Seasonal prediction of end-of-dry season watershed behavior in a highly interconnected alluvial watershed, northern California

Claire Kouba[1] and Thomas Harter[1]

[1]Department of Land, Air and Water Resources, University of California, Davis, One Shields Avenue, Davis, CA, United States

Correspondence: Claire Kouba (cmkouba@ucdavis.edu)

Abstract. In undammed watersheds in Mediterranean climates, the timing and abruptness of the transition from the dry season to the wet season have major implications for aquatic ecosystems. Of particular concern in many coastal areas is whether this transition can provide sufficient flows at the right time to allow passage for spawning anadromous fish, which is determined by dry season baseflow rates and the timing of the onset of the rainy season. In (semi-) ephemeral watershed systems, these functional flows also dictate the timing of full reconnection of the stream system. In this study, we propose methods to predict, approximately five months in advance, two key hydrologic metrics in the undammed rural Scott River watershed in northern California. The two metrics are intended to quantify the relative transition timing from the dry to the wet season and to characterize the severity of a dry year. The ability to predict these metrics in advance could support seasonal adaptive management. The first metric is the minimum 30-day dry season baseflow volume, V_{min}, which occurs at the end of the dry season (September-October) in this Mediterranean climate. The second metric is the cumulative precipitation, starting Sept. 1st, necessary to bring the watershed to a "full" or "spilling" condition (i.e. initiate the onset of wet season storm- or baseflows) after the end of the dry season, referred to here as P_{spill}. As potential predictors of these two metrics, we assess maximum snowpack, cumulative precipitation, the timing of the snowpack and precipitation, spring groundwater levels, spring river flows, reference ET, and a subset of these metrics from the previous water year. Though many of these predictors are correlated with the two metrics of interest, we find that the best prediction for both metrics is a linear combination of the maximum snowpack water content and total October-April precipitation. These two linear models could reproduce historic values of V_{min} and P_{spill} with an average model error (RMSE) of 1.4 Mm^3 / 30 days (19.4 cfs) and 25.4 mm (1 inch), corresponding to 49% and 37% of mean observed values, respectively. Although these predictive indices could be used by governance entities to support local water management, careful consideration of baseline conditions used as a basis for prediction is necessary.

1 Introduction

In regions that experience periodic drought, such as the western United States, spatially distributed indices summarizing hydroclimate or surface water supply conditions are often critical decision-support tools for water managers (e.g., Garen, 1993). An index can be forward-looking, such as those that forecast near-term seasonal water supplies (e.g., Null and Viers, 2013; Verley, 2020), or backward-looking, such as ones that evaluate drought severity (e.g., Palmer, 1965; Guttman, 1998; McKee et al., 1993; Wilhite and Glantz, 1985; Wilhite et al., 2000). In many western states, forward-looking seasonal indices are used extensively by water agencies to inform local adaptive management decisions, e.g. in Colorado (CDWR 2023), Idaho (NRCS 2023) and California (Null and Viers, 2013). In California the principal examples are the Sacramento Valley Index (SVI) and San Joaquin Index (SJI), which are seasonal forecasts used to determine water allocations from these watersheds through the State Water Project (DWR 2022). The state has more recently published a retroactive categorical water year type (WYT) dataset for sub-county level regions throughout California (DWR 2021).

These summary indices provide broad characterizations of anticipated or historically available water supplies, particularly in climates and geographies with significant snowpack, groundwater storage, and/or surface water storage capacity and a consummate delay – and possibly significant spatial transfer – between precipitation occurrence and water use. More detailed characteristics, including the hydrologic effects of water year type, climate change, human water use, and other factors, can be identified using the functional flows approach (e.g., Poff et al., 1997; Bunn and Arthington, 2002; Poff et al., 2010; Wheeler et al., 2018). The flows are "functional" because they serve an ecological purpose, such as wet season flood flows, needed to disperse cottonwood seeds (Mahoney and Rood, 1998) and fall pulse flows, needed to provide passage for spawning fall-run anadromous fish (Moyle, 2002). A California-specific functional flows framework has been developed to assess the degree of hydrologic alteration between current and unimpaired conditions (Yarnell et al., 2020; Patterson et al., 2020).

Unlike forward-looking indices such as the SVI and SJI, the functional flows approach (e.g. Yarnell et al., 2020) does not provide numerical flow predictions from recent seasonal hydrologic datasets. The existing forward-looking indices developed for Mediterranean climate, such as the ones used in California, Idaho and Colorado, referenced above, are widespread in stream systems with significant managed surface water storage, capable of bridging the temporal gap between the timing of precipitation and water use (e.g., DWR 2023). Under Mediterranean climate conditions, the completion of the wet season (winter, spring) defines the total available water supply, which is then managed through reservoirs for supply deliveries throughout the dry irrigation season (summer and fall, e.g., CDWR 2023). However, such indices have not been employed in basins without managed surface water storage, but significant seasonal snowpack or groundwater storage. Neither have such indices been developed specifically to manage environmental (instream) flow protection. Finally, such forward-looking indices to inform adaptive water supply management are lacking in smaller-scale geographic settings, especially for the mixed rain and snow-fed stream category of Patterson et al., 2020.

Here, we outline and test a novel approach to developing a forward-looking index that may be useful to inform water management decisions targeting environmental instream flow management in mixed rain and snow-fed watersheds without managed surface water storage under Mediterranean climates. Environmentally, a critical period in such systems is the summer

baseflow period at the end of the dry season, bracketed by the onset of winter stormflow (Peek et al., 2022). Summer baseflow conditions are primarily an expression of underlying groundwater storage, fed in turn by recharge from winter and spring precipitation, snowmelt, and runoff (Tarboton, 2003). During this period, native fish, particularly anadromous and/or salmonid fish, are highly vulnerable to below average low flow conditions (e.g., Van Kirk and Naman, 2008). Due to lack of surface water storage, such basins may seek early protective water management decisions to support environmental flows. These decisions must be made prior to the onset of the irrigation season but following the (near) completion of the wet season, which defines the bulk of annual water input to a Mediterranean watershed.

We utilize the Scott River watershed in northern California as an example Mediterranean climate stream system (without surface water storage) to outline our approach and to evaluate whether a statistically significant, forward-looking index can be defined to support environmental water management. Three periods of water use and climate forces have been proposed for the Scott River (e.g., by Pyschik, 2022): Eras 1, 2, and 3, ranging from 1942-1976, 1977-2000, and 2001-2021, respectively. These eras reflect changes in human management, such as the widespread installation of groundwater pumps in the region in the late 1970s (Tolley et al., 2019); and climate conditions, such as the shift in the Pacific Decadal Oscillation in the late 1970s (Francis et al., 1998) or the onset of a two-decade abnormally dry period in 2000 (Williams et al., 2020). These overlapping changes make it difficult to identify the cause of decade-scale changes in regional hydrology; therefore, the proposed predictive method of hydrologic behavior is agnostic as to the mechanism linking the predictors and hydrologic response.

We propose two decision-support metrics, both designed as quantitative forecasts. We additionally explore the significance of each metric in the context of functional flows. The first metric is V_{min}, the minimum 30-day dry season baseflow volume in a given water year, which typically occurs in September or October. The second is a prediction of the cumulative rainfall needed to wet up the watershed after the dry season such that subsequent rainfall results in significant, measurable runoff, or storm surge, events. This cumulative precipitation depth is referred to as P_{spill}. Both metrics have significance for environmental flows and – if predictable in advance – could support near-term (seasonal) adaptive management, similar to the SVI and SJI in California's Central Valley. Specifically, the magnitude of the minimum baseflow rate sets the spatial extent of the aquatic ecosystem during the dry season and influences rearing conditions for over-summering juvenile salmonids (Gorman, 2016). P_{spill} is intuitively related to the relative timing of flows necessary for fall-run salmon passage: under equivalent fall rainfall, a greater amount of rain needed to generate stormflow would be associated with a prolonged dry season. This type of prolonged dry season has delayed salmon access to spawning habitat in recent years (CDFW 2015). In the following, we first define V_{min} and P_{spill}. Next, we develop seasonal predictions for V_{min} and P_{spill} from environmental observations near the end of the wet season, about a half a year earlier. We then evaluate trends over time and consider the effects that climate change and changing water use patterns may have on the metrics considered in this study, and the decisions they support.

2 Methods

In this study we used linear regression modeling to predict watershed behavior at the end of the dry season (the response) using data available the previous spring (the predictors). The Scott River watershed (Figure 1) has a snow-influenced Mediterranean

climate, giving the river's annual hydrograph a characteristic high-flow season during the rainy winters, a gradual flow recession in the spring-summer as the snowpack melts, and a low-flow dry season after the snowpack is depleted (e.g., Figure 2). In the U.S. Geological Survey (USGS) National Hydrography Dataset, the Scott River watershed is denoted with the 8-digit Hydrologic Unit Code (HUC8) 18010208. Annual demand for agricultural and domestic use (estimated at 23 and 1.3 thousand acre-feet, respectively) (DWR 2004) are relatively stable in the Scott River system (although some reports of dry wells occur in dry years) (Siskiyou County 2021). A key management challenge is persistent low environmental flows during the dry season baseflow period. In dry years, the lowest annual flowrates can overlap with the spawning periods for fall-run anadromous fish, potentially restricting fish passage and imperiling the long-term viability of the Scott River fishery (Siskiyou County 2021). Post-1970s minimum dry season baseflows have been lower than pre-1977, and very low minima (< 10 cfs or 0.7 Mm^3 / 30 days) have been more frequent in the past two decades (see Results, Section 3.2.1), making the management of these flows more urgent.

This study focuses on the transition from the dry season to the wet season, which at times can straddle the conventional water year boundary of October 1st, and cumulative precipitation is used both as a predictor and as a response variable (P_{spill}). When it is a predictor, a traditional October 1st start date is used and it is summed as the cumulative precipitation of October-April, to facilitate an end-of-April prediction of fall conditions. When it is the response variable, cumulative precipitation is counted starting on September 1st of the preceding water year, to capture uncommon September precipitation. This September 1st start date is also used in some graphs of climate and flow data (e.g. in Section 3.2 below), to establish and visualize baseline dry season conditions. Additionally, all flows in this study are observed or simulated at the USGS Fort Jones streamflow gauge (station 11519500), a key monitoring location downstream of nearly all water use and cultivated land in the HUC8 watershed (Figure 1), with an observation record covering water years 1942-2021.

2.1 Scott River watershed precipitation-runoff behavior

To establish the context and hydrologic relevance of the two proposed predictive indices V_{min} and P_{spill}, a brief description of the behavior of the watershed is necessary. In an undammed catchment, the runoff response to one (or a series of) precipitation event(s) is dependent on multiple factors, including antecedent soil moisture conditions, the intensity and magnitude of the precipitation, and on the dampening and delay of runoff; the latter is due to interflow, snow storage and recharge to groundwater storage that later returns as stream baseflow (Tarboton, 2003). A threshold runoff response to individual storm events has been observed at the hillslope scale where soil directly overlays (relatively) impermeable bedrock: absent significant aquifer storage, subsurface flow increases dramatically after a quantifiable threshold of precipitation is reached (Tromp-Van Meerveld and McDonnell, 2006). The proposed mechanism is the filling and connecting of various distributed storage volumes, such as soil pores and microtopographic relief in the bedrock surface (Tromp-Van Meerveld and McDonnell, 2006). Recently this concept has been extended to the watershed or basin scale: relative to the beginning of a storm event, a much higher flow response is possible only when a critical number of storage volumes throughout a basin fill to a threshold level and become connected (McDonnell et al., 2021).
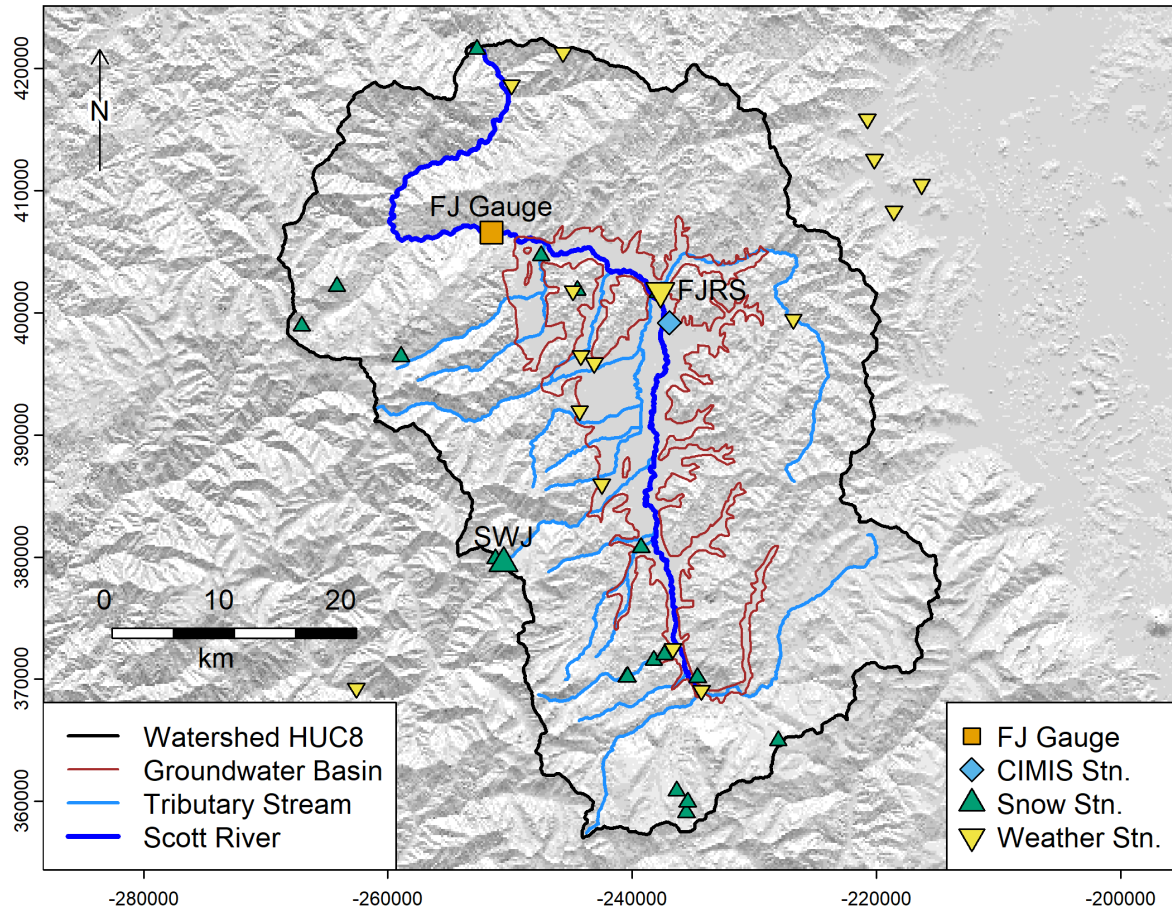
Figure 1. Scott River HUC8 watershed and groundwater basin boundaries, stream network, and key monitoring locations: the Fort Jones stream gauge (USGS ID 11519500), weather stations, snow observation locations, and a CIMIS station (used to estimate reference evapotranspiration). Selected locations are highlighted with an enlarged symbol and an abbreviated label.
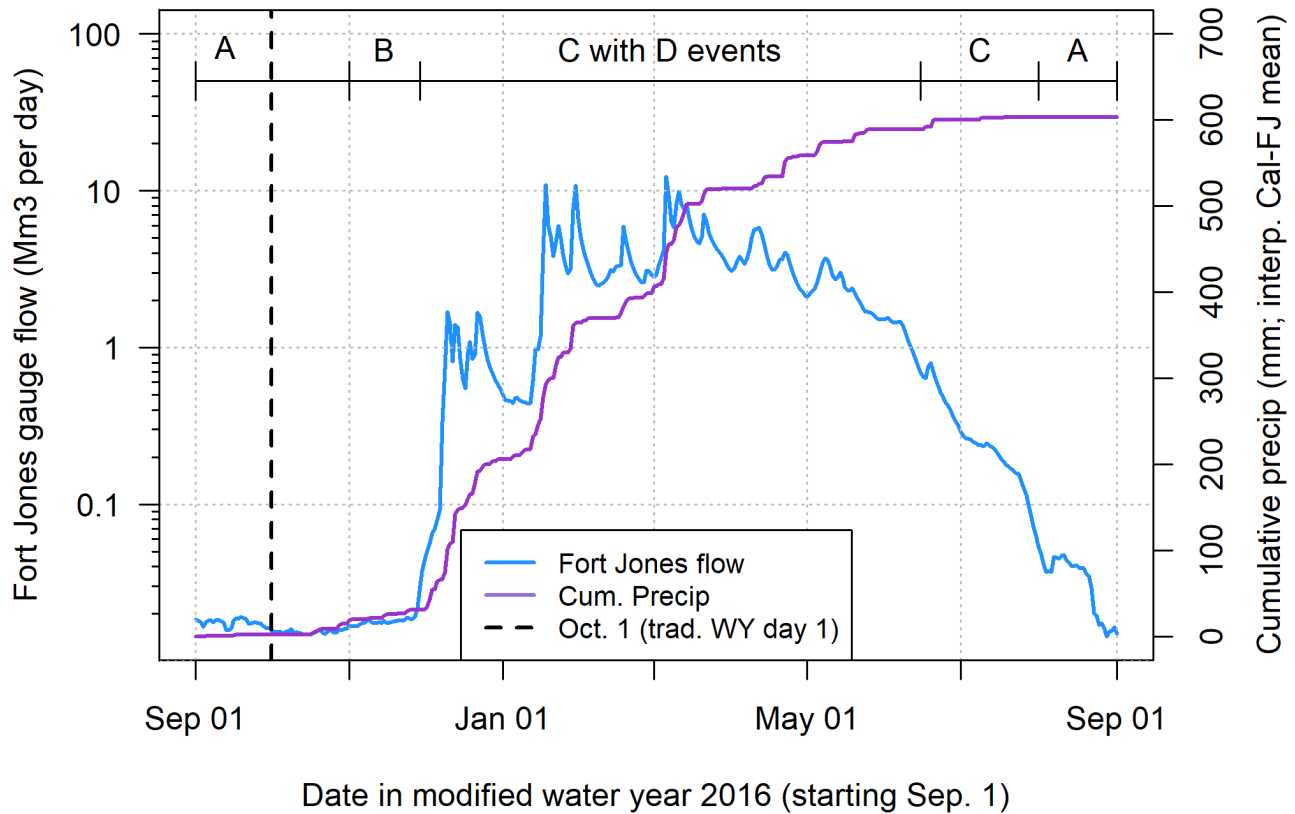
Figure 2. Illustration of four categories of Scott River watershed behavior. The hydrograph in the highlighted periods demonstrates the following watershed behavior: A, dry season baseflow – watershed draining from a medium-to-low storage level; B, moderate flow increase – muted flow surge response to new precipitation; C, winter baseflow and early spring recession – watershed draining from a high storage level; and D, winter stormflow – rapid flow surge response to new precipitation (storm spikes).

In this study we expand this concept of a basin-scale, threshold-based runoff response to the temporal scale of a season, rather than a single storm event. In this framework the condition of the Scott River watershed, as measured at the basin scale using the Fort Jones stream gauge, can be classified in four main categories. These categories are distinguished by current precipitation conditions and the volumetric proportion of the hydrologically connected reservoirs that are full of water (Table 1).

Water in the Scott River watershed is stored in five primary reservoirs (Harter and Hines, 2008):

- – snowpack
- – fractures in impermeable bedrock (uplands)
- – soil moisture/subflow
- – the alluvial groundwater aquifer (lowlands)
- – streams and surface water bodies

Accumulating snowpack is present only in the mountainous areas of the upper watershed and is limited to the winter and spring season. The alluvial aquifer is present only within the bounds of the groundwater basin underlying the flat valley floor. Water stored in fractured rock and soils/subflow emerges as springs in streambeds in the upper watershed (Mack, 1958) (Figure 1). In conditions with sufficiently high soil water content or groundwater elevations, soil moisture/subflow and groundwater become hydrologically connected to the surface water system. Conversely, water in the snowpack and fractured rock reservoirs is not hydrologically connected to major surface water bodies until it melts or descends lower into the watershed, effectively passing through the aquifer or soil to reach the stream. For convenience the soil moisture/subflow and aquifer reservoirs will be referred to as "connected" storage. Storage in streams and surface water bodies is negligible at temporal scales exceeding a few days.

2.1.1 Rainfall-runoff response, functional flows and Q_{spill}

In the absence of surface water reservoirs, it is useful to consider streamflow regimes from the perspective of natural water storage, as an intermediary between precipitation and stream flow at the outlet of the watershed. At the end of the dry season, the watershed is in a "draining from low storage" condition. Snow storage has not been available for several months, and groundwater storage is reaching its annual low point. This is reflected in a slowly declining or flat hydrograph, with a flowrate that has decreased for several months (Figure 2, first period A). As the dry season ends, the watershed begins receiving rain, and enters a condition of "filling from a low storage level". In this catchment, much of the earliest water entering the system is routed as recharge through the soil or the streambed to occupy space in the aquifer. Because groundwater moves more slowly through the watershed than surface water, the hydrograph at the Fort Jones gauge demonstrates a muted or delayed response to early rain events (Figure 2, period B).

At the onset of a new wet season, under average conditions, the flowrate of filling is greater than the flowrate of draining, and so the "filling from a low storage level" condition at the beginning of a rainy season is transient, lasting only until the filling process occupies enough aquifer and soil storage volume to produce a "full" condition. After the water storage in the basin reaches "full", if no more rain occurs, the watershed returns to its default "draining" condition, though from a higher

Table 1. Schematic of watershed behavior and functional flow components occurring during the transition from the dry season to the wet season in a Mediterranean climate; the categories are illustrated in an example annual hydrograph in Figure 2. Water storage level refers to the relative water content of the soil and aquifer within the watershed.

Water storage level	New precip. occurring?	Flow behavior description	Relevant functional flows
Low	No	(A) Watershed draining from a medium-to-low storage level (depleting snow storage filling groundwater storage via recharge; depleting groundwater storage)	Late spring recession and dry season baseflow
Low	Yes	(B) Watershed filling from a low storage level, with muted response to new precipitation (hillslope interflow, landscape recharge, rapid streambed recharge into alluvial groundwater basin, especially near the mountain front - downstream from the bedrock-alluvium transition)	Fall pulse flow or small/slow post-dry-season flow increase
High	No	(C) Watershed draining from a high storage level (hillslope interflow, snowmelt, groundwater discharge to streams)	Winter baseflow and early spring recession
High	Yes	(D) Watershed spilling from a high storage level, with rapid response to new precipitation	Winter stormflow

storage baseline than during the dry season, and with a higher draining flowrate (Figure 2, first period C). If there is additional
155 precipitation, the resulting surge in flow is much more rapid, reflecting a "spilling" condition (Figure 2, intermittent events D).

The precipitation and winter temperatures during the wet season produce an accumulation of snowpack, though in some years this can be reduced by warm periods and rain-on-snow events. Melting snowpack contributes subsurface flow and tributary streamflow to the lower watershed, producing a spring flow recession typically lasting from the last major precipitation event into the summer (Figure 2, second period C and second period A) and providing significant recharge via stream leakage near
160 the mountain front, at the margins of the alluvial basin. This process amounts to a net transfer of water in snow storage during winter and spring to water in groundwater storage in spring and summer.

Many of these phenomena are are reflected in the well-established elements of functional flows (Table 1). Winter stormflow is the functional flow metric corresponding to a "spilling" watershed. The spring recession can last for three to six months and its steepness is moderated by snowmelt. Because it bridges the high-storage and low-storage states, the early and late

spring recession appear in two different flow behavior categories (Table 1). Conversely, the flows that are here classified under "watershed filling from a low storage level" have no specifically discerning feature in the hydrograph and depend on year-to-year conditions. A discrete fall pulse flow does not occur in every water year, and no distinct metric has been proposed for the gradual post-dry-season flow increase.

Given the regular behavior observed during the dry season-to-wet season transition of the Fort Jones hydrograph (i.e., prolonged dry season baseflows followed by gradual flow increase and then storm surges), and the physical structure of this highly inter-connected basin, we propose that an operational flowrate threshold exists at the Fort Jones gauge that defines approximate transition to "full" or "spilling" basin condition (referred to here as Q_{spill}; see Results, Section 3.1).

2.1.2 Stream-aquifer interaction

In the groundwater basin (i.e., lower) portion of the watershed, the alluvial aquifer is the largest storage reservoir (Mack, 1958). Groundwater-surface water interactions drive Scott River flow behavior towards the end of the dry season, before the next rainy season begins, when snowpack is depleted and streamflow in many areas is sustained by groundwater discharge alone (Foglia et al., 2018). Discharge to streams from the alluvial aquifer occurs along the thalweg of the Scott River. In this highly interconnected system, groundwater discharge along the thalweg, at the annual scale, is balanced by landscape recharge and by recharge from stream reaches near the margins of the basin, overlying coarse alluvial fans (see discussion below).

We used the Scott Valley Integrated Hydrologic Model (SVIHM) (Tolley et al., 2019; Foglia et al., 2013a, b) to obtain the estimated volume of water exchanged monthly, in water years 1991-2018, between the surface stream network and the underlying aquifer. Streamflow gains from and streamflow losses to groundwater were integrated across the stream network to obtain a net monthly groundwater-surface water exchange value for the basin (Figure 3, panel A). These net monthly groundwater-to-stream flux values were then compared to simulated monthly flow volumes in the Scott River, measured at the Fort Jones gauge (Figure 3, panel B).

2.2 Observed response variables (V_{min} and P_{spill})

The Scott River is an undammed watershed, in which estimates of annual precipitation are an order of magnitude greater than the estimated water pumped or diverted for agriculture (Tolley et al., 2019). In this study we tested whether fundamental hydrologic characteristics, specifically the dry-season baseflow recession and the rainfall-runoff response to early wet season cumulative precipitation, can be predicted five to six months prior using observable hydroclimate data of the preceding wet season and antecedent climate and flow conditions prior to the water year. The first step was identification and quantification of relevant response variables describing these two processes.

2.2.1 Dry season baseflow quantities (V_{min}) and timing

Multiple integrated numerical indicators of dry season baseflows were evaluated for suitability as the response variables in this prediction exercise. Monthly flow volumes were preferred over a minimum daily flow value to reduce the influence of
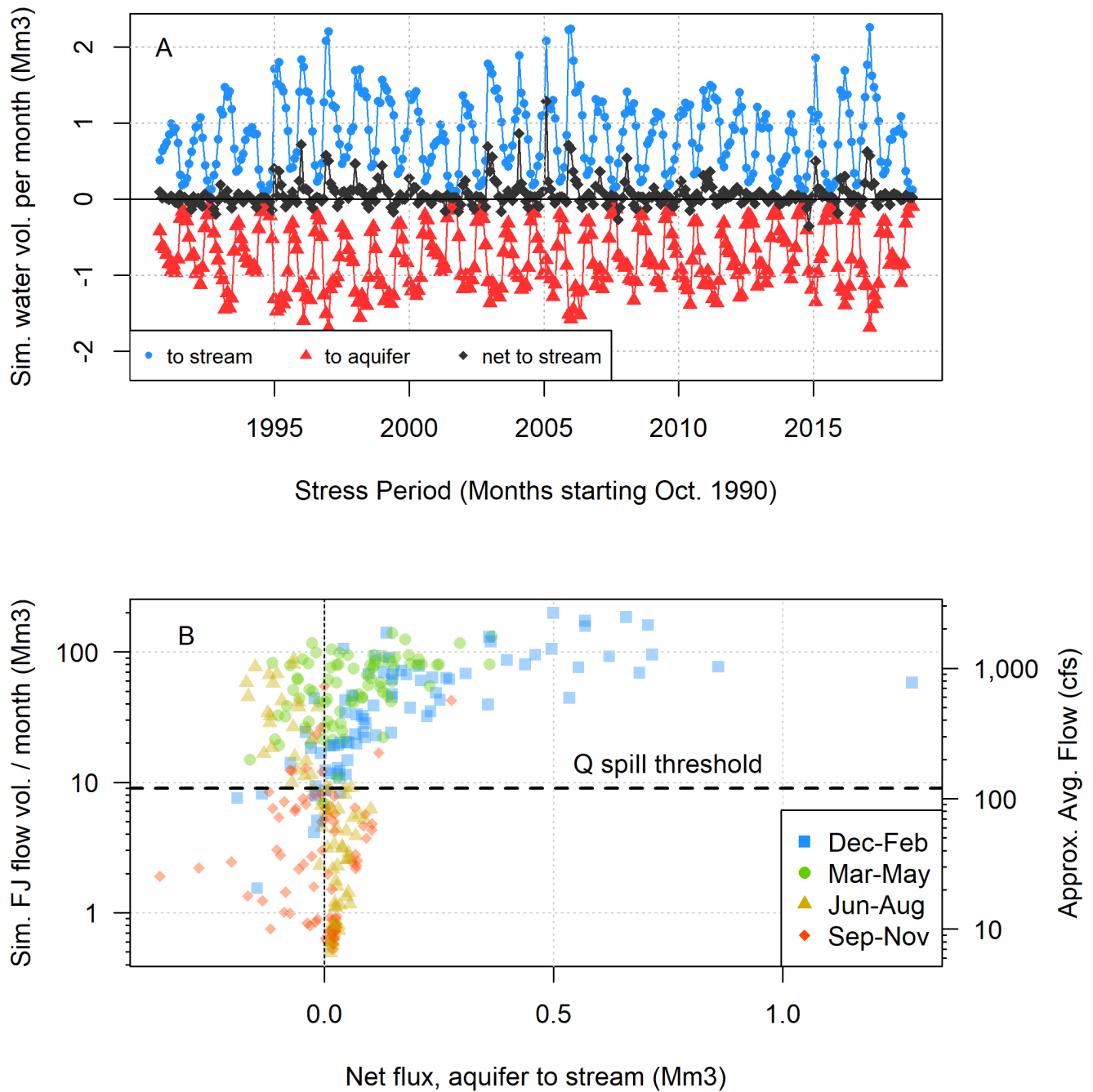
Figure 3. Both stream leakage and aquifer discharge increase in the rainy season, while net flux to the stream remains relatively close to 0 (panel A). Strong seasonal trends are evident in net flux to the stream (panel B; described further in text).

individual events that might affect flow on one or a small number of days, such as groundwater pumping or surface water diversions.

Historically, the rainy season in California tends to begin in October. By convention each water year begins on October 1st of the previous calendar year, and ends on September 30th. Matching this convention, in most years, the minimum-flow month for the Scott River is September; however, uncommon September storms can elevate flow volumes, and in some years with a late rainy season onset, the October flow volume may be lower. To capture these dynamics, we calculated a rolling 30-day sum of daily flow volumes in the period July-December for each calendar year to identify the 30-day period with the least flow volume, referred to as V_{min} (see Section 3.2). For consistency, each annual V_{min} value was assigned to the water year ending in September of that calendar year, even if the minimum flow window included days in October-December of the following water year.

2.2.2 Cumulative precipitation P_{spill}

P_{spill} was calculated for each water year as the cumulative rainfall at the end of a dry season, between September 1st and the first day that the Fort Jones gauge measured flow greater than Q_{spill} (see Results, Section 3.2). As stated above, conceptually, it is the amount of rainfall needed to "fill" the watershed such that streamflow at the outlet of the watershed responds rapidly to new precipitation without significant intervening storage delays. A dry season can have negative effects on an aquatic ecosystem if it produces extraordinarily low flows, but also if it lasts for an extraordinarily long time (e.g., delayed salmon habitat access, CDFW 2015).

In addition to the volumetric quantity P_{spill}, there could also be demand for seasonal predictions of the *timing* of onset of the coming rainy season. However, predicting the timing of the onset of the rainy season or of Q_{spill} would likely rely on uncertain long-term weather forecasts and is beyond the scope of this paper. In other words, due to randomness in rainfall timing, the exact dry season baseflow duration associated with a higher P_{spill} is highly variable and, hence, unpredictable.

2.3 Potential predictors and selected formulations

To evaluate candidate predictors of dry season baseflows, Pearson's correlation coefficient, R, was calculated between observed response variables V_{min} and P_{spill}, and the following categories of observed predictor data (see Results, Section 3.3):

1. Spring (March-May) water level observations in this case study in each of 54 individual wells (Figure A1).
2. Annual maximum snowpack water content at each individual snow monitoring station here at 20 stations aggregated by the California Data Exchange Center (CDEC; Figure 1).
3. Cumulative precipitation, October-April, at local weather stations (here: 12 stations within and 5 outside the watershed for a total of 17 NOAA stations; Figure 1). In these records, missing values (i.e., days with no recorded observation) are assumed to have 0 precipitation. Water years with more than 5 missing days are excluded from the predictor dataset.
4. Cumulative precipitation, October-April, of a composite precipitation record with no missing values, here representing the mean of the Callahan and Fort Jones NOAA weather stations (located at the southern and northern ends of the valley,

respectively), and referred to as "cal_fj_interp". To generate the composite record, missing values in the Callahan and Fort Jones station records were imputed based on observations at neighboring stations (see method in Foglia et al., 2013a).

5. The flow volumes observed at the watershed outlet, here the Fort Jones gauge (USGS ID 11519500; Figure 1) during the preceding March and April.

6. Cumulative reference evapotranspiration (ET_0), October-April, here using observations from Station No. 225 in the California Information Management Information System (CIMIS) network (2015-2021), or Spatial CIMIS estimates of ET_0 at the location of Station 225 (2002-2015) (Figure 1).

7. The timing (in Julian days) of the date of maximum snowpack measurement.

8. The timing (in Julian days) of the date of the volumetric center of the rainy season, calculated as the day the cumulative precipitation crossed 50% of the total.

9. The 1-year-lagged metrics of maximum snowpack, October-April cumulative precipitation, and April water levels (e.g., the October-April cumulative precipitation measured a full 17-23 months prior to a September minimum flow).

Individual measuring locations, such as wells or weather stations, were evaluated for sample size (i.e., years of data) and degree of relatedness with the two response variables. Relatedness of the monitoring locations with the highest R values in each category of monitoring observation are included in analysis results (see Results, Section 3.3).

2.3.1 Prediction formulae for V_{min} and P_{spill}

With a sample size of 80 years of dry season baseflow volumes, a one- or two-predictor model is best to avoid overfitting (James et al., 2013). Nonetheless, to thoroughly assess predictive potential of this dataset, three-predictor models were also considered.

Commonly, model diagnostics such as Akaike's Information Criterion (AIC) are used to evaluate the best of a set of competing statistical models (Burnham and Anderson, 2004). However, AIC and other information criteria methods can only be used when comparing models based on the same dataset (Burnham and Anderson, 2004). In the dataset for this study, variable record lengths and missing data points produced a situation in which the sample size is different for most of the combinations of predictors under consideration. Consequently, though AIC and additional diagnostics were calculated for all models (Tables A1 through A6), cross-validation was used in this study as the primary model selection technique.

To predict V_{min}, a set of six one-predictor models were generated using the observation location from each category with the highest R, and model fit was evaluated using Leave One Out Cross Validation (LOOCV) (James et al., 2013) (see Results, Section 3.3 and Table A1). For a dataset with n observations, the LOOCV error of a predictive model is obtained by recalculating the model coefficients n times, each time leaving out one observation, and comparing the resulting prediction to the single left-out observation. The root mean square of these n errors is the LOOCV error used to evaluate model performance in Results.

260   The single predictors with the lowest LOOCV error (other than evapotranspiration of a reference crop, ET_{ref}, which was excluded due to insufficient observation record length) were used to produce a set of four two-predictor models (see Results, Section 3.4 and Table A2) for V_{min}, including two that incorporate data from a full calendar year prior. A similar approach was used to assess two-predictor models for P_{spill}, though no predictors from a prior year were included, and several additional two-predictor combinations were evaluated. In both cases, the best-performing model took the following form:

$$Predicted_i = Int. + m_A * obs_{A,\,i} + m_B * obs_{B,\,i} \tag{1}$$

Where:

- $Predicted_i$ is the predicted value (either V_{min} or P_{spill}) in calendar year i (i.e., at the end of water year i).
- $obs_{A,\,i}$, $obs_{B,\,i}$ are the observed predictor values in October-April in water year i.
- $Int.$, m_A, m_B are the coefficients of the selected linear model.

3   Results

3.1   Scott River precipitation-runoff behavior

The quantity P_{spill} is correlated with both a lower minimum flow volume and a later river reconnection (Figure 4). This corroborates the hypothesis that P_{spill} is an indicator of the risk of a severe dry season.

    Visual inspection of 80 years of Fort Jones hydrograph behavior during the transition from the dry season to the rainy season (Figure 5, panel A) indicated that there were two distinct domains of flow: one in which flow is relatively flat (dry season baseflow), and one in which the flowrate is an order of magnitude higher, and it is highly responsive to rain events. The intermediate hydrologic state, "filling from low storage", was visible in some fall-winter hydrographs (e.g., Figure 2), but tended to last a relatively short time before the filling rate overwhelmed the draining rate and produced a responsive "spilling" condition.

    The approximate threshold between these two hydrologic periods, denoted as Q_{spill}, was 120 cfs (294 thousand m^3/day or approximately 9 Mm^3 per month). This value was determined by testing a range of potential Q_{spill} values to evaluate rainfall-runoff responses preceding and following the threshold (Figure A2).

3.2   Observed response variables

3.2.1   Dry season baseflow quantity and timing

Minimum 30-day dry season baseflow volumes, denoted here as V_{min}, ranged from 0.3 to 7.5 Mm^3 / 30 days (3.9 to 102 cfs), with one outlier value of 14 Mm^3 / 30 days (189.2 cfs) in 1984, when an early September storm followed a wet year in 1983 (Figure 6).
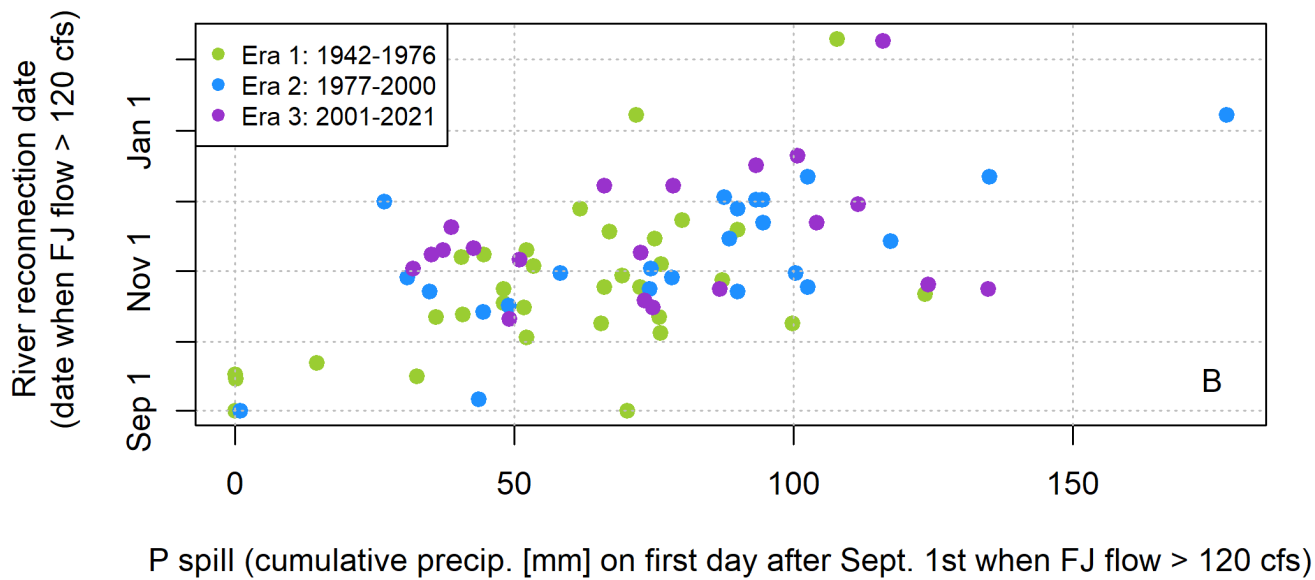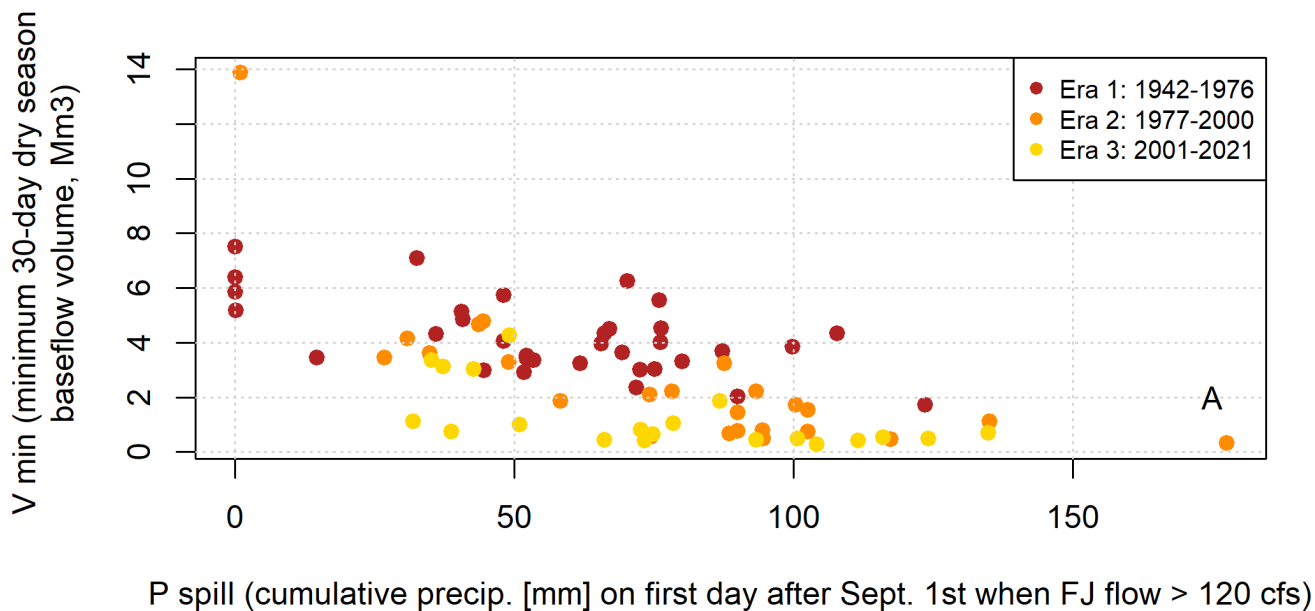
Figure 4. The quantity P spill (i.e., the amount of rainfall needed to 'fill' the watershed such that it 'spills', or responds rapidly to new precipitation) is correlated with both a lower minimum dry season baseflow volume (panel A) and a later date of river reconnection (panel B).
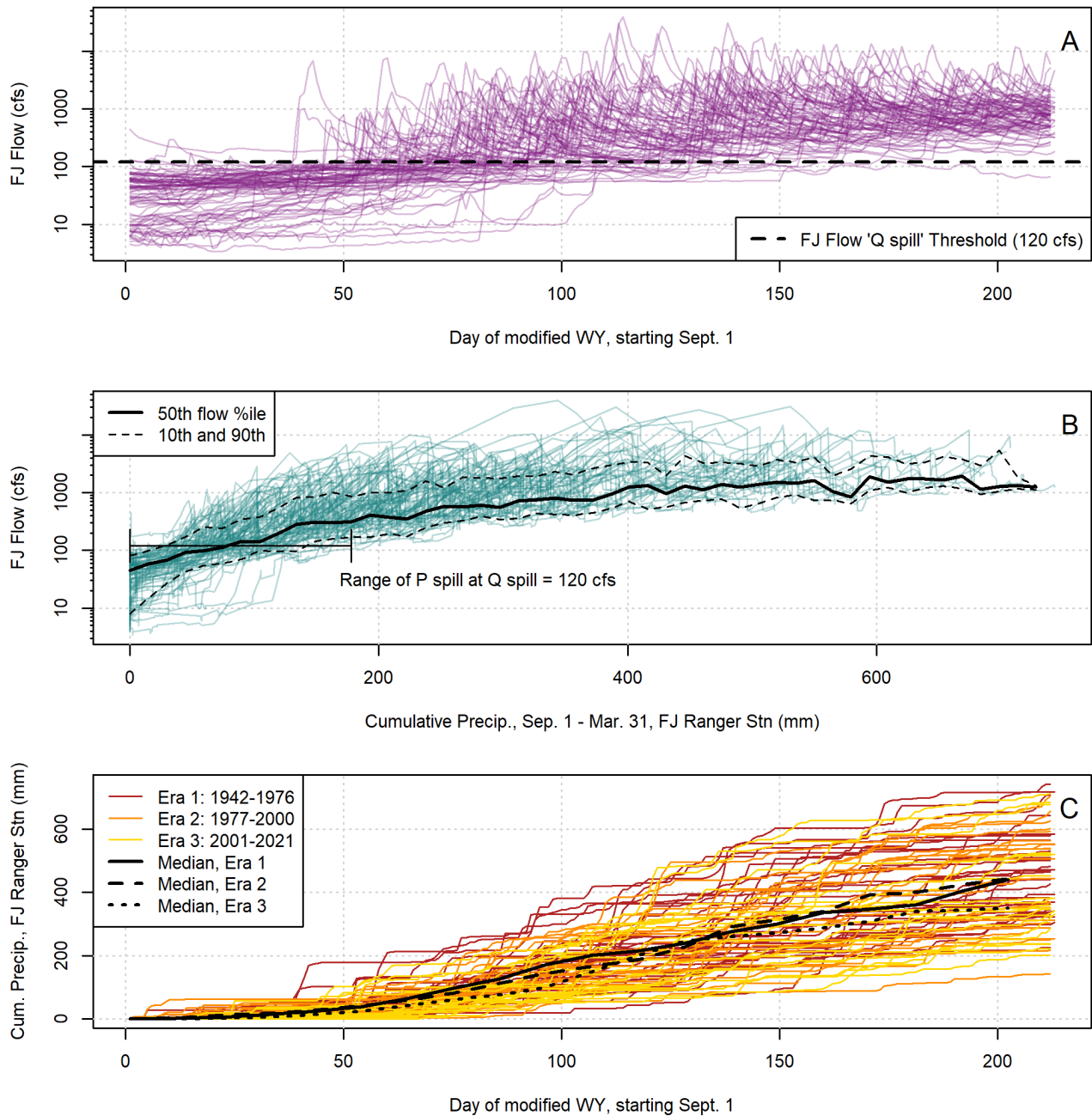
Figure 5. In all three panels, 80 years of data series from September 1 to March 31 are overplotted to illustrate dynamics during the transition from the dry to the wet season: observed Fort Jones hydrographs in Panel A; cumulative rainfall and Fort Jones flow values on fall and winter days in Panel B; and cumulative rainfall over time in Panel C.

As mentioned previously, three eras have been proposed for the Scott River flow record: Eras 1, 2, and 3, ranging from 1942-1976, 1977-2000, and 2001-2021, respectively. Matching other long-term declining flow trends in this watershed, the flows in August and September are relatively steady in Era 1, and they become more variable with significantly lower lows in Eras 2 and 3 (minima of 2.1, 0.35, and 0.33 Mm^3 / 30 days; or 28.6, 4.8 and 4.4 average cfs; in Eras 1, 2 and 3 respectively; Figure 6, panel A).

The timing of the midpoint of the 30-day minimum-flow period falls most commonly in September, though it has fluctuated over the last eight decades (Figure 6, panel B).

3.2.2 Cumulative fall precipitation and watershed response

Typically, flow at the Fort Jones gauge is low (i.e., less than the Q_{spill} threshold of 120 cfs) and stable through the end of the dry season (Figure 5, panel A). However, in some water years prior to the 1980s, the Fort Jones flowrate exceeded Q_{spill} on September 1st (Figure 5, panels A and B), indicating that even under persistent dry season draining conditions, under the climate and water use conditions of wet years in the mid-20th century, the Scott River remained responsive to new precipitation year-round. As a result, the range of P_{spill}, the cumulative precipitation necessary to reach Q_{spill}, is wide (0 to 178 mm, or 0 to 7 inches) (Figure 6, panel C). Mean P_{spill} values were 57, 79, and 76 mm (2.3, 3.1, and 3 inches) in Eras 1, 2 and 3, respectively.

3.3 Comparison with California DWR Water Year Type (WYT) category

The DWR water year type categories map well onto the two proposed hydrologic indices V_{min} and P_{spill} (Figure 7, panels A and B), which is to be expected, as both DWR WYT and the two proposed indices are based in part on cumulative precipitation data. However, there is less of an ability to identify a long-term trend in the DWR WYT index time series than in the time series of observed or predicted V_{min} or P_{spill} values. Likely causes include, first, the information lost when binning water years into five categories, and, secondly, the 30-year ranking window that would prevent a direct comparison of post-2000 WYTs with pre-1950s WYTs (Figure 7, panel C).

3.4 Predictor comparison for V_{min} and P_{spill}

The observations of spring flows, snowpack, valley floor precipitation, and groundwater elevation are positively correlated both within each category and to each other overall, which is unsurprising: wet years are associated with higher values in all of these categories. Groundwater wells with highest predictive power tend to have long records (e.g., n of 10 or greater years) and to be close to the Scott River; these results incorporate two wells proximate to the river, with record lengths of 43 and 57 years (Figure A1).

Both response variables are correlated with four categories of observations: spring flowrates, maximum snow water content, cumulative precipitation recorded at weather stations on or near the valley floor (October-April), and March-May water levels in some wells. Observations in these categories are positively correlated with V_{min} and negatively correlated with P_{spill}. The
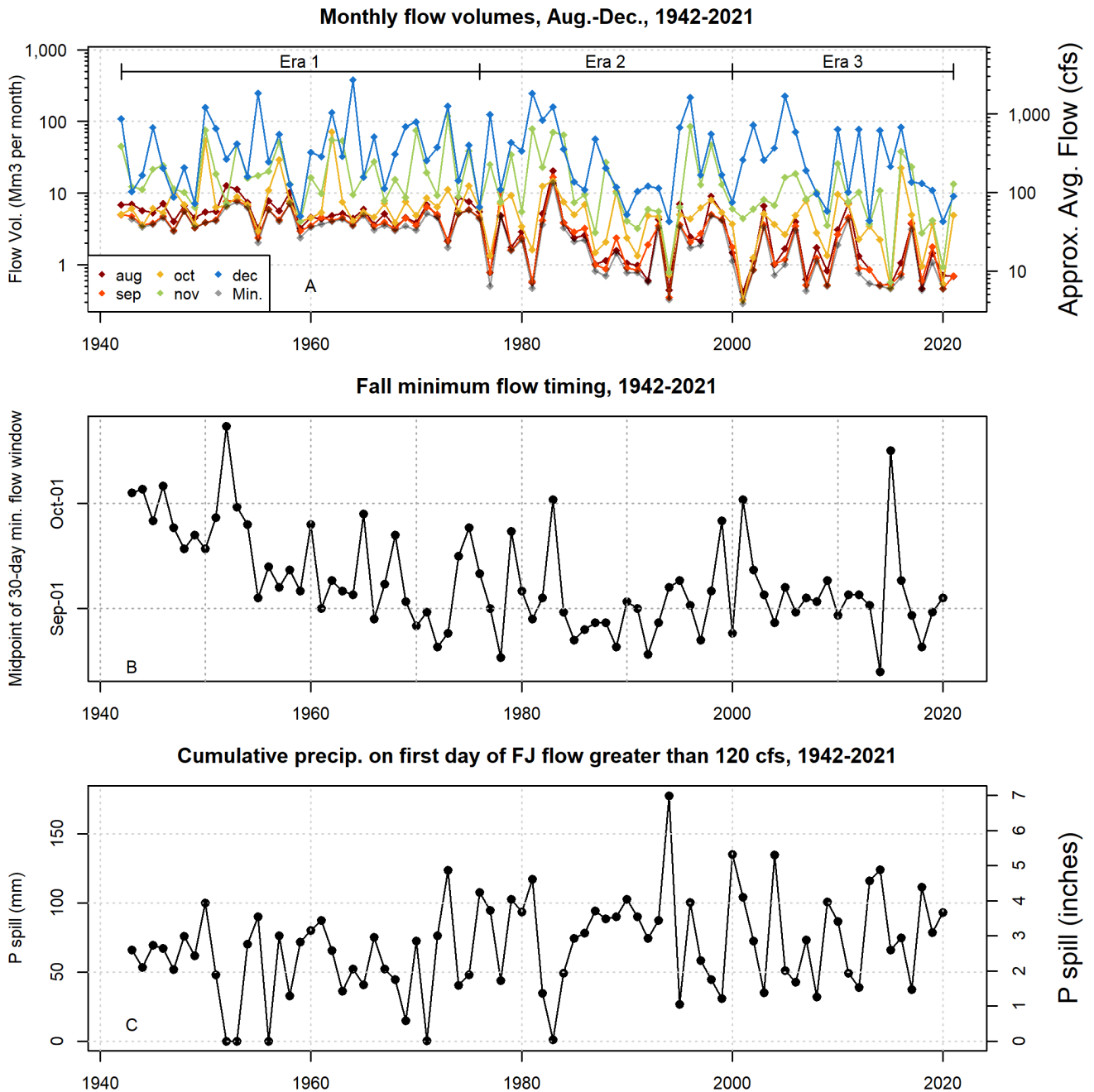
Figure 6. Panel A: FJ Gauge flow volume, by year, aggregated to monthly time windows in the late summer, fall, and early winter. Eras are noted that correspond to various management and climate factors. Panel B: The timing of dry season minimum flows has ranged from late August to mid-October over the past 8 decades. Panel C: P spill has trended upward over the period of record.
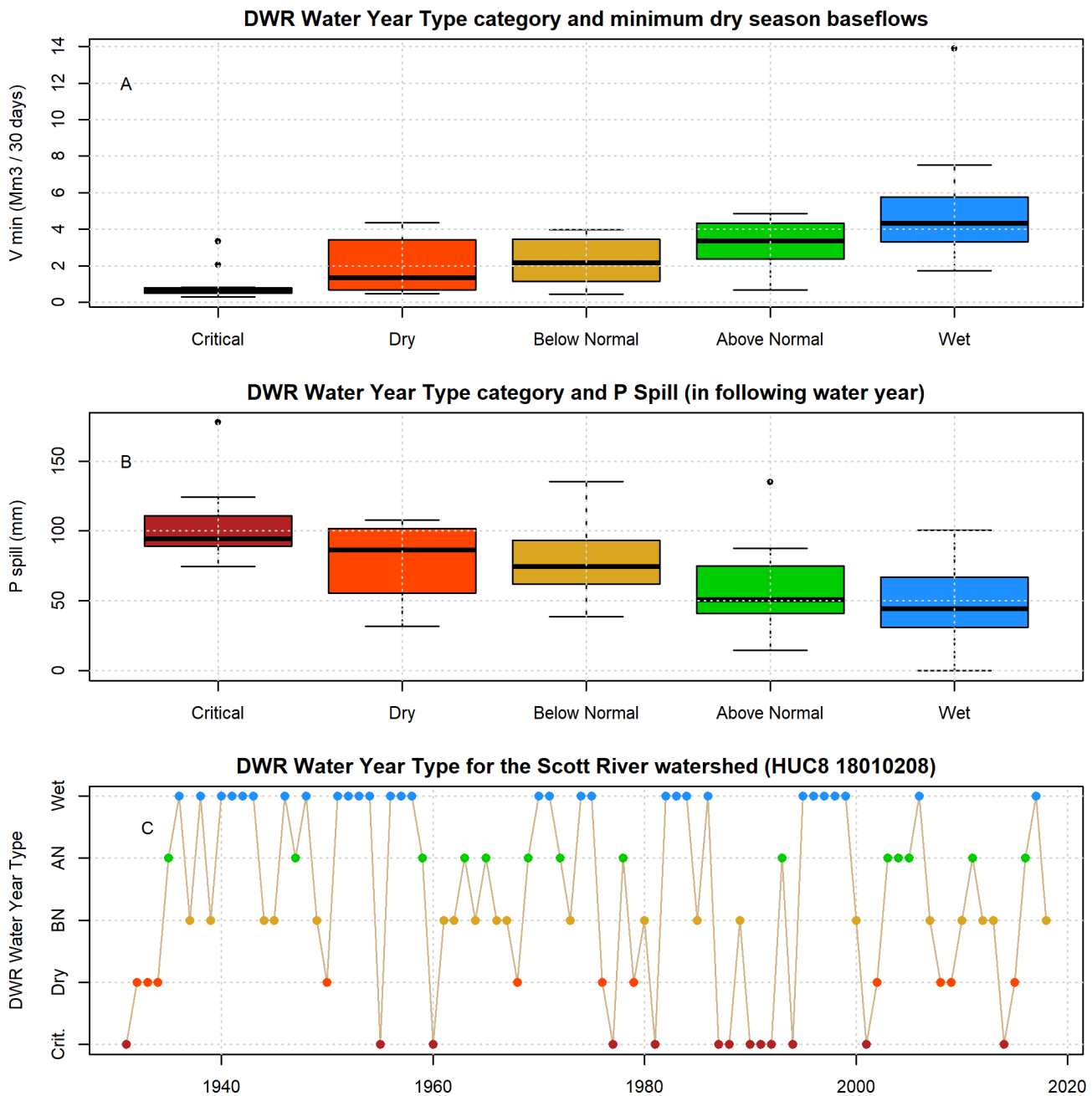
Figure 7. DWR Water Year Type indices over time and compared to the two metrics of hydrologic conditions developed in this study: minimum 30-day dry season baseflow volume (V min) and the amount of precipitation necessary to produce 120 cfs flow in the Scott River (P spill).
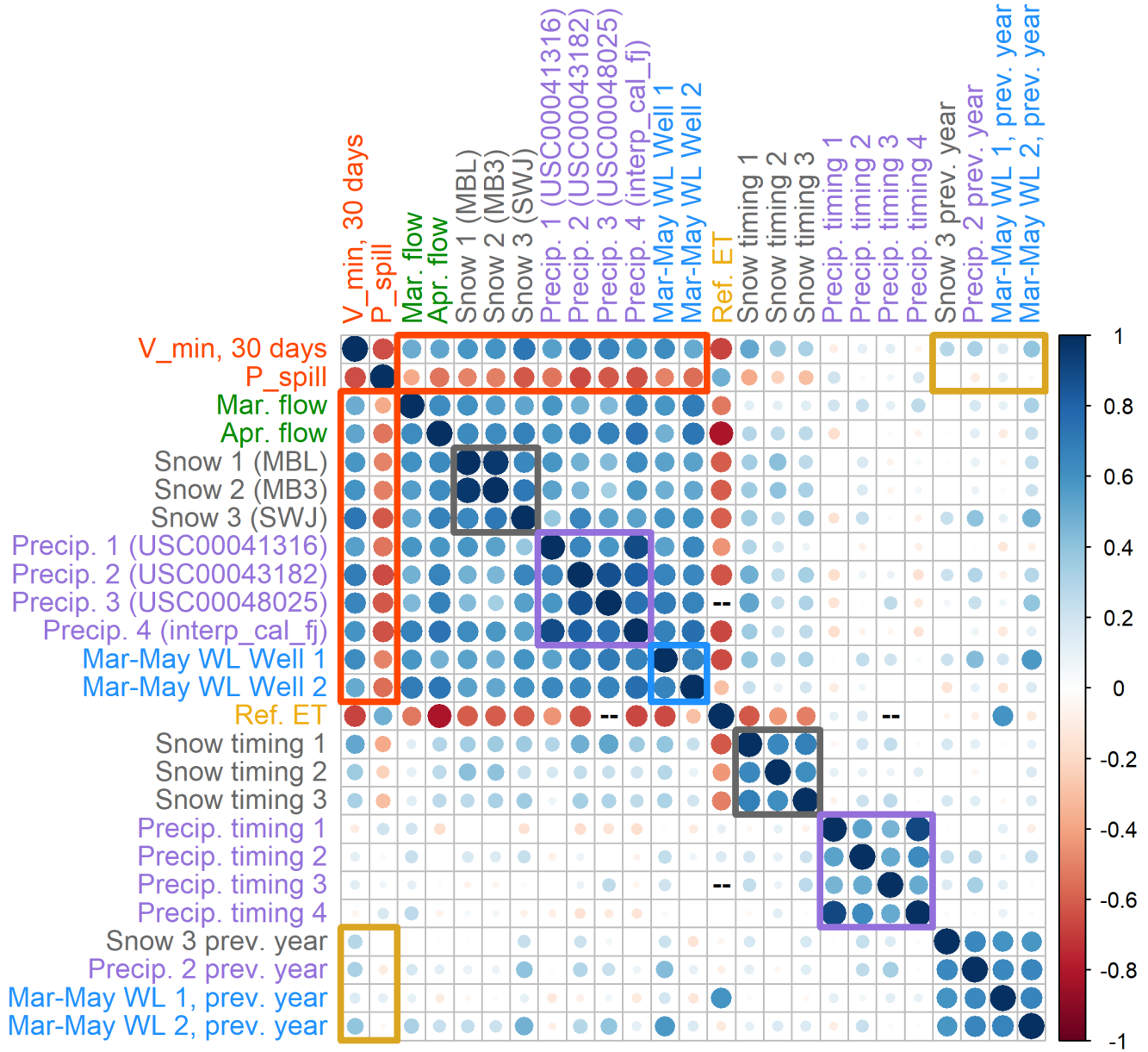
Figure 8. Correlation coefficient matrix of two response variables, minimum 30-day dry season baseflow volumes (V min) and cumulative precipitation necessary to produce 120 cfs in the Scott River (P spill), with various possible predictor metrics. Gray, purple, and blue squares highlight the inter-category correlation coefficients of snowpack metrics, Oct-April cumulative precipitation, and March-May groundwater elevation measurements. Red rectangles highlight the predictors with the greatest absolute correlation coefficient values with V min and P spill, respectively. Yellow rectangles highlight correlations with previous-year hydroclimate quantities.

correlation coefficient, R, of these response-predictor relationships range from 0.5 to 0.73 for V_{min} and from -0.38 to -0.66 for P_{spill} (Figure 8).

In contrast to these four categories, October-April cumulative ET_0 is negatively correlated with V_{min} and positively correlated with P_{spill} (R of -0.68 and 0.48 for V_{min} and P_{spill}, respectively). October-April cumulative ET_0 is also negatively correlated with snow, precipitation, and groundwater level indicators. This could be because ET can remove a significant volume of water from the watershed, or because years with more rainy or stormy days accumulate less total insolation and atmospheric water demand. Additionally, the relatively high correlation (R) for between ET_0 and the two response variables may be due to a small sample size, as all available ET_0 observations or estimates were collected in 2002 or later (i.e., in Era 3; Figure A3).

Some meteorological timing was evaluated in addition to quantities. Both response variables are moderately-to-weakly correlated with snow timing (i.e., the Julian day of the maximum measured snowpack in a given year; R of 0.33 to 0.52 and -0.22 to -0.38 for V_{min} and P_{spill}, respectively), but no significant correlation is evident between the response variables and precipitation timing (Figure 8).

Finally, a subset of observations from the previous water year were included in the correlation matrix to test for multi-year influence on the response variables. These previous-year metrics had very little predictive power regarding minimum flows (correlated with V_{min} with an R of 0.29 to 0.33) and virtually none regarding P_{spill} (correlated with P_{spill} with an R of -0.01 to -0.1).

3.5   Predicted values of V_{min}

3.5.1 V_{min} predictor assessment and prediction formula

Six hydroclimate data categories contained at least one measurement station with an R value greater than 0.5 between a data record and observed V_{min} values. The single monitoring location in each of these categories with the highest R value was selected for further analysis (Figure A3). (Although an R value of 0.5 is not considered a strong correlation on its own, it was selected as an inclusion threshold to retain any predictors that could provide useful information in combination with other predictors.) Out of this set of six hydroclimate data records, the maximum snowpack and October-April cumulative precipitation produce the lowest model errors (LOOCV errors of 2.74 and 2.72 Mm^3, respectively; Figure A3, top two panels; Table A1). Among combinations of two predictors, a linear combination of maximum snowpack and cumulative precipitation improved on the best single-predictor model, with an LOOCV error of 2.29 Mm^3 (Figure A4, upper left panel; Table A2). No three-predictor models produced lower LOOCV or higher R^2 values than the best two-predictor models (Table A3), so no visualizations were made of three-predictor model results.

Among the two-predictor models evaluated was a combination of maximum snowpack water content and the timing of the maximum measurement (Figure A4, top right panel). This produced a slightly larger error (2.78 Mm^3) than the single-predictor model with maximum snowpack water content alone (2.74 Mm^3; Figure A3, middle left panel), indicating that the timing of maximum snow accumulation is either relatively unimportant in generating dry season baseflows or that the actual timing of snowpack maximum is not captured in temporally sparse snow course measurements.

Additionally, two models featuring a partial one-year holdover were evaluated, to test the validity of this component of the methodology of DWR's Water Year Type index. In both cases, the addition of the climate data from the previous year produced a very small change in model error (Figure A4, two lower panels), indicating that in the Scott Valley context, the previous year's climate may have a minor influence on dry season flows relative to the immediately preceding rainy season.

Based on these results, the model selected to provide the best V_{min} prediction formulation was a linear combination of snowpack maximum from the Swampy John (SWJ) snow station (with data collected by CDEC) and cumulative October-April precipitation from the Fort Jones Ranger Station (FJRS) weather station (with data collected by NOAA) as follows:

$$V_{min,\,i} = -1.33 + 0.0053 * FJRS_i + 0.0027 * SWJ_i \tag{2}$$

Where:

- $V_{min,\,i}$ is the predicted value of minimum 30-day dry season baseflows in calendar year i (i.e., at the end of water year i) (million m^3 or Mm^3)
- SWJ_i is the maximum snow water content recorded at the Swampy John snow course (CDEC station ID SWJ or 285) in water year i (millimeters)
- $FJRS_i$ is the cumulative precipitation, recorded October-April of water year i, measured at the Fort Jones Ranger Station (NOAA station ID USC00043182) (millimeters)

Diagnostics for the linear regression models assessed in the V_{min} analysis are included in the Appendix (Figures A3 through A5 and Tables A1 through A3).

3.5.2 Predicted and observed V_{min} over time

The V_{min} formulation proposed above predicts minimum 30-day dry season baseflows with a model error of 2.29 Mm^3 per 30 days (31.2 cfs), and a root mean squared error of 1.4 Mm^3 (19.4 cfs). This RMSE indicates substantial uncertainty in any single year's prediction: it corresponds to 49% of the mean V_{min} value, 2.9 Mm^3 (40 cfs).

Matching the historical trends of decreasing snowpack, the observed and predicted V_{min} values show a downward trend over time (Figure A5, top panel). An outlier in the year 1984 reflects extremely high minimum dry season baseflows, relative to the predicted values and the overall distribution. In that year, a relatively high-baseflow season was followed by an early September storm. The model residual (predicted minus observed flow volumes) for 1984 is also an outlier, indicating that the model has a sufficient sample size to not be overwhelmed by this extreme value produced by an exceedingly rare sequence of events (Figure A5, middle panel).

The predictive V_{min} model is based on observations from the full record, but three additional models were generated based on only the observations from each period: Eras 1, 2, and 3, respectively. Residuals based on Era 1 data are similar to those of the full record, with a slight but systematic overprediction in Era 3; Era 2 residuals tend to overpredict in Era 1 more than the full record; and Era 3 residuals offer better performance in Era 3 than the full record, but produce significant systematic underpredictions pre-2000 V_{min} values (Figure A5, middle panel).

3.6 Predicted values of P_{spill}

3.6.1 P_{spill} predictor assessment and prediction formula

The two single predictors of P_{spill} (with the largest R or least error or both) were October-April cumulative precipitation and the maximum snowpack (Figure A6, top two panels), the same parameters that provided the best prediction V_{min}. The LOOCV model errors are 850 mm and 718 mm, respectively. As in the V_{min} model development, ET_0 was excluded from consideration due to short record length. The best two-predictor model was the combination of the two best single predictors, with an LOOCV error of 697 mm (Figure A7, upper left panel). No three-predictor models produced lower LOOCV or R^2 values (Table A6), so no visualizations were made of three-predictor model results.

Several combinations of correlated observation categories produced comparable model results, such as spring water levels with maximum snow, maximum snow timing, and cumulative precipitation (Figure A7, upper right and two middle panels). However, not all combinations of co-correlated data produced reasonable predictors; a model with a linear combination of maximum snowpack timing and March flow volumes performed relatively poorly (LOOCV error of 1,087 mm; Figure A7, lower right panel).

Based on these results, the model selected as the P_{spill} formulation for a given water year was a linear combination of the same observation records as V_{min}: snowpack maximum from the SWJ snow station (with data collected by CDEC) and cumulative October-April precipitation from the Fort Jones Ranger Station weather station (station ID USC00043182, with data collected by NOAA).

$$P_{spill,\,i} = 128 - 0.095 * FJRS_i - 0.028 * SWJ_i \tag{3}$$

Where:

- $P_{spill,\,i}$ is the predicted value of cumulative rainfall at the end of the dry season, starting Sep. 1, on the first day that the Fort Jones gauge records flow greater than or equal to 120 cfs in calendar year i (i.e., at the end of water year i) (millimeters)
- SWJ_i is the maximum snow water content recorded at the Swampy John snow course (CDEC station ID SWJ or 285) in water year i (millimeters)
- $FJRS_i$ is the cumulative precipitation, recorded October-April of water year i, measured at the Fort Jones Ranger Station (NOAA station ID USC00043182) (millimeters)

Diagnostics for the linear regression models assessed in thu P_{spill} analysis are included in the Appendix (Figures A6 through A8 and Tables A4 through A6).

3.6.2 Predicted and observed P_{spill} over time

The P_{spill} estimate formulation proposed above predicts P_{spill} values with a model LOOCV error of 697 mm (27.4 inches), and a root mean squared error of 25.4 mm (1 inch). This RMSE indicates substantial uncertainty in any single year's prediction:
415   it corresponds to 37% of the mean P_{spill} value, 68.8 mm.

    Matching the historical trends of decreasing snowpack, the observed and predicted P_{spill} values show an increasing trend over time (Figure A8, top panel). A high outlier in calendar year 1994 (in early water year 1995) was caused by a dry water year 1994 followed by a series of small storms in November and December, none of which produced 120 cfs of flow, followed by a much larger storm on January 8th-9th of 1995 in which the river flow jumped to 600 and then 7,500 cfs in two days.

420   The predictive P_{spill} model is based on observations from the full record. As for V_{min}, three additional models were generated based on only the observations from each period: Eras 1, 2, and 3, respectively. Residuals based on Era 1 tend to underpredict Eras 2 and 3 more than the full-record model; Era 2 residuals tend to overpredict in Eras 1 and 3 more than the full record; and Era 3 residuals have a slightly higher tendency to underpredict than the full record, but overall are fairly similar to the full-record residuals (Figure A5, lower panel).

425   The linear coefficients for the two prediction equations and their standard errors are summarized in Table 2.

Table 2. Summary of linear model coefficients for V min and P spill.

	b	m_SWJ	m_FJRS
V min prediction	-1.3	0.0027	0.0053
V min std. error	0.4	0.0006	0.0013
P spill prediction	128.1	-0.0947	-0.0283
P spill std. error	7.9	0.0228	0.0103

4 Discussion

4.1 Scott River watershed behavior and long-term planning

The forward-looking seasonal predictions in this study are possible because of predictable hydrologic relationships between early-season inputs (precipitation) and late-season outputs (surface flow and ET), as is typical for many Mediterranean climate
430   watersheds. In this watershed there are reasonably consistent rainfall-runoff relationships: the general shape of the relationship between cumulative precipitation and river flow at onset of the rainy season is preserved in dry and wet water years (Figure 5, panel B). This consistency is also reflected in stormflow behavior occurring after precipitation fills the watershed, a condition in which the Scott River exceeds a flowrate of Q_{spill} at the Fort Jones stream gauge.

    A Q_{spill} threshold of 120 cfs was identified by an analysis of rainfall-runoff responses and a visual inspection of fall
435   hydrograph behavior (Figures A2 and 5). It also matches information from local stakeholders. Many tributary streams on the valley floor run dry during the summer and fall, and some tributary streams respond more quickly to fall precipitation than

others. Generally, the timing of all tributaries reaching flowing status corresponds with the Fort Jones gauge reaching 100 cfs (Sommarstrom, 2020).

Simulated estimates of stream-aquifer exchange corroborate these precipitation-flow relationships. Dry season baseflow (V_{min}) and the onset of wet season flow (framed in terms of P_{spill}) are both influenced by net groundwater discharge from the aquifer. One interpretation of the high frequency of near-0 net monthly stream-aquifer flux values (Figure 3, panel B) is that the high degree of connectivity between the streams and the aquifer in the Scott River system produces balancing counter-forces in response to hydrologic stresses on the system, such as large recharge events. This balancing tendency can be temporarily overwhelmed by large precipitation pulses, but high-flow conditions quickly reduce the volume of water in the surface water system, returning the Scott River to a baseline of relatively-balanced stream-to-aquifer and aquifer-to-stream fluxes. This dynamic also reflects the small size of the available aquifer storage, relative to the amount of precipitation received by the watershed in a given water year (DWR 2004), leading to net groundwater contributions that are one to two orders of magnitude smaller than streamflow during "spilling" conditions (Figure 3, panel B) and two or more orders of magnitude difference between dry season low flows (almost exclusively from groundwater contributions) and wet season high flows.

These limitations in natural water storage preclude the use of wet year surplus water for dry years and dictate that water management focus on within-year projects and management actions that either carry over some of the wet season flow to enhance dry season flow (V_{min}) that possibly reduce P_{spill} via interim aquifer storage, or that reduce surface water diversions and groundwater pumping, or a combination of both. These conditions also reduce resiliency to climate change.

4.2 V_{min} and P_{spill} prediction utility

Though various methods exist to qualitatively predict, in the spring, the severity of the coming low-flow season in the Scott River watershed, a quantitative short-term forecasting index could support more rigorous thresholds for adaptive management. To this end we developed two linear equations for predicting minimum dry season baseflows about five months in advance, by taking an inventory of relevant seasonal hydrologic inputs in April. The V_{min} and P_{spill} predictions could be used to support decision making processes in the late spring timeframe on potential regulatory actions and cropping decisions affecting the growing season.

There are several methods in current use. Observations at existing monitoring locations, such as weather stations and long-term snow course records, have been used as ad-hoc hydrologic indices. Historical adaptive management decisions in the Scott River watershed, such as planning to purchase surface water rights leases, have relied on individual monitoring observations, such as percent of snowpack relative to average conditions, or the Fort Jones flow in the spring (e.g., SRWT 2018). Additionally, DWR has effectively extended the methodology of the SVI and SJI metrics to all of California by publishing a categorical water year type (WYT) index for all its major watersheds (to the HUC8 level) (DWR 2021). This metric quantifies meteoric drought and relies only on precipitation data, so as to be comparable across the state. Matching SVI and SJI methodology, it can be calculated at multiple points in each spring, with a final determination in May, but in the case of Scott Valley it has been used to classify WYTs only retroactively through 2018. As previously mentioned it is a relatively complex metric with provisions

470   including a partial one-year holdover (i.e., dry conditions in the previous year will make a dry-type categorization more likely the following year), and non-stationary index thresholds, with a 30-year ranking window.

The quantitative prediction methods proposed in this study map well onto the existing DWR WYT index (Figure 7), which serves as a validation of this approach. However, the V_{min} and P_{spill} metrics preserve more detail. The primary advantages of the proposed method over the DWR WYT and other previous methods of gauging near-term hydrologic conditions is that

475   it is tailored to local hydroclimate data and is interpretable as a numeric prediction of fall streamflow conditions. This could be used to inform regulatory actions that increase fall environmental flows, or – for surface water diverters – to plan for low-flow conditions. However, each seasonal prediction should be accompanied by its measure of uncertainty: the RMSE of this predictive model is 49% of the mean value for V_{min}, 2.9 Mm^3 (40 cfs).

Though it also serves as an indication of the severity of a water year, the immediate seasonal utility of the second predicted

480   metric, P_{spill}, may be less than for that of minimum dry season baseflows. Management decisions such as the last possible date to keep a temporary stream gauge installed in a river, without risk of it being washed out, could be informed by a P_{spill} prediction when combined with weather forecasts in the fall. This prediction also carries substantial uncertainty (RMSE of 37% of the mean P_{spill} value, 68.8 mm).

4.3   Management implications of best-performing predictors

485   As described in Results, the linear models that best predicted observed values of V_{min} and P_{spill} were both based on the same two observation locations (the SWJ snow course and the FJRS weather station; Figure 1), both with lengthy observation records. One interpretation of these results is that the climate inputs produce a predictable fall watershed response, and that human management decisions have a negligible influence on fall river flow. However, model simulations suggest that the timing and magnitude of fall flow increases can be meaningfully influenced by human water use (e.g., scenarios in the Siskiyou County

490   GSP) (Siskiyou County 2021).

Multiple possible explanations could reconcile these two pieces of seemingly contradictory evidence. First, random variability in human water use could be a contributing factor to the error of the predictive models of fall-season hydrologic behavior. Alternatively, human water uses could be so consistent in response to wet or dry season conditions that these water uses could be implicitly incorporated into the predictive models. If adaptive management actions (potentially including events as diverse

495   as regulatory curtailments or individual cropping choices) are carefully recorded in the future, they could be compared to residuals of the climate-based predictive models to evaluate whether any signal of a response to human interventions can be observed.

Potentially, these seasonal predictions could be extended to additional watersheds (albeit ones with abundant available hydroclimate data). However, the feasibility of this geographic generalization is beyond the scope of this study and should

500   be investigated in future work.

4.4 Influences of climate change on predictive indices

Both predictions (using the full record of hydrologic data) assume some degree of hydroclimate stationarity, in that they use historical snowpack- and precipitation-runoff relationships to predict modern runoff. In one sense, a longer-term record can be an asset, in that it provides context for the severity of the dry periods of the past two decades. In another sense it is a liability for prediction accuracy: for example, the predicted V_{min} values based on the full record appear to systematically overpredict V_{min} in the most recent era (2001-2021; Figure A4, top left panel, and Figure A5, middle panel). This suggests that factors not captured in these climate data, such as warmer air temperatures, changing upland vegetation and evapotranspiration dynamics, and possibly unknown changes in water use, may be altering the relationship between the spring water supply and dry season baseflow rates. However, even with detailed records of water use and management actions, disentangling the influences of hydroclimate and human management on streamflow conditions remains a complex challenge.

5 Conclusions

This study proposed two locally-tailored hydrologic decision-support metrics for the Scott River watershed in northern California. Both metrics have been shown to be predictable quantities using snowpack and cumulative precipitation data from October-April: the first metric is the minimum 30-day flow volume in a given water year, referred to as V_{min}, which typically occurs in September or October. The second metric is the cumulative rainfall needed to "fill" the watershed after the end of the dry season to a "spilling" condition that responds quickly to precipitation events, referred to as P_{spill}. Both indices can be calculated at the end of April to support near-term (seasonal) adaptive management regarding the growing season (summer through the fall), similar to the SVI and SJI in California's Central Valley and other indices used throughout the western United States. However, climate change may reduce the predictive accuracy of indices based on long-term data records, and updates based on smaller numbers of more recent water years should be considered periodically.

The management choices facing local managers in this basin are difficult to quantify and summarize, as is the case in basins throughout the western U.S. and other semi-arid and arid regions globally. Locally-derived summary metrics, tailored to regional hydrologic dynamics, have provided and will continue to provide tools for supporting those choices and communicating them to diverse stakeholders and the general public.

 6 Appendix

A.1 Groundwater Well Data

Groundwater elevation measurements collected in the spring were evaluated as a potential predictor of river flow behavior the following fall. The length of the records and number of spring-season measurements were variable, limiting the number of wells that could be used as predictor sites. Two wells were selected for the final predictor evaluation.

Correlation coefficients between April groundwater elevations and subsequent September flow volume
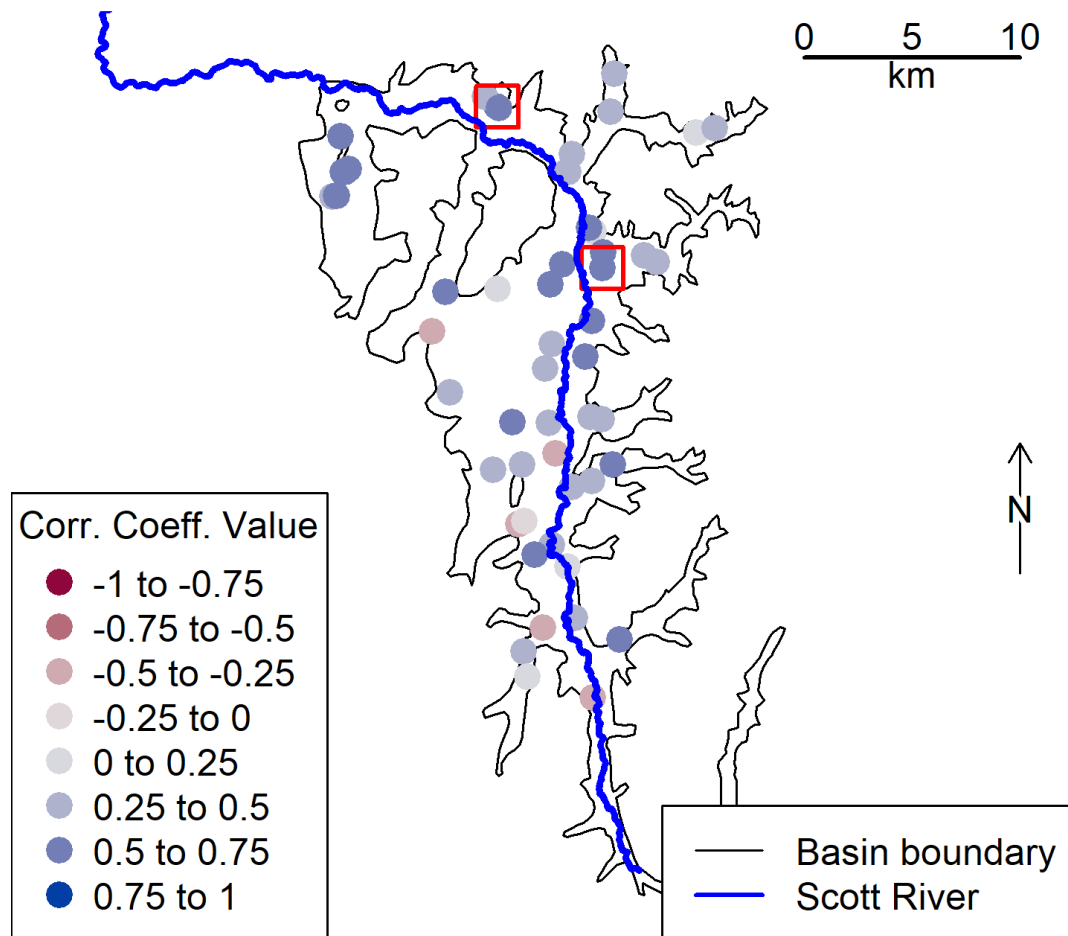
Figure A1. Boundary of the groundwater basin (corresponding approximately to the extent of the flat valley floor in the Scott River watershed) and selected well locations. Colors correspond to the correlation coefficients between April groundwater elevations and September flow volume. The wells included in the predictor comparison are highlighted with a red outer square. 54 wells had enough spring-season water level measurements to be included in this correlation exercise, though some wells are so close together that their symbols overplot on this map.

A.2 Selection of the Q_{spill} threshold value

The Q_{spill} value of 120 cfs (294 thousand m³/day or approximately 9 Mm³ per month) was determined by testing a range of potential Q_{spill} values (10 to 500 cfs [24 to 1223 thousand m³/day]) as a dividing threshold between dry- and wet-season flow behavior. Specifically, the threshold, Q_{spill}, was here operationally defined as the flow rate that generated the largest, difference in rainfall-runoff response on either side of the dry season-wet season divide (Figure A2, Panel A). The rainfall-runoff response was measured as the ratio dQ/dP, where dQ is the change in runoff and dP is the cumulative precipitation over the same moving 30-day period. Hence, the objective of this analysis was to find Q_{spill} such that the difference between the two dQ/dP values for the 30 days immediately prior to reaching Q_{spill} and the 30 days immediately following the day when Q_{spill} reaches an initial maximum.

For the Scott Valley watershed, the difference between the wet and dry season rainfall-runoff responses reaches an initial plateau (113 cfs [275 m³/day] of increased flow per mm of precipitation) for spill thresholds from 105-135 cfs (256-330 10³ m³/day) (Figure A2, Panel A). This indicates that in a large number of water years, flows in the range of 105-135 cfs (257 to 330 thousand m³/day) range are preceded by a dry season flow-response regime and followed by a distinct, flashier flow regime regime. Though higher wet-dry flow response differences were calculated at higher threshold values (i.e., up to 500 cfs [1,223 thousand m³/day]), these progressively higher wet-season flow responses likely reflect the falling limb of individual large storms that over-fill the watershed rather than separating filling from spilling behavior.

Additionally, in visual inspection of 78 years of Fort Jones hydrographs, the 120 cfs (294 10³ m³/day) Q_{spill} threshold generated a plausible breakpoint between dry and wet season river behavior in all water years (e.g., water year 2018 in Figure A2, Panel B). Furthermore, this value corroborates local observations that an approximate value of 100 cfs represents "fully connected" river conditions (see Discussion section).
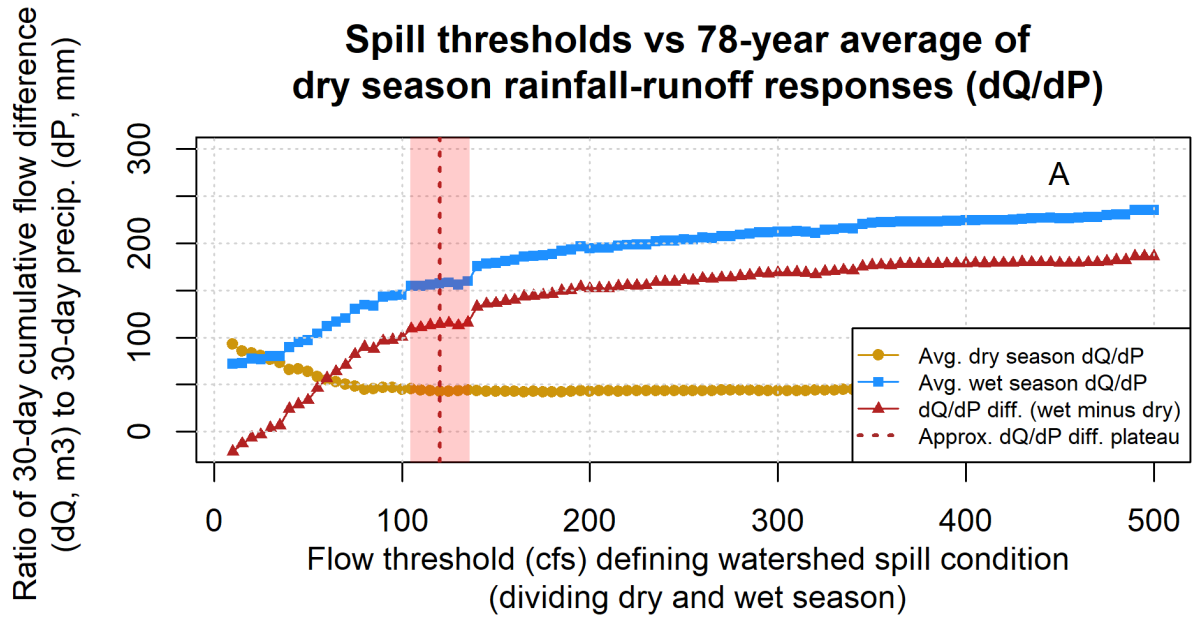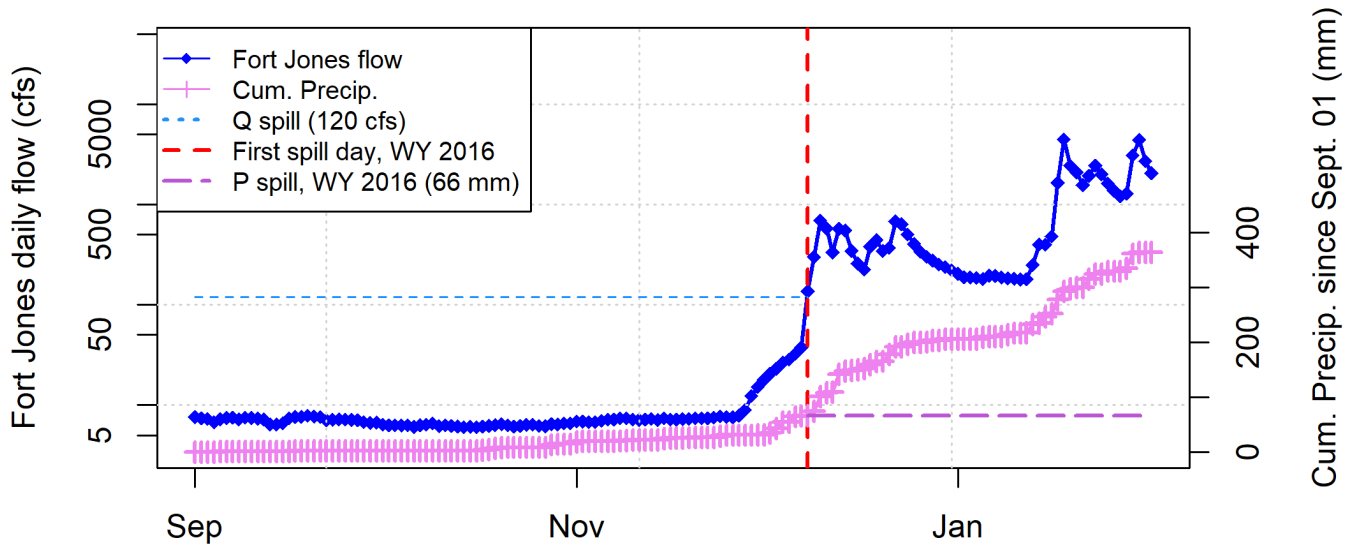
Figure A2. This analysis (Panel A) aimed to identify the flow threshold that approximately defines the boundary between filling (i.e. dry season) and spilling (i.e. wet season) flow behavior at the Fort Jones gauge. For each threshold value, for each water year, a rainfall-runoff response was calculated before and after the flow threshold. The rainfall-runoff response consisted of the 30-day cumulative flow difference (dQ) per 30-day cumulative rainfall difference (dP). 120 cfs was selected as the threshold value dividing the dry and wet seasons (e.g., Panel B).

 ## A.3   Model selection criteria - V_{min}

Diagnostics used to select the predictive models for V_{min} are shown below and discussed in Results. Predictors are abbreviated in tables and described briefly in Table A1; for more information on potential predictors see Section 2.3.

Table A1. Linear model diagnostics for one-predictor models of minimum fall flows (V min).

Predictor ID	Predictor Descrip.	n	Log Likelihood	AIC	LOOCV	R squared
SWJ_max_wc_mm	Snow maximum	70	-131	269	2.7	0.53
USC00043182_oct_apr_mm	Oct.-Apr. Precip.	75	-142	290	2.7	0.49
SWJ_jday_of_max	Snow maximum timing	70	-154	314	5.1	0.11
springWL_415635N1228315W001	March-May WLs	50	-100	206	3.6	0.39
et0_oct_apr	ET Ref.	17	-21	48	0.9	0.46
mar_flow	March flow vol.	78	-162	330	4.1	0.25

Table A2. Linear model diagnostics for two-predictor models of V min. See table of one-predictor models for description of predictor IDs. Reference ET was not included in two- and three-predictor models due to an insufficient sample size.

Predictor 1	Predictor 2	n	Log Likelihood	AIC	LOOCV	R squared
SWJ_max_wc_mm	USC00043182_oct_apr_mm	67	-119	246	2.3	0.62
SWJ_max_wc_mm	SWJ_jday_of_max	70	-131	270	2.8	0.53
SWJ_max_wc_mm	springWL_415635N1228315W001	50	-91	191	2.7	0.57
SWJ_max_wc_mm	mar_flow	70	-128	264	2.6	0.57
USC00043182_oct_apr_mm	SWJ_jday_of_max	67	-127	263	2.9	0.52
USC00043182_oct_apr_mm	springWL_415635N1228315W001	47	-91	190	3.3	0.48
USC00043182_oct_apr_mm	mar_flow	75	-139	286	2.8	0.52
SWJ_jday_of_max	springWL_415635N1228315W001	50	-97	203	3.3	0.45
SWJ_jday_of_max	mar_flow	70	-141	291	3.8	0.37
springWL_415635N1228315W001	mar_flow	50	-94	197	3.1	0.51
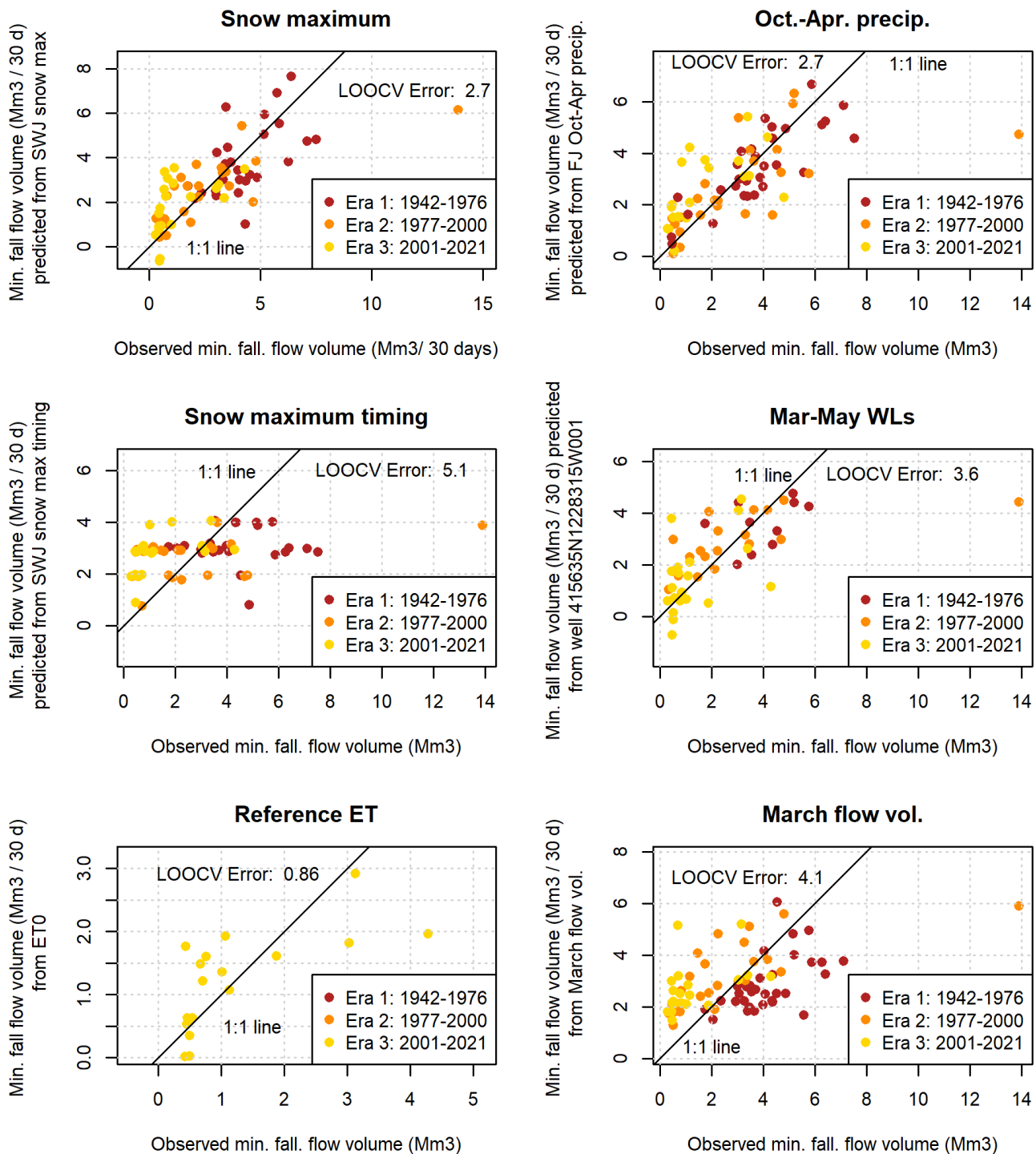
Figure A3. Single-predictor models of minimum 30-day dry season baseflows in the Scott River.
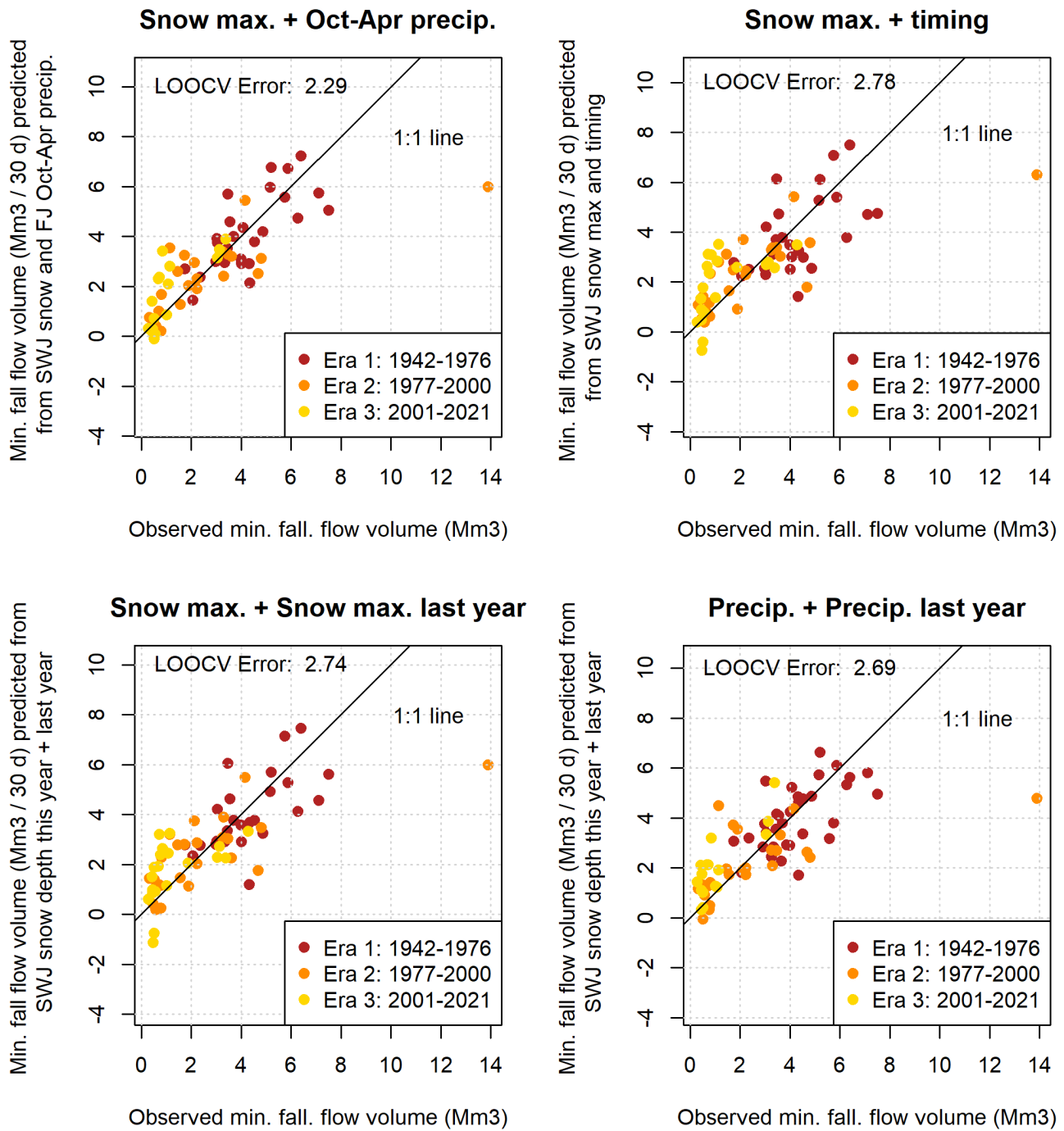
Figure A4. Two-predictor models of minimum 30-day dry season baseflows in the Scott River.
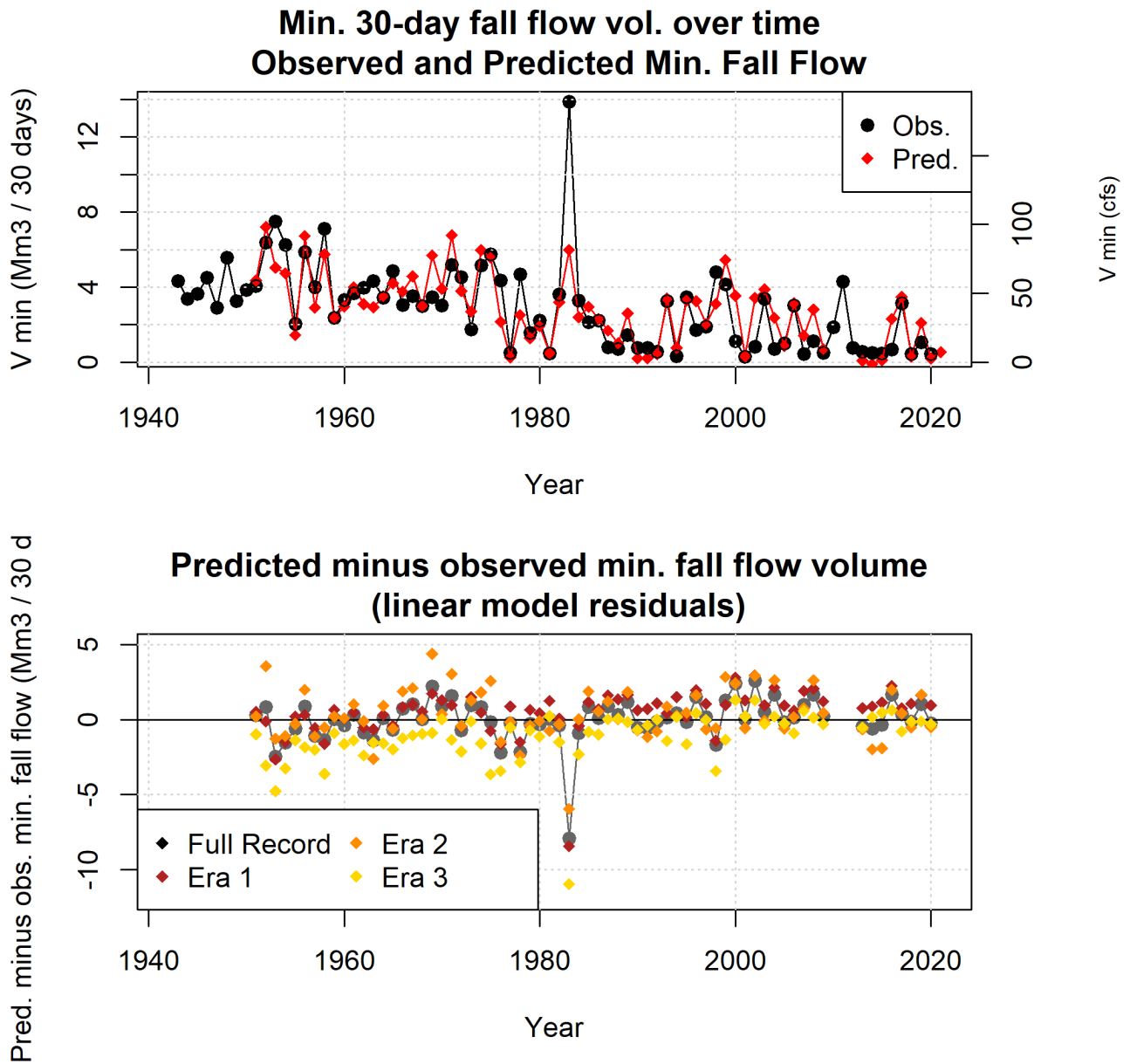
Figure A5. Observed and predicted minimum 30-day dry season baseflows both trend downward between the three eras of the period of record (top panel). The predicted-minus-observed difference (residual) over time also reflects this trend, underpredicting minimum flows pre-1977 and overpredicting them post-2000 (middle panel). The predictive model is based on observations from the full record, but three additional models were generated based on only the observations from Eras 1, 2, and 3. Residuals based on Era 1 data are similar to those of the full record; Era 2 residuals tend to overpredict more than the full record; and Era 3 residuals show better performance post-2000 than the full record, but significant underprediction pre-2000.

Table A3. Linear model diagnostics for three-predictor models of minimum fall flows (V min). See table of one-predictor models for description of predictor IDs. Reference ET was not included in two- and three-predictor models due to an insufficient sample size.

Predictor 1	Predictor 2	Predictor 3	n	Log Like.	AIC	LOOCV	R squared
SWJ_max_wc_mm	USC00043182_oct_apr_mm	SWJ_jday_of_max	67	-119	247	2.3	0.63
SWJ_max_wc_mm	USC00043182_oct_apr_mm	springWL_415635N1228315W001	47	-86	181	2.8	0.59
SWJ_max_wc_mm	USC00043182_oct_apr_mm	mar_flow	67	-117	244	2.3	0.64
SWJ_max_wc_mm	SWJ_jday_of_max	springWL_415635N1228315W001	50	-91	192	2.7	0.58
SWJ_max_wc_mm	SWJ_jday_of_max	mar_flow	70	-127	265	2.7	0.58
SWJ_max_wc_mm	springWL_415635N1228315W001	mar_flow	50	-89	187	2.6	0.61
USC00043182_oct_apr_mm	SWJ_jday_of_max	springWL_415635N1228315W001	47	-90	190	3.3	0.51
USC00043182_oct_apr_mm	SWJ_jday_of_max	mar_flow	67	-123	256	2.8	0.58
USC00043182_oct_apr_mm	springWL_415635N1228315W001	mar_flow	47	-87	184	3.0	0.56
SWJ_jday_of_max	springWL_415635N1228315W001	mar_flow	50	-91	191	2.9	0.58

A.4 Model selection criteria - P_{spill}

Diagnostics used to select the predictive models for P_{spill} are shown below and discussed in Results. Predictors are abbreviated in tables and described briefly in Table A4; for more information on potential predictors see Section 2.3.

Table A4. Linear model diagnostics for one-predictor models of P spill.

Predictor ID	Predictor Descrip.	n	Log Likelihood	AIC	LOOCV	R squared
SWJ_max_wc_mm	Snow maximum	70	-333	673	850	0.38
USC00043182_oct_apr_mm	Oct.-Apr. Precip.	75	-351	708	718	0.43
SWJ_jday_of_max	Snow maximum timing	70	-347	699	1243	0.09
springWL_415635N1228315W001	March-May WLs	50	-245	495	1123	0.24
et0_oct_apr	ET Ref.	17	-81	167	932	0.23
mar_flow	March flow vol.	78	-380	767	1061	0.14
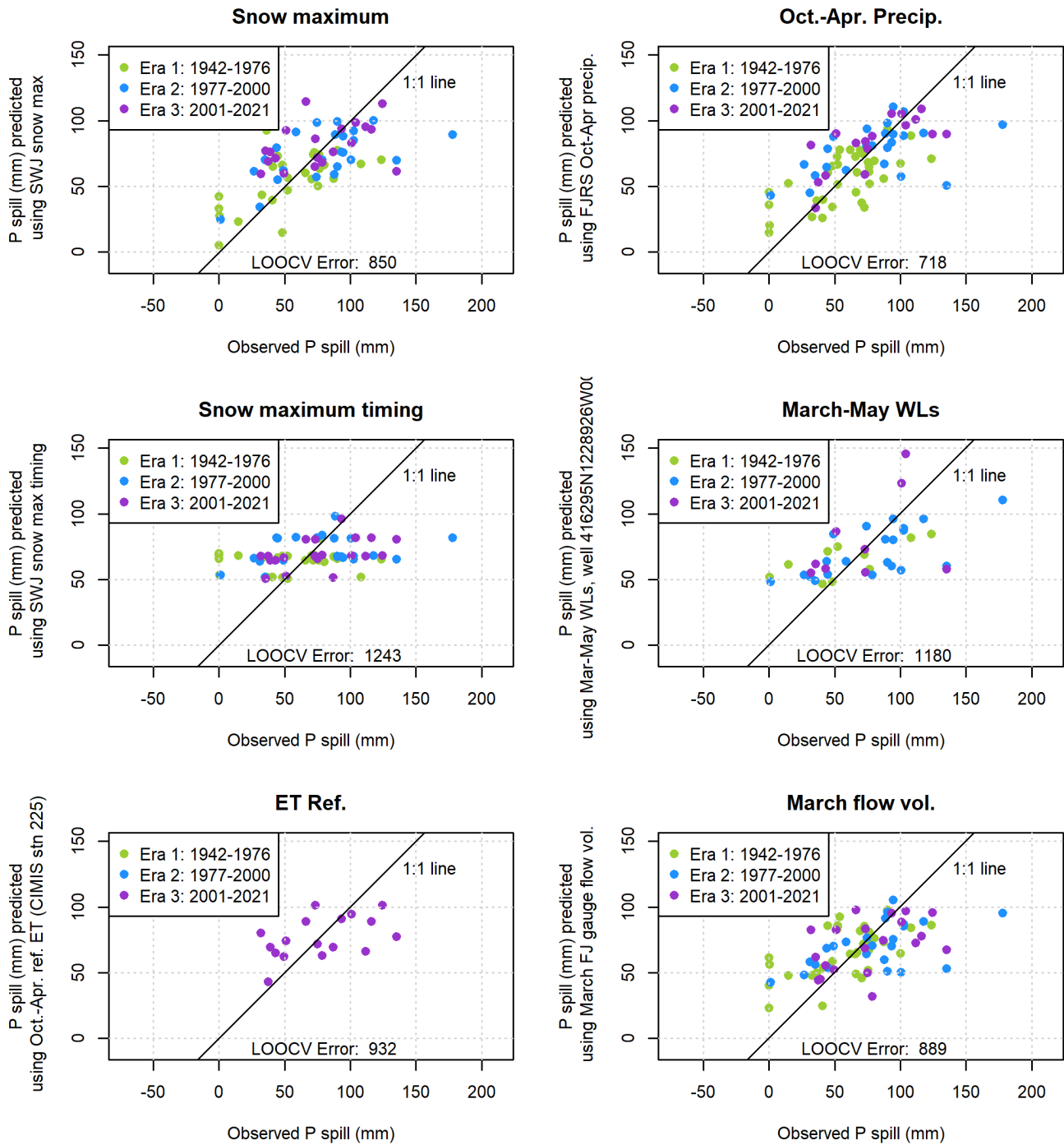
Figure A6. Single-predictor models of P spill, the cumulative precipitation after the dry season needed to generate 120 cfs of flow in the Scott River.
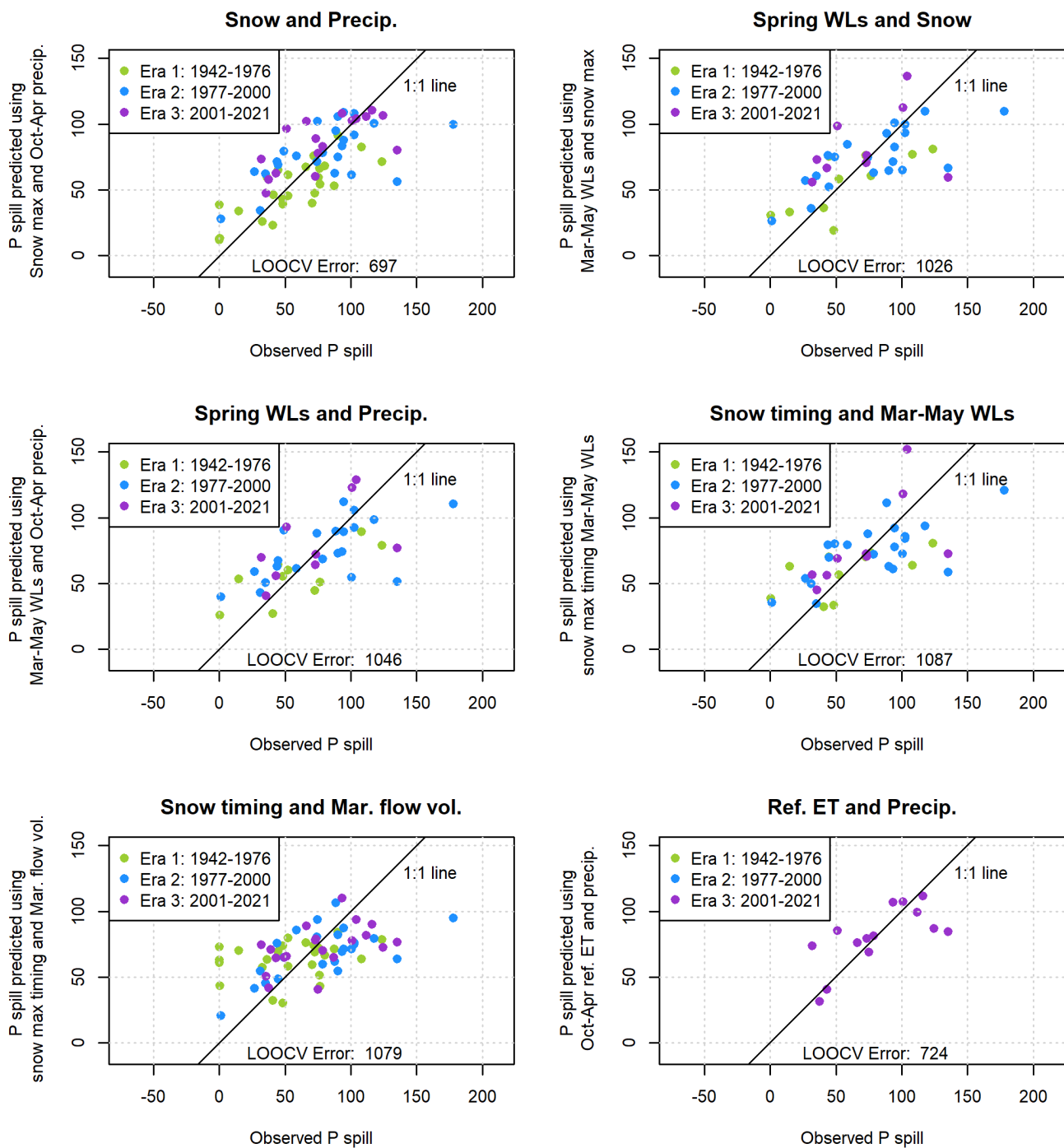
Figure A7. Two-predictor models of P spill, the cumulative precipitation after the dry season needed to generate 120 cfs of flow in the Scott River.
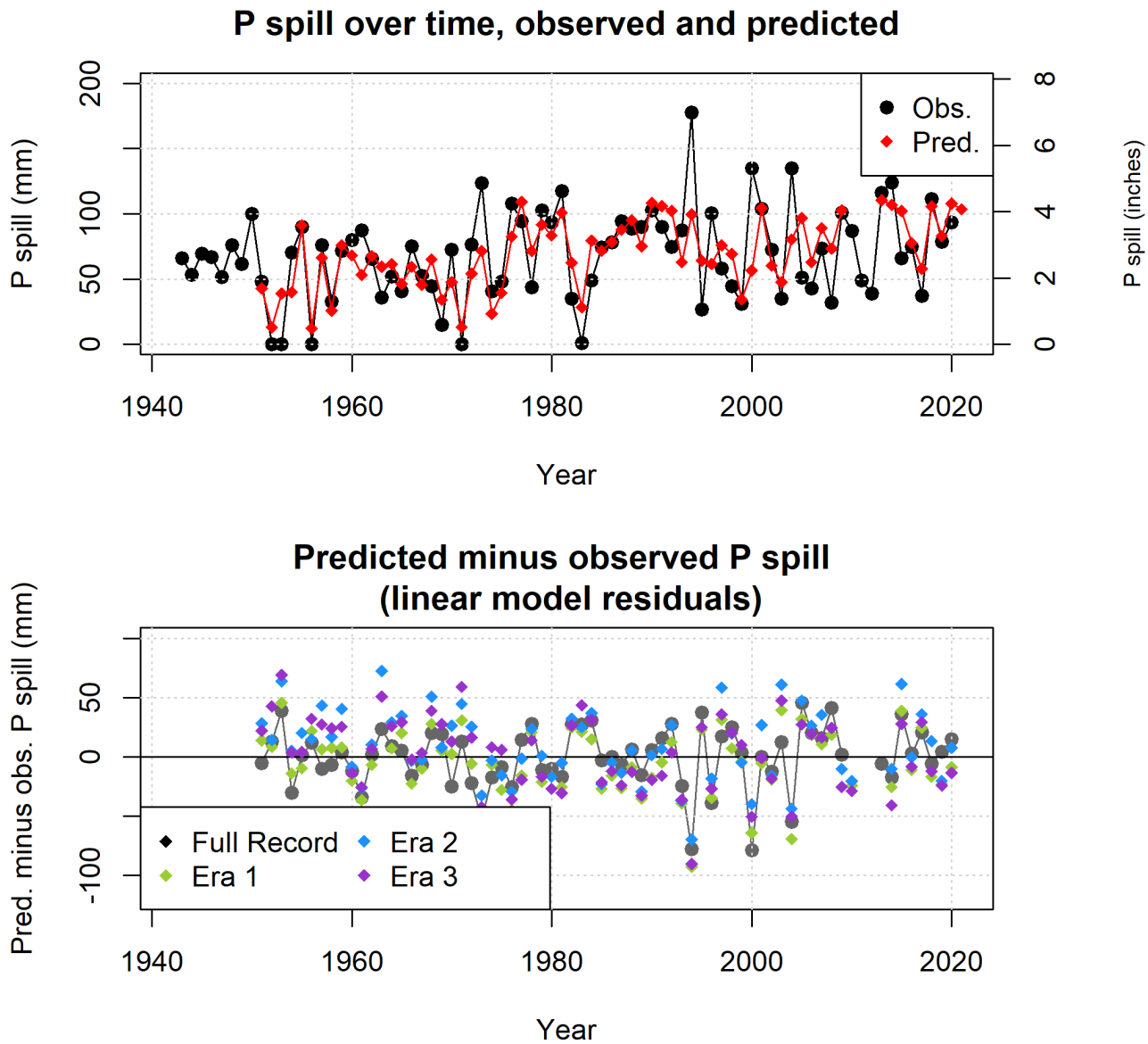
Figure A8. Observed and predicted values of P spill (panel A) indicate a worse model fit for the P spill prediction than for minimum 30-day dry season baseflows. Serious overprediction in Era 1 is followed by more mixed over- and under-prediction in Eras 2 and 3 (bottom panel). The overall P spill model is based on observations from the full record, but three additional models were generated based on only the observations from Eras 1, 2, and 3. Residuals based on Era 1 data are generally lower than those from Eras 2 or 3 or from the full record.

Table A5. Linear model diagnostics for two-predictor models of P spill. See table of one-predictor models for description of predictor IDs. Reference ET was not included in two- and three-predictor models due to an insufficient sample size.

Predictor 1	Predictor 2	n	Log Likelihood	AIC	LOOCV	R squared
SWJ_max_wc_mm	USC00043182_oct_apr_mm	67	-312	631	697	0.51
SWJ_max_wc_mm	SWJ_jday_of_max	70	-333	674	870	0.39
SWJ_max_wc_mm	springWL_415635N1228315W001	50	-240	487	951	0.38
SWJ_max_wc_mm	mar_flow	70	-333	674	853	0.39
USC00043182_oct_apr_mm	SWJ_jday_of_max	67	-315	637	760	0.47
USC00043182_oct_apr_mm	springWL_415635N1228315W001	47	-224	457	920	0.43
USC00043182_oct_apr_mm	mar_flow	75	-351	709	728	0.44
SWJ_jday_of_max	springWL_415635N1228315W001	50	-243	493	1068	0.30
SWJ_jday_of_max	mar_flow	70	-341	690	1079	0.23
springWL_415635N1228315W001	mar_flow	50	-242	493	1067	0.31

Table A6. Linear model diagnostics for three-predictor models of P spill. See table of one-predictor models for description of predictor IDs. Reference ET was not included in two- and three-predictor models due to an insufficient sample size.

Predictor 1	Predictor 2	Predictor 3	n	Log Like.	AIC	LOOCV	R squared
SWJ_max_wc_mm	USC00043182_oct_apr_mm	SWJ_jday_of_max	67	-311	633	712	0.52
SWJ_max_wc_mm	USC00043182_oct_apr_mm	springWL_415635N1228315W001	47	-222	455	874	0.47
SWJ_max_wc_mm	USC00043182_oct_apr_mm	mar_flow	67	-312	633	714	0.51
SWJ_max_wc_mm	SWJ_jday_of_max	springWL_415635N1228315W001	50	-239	488	973	0.39
SWJ_max_wc_mm	SWJ_jday_of_max	mar_flow	70	-332	675	872	0.40
SWJ_max_wc_mm	springWL_415635N1228315W001	mar_flow	50	-239	488	955	0.40
USC00043182_oct_apr_mm	SWJ_jday_of_max	springWL_415635N1228315W001	47	-224	457	936	0.44
USC00043182_oct_apr_mm	SWJ_jday_of_max	mar_flow	67	-314	638	768	0.48
USC00043182_oct_apr_mm	springWL_415635N1228315W001	mar_flow	47	-223	457	923	0.45
SWJ_jday_of_max	springWL_415635N1228315W001	mar_flow	50	-240	490	1008	0.37

A.5 Diagnostic plots for selected models

Standard diagnostic plots for the selected predictive models for V_{min} and P_{spill}. For V_{min}, these diagnostic plots highlight outlier record 42, which corresponds to the year 1983, when an early September storm followed a wet year. For P_{spill}, three lesser outliers are highlighted, corresponding to water years 1994, 2000 and 2004, in which the three highest values of P_{spill} were observed. These outliers represent the basin in extreme hydrologic conditions, so are retained in the dataset even though they exert disproportionate leverage over the predictive models.
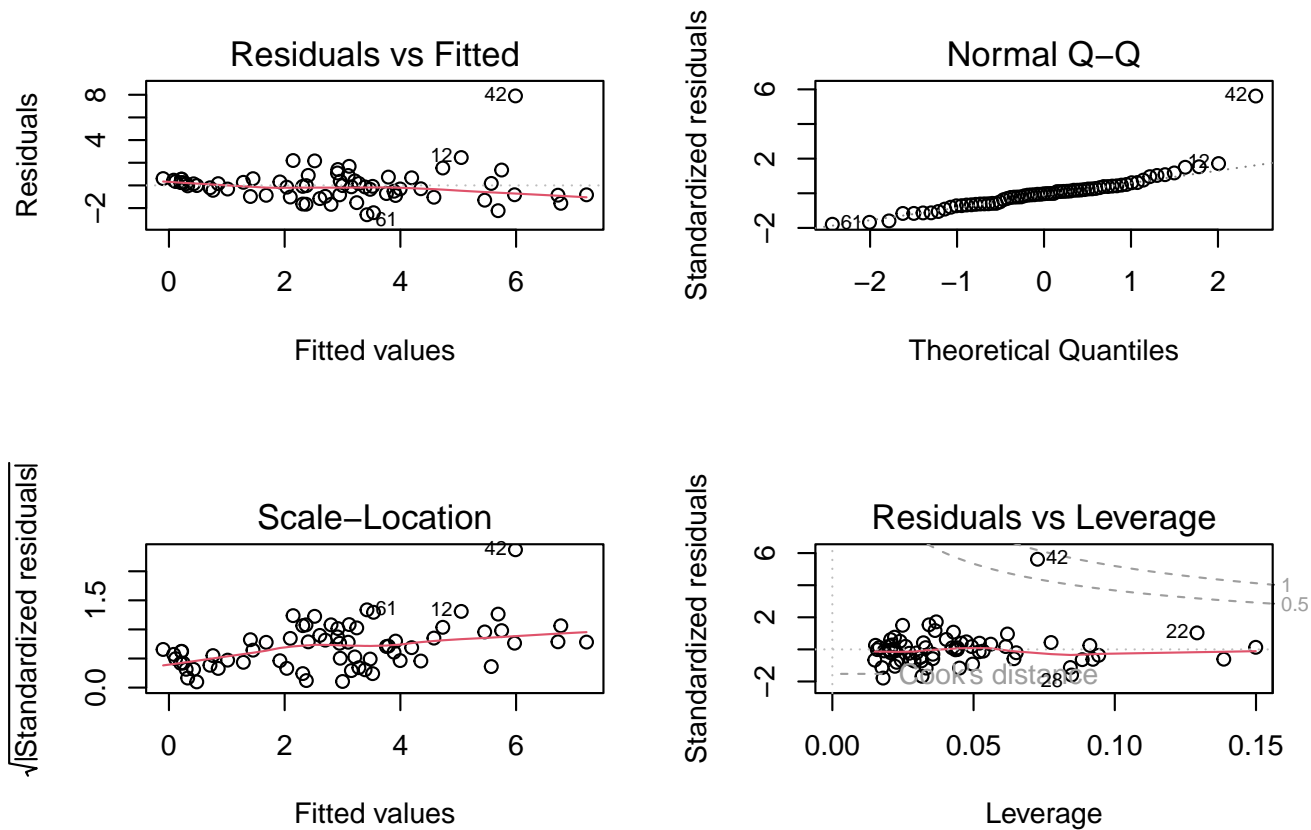
Figure A9. Diagnostic plots for the selected predictive models of V min.

. Analyses and figures in this manuscript were drafted in RMarkdown. The RMarkdown scripts are available on the corresponding author's GitHub page. All data used in this manuscript are publicly available on local, state or federal data portals.

. The authors declare no competing interests.

565 . This manuscript emerged from dissertation work funded by Siskiyou County SGMA planning grants, with funding from California water bonds.
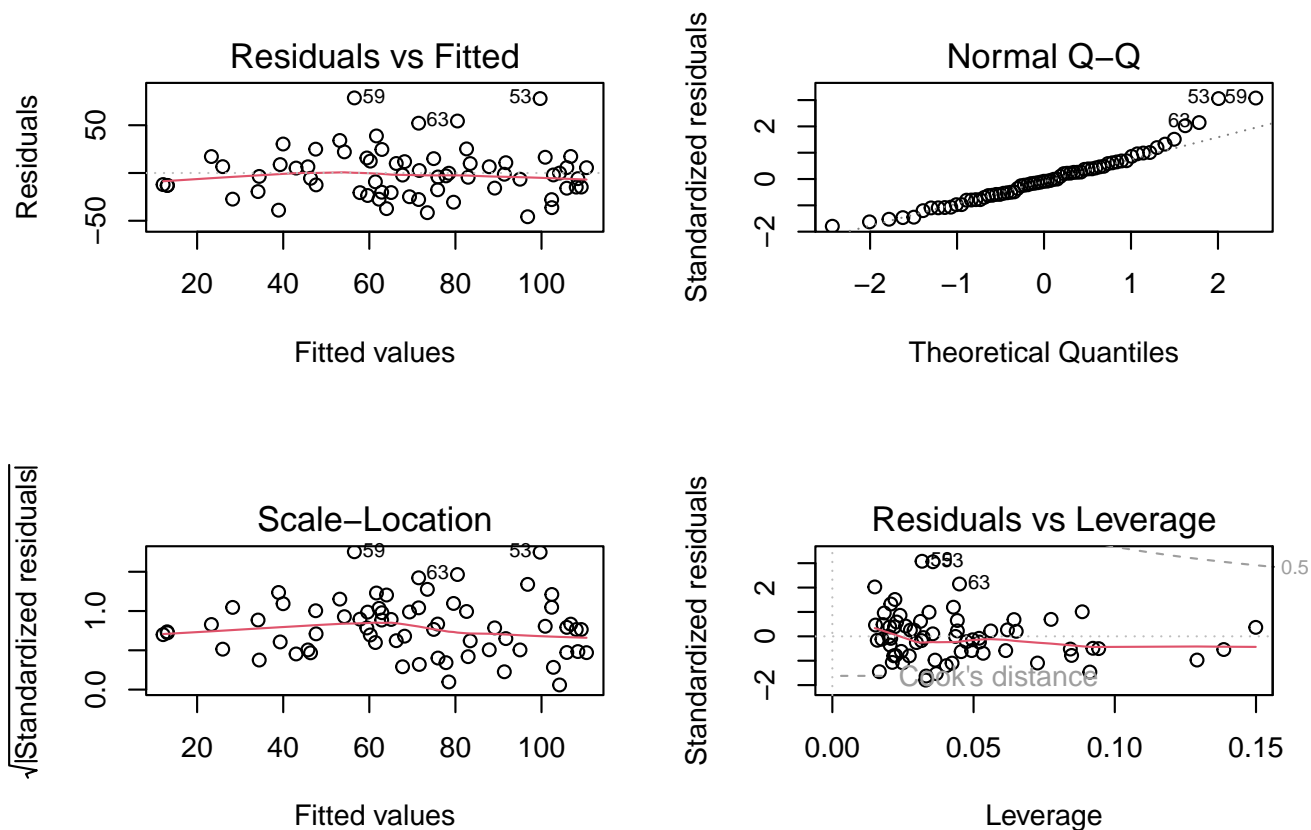
Figure A10. Diagnostic plots for the selected predictive models of P spill.

References

Bunn, S. E. and Arthington, A. H.: Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, Environmental Management, 30, 492–507, https://doi.org/10.1007/s00267-002-2737-0, 2002.

570 Burnham, K. P. and Anderson, D. R.: Multimodel inference: A Practical Information-Theoretic Approach, 2004.

California Department of Fish and Wildlife (CDFW): Cooperative Report of the Scott River Coho Salmon Rescue and Relocation Effort: 2014 Drought Emergency, Tech. Rep. August, California Department of Fish and Wildlife (CDFW), https://www.fs.usda.gov/Internet/ FSE{_}DOCUMENTS/stelprd3850544.pdf, 2015.

California Department of Water Resources (DWR): Bulletin 118: Scott River Valley Groundwater Basin, Tech. rep., https://water.ca.gov/

575 LegacyFiles/pubs/groundwater/bulletin{_}118/basindescriptions/1-5.pdf, 2004.

California Department of Water Resources (DWR): Sustainable Groundwater Management Act - Water Year Type Dataset Development Report, Tech. Rep. January, Department of Water Resources - Sustainable Groundwater Management Office, 2021.

California Department of Water Resources (DWR): Notice to State Water Project Contractors: 2022 State Water Project Table A Allocation Decrease from 15 to 5 Percent, https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/State-Water-Project/Management/ SWP-Water-Contractors/Files/22-03-2022-SWP-Allocation-Decrease-5-Percent-031822.pdf, 2022.

California Department of Water Resources (DWR): State Water Project, https://water.ca.gov/programs/state-water-project, 2023.

Colorado Division of Water Resources: Drought & Surface Water Supply Index, https://dwr.colorado.gov/services/water-administration/ drought-and-swsi, 2023.

Foglia, L., McNally, A., Hall, C., Ledesma, L., Hines, R., and Harter, T.: Scott Valley Integrated Hydrologic Model : Data Collection , Analysis , and Water Budget, Tech. Rep. April, University of California, Davis, http://groundwater.ucdavis.edu/files/165395.pdf, 2013a.

Foglia, L., McNally, A., and Harter, T.: Coupling a spatiotemporally distributed soil water budget with stream-depletion functions to inform stakeholder-driven management of groundwater-dependent ecosystems, Water Resources Research, 49, 7292–7310, https://doi.org/10.1002/wrcr.20555, 2013b.

Foglia, L., Neumann, J., Tolley, D. G., Orloff, S. B., Snyder, R. L., and Harter, T.: Modeling guides groundwater management in a basin with river–aquifer interactions, California Agriculture, 72, 84–95, 2018.

Francis, R. C., Hare, S. R., Hollowed, A. B., and Wooster, W. S.: Effects of interdecadal climate variability on the oceanic ecosystems of the NE Pacific, Fisheries Oceanography, 7, 1–21, 1998.

Garen, D. C.: REVISED SURFACE-WATER SUPPLY INDEX FOR WESTERN UNITED STATES, Journal of Water Resources Planning and Management, 119, 437–454, https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/State-Water-Project/Management/ SWP-Water-Contractors/Files/22-03-2022-SWP-Allocation-Decrease-5-Percent-031822.pdf, 1993.

Gorman, M. P.: Juvenile survival and adult return as a functional of freshwater rearing life history for coho salmon in the Klamath River basin, Master's thesis, Humboldt State University, https://digitalcommons.humboldt.edu/cgi/viewcontent.cgi?article=1005{&}context= etd, 2016.

Guttman, N. B.: COMPARING THE PALMER DROUGHT INDEX AND THE STANDARDIZED PRECIPITATION INDEX ' ties of the PDSI and its variations have been the referenced studies show that the intended, Journal Of The American Water Resources Association, 34, 113–121, https://doi.org/10.1111/j.1752-1688.1998.tb05964.x, 1998.

Harter, T. and Hines, R.: Scott Valley Community Groundwater Study Plan, Tech. rep., Groundwater Cooperative Extension Program University of California, Davis, Davis. CA, http://groundwater.ucdavis.edu/files/136426.pdf, 2008.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning, Springer Science+Business Media, New York, 7th edn., https://doi.org/10.1007/978-1-4614-7138-7, 2013.

Mack, S.: Geology and Ground-Water Features of Scott Valley Siskiyou County, California., Tech. rep., Geological Survey Water-Supply Paper 1462, https://pubs.usgs.gov/wsp/1462/report.pdf, 1958.

Mahoney, J. M. and Rood, S. B.: Streamflow requirements for cottonwood seedling recruitment-An integrative model, Wetlands, 18, 634–645, https://doi.org/10.1007/BF03161678, 1998.

McDonnell, J. J., Spence, C., Karran, D. J., van Meerveld, H. J., and Harman, C. J.: Fill-and-Spill: A Process Description of Runoff Generation at the Scale of the Beholder, Water Resources Research, 57, 1–13, https://doi.org/10.1029/2020WR027514, 2021.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: Proceedings of the Eighth Conference on Applied Climatology, January, p. 5, Anaheim, California, 1993.

Moyle, P.: Coho Salmon, Oncorhynchus kisutch (Walbaum), in: Inland Fishes of California, pp. 245–251, 2002.

615 Natural Resources Conservation Service (NRCS): Idaho Water Supply Outlook Report, Tech. rep., USDA, Boise, ID, https://www.wcc.nrcs. usda.gov/ftpref/support/states/ID/wy2023/wsor/borid623.pdf, 2023.

Null, S. E. and Viers, J. H.: In bad waters: Water year classification in nonstationary climates, Water Resources Research, 49, 1137–1148, https://doi.org/10.1002/wrcr.20097, 2013.

Palmer, W. C.: Meteolorogical Drought, Research Paper No. 45, Tech. rep., US Weather Bureau, Washington, D.C., 1965.

620 Patterson, N. K., Lane, B. A., Yarnell, S. M., Qiu, Y., Sandoval-Solis, S., and Pasternack, G. B.: A hydrologic feature detection algorithm to quantify seasonal components of flow regimes, Journal of Hydrology, 585, 2020.

Peek, R., Irving, K., Yarnell, S. M., Lusardi, R., Stein, E. D., and Mazor, R.: Identifying Functional Flow Linkages Between Stream Alteration and Biological Stream Condition Indices Across California, Frontiers in Environmental Science, 9, 1–14, https://doi.org/10.3389/fenvs.2021.790667, 2022.

625 Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C.: A paradigm for river conservation and restoration, BioScience, 47, 769–784, https://doi.org/10.2307/1313099, 1997.

Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C., Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., O'Keeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E., and Warner, A.: The ecological limits of hydrologic alteration (ELOHA): A new framework for developing regional environmental flow standards, Freshwater

630 Biology, 55, 147–170, https://doi.org/10.1111/j.1365-2427.2009.02204.x, 2010.

Pyschik, J.: Assessing Climate Impacts Against Groundwater Pumping Impacts on Stream Flow with Statistical Analysis, Master's, Albert Ludwigs University Freiburg, 2022.

Scott River Water Trust (SRWT): 2017 Monitoring Report, Tech. Rep. June, https://www.scottwatertrust.org/blank, 2018.

Siskiyou County Flood Control and Water Conservation District: Scott Valley Groundwater Sustainability Plan, Tech. rep., Siskiyou County

635 Flood Control and Water Conservation District, https://www.co.siskiyou.ca.us/naturalresources/page/scott-valley-gsp-chapters, 2021.

Sommarstrom, S.: Email communication regarding connectivity of Scott River tailings reach, Nov. 18, 2020, 2020.

Tarboton, D. G.: Rainfall-Runoff Processes., http://www.engineering.usu.edu/dtarb/rrp.html, 2003.

Tolley, D. G., Foglia, L., and Harter, T.: Sensitivity Analysis and Calibration of an Integrated Hydrologic Model in an Irrigated Agricultural Basin with a Groundwater-Dependent Ecosystem, Water Resources Research, 55, https://doi.org/10.1029/2018WR024209, 2019.

640 Tromp-Van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resources Research, 42, 1–11, https://doi.org/10.1029/2004WR003800, 2006.

Van Kirk, R. W. and Naman, S. W.: Relative effects of climate and water use on base-flow trends in the lower Klamath Basin, Journal of the American Water Resources Association, 44, 1035–1052, https://doi.org/10.1111/j.1752-1688.2008.00212.x, 2008.

Verley, F.: Lessons from Twenty Years of Local Volumetric Groundwater Management: The Case of the Beauce Aquifer, Central France, in:

645 Sustainable Groundwater Management: A Comparative Analysis of French and Australian Policies and Implications to Other Countries, edited by Rinaudo, J.-D., Holley, C., Barnett, S., and Montginoul, M., pp. 93–108, Springer Nature Switzerland AG, Cham, Switzerland, https://doi.org/10.1142/s2382624x22800017, 2020.

Wheeler, K., Wenger, S. J., and Freeman, M. C.: States and rates: Complementary approaches to developing flow-ecology relationships, Freshwater Biology, 63, 906–916, https://doi.org/10.1111/fwb.13001, 2018.

650 Wilhite, D. A. and Glantz, M. H.: Understanding: The drought phenomenon: The role of definitions, Water International, 10, 111–120, https://doi.org/10.1080/02508068508686328, 1985.

Wilhite, D. A., Hayes, M. J., and Svoboda, M.: Drought Monitoring and Assessment : Status and Trends in the United States, Drought
Mitigation Center Faculty Publications, 76, http://digitalcommons.unl.edu/droughtfacpub/76, 2000.

Williams, A., Cook, E., Smerdon, J., Cook, B., Abatzoglou, J., Bolles, K., Baek, S., Badger, A., and Livne, B.: Large contribution from
anthropogenic warming to an emerging North American megadrought, Science, In press, 314–318, 2020.

Yarnell, S. M., Stein, E. D., Webb, J. A., Grantham, T., Lusardi, R. A., Zimmerman, J., Peek, R. A., Lane, B. A., Howard, J., and Sandoval-
Solis, S.: A functional flows approach to selecting ecologically relevant flow metrics for environmental flow applications, River Research
and Applications, 36, 318–324, https://doi.org/10.1002/rra.3575, 2020.

655