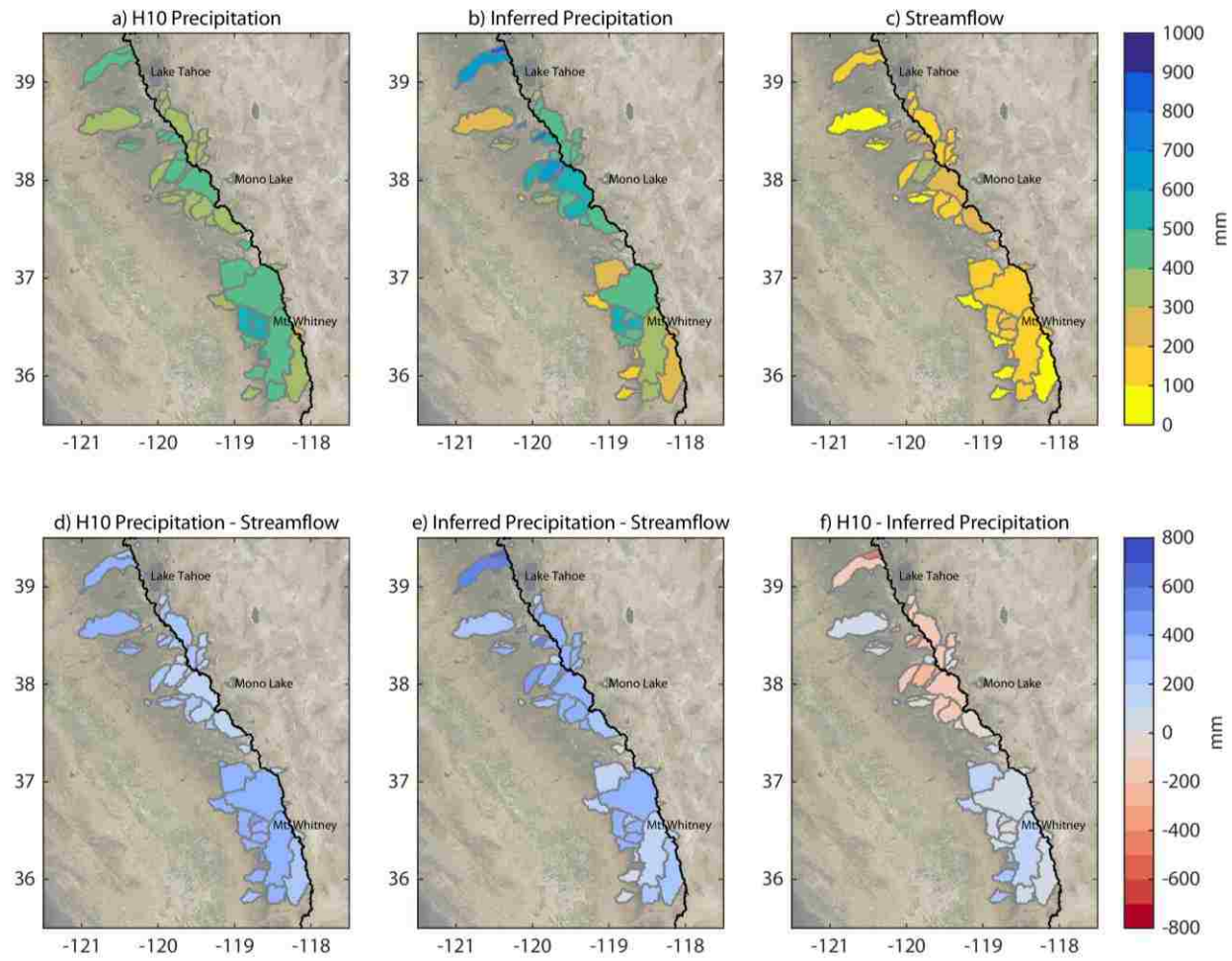
Figure 5.4. Same as Figure 5.3, but for water year 1977. Note different color scale.
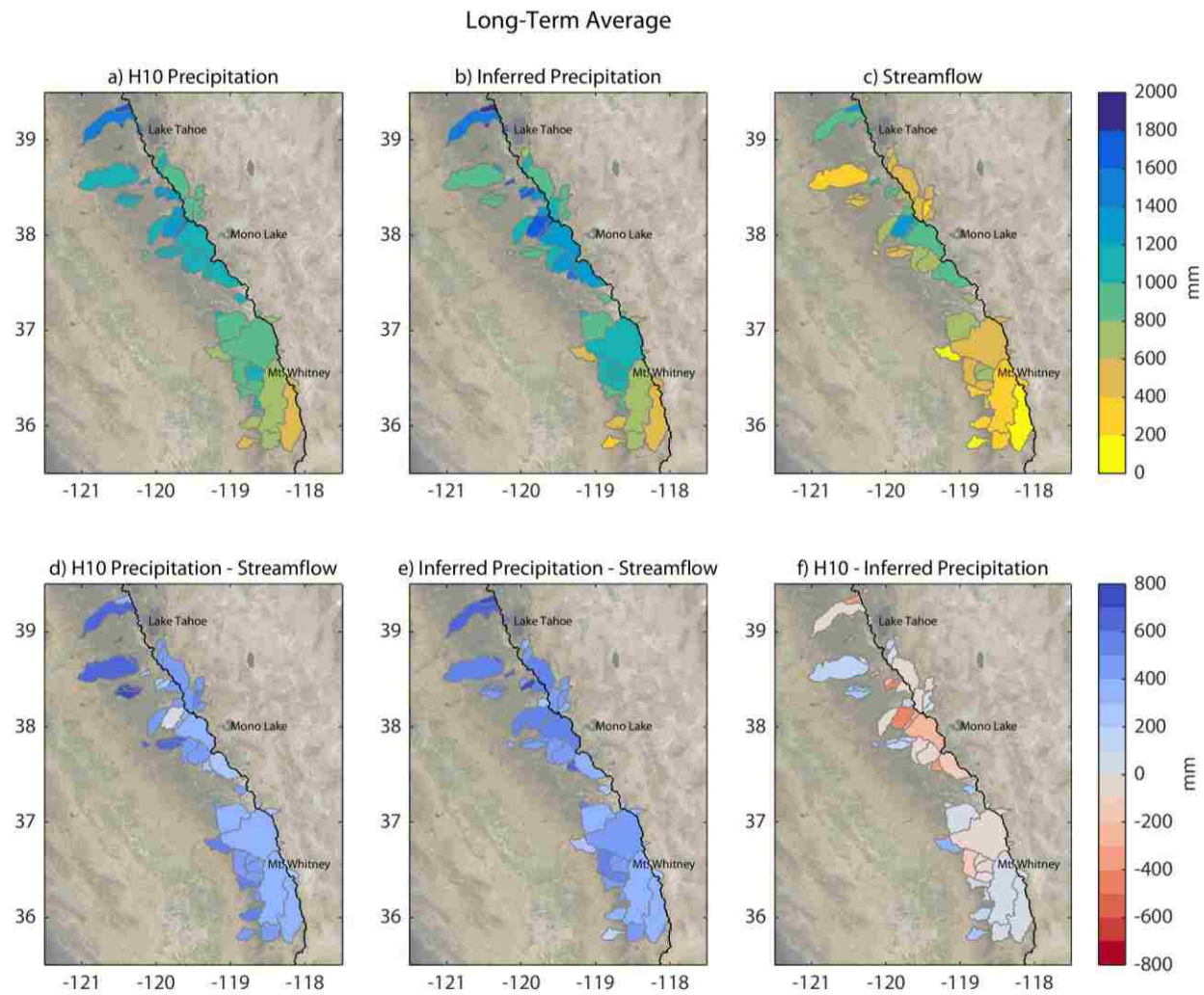
Figure 5.5. Same as Figures 5.3 and 5.4, but for basin-mean averages of all years in which streamflow is available for each basin.
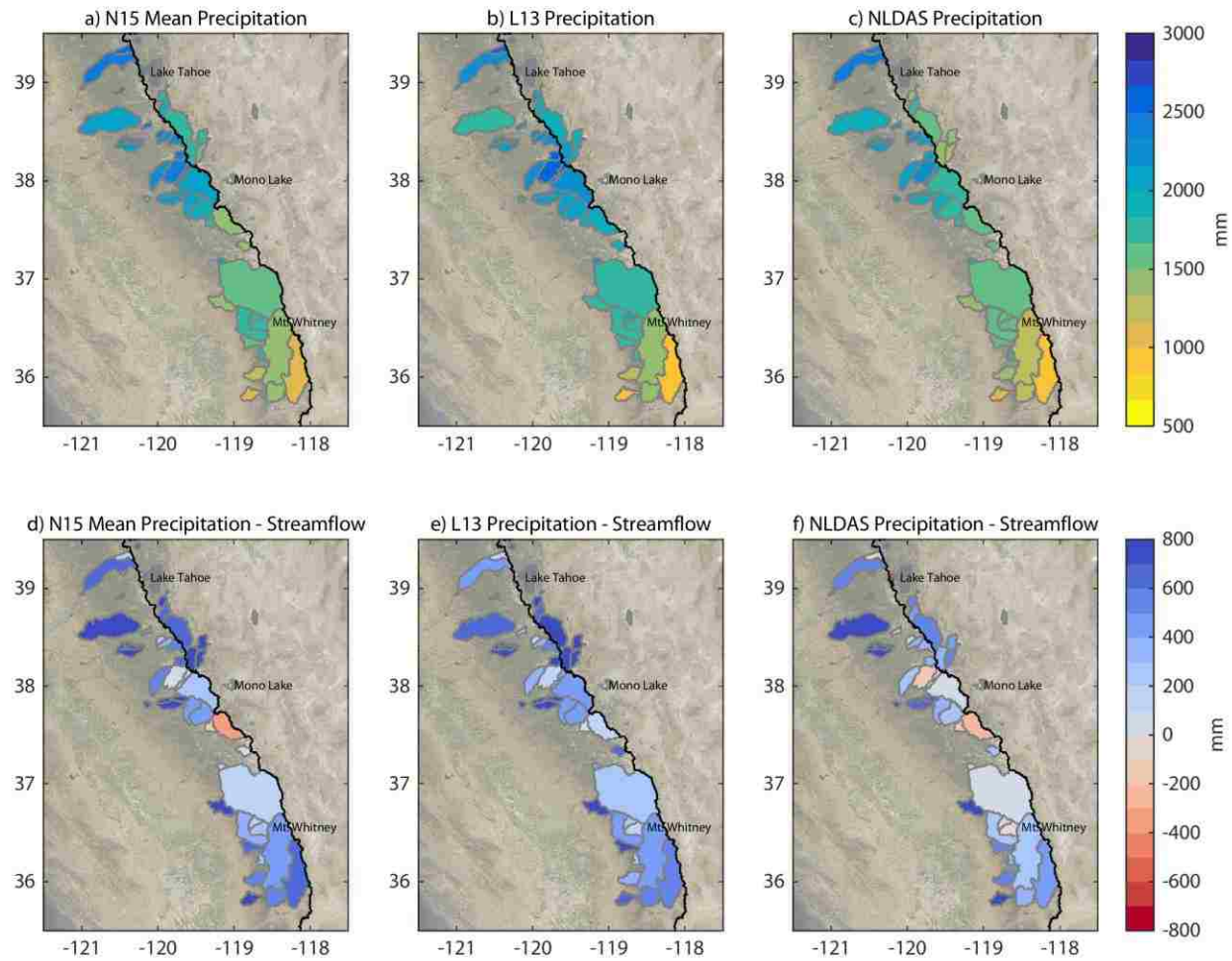
Figure 5.6. Top row: water year 1983 basin-mean precipitation for three other gridded datasets (N15, L13 and NLDAS). Bottom row: precipitation minus streamflow for each of the same datasets.
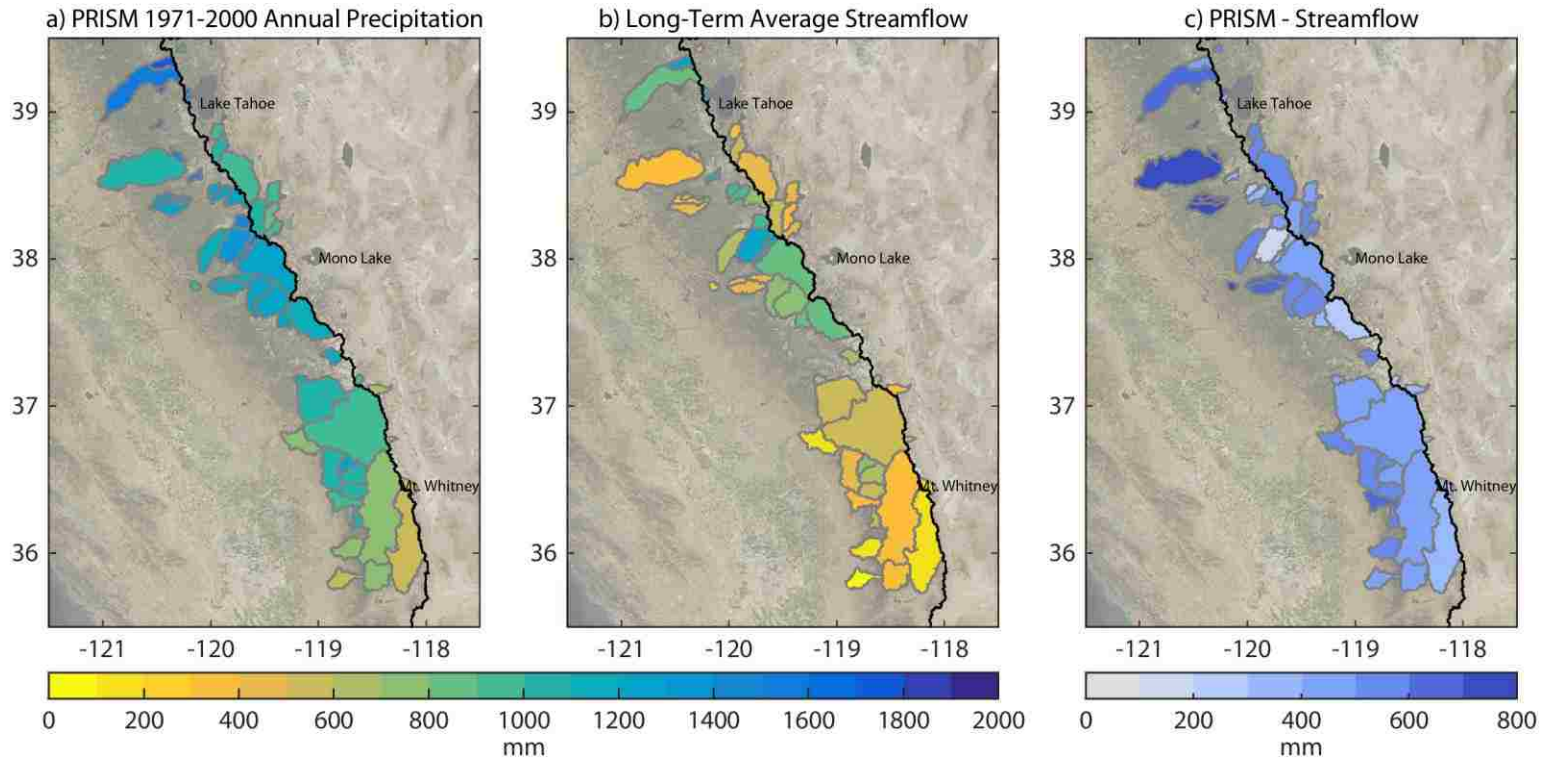
Figure 5.7. PRISM 1971-2000 precipitation normal for each streamflow basin (a), streamflow (b), and PRISM precipitation minus streamflow (c).
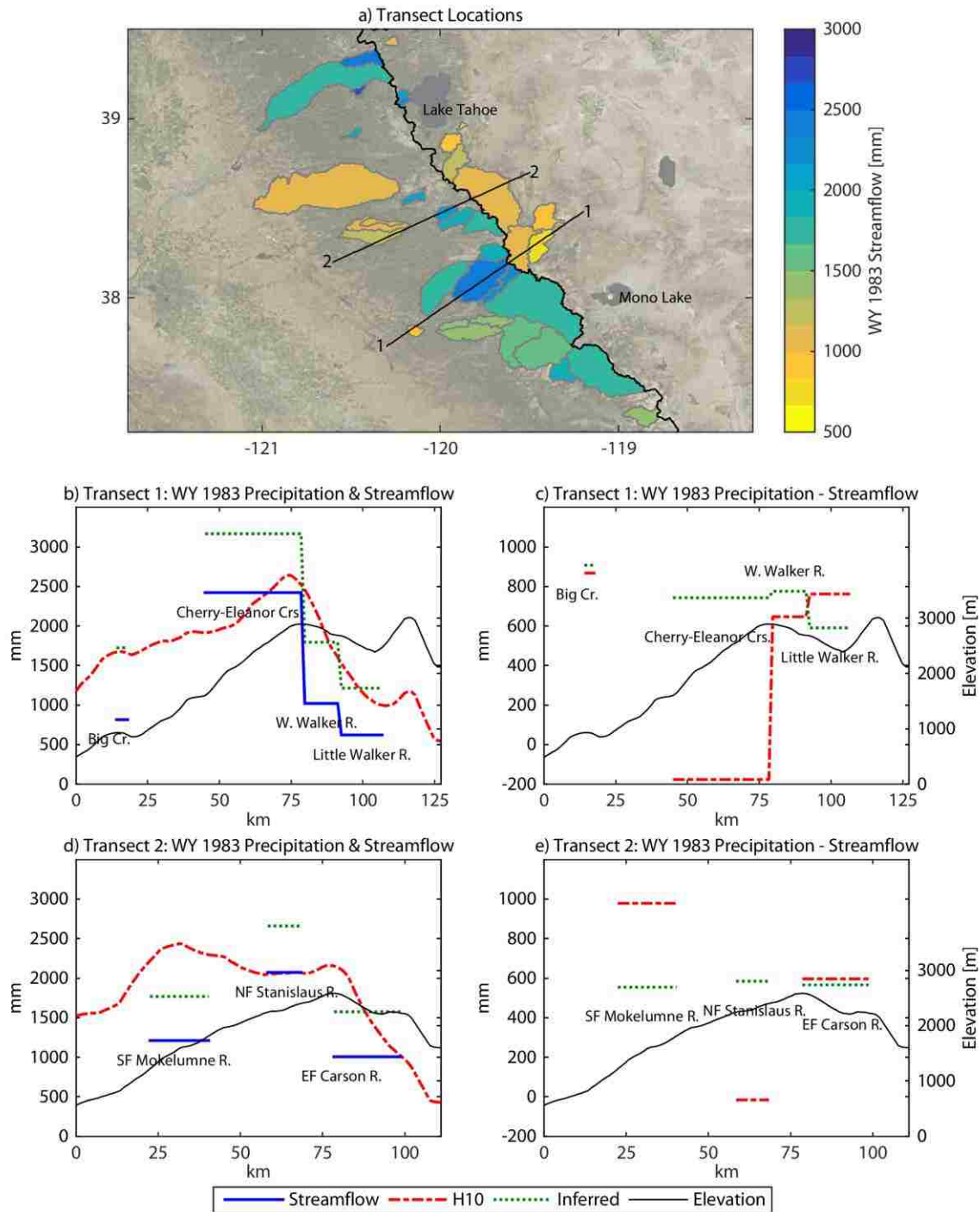
Figure 5.8. Cross-Sierra Nevada patterns of streamflow and precipitation. a) Inset of 1983 streamflow map, showing locations of two transects that pass from southwest to northeast across the range. b), d) Transects of water year 1983 streamflow and H10 and inferred precipitation, along with elevation. c), e) Precipitation (H10 and inferred) minus streamflow for the same transects. Streamflow, inferred precipitation, and precipitation minus streamflow are shown as the basin-mean for all points within the basin.

149
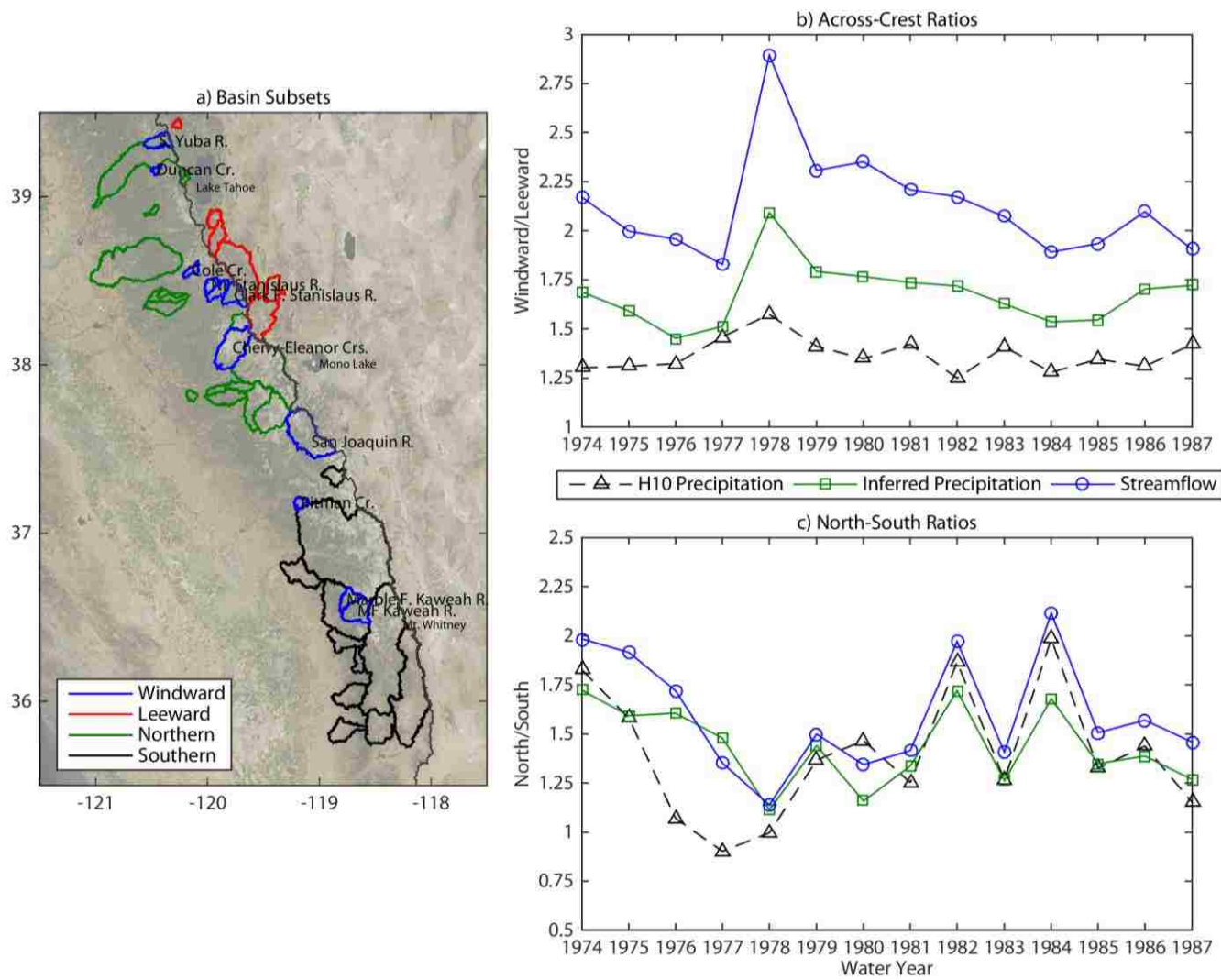
Figure 5.9. a) Map of basins indicating windward, leeward, northern and southern groups. Time series of windward-leeward (b) and north-south (c) ratios of precipitation and streamflow over water years 1974-1987.

150
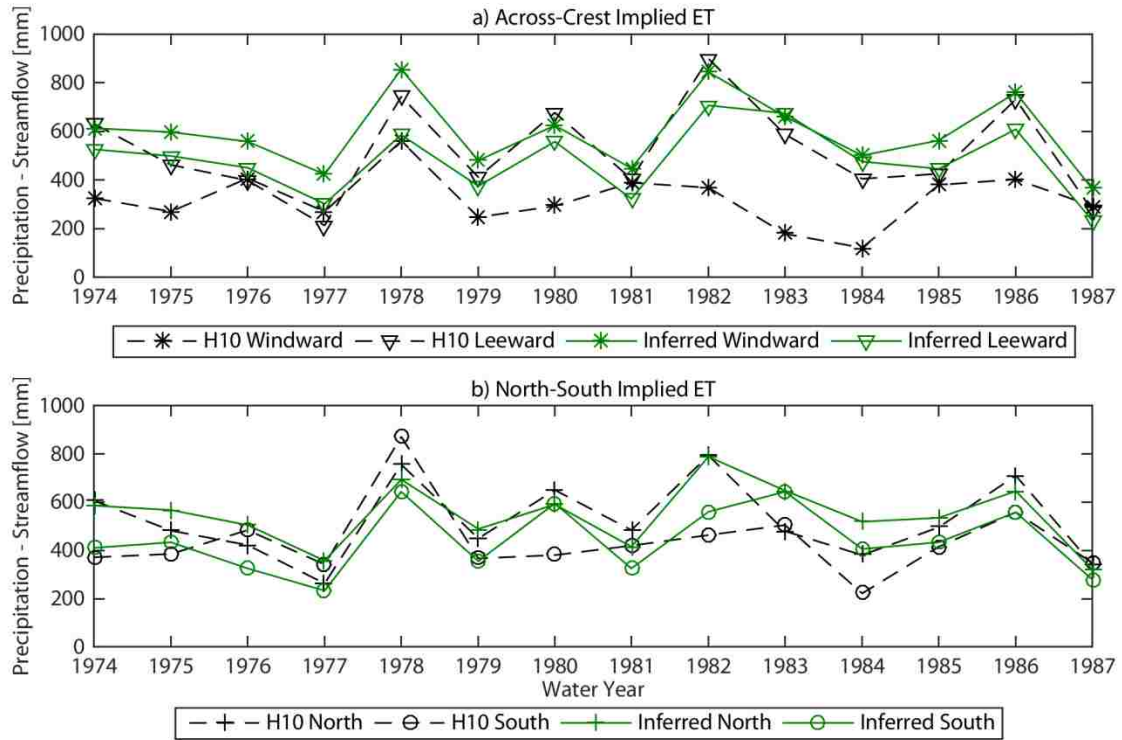
Figure 5.10. a) Time series of precipitation minus streamflow (implied ET) for the windward and leeward basins from H10 and inferred precipitation. b) Time series of implied ET for the northern and southern basins from H10 and inferred precipitation.
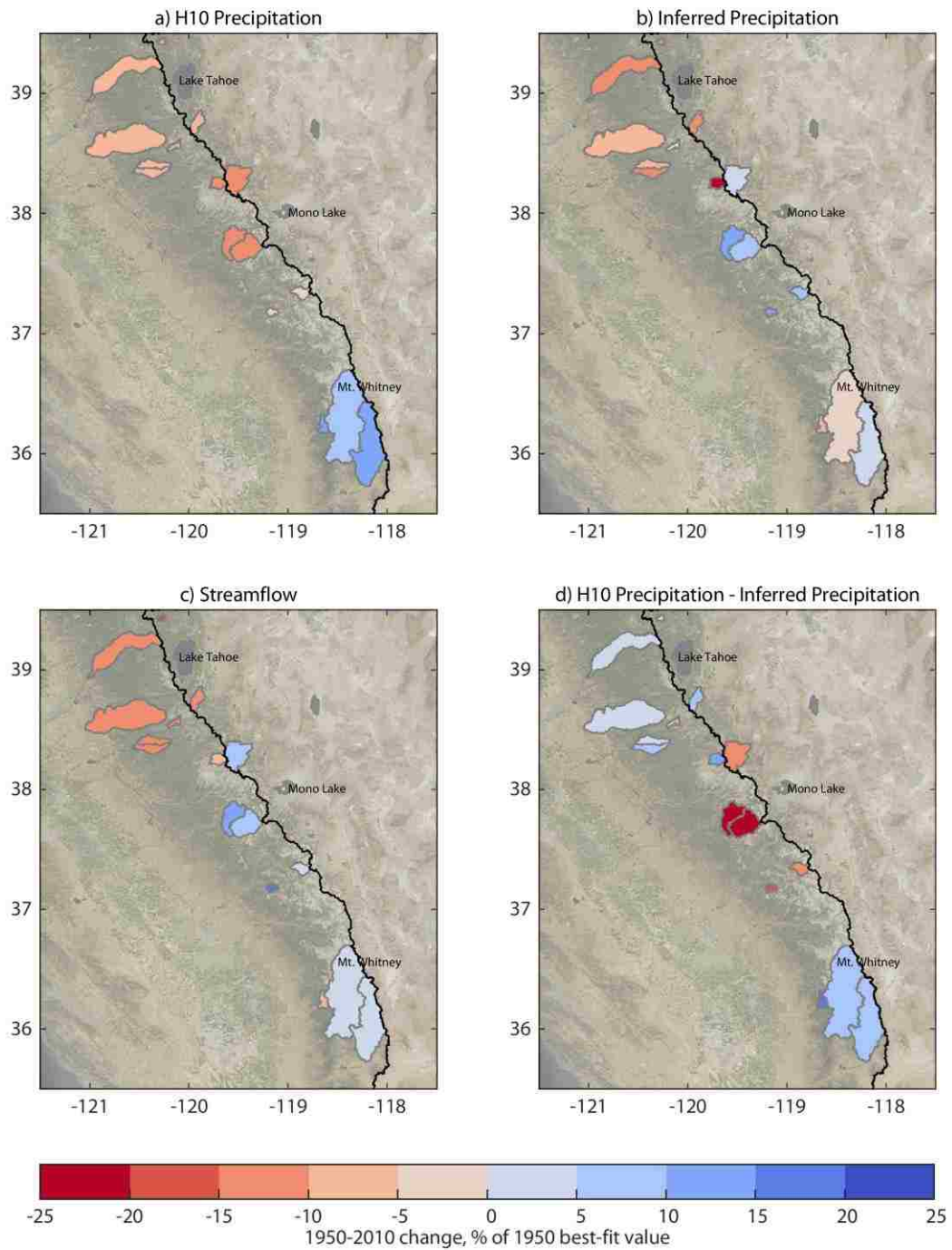
Figure 5.11. 1950-2010 trends in precipitation and streamflow over the Sierra Nevada, expressed as the 1950-2010 change in the best-fit line as a percentage of the 1950 best-fit value. a) Trends in basin-mean precipitation from the H10 dataset. b) Trends in inferred precipitation. c) Trends in streamflow. d) Differences between H10 and inferred precipitation trends.

152
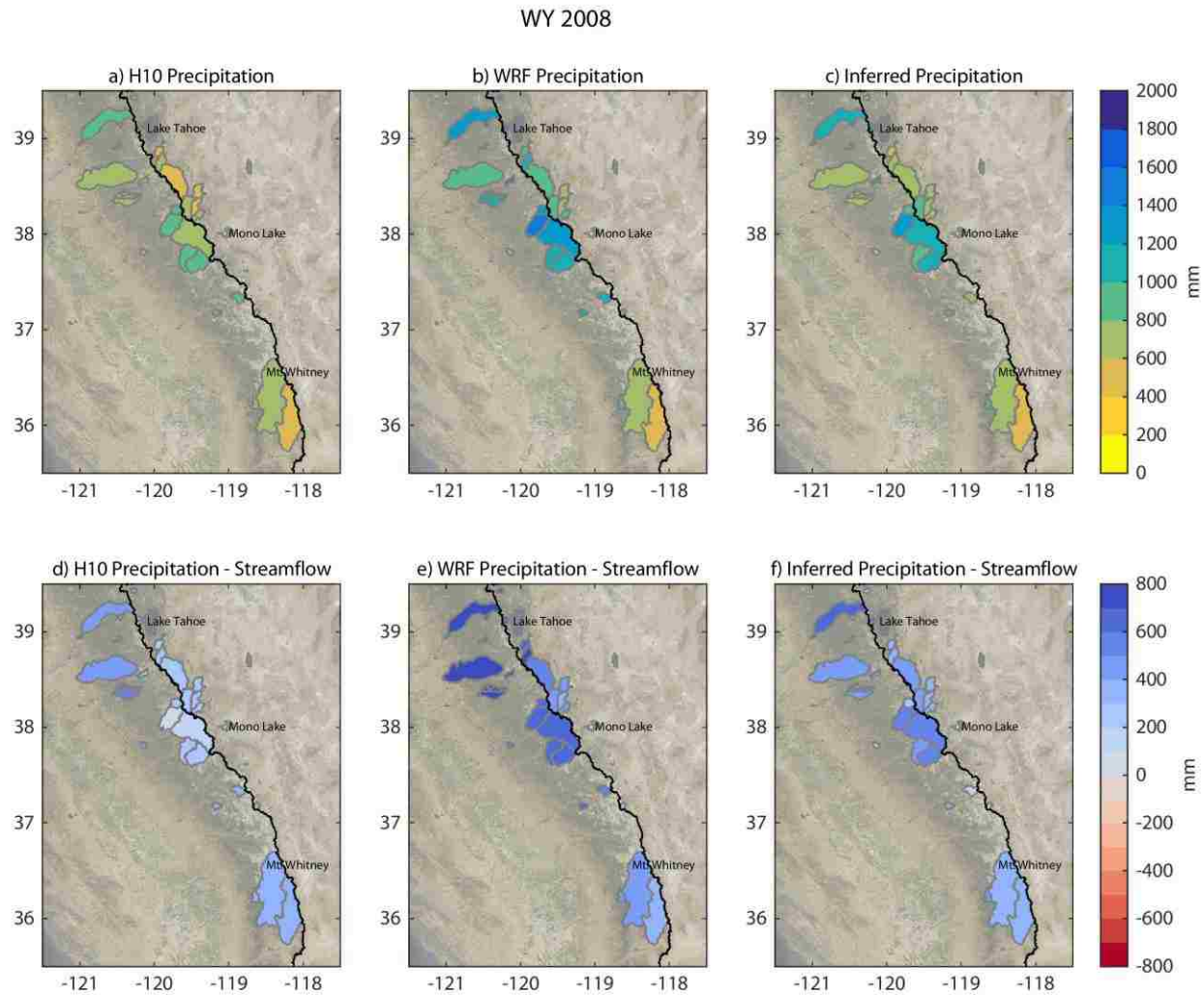
Figure 5.12. Top row: water year 2008 basin-mean precipitation from three sources: H10, WRF, and inferred from streamflow. Bottom row: precipitation minus streamflow for each of the same three datasets.
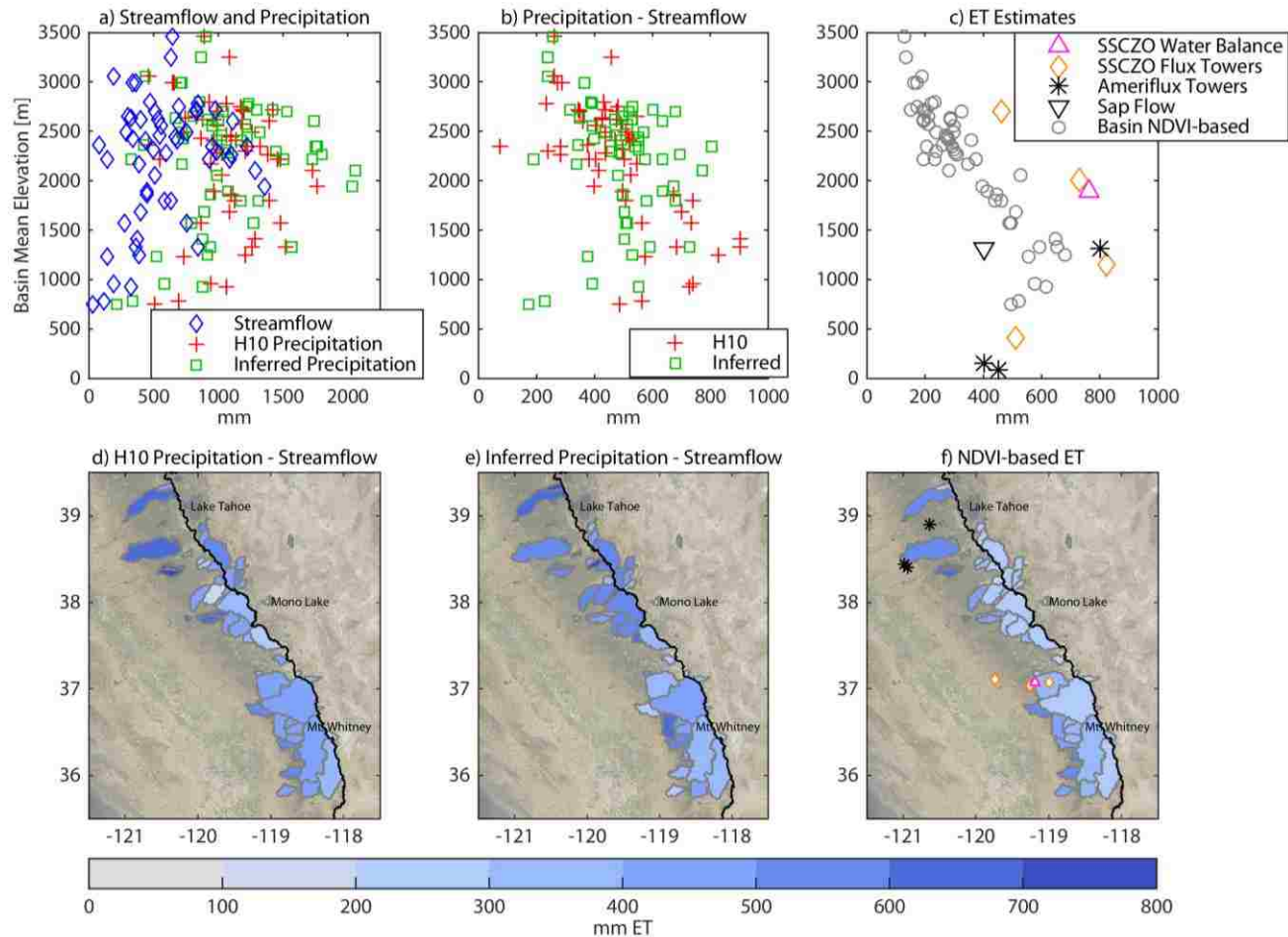
Figure 5.13. Patterns of precipitation, streamflow and ET over the Sierra Nevada domain. a) Streamflow and precipitation vs. mean elevation for each basin. b) H10 and inferred precipitation minus streamflow (implied actual ET) for each basin. c) Elevational pattern of Sierra Nevada ET estimates: Southern Sierra Critical Zone Observatory flux towers [*Goulden et al.*, 2012] and water balance [*Bales et al.*, 2011] estimates of ET; Ameriflux tower ET estimates [*Kurpius et al.*, 2003; *Ichii et al.*, 2009]; ET based on MODIS NDVI regression [*Goulden et al.*, 2012] for each basin. d) – f) Maps of basin-mean ET estimates from H10 precipitation, inferred precipitation, and MODIS NDVI regression, respectively. Flux tower locations are shown in f).

154

References

Adam, J., E. Clark, and D. Lettenmaier (2006), Correction of global precipitation products for orographic effects, *J. Clim.*, *19*(2005), 15–38, doi:http://dx.doi.org/10.1175/JCLI3604.1.

Aguado, E. (1990), Elevational and latitudinal patterns of snow accumulation departures from normal in the Sierra Nevada, *Theor. Appl. Climatol.*, *42*(3), 177–185.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), *Crop evapotranspiration: Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*, Rome.

Anders, A. M., G. H. Roe, D. R. Durran, and J. R. Minder (2007), Small-Scale Spatial Gradients in Climatological Precipitation on the Olympic Peninsula, *J. Hydrometeorol.*, *8*(5), 1068–1081, doi:10.1175/JHM610.1.

Anderson, E. (2006), Snow accumulation and ablation model–SNOW-17, *On-line Doc.*, (January).

Anderson, S. W., G. L. Rockwell, M. F. Friebel, and M. D. Webster (1996), *Water Resources Data - California, Water Year 1996, Vol. 4 - U.S. Geological Survey Water-Data Report*.

Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: a quantitative analysis, *Hydrol. Earth Syst. ...*, 913–921.

Bales, R. C., N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier (2006), Mountain hydrology of the western United States, *Water Resour. Res.*, *42*(8), W08432, doi:10.1029/2005WR004387.

Bales, R. C., J. W. Hopmans, A. T. O'Geen, M. Meadows, P. C. Hartsough, P. Kirchner, C. T. Hunsaker, and D. Beaudette (2011), Soil Moisture Response to Snowmelt and Rainfall in a Sierra Nevada Mixed-Conifer Forest, *Vadose Zo. J.*, *10*(3), 786, doi:10.2136/vzj2011.0001.

Barnett, T. P., J. C. Adam, and D. P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions., *Nature*, *438*(7066), 303–9, doi:10.1038/nature04141.

Beven, K., and A. Binley (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, *6*, 279–298, doi:10.1002/hyp.3360060305.

Bristow, K., and G. Campbell (1984), On the relationship between incoming solar radiation and daily maximum and minimum temperature, *Agric. For. Meteorol.*, *31*(0427), 159–166.

Caldwell, P., H.-N. S. Chin, D. C. Bader, and G. Bala (2009), Evaluation of a WRF dynamical downscaling simulation over California, *Clim. Change*, *95*(3-4), 499–521, doi:10.1007/s10584-009-9583-5.

Christensen, L., C. Tague, and J. Baron (2008), Spatial patterns of simulated transpiration response to climate variability in a snow dominated mountain ecosystem, *Hydrol. Process.*, *3588*(January), 3576–3588, doi:10.1002/hyp.

Church, J. E. (1933), Snow surveying: its principles and possibilities, *Geogr. Rev.*, *23*(4), 529–563, doi:10.2307/209242.

Clark, M., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, *J. Hydrometeorol.*, *7*(2000), 3–22, doi:http://dx.doi.org/10.1175/JHM474.1.

Clark, M. P., A. G. Slater, D. E. Rupp, R. a. Woods, J. a. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, *44*, W00B02, doi:10.1029/2007WR006735.

Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. a. Woods (2011a), Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, *Hydrol. Process.*, *25*(4), 523–543, doi:10.1002/hyp.7902.

Clark, M. P., J. Hendrikx, A. G. Slater, D. Kavetski, B. Anderson, N. J. Cullen, T. Kerr, E. Örn Hreinsson, and R. a. Woods (2011b), Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, *Water Resour. Res.*, *47*(7), doi:10.1029/2011WR010745.

Cristea, N., S. Kampf, and S. Burges (2013), Revised coefficients for Priestley-Taylor and Makkink-Hansen equations for estimating daily reference evapotranspiration, *J. Hydrol. Eng.*, *18*(October), 1289–1300, doi:10.1061/(ASCE)HE.1943-5584.0000679.

Daly, C., R. Neilson, and D. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *Int. J. Climatol.*, *28*(15), 2031–2064, doi:10.1002/joc.1688.

Deems, J. S., S. R. Fassnacht, and K. J. Elder (2008), Interannual Consistency in Fractal Snow Depth Patterns at Two Colorado Mountain Sites, *J. Hydrometeorol.*, *9*(5), 977–988, doi:10.1175/2008JHM901.1.

Deems, J. S., T. H. Painter, and D. C. Finnegan (2013), Lidar measurement of snow depth: a review, *J. Glaciol.*, *59*(215), 467–479, doi:10.3189/2013JoG12J154.

Dettinger, M. (2014), Impacts in the third dimension, *Nat. Geosci.*, *7*(3), 166–167, doi:10.1038/ngeo2096.

Dettinger, M., K. Redmond, and D. Cayan (2004), Winter orographic precipitation ratios in the Sierra Nevada-Large-scale atmospheric circulations and hydrologic consequences, *J. Hydrometeorol.*, *5*(1992), 1102–1116.

Dudhia, J. (1989), Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model, *J. Atmos. Sci.*, *46*(20), 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Evin, G., M. Thyer, D. Kavetski, D. McInerney, and G. Kuczera (2014), Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for

error autocorrelation and heteroscedasticity, *Water Resour. Res.*, *50*(3), 2350–2375, doi:10.1002/2013WR014185.

Finger, D., M. Vis, M. Huss, and J. Seibert (2015), The value of data availability versus model complexity for estimating snow , glacier and rain contribution to runoff in mountain streams, *Water Resour. Res.*, *51*, 1–20, doi:10.1002/2014WR015712.

Flint, A. L., L. E. Flint, and M. D. Dettinger (2008), Modeling Soil Moisture Processes and Recharge under a Melting Snowpack, *Vadose Zo. J.*, *7*(1), 350, doi:10.2136/vzj2006.0135.

Frei, C., and C. Schar (1998), A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, *18.8*, 873–900.

Goulden, M. L., and R. C. Bales (2014), Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion, *Proc. Natl. Acad. Sci.*, *111*(39), 14071–14075, doi:10.1073/pnas.1319316111.

Goulden, M. L., R. G. Anderson, R. C. Bales, a. E. Kelly, M. Meadows, and G. C. Winston (2012), Evapotranspiration along an elevation gradient in California's Sierra Nevada, *J. Geophys. Res. Biogeosciences*, *117*(1), 1–13, doi:10.1029/2012JG002027.

Gustafson, J. R., P. D. Brooks, N. P. Molotch, and W. C. Veatch (2010), Estimating snow sublimation using natural chemical and isotopic tracers across a gradient of solar radiation, *Water Resour. Res.*, *46*(September), 1–14, doi:10.1029/2009WR009060.

Gutmann, E., T. Pruitt, and M. Clark (2014), An intercomparison of statistical downscaling methods used for water resource assessments in the United States, *Water Resour. Reseach*, 7167–7186, doi:10.1002/2014WR015559.Received.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson (2012), A Comparison of Statistical and Dynamical Downscaling of Winter Precipitation over Complex Terrain, *J. Clim.*, *25*(1), 262–281, doi:10.1175/2011JCLI4109.1.

Hamlet, A., and D. Lettenmaier (2005), Production of Temporally Consistent Gridded Precipitation and Temperature Fields for the Continental United States, *J. Hydrometeorol.*, *6*, 330–336.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier (2005), Effects of temperature and precipitation variability on snowpack trends in the sestern United States, *J. Clim.*, *18*(21), 4545–4561, doi:10.1175/JCLI3538.1.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier (2007), Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the western United States, *J. Clim.*, *20*(8), 1468–1486, doi:10.1175/JCLI4051.1.

Hamlet, A. F. et al. (2010), *Final report for the Columbia Basin Climate Change Scenarios Project, Pacific Northwest (PNW) Hydroclimate Scenarios Project 2860*.

Henn, B., M. P. Clark, D. Kavetski, and J. D. Lundquist (2015), Estimating mountain basin-mean precipitation from streamflow using Bayesian inference, *Water Resour. Res.*, *51*, doi:10.1002/2014WR016736.

Hong, S.-Y., Y. Noh, and J. Dudhia (2006), A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, *Mon. Weather Rev.*, *134*(9), 2318–2341, doi:10.1175/MWR3199.1.

Hood, J. L., and M. Hayashi (2015), Characterization of snowmelt flux and groundwater storage in an alpine headwater basin, *J. Hydrol.*, *521*, 482–497, doi:10.1016/j.jhydrol.2014.12.041.

Hughes, M., A. Hall, and R. G. Fovell (2009), Blocking in Areas of Complex Topography, and Its Influence on Rainfall Distribution, *J. Atmos. Sci.*, *66*(2), 508–518, doi:10.1175/2008JAS2689.1.

Hughes, M., P. J. Neiman, E. Sukovich, and M. Ralph (2012), Representation of the Sierra Barrier Jet in 11 years of a high-resolution dynamical reanalysis downscaling compared with long-term wind profiler observations, *J. Geophys. Res. Atmos.*, *117*(17), 1–18, doi:10.1029/2012JD017869.

Ichii, K., W. Wang, H. Hashimoto, F. Yang, P. Votava, A. R. Michaelis, and R. R. Nemani (2009), Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California, *Agric. For. Meteorol.*, *149*(11), 1907–1918, doi:10.1016/j.agrformet.2009.06.019.

Jankov, I., J.-W. Bao, P. J. Neiman, P. J. Schultz, H. Yuan, and A. B. White (2009), Evaluation and Comparison of Microphysical Algorithms in ARW-WRF Model Simulations of Atmospheric River Events Affecting the California Coast, *J. Hydrometeorol.*, *10*(4), 847–870, doi:10.1175/2009JHM1059.1.

Jeton, A. E., M. D. Dettinger, and J. L. Smith (1996), Potential Effects of Climate Change on Streamflow , Eastern and Western Slopes of the Sierra Nevada , California and Nevada, , *U.S. GEOLO*(95-4260).

Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian (2013), A comprehensive change detection method for updating the National Land Cover Database to circa 2011, *Remote Sens. Environ.*, *132*, 159–175, doi:10.1016/j.rse.2013.01.012.

Kain, J. S. (2004), The Kain–Fritsch Convective Parameterization: An Update, *J. Appl. Meteorol.*, *43*(1), 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Kattelmann, R., and K. Elder (1991), Hydrologic characteristics and water-balance of an alpine basin in the Sierra-Nevada, *Water Resour. Res.*, *27*(7), 1553–1562, doi:10.1029/90WR02771.

Kavetski, D., S. W. Franks, and G. Kuczera (2003), Confronting Input Uncertainty in Environmental Modelling, in *Calibration of Watershed Models*, edited by Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte, pp. 49–68, Washington, D.C.: American Geophysical Union.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*(3), 1–9, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in

hydrological modeling: 2. Application, *Water Resour. Res.*, *42*(3), W03408, doi:10.1029/2005WR004376.

Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, *45*(2), 1–34, doi:10.1029/2008WR006912.

Koskela, J. J., B. W. F. Croke, H. Koivusalo, a. J. Jakeman, and T. Kokkonen (2012), Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment, *Water Resour. Res.*, *48*(11), n/a–n/a, doi:10.1029/2011WR011773.

Kuczera, G., and M. Mroczkowski (1998), Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, *34*(6), 1481, doi:10.1029/98WR00496.

Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, *331*(1-2), 161–177, doi:10.1016/j.jhydrol.2006.05.010.

Kurpius, M. R., J. a. Panek, N. T. Nikolov, M. McKay, and a. H. Goldstein (2003), Partitioning of water flux in a Sierra Nevada ponderosa pine plantation, *Agric. For. Meteorol.*, *117*(3-4), 173–192, doi:10.1016/S0168-1923(03)00062-5.

Leavesley, G., R. Lichty, B. Troutman, and L. Saindon (1983), *Precipitation-runoff modeling system: User's manual*.

Leydecker, A., and J. M. Melack (1999), Evaporation from snow in the central Sierra Nevada of California, *Nord. Hydrol.*, *30*(2), 81–108.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, *99*(D7), 14415–14428.

Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, *43*(7), 1–18, doi:10.1029/2006WR005756.

Livneh, B., E. a. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier (2013), A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions*, *J. Clim.*, *26*(23), 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

Livneh, B., J. S. Deems, D. Schneider, J. J. Barsugli, and N. P. Molotch (2014), Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain, *Water Resour. Res.*, *50*(11), 8589–8610, doi:10.1002/2014WR015442.

Luce, C., and D. Tarboton (1998), The influence of the spatial distribution of snow on basin-averaged snowmelt, *Hydrol. Process.*, *12*, 1671–1683.

Luce, C. H., J. T. Abatzoglou, and Z. a Holden (2013), The Missing Mountain Water: Slower Westerlies Decrease Orographic Enhancement in the Pacific Northwest USA., *Science (80-.).*, *1360*, doi:10.1126/science.1242335.

Lundquist, J., D. Cayan, and M. Dettinger (2003), Meteorology and hydrology in yosemite

national park: A sensor network application, *Inf. Process. Sens. ...*, 518–528.

Lundquist, J., B. Huggett, H. Roop, and N. Low (2009), Use of spatially distributed stream stage recorders to augment rain gages by identifying locations of thunderstorm precipitation and distinguishing rain from snow, *Water Resour. Res.*, *45*, 1–7, doi:10.1029/2008WR006995.

Lundquist, J. D., and S. P. Loheide (2011), How evaporative water losses vary between wet and dry water years as a function of elevation in the Sierra Nevada, California, and critical factors for modeling, *Water Resour. Res.*, *47*, 1–13, doi:10.1029/2010WR010050.

Lundquist, J. D., J. R. Minder, P. J. Neiman, and E. Sukovich (2010), Relationships between Barrier Jet Heights, Orographic Precipitation Gradients, and Streamflow in the Northern Sierra Nevada, *J. Hydrometeorol.*, *11*(5), 1141–1156, doi:10.1175/2010JHM1264.1.

Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman (2015), High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets across the Sierra Nevada, California, *J. Hydrometeorol.*, *16*(4), 1773–1792, doi:10.1175/JHM-D-15-0019.1.

Lutz, J. a., J. W. van Wagtendonk, and J. F. Franklin (2010), Climatic water deficit, tree species ranges, and climate change in Yosemite National Park, *J. Biogeogr.*, *37*(5), 936–950, doi:10.1111/j.1365-2699.2009.02268.x.

Maurer, E., A. Wood, J. Adam, and Lettenmaier (2002), A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States*, *J. Clim.*, *15*, 3237–3251.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011a), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, *400*(1-2), 83–94, doi:10.1016/j.jhydrol.2011.01.026.

McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. a. Woods (2011b), Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Process.*, *25*(4), 511–522, doi:10.1002/hyp.7841.

Meromy, L., N. P. Molotch, T. E. Link, S. R. Fassnacht, and R. Rice (2013), Subgrid variability of snow water equivalent at operational snow stations in the western USA, *Hydrol. Process.*, *27*(May 2012), 2383–2400, doi:10.1002/hyp.9355.

Milly, P. C. D., and K. A. Dunne (2002), Macroscale water fluxes 1. Quantifying errors in the estimation of basin mean precipitation, *Water Resour. Res.*, *38*(10), 1–14, doi:10.1029/2001WR000759.

Minder, J. R., D. R. Durran, G. H. Roe, and A. M. Anders (2008), The climatology of small-scale orographic precipitation over the Olympic Mountains: Patterns and processes, *Q. J. R. Meteorol. Soc.*, *134*(633), 817–839, doi:10.1002/qj.258.

Mizukami, N., and M. B. Smith (2012), Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling, *J. Hydrol.*, *428-429*, 129–141, doi:10.1016/j.jhydrol.2012.01.030.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. a. Clough (1997), Radiative

transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, *102*(D14), 16663, doi:10.1029/97JD00237.

Le Moine, N., F. Hendrickx, J. Gailhard, R. Garcon, and F. Gottardi (2015), Hydrologically-aided interpolation (HAI) of daily precipitation and temperature fields in a meso-scale Alpine catchment, *J. Hydrometeorol.*, doi:dx.doi.org/10.1175/JHM-D-14-0162.1.

Molotch, N. P., and R. C. Bales (2005), Scaling snow observations from the point to the grid element: Implications for observation network design, *Water Resour. Res.*, *41*, 1–16, doi:10.1029/2005WR004229.

Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes, *Mon. Weather Rev.*, *137*(3), 991–1007, doi:10.1175/2008MWR2556.1.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier (2005), Declining Mountain Snowpack in Western North America, *Bull. Am. Meteorol. Soc.*, *86*(1), 39–49, doi:10.1175/BAMS-86-1-39.

Newman, A. J., M. P. Clark, J. Craig, B. Nijssen, A. Wood, E. Gutmann, N. Mizukami, L. Brekke, and J. R. Arnold (2015), Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States, *J. Hydrometeorol.*, *16*(6), 2481–2500, doi:10.1175/JHM-D-15-0026.1.

NRCS (2007), *Soil Survey of Yosemite National Park, California*.

Ogden, N. H., R. Milka, C. Caminade, and P. Gachon (2014), Recent and projected future climatic suitability of North America for the Asian tiger mosquito Aedes albopictus, *Parasit. Vectors*, *7*(1), 532, doi:10.1186/s13071-014-0532-4.

Painter, T. H., K. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier (2009), Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, *Remote Sens. Environ.*, *113*(4), 868–879, doi:10.1016/j.rse.2009.01.001.

Painter, T. H. et al. (2015), The Airborne Snow Observatory: scanning lidar and imaging spectrometer fusion for mapping snow water equivalent and snow albedo, *Remote Sens. Environ.*

Pavelsky, T. M., S. Sobolowski, S. B. Kapnick, and J. B. Barnes (2012), Changes in orographic precipitation patterns caused by a shift from snow to rain, *Geophys. Res. Lett.*, *39*(18), L18706, doi:10.1029/2012GL052741.

Pelletier, P. M. (1988), Uncertainties in the single determination of river discharge: a literature review, *Can. J. Civ. Eng.*, *15*(5), 834–850.

Raleigh, M. S., and J. D. Lundquist (2012), Comparing and combining SWE estimates from the SNOW-17 model using PRISM and SWE reconstruction, *Water Resour. Res.*, *48*(1), W01506, doi:10.1029/2011WR010542.

Ralph, F. M. et al. (2005), Improving Short-Term (0–48 h) Cool-Season Quantitative Precipitation Forecasting: Recommendations from a USWRP Workshop, *Bull. Am.*

Meteorol. Soc., *86*(11), 1619–1632, doi:10.1175/BAMS-86-11-1619.

Rasmussen, R. et al. (2011a), High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate, *J. Clim.*, *24*(12), 3015–3048, doi:10.1175/2010JCLI3985.1.

Rasmussen, R. et al. (2011b), The NOAA/FAA/NCAR Winter Precipitation Test Bed: How Well Are We Measuring Snow?, *Bull. Am. Meteorol. Soc.*, (june), 811–829, doi:10.1175/BAMS-D-11-00052.1.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, *46*(5), W05521, doi:10.1029/2009WR008328.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, *47*(11), W11516, doi:10.1029/2011WR010643.

Rice, R., and R. C. Bales (2010), Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California, *Water Resour. Res.*, *46*, 1–13, doi:10.1029/2008WR007318.

Rice, R., R. C. Bales, T. H. Painter, and J. Dozier (2011), Snow water equivalent along elevation gradients in the Merced and Tuolumne River basins of the Sierra Nevada, *Water Resour. Res.*, *47*(May), 1–11, doi:10.1029/2010WR009278.

Rittger, K., A. Kahl, and J. Dozier (2011), Topographic Distribution of Snow Water Equivalent in the Sierra Nevada, in *Paper Presented at Western Snow Conference 2011*.

Rockwell, G. L., S. W. Anderson, and P. D. Hayes (1996a), *Water Resources Data - California, Water Year 1996*.

Rockwell, G. L., P. D. Hayes, and J. Agajanian (1996b), *Water Resources Data - California, Water Year 1996, Vol. 1 - U.S. Geological Survey Water-Data Report*.

Roe, G. H. (2005), Orographic Precipitation, *Annu. Rev. Earth Planet. Sci.*, *33*(1), 645–671, doi:10.1146/annurev.earth.33.092203.122541.

Rose, K., R. Graham, and D. Parker (2003), Water source utilization by Pinus jeffreyi and Arctostaphylos patula on thin soils over bedrock, *Oecologia*, *134*(1), 46–54, doi:10.1007/s00442-002-1084-4.

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

Sieck, L. C., S. J. Burges, and M. Steiner (2007), Challenges in obtaining reliable measurements of point rainfall, *Water Resour. Res.*, *43*(1), doi:10.1029/2005WR004519.

Siler, N., G. Roe, and D. Durran (2013), On the Dynamical Causes of Variability in the Rain-Shadow Effect: A Case Study of the Washington Cascades, *J. Hydrometeorol.*, *14*(1), 122–

139, doi:10.1175/JHM-D-12-045.1.

Skamarock, W. C., and J. B. Klemp (2008), A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, *J. Comput. Phys.*, *227*(7), 3465–3485, doi:10.1016/j.jcp.2007.01.037.

Skamarock, W. C., J. B. Klemp, J. Dudhi, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers (2008), A Description of the Advanced Research WRF Version 3, *Tech. Rep.*, (June), 113, doi:10.5065/D6DZ069T.

Slack, J. R., and J. M. Landwehr (1992), Hydro-Climatic Data Network (HCDN): A U. S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988, *U.S. Geol. Surv. Open-File Rep.*, *92-129*.

Stephenson, N. L. (1998), Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales, *J. Biogeogr.*, *25*, 855–870, doi:10.1046/j.1365-2699.1998.00233.x.

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca (2004), Implementation and verification of the unified noah land surface model in the WRF model, in *Bulletin of the American Meteorological Society*, pp. 2165–2170.

Thornton, P. E., and S. W. Running (1999), An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation, *Agric. For. Meteorol.*, *93*(4), 211–228, doi:10.1016/S0168-1923(98)00126-9.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*(3), W00B14, doi:10.1029/2008WR006825.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, 1–15, doi:10.1029/2007WR006720.

Wayand, N. E., A. F. Hamlet, M. Hughes, S. I. Feld, and J. D. Lundquist (2013), Intercomparison of Meteorological Forcing Data from Empirical and Mesoscale Model Sources in the North Fork American River Basin in Northern Sierra Nevada, California, *J. Hydrometeorol.*, *14*(3), 677–699, doi:10.1175/JHM-D-12-0102.1.

Weingartner, R., D. Viviroli, and B. Schädler (2007), Water resources in mountain regions: A methodological approach to assess the water balance in a highland-lowland-system, *Hydrol. Process.*, *21*(5), 578–585, doi:10.1002/hyp.6268.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier (1994), A distributed hydrology-vegetation model for complex terrain, *Water Resour. Res.*, *30*(6), 1665–1679.

Xia, Y. et al. (2012), Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, *117*, D03110, doi:10.1029/2011JD016051.

Curriculum Vitae
Brian M. Henn, P.E., Ph.D. Candidate

Department of Civil and Environmental Engineering   bhenn@u.washington.edu
University of Washington
201 More Hall, Box 352700
Seattle, WA 98195-2700

Educational History

Ph.D., Civil and Environmental Engineering – Water Resources and Hydrology
   University of Washington, 2015

M.S., Civil and Environmental Engineering – Environmental Fluid Mechanics and Hydrology
   Stanford University, 2006

B.S.E., Civil and Environmental Engineering, *cum laude*; Certificate, Woodrow Wilson School of Public and International Affairs
   Princeton University, 2005

Employment History

University of Washington, Seattle, Wash.: Graduate Student Researcher
   September 2010 – present

Hazen and Sawyer, P.C., New York, N.Y.: Engineer
   September 2006 – July 2010

Awards and Fellowships

Runner-up, Dr. J. E. Church Student Paper Award, Western Snow Conference, 2015: Snow pillows, LiDAR, and streamgauges: Assimilating observations into basin water balance

Valle Scholarship, University of Washington, 2014

Rod Sakrison Memorial Fellowship, American Water Resources Association, Washington Section, 2014

Visiting Student, National Center for Atmospheric Research, Boulder, Colo., Feb. 2013

National Defense Science and Engineering Graduate Fellowship, 2010-2013

Research and Publications

Peer-Reviewed Publications:

Henn, B., M. P. Clark, D. Kavetski, and J. D. Lundquist, 2015: Estimating mountain basin-mean precipitation from streamflow using Bayesian inference, *Water Resources Research*, **51**, doi:10.1002/2014WR016736.

Lundquist, J. D., M. Hughes, **B. Henn**, E. Gutmann, B. Livneh, J. Dozier and P. Neiman, 2015: High-elevation precipitation patterns: Using snow measurements to diagnose when and why gridded datasets succeed or fail across the Sierra Nevada, California. *Journal of Hydrometeorology,* **16**, 1773-1792, doi:10.1175/JHM-D-15-0019.1.

Henn, B., Q. Cao, D. P. Lettenmaier, C. S. Magirl, C. Mass, J. B. Bower, M. St. Laurent, Y. Mao and S. Perica, 2015: Hydroclimatic conditions preceding the March 2014 Oso Landslide. *Journal of Hydrometeorology*, **16**, 1243-1249, doi:10.1175/JHM-D-15-0008.1.

Henn, B., J. D. Lundquist, M. S. Raleigh and A. Fisher, 2013: A comparison of methods for filling gaps in hourly near-surface air temperature data. *Journal of Hydrometeorology*, **14**, 929-945, doi:10.1175/JHM-D-12-027.1.

Raleigh, M. S**.**, K. Rittger, C. E. Moore, **B. Henn**, J. A. Lutz and J. D. Lundquist, 2013: Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada. *Remote Sensing of Environment*, **128**, 44-57, doi:10.1016/j.rse.2012.09.016.

Posters and Presentations:

Henn, B., M. P. Clark, D. Kavetski, B. McGurk, T. H. Painter and J. D. Lundquist: Snow pillows, LiDAR, satellites and streamgauges: Assimilating observations of snow and streamflow into basin water balance. Poster, *Oregon State University Water Research Symposium*, Corvallis, OR, Apr. 27, 2015.

Henn, B., M. P. Clark, D. Kavetski, B. McGurk, T. H. Painter and J. D. Lundquist: Snow pillows, LiDAR, and streamgauges: Assimilating observations into basin water balance. Presentation, *83rd Western Snow Conference*, Grass Valley, CA, Apr. 22, 2015.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Inferring mountain basin precipitation from streamflow observations using Bayesian model calibration. Poster, *American Geophysical Union Fall Meeting*, San Francisco, CA, Dec. 15-19, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Using models and indirect observations to map precipitation across mountain basins. *University of Washington ESRM Seminar*, Seattle, WA, Oct. 28, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Doing hydrology backward to estimate mountain precipitation patterns from streamflow. *16th AMS Conference on Mountain Meteorology*, San Diego, CA, Aug. 17, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Doing hydrology backward: Estimating mountain precipitation patterns from streamflow. Presentation, *Oregon State University Water Research Symposium*, Corvallis, OR, May 12, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Doing hydrology backward: Estimating mountain precipitation patterns from streamflow. Presentation, *82nd Western Snow Conference*, Durango, CO, Apr. 16, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Better precipitation estimation in mountain watersheds using streamflow observations. Presentation, *University of Washington Land Surface Hydrology Seminar*, Seattle, WA, Feb. 12, 2014.

Henn, B., M. P. Clark, D. Kavetski and J. D. Lundquist: Better precipitation estimation in mountain watersheds using streamflow and snowpack observations. Presentation, *2013 American Water Resources Association Annual Conference*, Portland, OR, Nov. 4, 2013.

Henn, B., M.S. Raleigh and J.D. Lundquist: To what extent does snow cover influence near-surface air temperature in complex terrain? Poster, *American Geophysical Union Fall Meeting*, San Francisco, CA, Dec. 3-7, 2012.

Henn, B.: Evaluation of different modes of Sierra Nevada snow observation. Presentation, *California Cooperative Snow Surveys Meeting*, Sequoia National Park, CA, Nov. 8, 2012.

Henn, B., J. D. Lundquist, M. S. Raleigh and A. Fisher: Estimation of hourly near-surface air temperature in complex terrain: Influence of elevation, cold air pools and snow cover. Poster, *Graduate Climate Conference*, Pack Forest, WA, Oct. 26-28, 2012.

Henn, B., J. D. Lundquist, M. S. Raleigh and A. Fisher: Estimation of hourly near-surface air temperature in complex terrain: Influence of elevation, cold air pools and snow cover. Poster, *American Meteorological Society 15th Conference on Mountain Meteorology*, Steamboat Springs, CO, Aug. 20-24, 2012.

Henn, B., J. D. Lundquist, M. S. Raleigh and A. Fisher, 2011: Filling missing temperature data in complex terrain: a comparison of EOF-based and traditional techniques. Poster, *American Geophysical Union Fall Meeting*, San Francisco, CA, Dec. 5-9, 2011.

Other Publications:

Garin, J., D. Gumb, B. Henn and S. Mehrotra, 2009: Hydrologic and hydraulic modeling paradigm for green stormwater practices. *Clearwaters*, magazine of the New York Water Environment Association, **39**, no. 4.

Henn, B., 2005: The Harry's Brook Environmental Laboratory: Initial hydrologic studies. Princeton University Senior Thesis.

Reviewer for Research Journals:

Journal of Hydrology

Journal of Hydrometeorology
Remote Sensing of Environment
Journal of Geophysical Research – Atmospheres
Journal of Atmospheric and Solar-Terrestrial Physics
Journal of Terrestrial, Atmospheric and Oceanic Sciences (Taiwan)

Teaching Experience

Instructor, CEE 491: Deterministic Systems, University of Washington, Fall 2013

Writing Tutor, Center for Learning and Undergraduate Enrichment, University of Washington
September 2012 – present

Hydrology Instructor, Civil Engineering P.E. Review Course, University of Washington
September 2012 – present

Grader, CEE 491: Deterministic Systems, University of Washington, Fall 2012

Professional Certifications and Memberships

Professional Engineer (Civil), Calif., 2009 – present

Member, American Geophysical Union, 2011 – present

Member, American Water Resources Association, 2012 – present

Member, American Meteorological Society, 2012 – present

Leadership and Service

Officer, UW Chapter of the American Water Resources Association, 2011 – present

Public Outreach Volunteer, University of Washington Program on Climate Change, 2011 – present

Guatemala Monitoring Team Member, Water for People, 2008: Documented and evaluated more than 20 existing Water for People drinking water and sanitation projects.

Writer and News Editor, *The Daily Princetonian*, 2001 – 2005: Wrote and edited articles for and managed production of a daily student newspaper

Freshman Orientation Trip Leader, Princeton University Outdoor Action, 2002 – 2005: Organized and led week-long freshman pre-orientation backpacking trips in 2002, 2003 and 2004.

Eagle Scout, Troop 11, Piedmont, Calif., 2001