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

Claire Kouba1 and Thomas Harter1

1Department 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, t wo key hydrologic metrics in the undammed rural Scott River watershed (HUC8 18010208) in northern California. Both metrics are intended to quantify the transition from the dry to the wet season, to characterize the severity of a dry year and support seasonal adaptive management. The first metric is the minimum 30-day dry season baseflow volume, *Vmin*, 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 *Pspill*. As potential predictors of these two values, 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. We find that, though many of these predictors are correlated with the two metrics of interest, of the predictors considered here, 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 *Vmin* and *Pspill* with an average model error (RMSE) of 1.4 Mm3 / 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.

*.* The authors retain copyright for this publication.

## Introduction

In regions that experience periodic drought, such as the western United States, indices summarizing hydroclimate or surface water supply conditions are often critical decision-support tools for water managers (e.g., [Garen,](#_bookmark34) [1993).](#_bookmark34) An index can be forward-looking, such as those that forecast near-term seasonal water supplies (e.g., [Null and Viers,](#_bookmark45) [2013;](#_bookmark45) [Verley,](#_bookmark57) [2020),](#_bookmark57) or backward-looking, such as ones that evaluate drought severity (e.g., [Palmer,](#_bookmark46) [1965;](#_bookmark46) [Guttman,](#_bookmark36) [1998;](#_bookmark36) [McKee et](#_bookmark40) al., [1993;](#_bookmark40) [Wilhite](#_bookmark58) [and Glantz,](#_bookmark58) [1985;](#_bookmark58) [Wilhite et al.,](#_bookmark59) [2000).](#_bookmark59) In many western states, forward-looking seasonal indices are used extensively by water agencies to inform adaptive management decisions, e.g. in Colorado (CDWR [2023),](#_bookmark28) Idaho (NRCS [2023)](#_bookmark43) and California [(Null and Viers,](#_bookmark45) [2013).](#_bookmark45) 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 through the State Water Project (DWR [2022).](#_bookmark29) The state has more recently published a retroactive categorical water year type (WYT) dataset for sub-county level regions throughout California (DWR [2021).](#_bookmark26)

These summary indices provide broad characterizations of a water year type. 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.,](#_bookmark47) [1997;](#_bookmark47) [Bunn and Arthington,](#_bookmark23) [2002;](#_bookmark23) [Poff et al.,](#_bookmark49) [2010;](#_bookmark49) [Wheeler et al.,](#_bookmark60) [2018).](#_bookmark60) The flows are “functional” because they serve an ecological purpose, such as wet season flood flows, needed to disperse cottonwood seeds [(Mahoney and Rood,](#_bookmark41) [1998)](#_bookmark41) and fall pulse flows, needed to provide passage for spawning fall-run anadromous fish [(Moyle,](#_bookmark44) [2002).](#_bookmark44) A California-specific functional flows framework has been developed to assess the degree of hydrologic alteration between current and unimpaired conditions [(Yarnell et al.,](#_bookmark61) [2020;](#_bookmark61) [Patterson et al.,](#_bookmark48) [2020).](#_bookmark48)

In this study, to test the utility of locally-tailored predictive methods for hydrologic indices, we focus on a single basin, the Scott River watershed in northern California. Three periods of water use and climate forces have been proposed for the Scott River (e.g., by [Pyschik,](#_bookmark50) [2022):](#_bookmark50) 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.,](#_bookmark53) [2019);](#_bookmark53) and climate conditions, such as the shift in the Pacific Decadal Oscillation in the late 1970s [(Francis et al.,](#_bookmark32) [1998)](#_bookmark32) or the onset of a two-decade abnormally dry period in 2000 [(Williams et al.,](#_bookmark62) [2020).](#_bookmark62) 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 first review the hydrologic indices and methods currently used in decision-making, such as agricultural cropping choices or regulatory water use restrictions, and propose two additional 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 *Vmin*, 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 clear runoff, or storm surge, events. This cumulative precipitation depth is referred to as *Pspill*. Both of these metrics have significance for environmental flows and 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 oversummering juvenile salmonids [(Gorman,](#_bookmark35) [2016).](#_bookmark35) *Pspill*, meanwhile, is intuitively related to the 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).](#_bookmark24) After defining and developing seasonal predictions for *Vmin* and *Pspill*, 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.

## 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 (which in the U.S. Geological Survey [USGS] National Hydrography Dataset is denoted with the 8-digit Hydrologic Unit Code [HUC8] 18010208) [(Figure](#_bookmark0) 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](#_bookmark1) 2). Annual demand for agricultural and domestic use (estimated at 23 and 1.3 thousand acre-feet, respectively) (DWR [2004)](#_bookmark27) are relatively reliable in the Scott River system (although some reports of dry wells occur in dry years) (Siskiyou County [2021),](#_bookmark51) and 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).](#_bookmark51) Post-1970s minimum dry season baseflows have been lower than pre-1977, and very low minima (< 10 cfs or 0.7 Mm3 / 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 (*Pspill*). 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, to capture uncommon September precipitation, cumulative precipitation is counted starting on September 1st of the preceding water year. 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](#_bookmark0) 1), with an observation record covering water years 1942-2021.

## Scott River watershed precipitation-runoff behavior

To establish the context and meaning of the two proposed predictive indices *Vmin* and *Pspill*, 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 the volume of water in aquifer storage [(Tarboton,](#_bookmark54) [2003).](#_bookmark54) At a hillslope scale, in areas where soil directly overlays (relatively) impermeable bedrock and aquifer storage is not appreciably present, a threshold runoff response to individual storm events has been observed: after a certain quantity of rainfall, subsurface flow increases dramatically [(Tromp-Van Meerveld and](#_bookmark55) [McDonnell,](#_bookmark55) [2006).](#_bookmark55) 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,](#_bookmark55) [2006).](#_bookmark55) 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.,](#_bookmark42) [2021).](#_bookmark42)

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 observed at a regional 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](#_bookmark2) 1).

The term “hydrologically connected reservoirs” is defined here functionally as storage that is sufficiently connected to the stream system as to influence stream reactions to storm inflows on an hourly or daily timescale. In this study area only soil water and groundwater meet this definition. More generally, water in the Scott River watershed is stored in five primary reservoirs [(Harter and Hines,](#_bookmark37) [2008):](#_bookmark37)

* snowpack
* fractures in impermeable bedrock
* soil moisture/subflow
* the alluvial groundwater aquifer
* streams and surface water bodies

Accumulating snowpack is present only in the mountainous areas of the upper watershed, while the alluvial aquifer is present only within the bounds of the groundwater basin underlying the flat valley floor; water stored in fractured rock emerges as springs in the upper watershed [(Mack,](#_bookmark38) [1958)](#_bookmark38) [(Figure](#_bookmark0) 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.

## Rainfall-runoff response, functional flows and *Qspill*

At the end of the dry season, the watershed is in a “draining from low storage” condition, which is reflected in a slowly declining or flat hydrograph, with a flowrate that has decreased for several months [(Figure 2,](#_bookmark1) 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

**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

Late spring recession and dry season baseflow

Low Yes (B) Watershed filling from a low

storage level, with muted response to new precipitation

Fall pulse flow or small/slow post-dry- season flow increase

High No (C) Watershed draining from a high

storage level

Winter baseflow and early spring recession

High Yes (D) Watershed spilling from a high

storage level, with rapid response to new precipitation

Winter stormflow

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,](#_bookmark1) 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 storage baseline than during the dry season, and with a higher draining flowrate [(Figure 2,](#_bookmark1) first period C). If there is additional precipitation, the resulting surge in flow response is much more rapid, reflecting a “spilling” condition [(Figure 2,](#_bookmark1) 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,](#_bookmark1) second period C and second period A).

Many of the phenomena described in the above paragraphs have been characterized as various types of functional flow [(Table](#_bookmark2) 1). Winter stormflow is the obvious 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](#_bookmark2) 1). Conversely, the flows classified under “watershed filling from a low storage level” are somewhat ambiguous and dependent on year-to-year conditions, since a discrete fall pulse flow does not occur in every water year, and no distinct metric has been proposed for a more gradual post-dry-season flow increase.

Given the regular behavior observed during the dry season-to-wet season transition in the Fort Jones hydrograph, and the physical structure of this highly inter-connected basin, we expected to find a flowrate threshold at the Fort Jones gauge approximately defining the lower limit of the “full” or “spilling” basin condition (referred to here as *Qspill*; see Results, section 3.1).

## Stream-aquifer interaction

In the groundwater basin portion of the watershed, the alluvial aquifer is the largest storage reservoir [(Mack,](#_bookmark38) [1958).](#_bookmark38) 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 (F[oglia et al.,](#_bookmark33) [2018).](#_bookmark33) Discharge to streams from the alluvial aquifer occurs along the thalweg of the Scott River. In this highly-interconnected system, groundwater discharge in one reach of the river is typically approximately balanced out by infiltration through the streambed to the aquifer, much of it occurring on the upper alluvial fans of the tributary streams (see discussion below).

We used the Scott Valley Integrated Hydrologic Model (SVIHM) [(Tolley et al.,](#_bookmark53) [2019;](#_bookmark53) [Foglia et al.,](#_bookmark30) [2013a,](#_bookmark30) [b)](#_bookmark31) to obtain the estimated volume of water exchanged monthly, in water years 1991-2018, between the surface stream network and the underlying aquifer. All positive fluxes and all negative fluxes (corresponding to gaining and losing stream reaches) were summed independently and then added to create a net value for each month in the simulation period [(Figure 3,](#_bookmark3) 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,](#_bookmark3) panel B).

* 1. **Observed response variables (*Vmin* and *Pspill*)**

The Scott River is an undammed watershed, in which estimates of annual precipitation are an order of magnitude greater than the estimated combined volume of water stored in surface water bodies or aquifers and water pumped or diverted for agriculture [(Tolley et al.,](#_bookmark53) [2019).](#_bookmark53) In this study we tested whether fundamental hydrologic characteristics, specifically dry-season draining behavior and rainfall-runoff response to early wet season cumulative precipitation, can be predicted using observable hydroclimate data. The first step was quantification of the two response variables.

## Dry season baseflow quantities (*Vmin*) and timing

Multiple numerical summaries 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 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, and so 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, for each calendar year, we calculated a rolling 30-day sum of daily flow volumes in the period July-December to identify the 30-day period with the minimum flow, referred to as *Vmin* (see Section 3.2). For consistency, each annual *Vmin* value was assigned to the water year ending in September of that calendar year, even if the minimum flow window included days in October of the following water year.

## Cumulative precipitation *Pspill*

*Pspill* was calculated for each water year as the cumulative rainfall at the end of a dry season, starting September 1st, recorded on the first day that the Fort Jones gauge measured flow greater than *Qspill* (see Results, Section 3.2). As stated above, conceptually, it is the amount of rainfall needed to “fill” the watershed such that it responds rapidly to new precipitation.

A dry season can have negative effects on an aquatic ecosystem if it produces extraordinarily low flows or if it lasts for an extraordinarily long time (e.g., delayed salmon habitat access documented in CDFW [2015).](#_bookmark24) The quantity *Pspill* is correlated with both a lower minimum flow volume and a later river reconnection [(Figure](#_bookmark4) 4). If predicted in advance, a forecasted value of *Pspill* would be an indicator of the risk of a severe dry season.

In addition to the volumetric quantity *Pspill*, 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 *Qspill* 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 *Pspill* is highly variable and, hence, unpredictable.

## Potential predictors and selected formulations

To evaluate candidate predictors of dry season baseflows, Pearson’s correlation coefficient, *R*, was calculated between observed

response variables *Vmin* and *Pspill*, and the following categories of observed predictor data (see Results, Section 3.3):

1. Spring (March-May) water level observations in each of 54 individual wells [(Figure](#_bookmark10) A1).
2. Annual maximum snowpack water content at each individual snow monitoring station at 20 stations aggregated by the California Data Exchange Center (CDEC; [Figure](#_bookmark0) 1).
3. Cumulative precipitation, October-April, at each weather station within the watershed, and five outside the watershed (total of 17 NOAA stations; [Figure](#_bookmark0) 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, 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 interpolated based on observations at neighboring stations (see method in [Foglia et al.,](#_bookmark30) [2013a).](#_bookmark30)
5. The flow volumes observed at the Fort Jones gauge (USGS ID 11519500; [Figure](#_bookmark0) 1) during the preceding March and April.
6. Cumulative reference evapotranspiration (ET0), October-April, using observations from Station No. 225 in the California Information Management Information System (CIMIS) network (2015-2021), or Spatial CIMIS estimates of ET0 at the location of Station 225 (2002-2015) [(Figure](#_bookmark0) 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 *Vmin* and *Pspill***

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.,](#_bookmark39) [2013).](#_bookmark39) 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,](#_bookmark25) [2004).](#_bookmark25) However, AIC and other information criteria methods can only be used when comparing models based on the same dataset [(Burnham and Anderson,](#_bookmark25) [2004).](#_bookmark25) 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](#_bookmark12) through [A6),](#_bookmark22) cross-validation was used in this study as the primary model selection technique.

To predict *Vmin*, 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.,](#_bookmark39) [2013)](#_bookmark39) (see Results, Section 3.3 and [Table](#_bookmark12) 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.

The single predictors with the lowest LOOCV error (other than evapotranspiration of a reference crop, [*ETref* ], 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)](#_bookmark13) for *Vmin*, including two that incorporate data from a full calendar year prior, to test the validity of the DWR Water Year Type index method in this local setting. A similar approach was used to assess two-predictor models for *Pspill*, 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:

*Predictedi* = *Int.* + *mA* ∗ *obsA, i* + *mB* ∗ *obsB, i* (1)

Where:

* *Predictedi* is the predicted value (either *Vmin* or *Pspill*) in calendar year *i* (i.e., at the end of water year *i*).
* *obsA, i, obsB, i* are the observed predictor values in October-April in water year *i*.
* *Int., mA, mB* are the coefficients of the selected linear model.

## Results

* 1. **Scott River precipitation-runoff behavior**

Visual inspection of 80 years of Fort Jones hydrograph behavior during the transition from the dry season to the rainy season

[(Figure 5,](#_bookmark5) 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](#_bookmark1) 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 boundary between these domains, denoted as *Qspill*, was 120 cfs (294 thousand m3/day or approximately 9 Mm3 per month). This value was determined by testing a range of potential *Qspill* values to evaluate rainfall-runoff responses preceding and following the threshold [(Figure](#_bookmark11) A2).

## Observed response variables

* + 1. **Dry season baseflow quantity and timing**

Minimum 30-day dry season baseflow volumes, denoted here as *Vmin*, ranged from 0.3 to 7.5 Mm3 / 30 days, with one outlier value of 13.9 Mm3 / 30 days in 1984, when an early September storm followed a wet year in 1983 [(Figure](#_bookmark6) 6).

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 (minimum of 2.1 Mm3 / 30 days [28.6 average cfs] in 1942-1976 and 0.3 Mm3 / 30 days [4.4 average cfs] in 1977-2021; [Figure 6,](#_bookmark6) 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,](#_bookmark6) panel B).

## Cumulative fall precipitation and watershed response

In some water years prior to the 1980s, the Fort Jones flowrate exceeded *Qspill* on September 1st [(Figure 5,](#_bookmark5) 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 *Pspill*, the cumulative precipitation necessary to reach *Qspill*, is wide (0 to 178 mm, or 0 to 7 inches) [(Figure 6,](#_bookmark6) panel C). Mean *Pspill* values were 57, 79, and 76 mm (2.3, 3.1, and 3 inches) in Eras 1, 2 and 3, respectively.

* 1. **Predictor comparison for *Vmin* and *Pspill***

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).](#_bookmark10)

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 *Vmin* and negatively correlated with *Pspill*. The correlation coefficient, *R*, of these response-predictor relationships range from 0.5 to 0.73 for *Vmin* and from -0.38 to -0.66 for *Pspill* [(Figure 7).](#_bookmark7)

In contrast to these four categories, October-April cumulative ET0 is negatively correlated with *Vmin* and positively correlated

with *Pspill* (*R* of -0.68 and 0.48 for *Vmin* and *Pspill*, respectively). October-April cumulative ET0 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 absolute values of *R* for between ET0 and the two response variables may be due to a small sample size, as all available ET0 observations or estimates were collected in 2002 or later (i.e., in Era 3; [Figure](#_bookmark14) 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 *Vmin* and *Pspill*, respectively), but no significant correlation is evident between the response variables and precipitation timing [(Figure 7).](#_bookmark7)

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 *Vmin* with an *R* of 0.29 to 0.33) and virtually none regarding *Pspill* (correlated with *Pspill* with an *R* of -0.01 to -0.1).

## Predicted values of *Vmin*

Diagnostics for the linear regression models assessed in the *Vmin* analysis are included in the Appendix (Figures [A3](#_bookmark14) through [A5](#_bookmark16) and Tables [A1](#_bookmark12) through [A3).](#_bookmark17)

## 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 *Vmin* values. The single monitoring location in each category with the highest *R* value was selected for further analysis [(Figure](#_bookmark14) 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 Mm3, respectively; [Figure A3,](#_bookmark14) top two panels; [Table](#_bookmark12) A1). In 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 Mm3 [(Figure A4,](#_bookmark15) upper left panel; [Table](#_bookmark13) A2). No three-predictor models

produced lower LOOCV or *R*2 values [(Table](#_bookmark17) 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](#_bookmark15) A4, top right panel). This produced a slightly larger error (2.78 Mm3) than the single-predictor model with maximum snowpack water content alone (2.74 Mm3; [Figure](#_bookmark14) 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,](#_bookmark15) 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 as the *Vmin* 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:

*Vmin, i* = −1*.*33 + 0*.*0053 ∗ *F JRSi* + 0*.*0027 ∗ *SWJi*

Where:

(2)

* *Vmin, i* is the predicted value of minimum 30-day dry season baseflows in calendar year *i* (i.e., at the end of water year
  + - 1. (million m3 or Mm3)
* *SWJi* is the maximum snow water content recorded at the Swampy John snow course (CDEC station ID SWJ or 285) in water year *i* (millimeters)
* *FJRSi* is the cumulative precipitation, recorded October-April of water year *i*, measured at the Fort Jones Ranger Station (NOAA station ID USC00043182) (millimeters)

## Predicted and observed *Vmin* over time

The *Vmin* formulation proposed above predicts minimum 30-day dry season baseflows with a model error of 2.29 Mm3 per 30 days (31.2 cfs), and a root mean squared error of 1.4 Mm3 (19.4 cfs). This RMSE indicates substantial uncertainty in any single year’s prediction: it corresponds to 49% of the mean *Vmin* value, 2.9 Mm3 (40 cfs).

Matching the historical trends of decreasing snowpack, the observed and predicted *Vmin* values show a downward trend over time [(Figure A5,](#_bookmark16) 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 this year is also an outlier, indicating that the model has a sufficient sample size to not be overwhelmed by this extreme value produced by an extremely uncommon sequence of events [(Figure A5,](#_bookmark16) middle panel).

The predictive *Vmin* 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 [(Figure A5,](#_bookmark16) middle panel).

* 1. **Predicted values of *Pspill***

Diagnostics for the linear regression models assessed in thu *Pspill* analysis are included in the Appendix [(Figures](#_bookmark19) A6 [throug](#_bookmark21)h A8 and [T](#_bookmark18)ables A4 [through](#_bookmark22) A6).

## Predictor assessment and prediction formula

The results of the predictor assessment for the *Pspill* prediction formula were similar to those for *Vmin*, in that the two best single predictors were October-April cumulative precipitation and maximum snowpack [(Figure A6,](#_bookmark19) top two panels), with LOOCV model errors of 850 and 718 mm, respectively. (*ET*0 was once again excluded from consideration based on a short record length.) Again similar to *Vmin*, the best two-predictor model was the combination of the two best single predictors, with an LOOCV error of 697 mm [(Figure A7,](#_bookmark20) upper left panel). No three-predictor models produced lower LOOCV or *R*2 values (**??**), 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,](#_bookmark20) 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,](#_bookmark20) lower right panel).

Based on these results, the model selected as the *Pspill* formulation for a given water year was a linear combination of the same observation records as *Vmin*: 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).

*Pspill, i* = 128 − 0*.*095 ∗ *FJRSi* − 0*.*028 ∗ *SWJi* (3)

Where:

* *Pspill, 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)
* *SWJi* is the maximum snow water content recorded at the Swampy John snow course (CDEC station ID SWJ or 285) in water year *i* (millimeters)
* *FJRSi* is the cumulative precipitation, recorded October-April of water year *i*, measured at the Fort Jones Ranger Station (NOAA station ID USC00043182) (millimeters)

## Predicted and observed *Pspill* over time

The *Pspill* estimate formulation proposed above predicts *Pspill* 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: it corresponds to 37% of the mean *Pspill* value, 68.8 mm.

Matching the historical trends of decreasing snowpack, the observed and predicted *Pspill* values show an increasing trend over time [(Figure A8,](#_bookmark21) 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.

The predictive *Pspill* 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 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,](#_bookmark16) lower panel).

The linear coefficients for the two prediction equations and their standard errors are summarized in [Table 2.](#_bookmark9)

## 3.6 Comparison with California DWR Water Year Type (WYT) category

The DWR water year type categories map fairly well onto the two proposed hydrologic indices *Vmin* and *Pspill* [(Figure 8,](#_bookmark8) panels A and B), which is to be expected, as both DWR WYT and the two proposed indices are based in part on cumulative

**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 |

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 *Vmin* or *Pspill* values. Likely causes include the information lost when binning water years into five categories, and the 30-year ranking window that would prevent a direct comparison of post-2000 WYTs with pre-1950s WYTs [(Figure 8,](#_bookmark8) panel C).

## Discussion

* 1. **Scott River watershed behavior and long-term planning**

The forward-looking seasonal predictions in this study are possible only because of predictable hydrologic relationships between early-season inputs (precipitation) and late-season outputs (surface flow and ET). 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,](#_bookmark5) 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 *Qspill* at the Fort Jones stream gauge.

A *Qspill* threshold of 120 cfs was identified by an analysis of rainfall-runoff responses and a visual inspection of fall

hydrograph behavior (Figures [A2](#_bookmark11) and [5),](#_bookmark5) and 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,](#_bookmark56) [2020).](#_bookmark56)

Simulated estimates of stream-aquifer exchange corroborate these precipitation-flow relationships. Dry season baseflow (*Vmin*) and the onset of wet season flow (framed in terms of *Pspill*) 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](#_bookmark3) (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 nearly-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).](#_bookmark27)

The resulting water storage limitations mean that multi-year planning, such as the long-term GSP projects, may be impossible in the Scott River watershed without making assumptions about how much it will rain (i.e., the future climate predictions in the Siskiyou County GSP) (Siskiyou County [2021).](#_bookmark51) If those assumptions are not fulfilled by future climate, year-by-year adaptive management may be necessary to achieve management outcomes.

* 1. ***Vmin* and *Pspill* 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, effectively taking an inventory in each April of relevant hydrologic inputs. It could be used to support decisions made in the late spring timeframe regarding the growing season, such as potential regulatory actions and some farmer cropping decisions. 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).](#_bookmark52) 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).](#_bookmark26) 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 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 [8),](#_bookmark8) which serves as a validation of this approach. However, the *Vmin* and *Pspill* 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 it is tailored to local hydroclimate data and is interpretable as a numeric prediction of fall conditions. This could be used to inform regulatory actions in an attempt to increase fall environmental flows, or for surface water diverters to plan for low-flow conditions. However, each seasonal prediction should be accompanied by its uncertainty: the RMSE of this predictive model is 49% of the mean value for *Vmin*, 2.9 Mm3 (40 cfs).

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

metric, *Pspill*, 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 *Pspill*

prediction when combined with weather forecasts in the fall. This prediction also carries substantial uncertainty (RMSE of 37% of the mean *Pspill* value, 68.8 mm).

## Management implications of best-performing predictors

As described in Results, the linear models that best predicted observed values of *Vmin* and *Pspill* were both based on the same two observation locations (the SWJ snow course and the FJRS weather station; [Figure](#_bookmark0) 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 GSP) (Siskiyou County [2021).](#_bookmark51)

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 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 be investigated in future work.

## Influences of climate change on predictive indices

Both predictions (using the full record of hydrologic data) assume some degree of hydroclimate stationarity, in that it uses 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 *Vmin* values based on the full record appear to systematically overpredict *Vmin* in the most recent era (2001-2021; [Figure A4,](#_bookmark15) top left panel, and [Figure A5,](#_bookmark16) 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.

## Conclusions

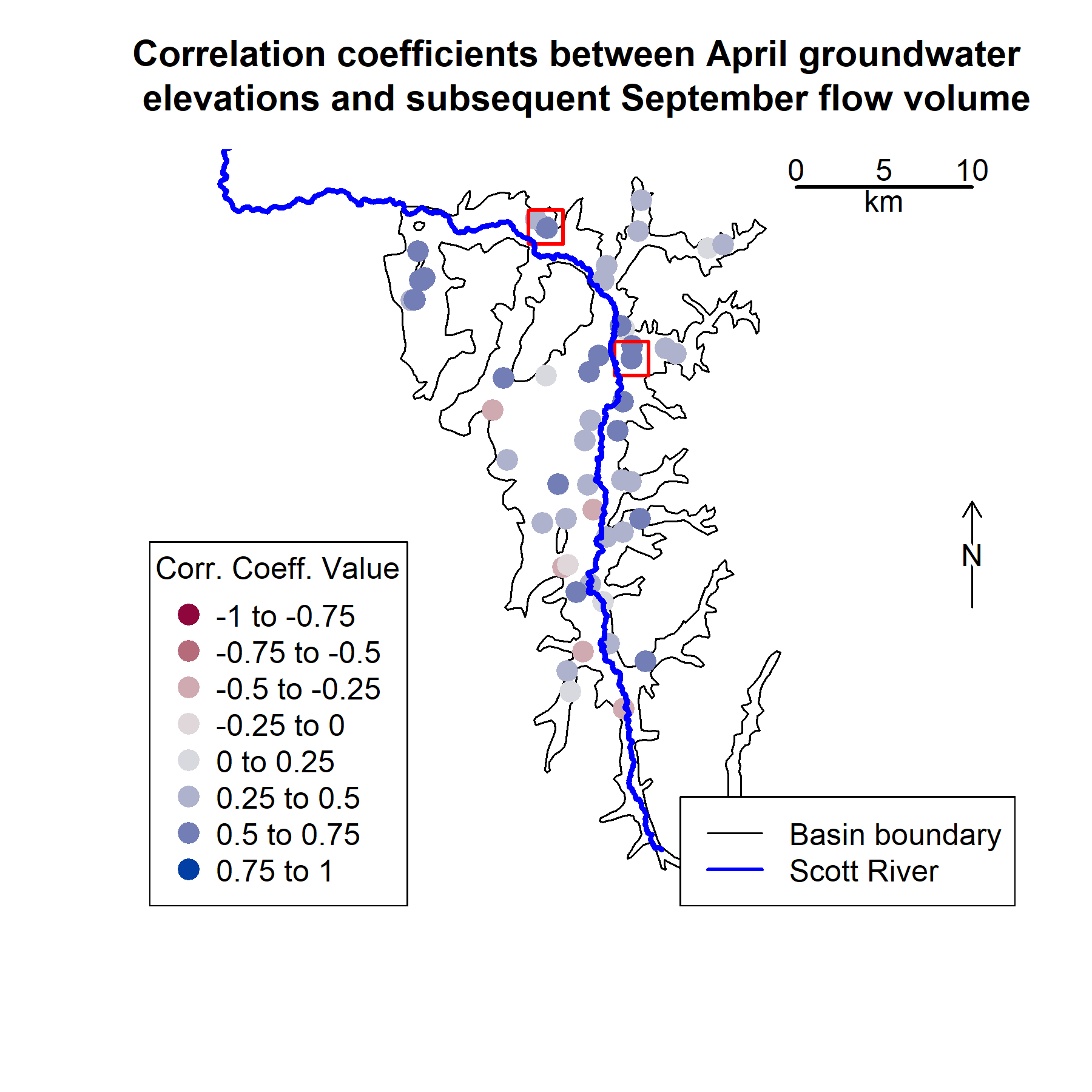
This study proposed two locally-tailored hydrologic decision-support metrics for the Scott River watershed in northern California. Both use snowpack and cumulative precipitation data from October-April to predict the quantity of interest: the first is the minimum 30-day flow volume in a given water year, referred to as *Vmin*, which typically occurs in September or October. The second index 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 *Pspill*. Both indices can be calculated at the end of April to support near-term (seasonal) adaptive management regarding the growing season or 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 California 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.

## Appendix

* 1. **Groundwater Well Data**

490 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.



**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.

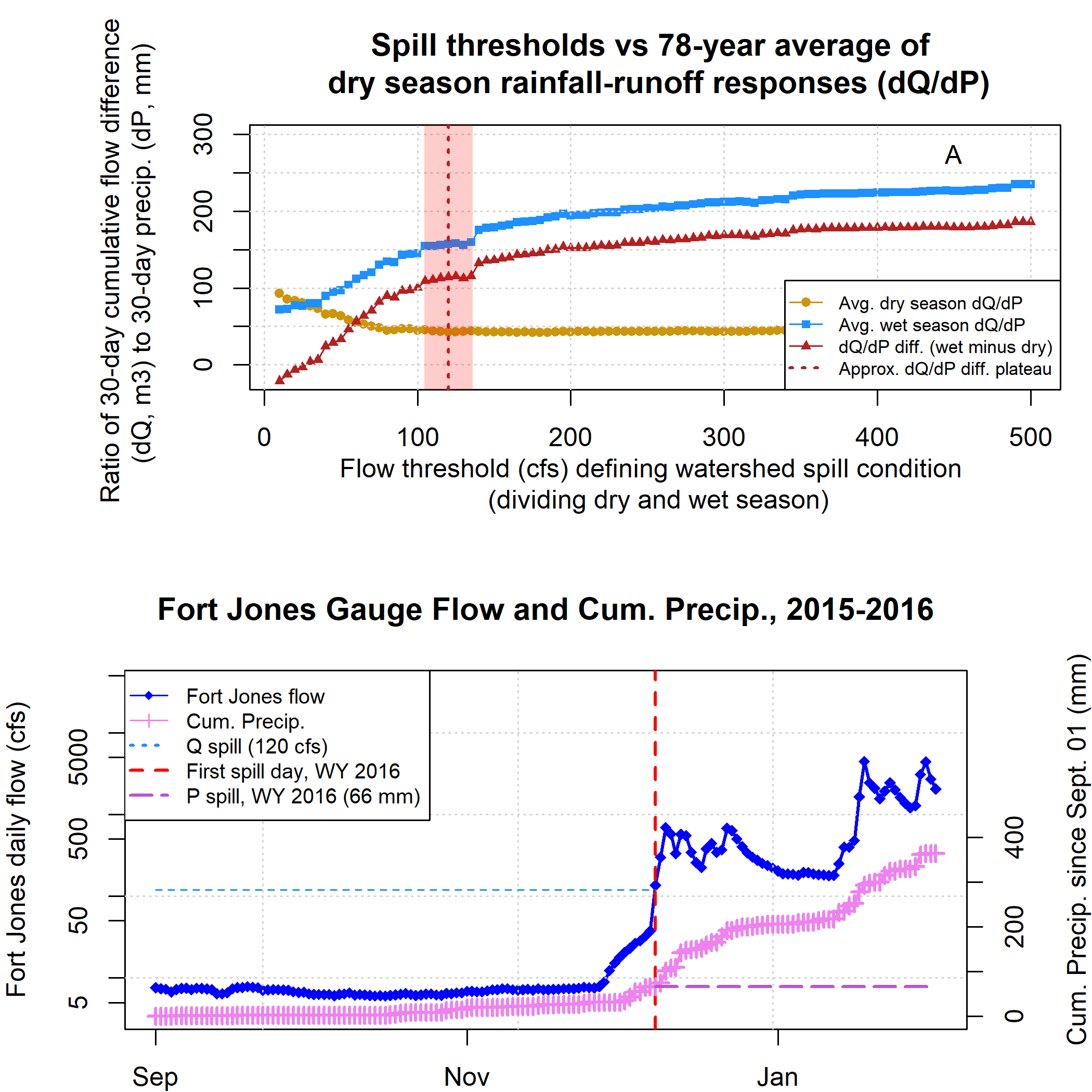
## Selection of the *Qspill* threshold value

The *Qspill* value of 120 cfs (294 thousand m3/day or approximately 9 Mm3 per month) was determined by testing a range of potential *Qspill* values (10 to 500 cfs [24 to 1223 thousand m3/day]) as a dividing threshold between dry- and wet-season flow behavior, and calculating the difference in rainfall-runoff response on either side of the dry season-wet season divide [(Figure A2,](#_bookmark11) Panel A). The objective of this analysis was to maximize the difference between these two *dQ/dP* values: to

identify the threshold that reflected a maximally different rainfall-runoff response before and after the threshold was reached. The difference between the wet and dry season rainfall-runoff responses is approximately the same (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,](#_bookmark11) 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 ) *Qspill* threshold generated a plausible breakpoint between dry and wet season river behavior in all water years (e.g., water year 2018 in [Figure A2,](#_bookmark11) 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).

## Model selection criteria - *Vmin*

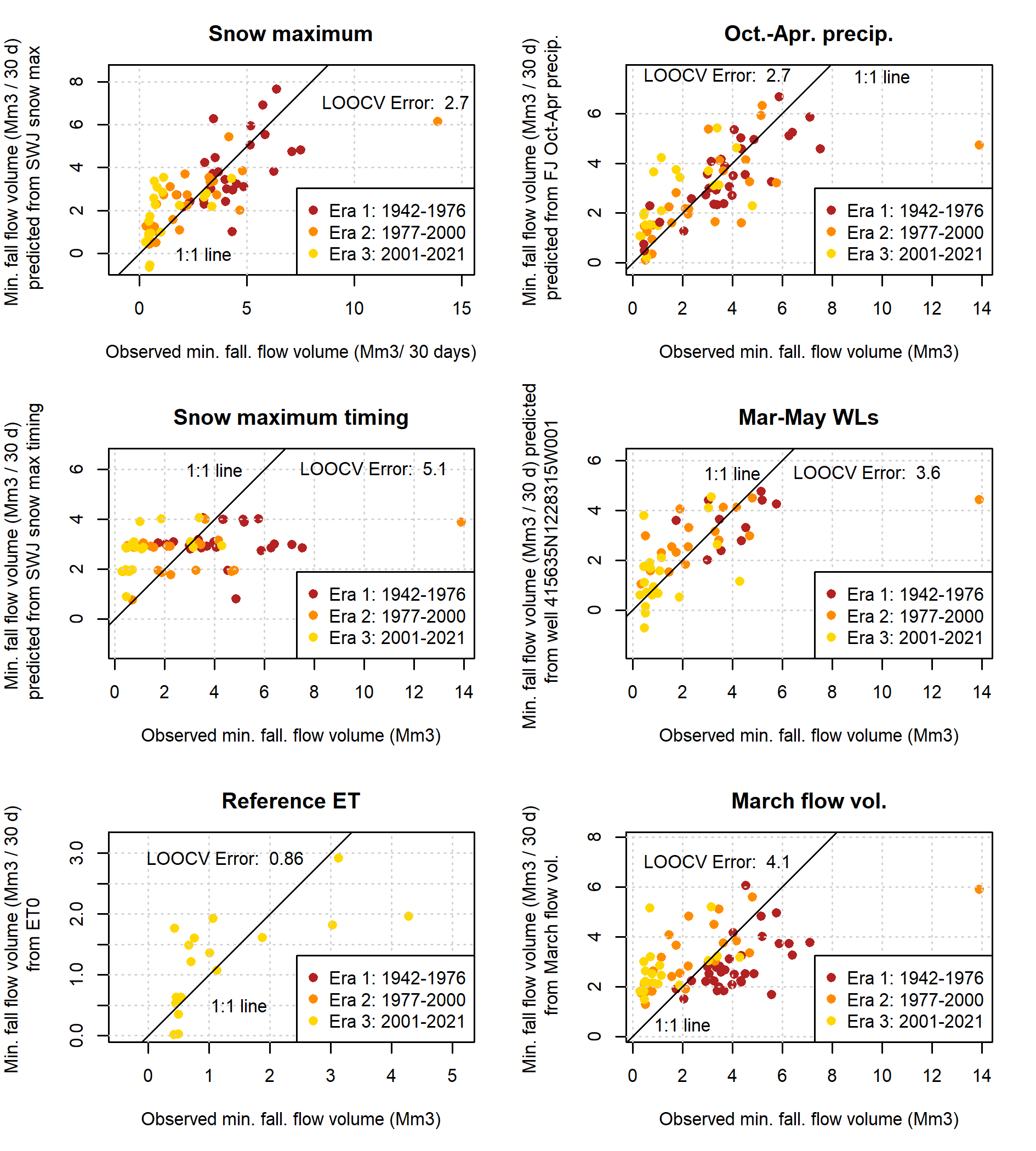
Diagnostics used to select the predictive models for *Vmin* are shown below and discussed in Results. Predictors are abbreviated in tables and described briefly in [Table A1;](#_bookmark12) for more information on potential predictors see Section 2.3.

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

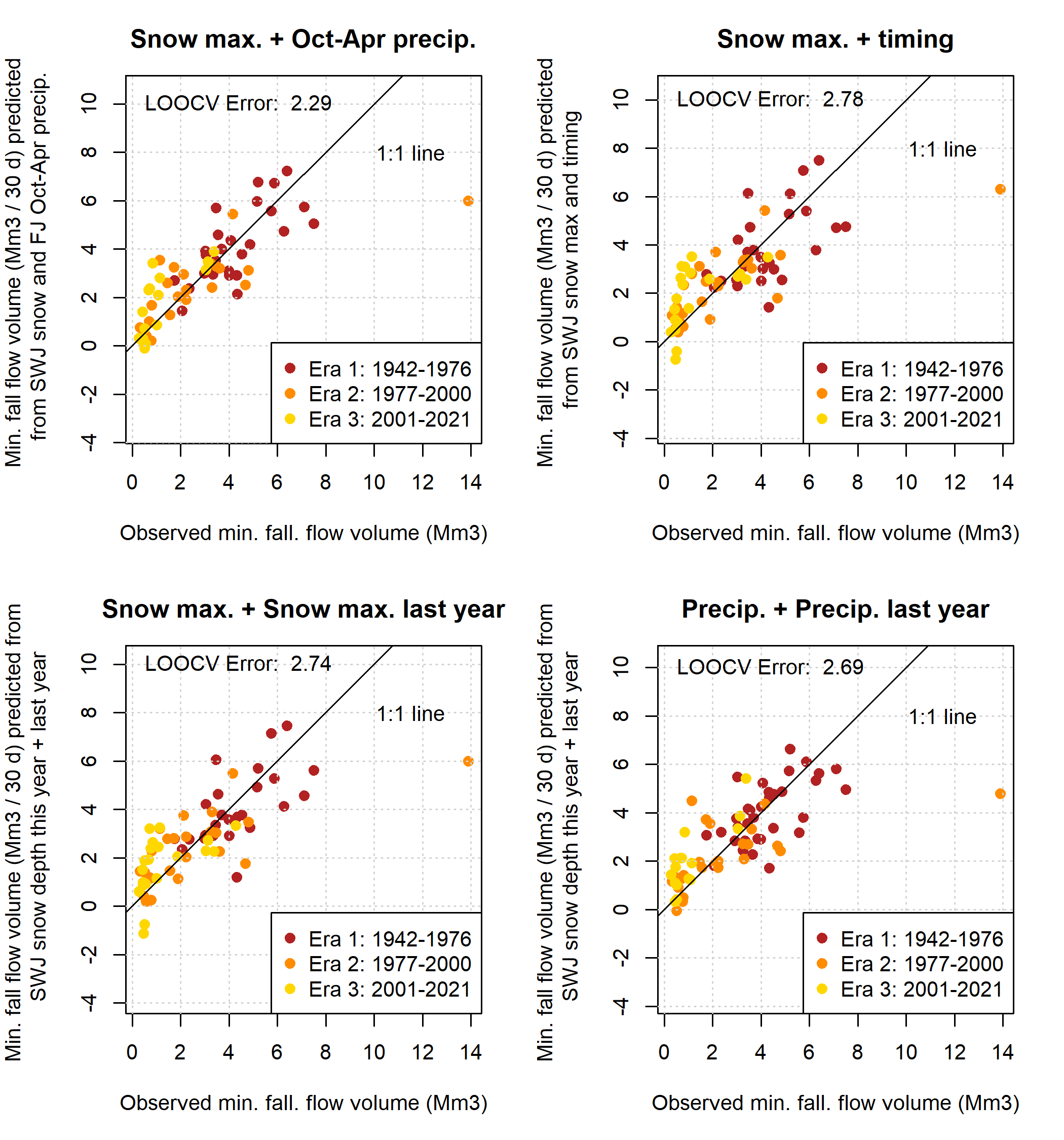
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predictor ID | Predictor Descrip. | n | Log Likelihood | AIC | LOOCV | R squared |
| SWJ\_max\_wc\_mm | Snow maximum | 70 | -131 | 269 | 2.7 | 0.53 |
| USC00043182\_oct\_apr\_mm | Oct.-Apr. Precip. | 75 | -142 | 290 | 2.7 | 0.49 |
| 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.

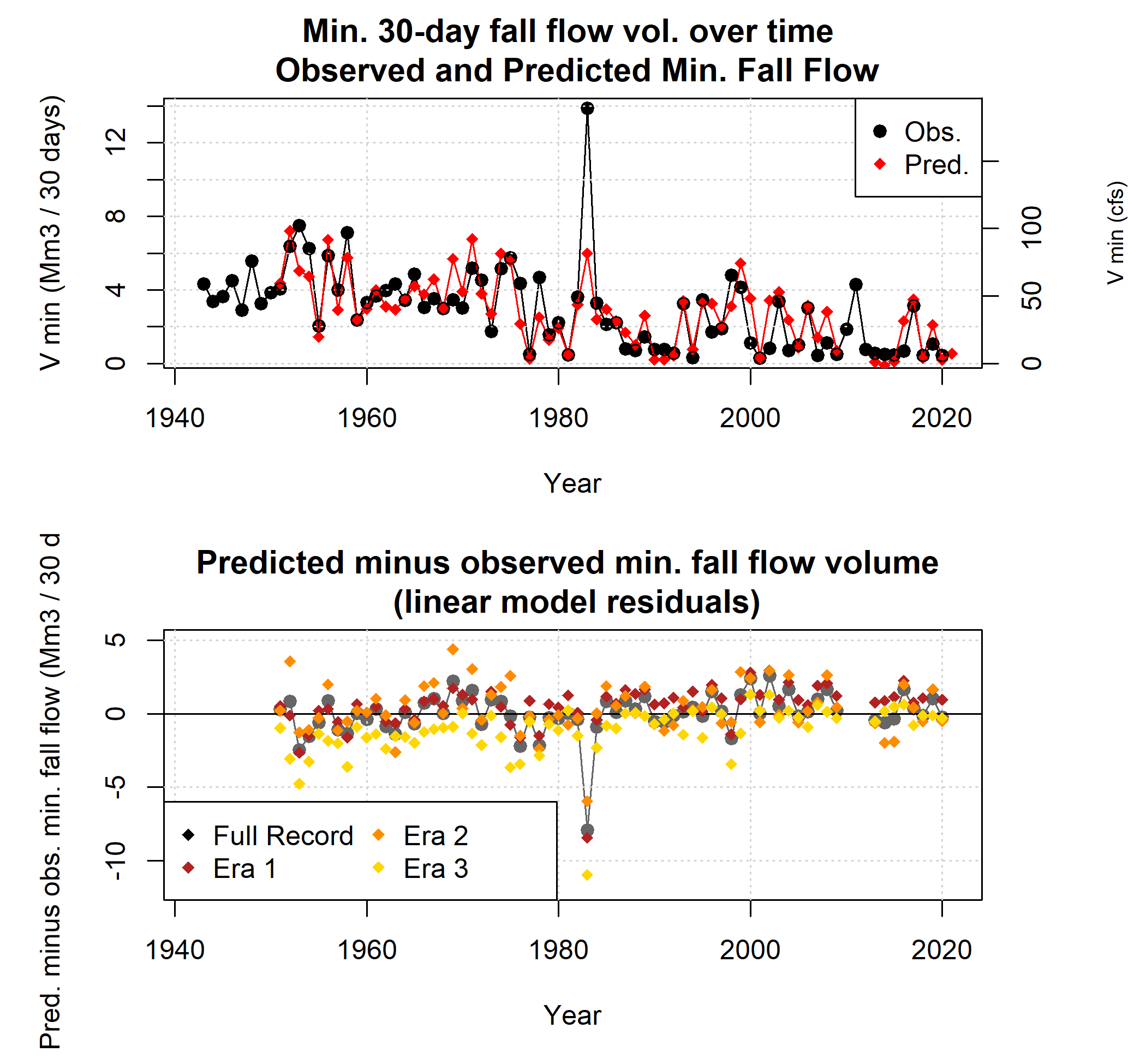
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predictor 1 | Predictor 2 | n | Log Likelihood | AIC | LOOCV | R squared |
| SWJ\_max\_wc\_mm | USC00043182\_oct\_apr\_mm | 67 | -119 | 246 | 2.3 | 0.62 |
| SWJ\_max\_wc\_mm | SWJ\_jday\_of\_max | 70 | -131 | 270 | 2.8 | 0.53 |
| SWJ\_max\_wc\_mm | 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.

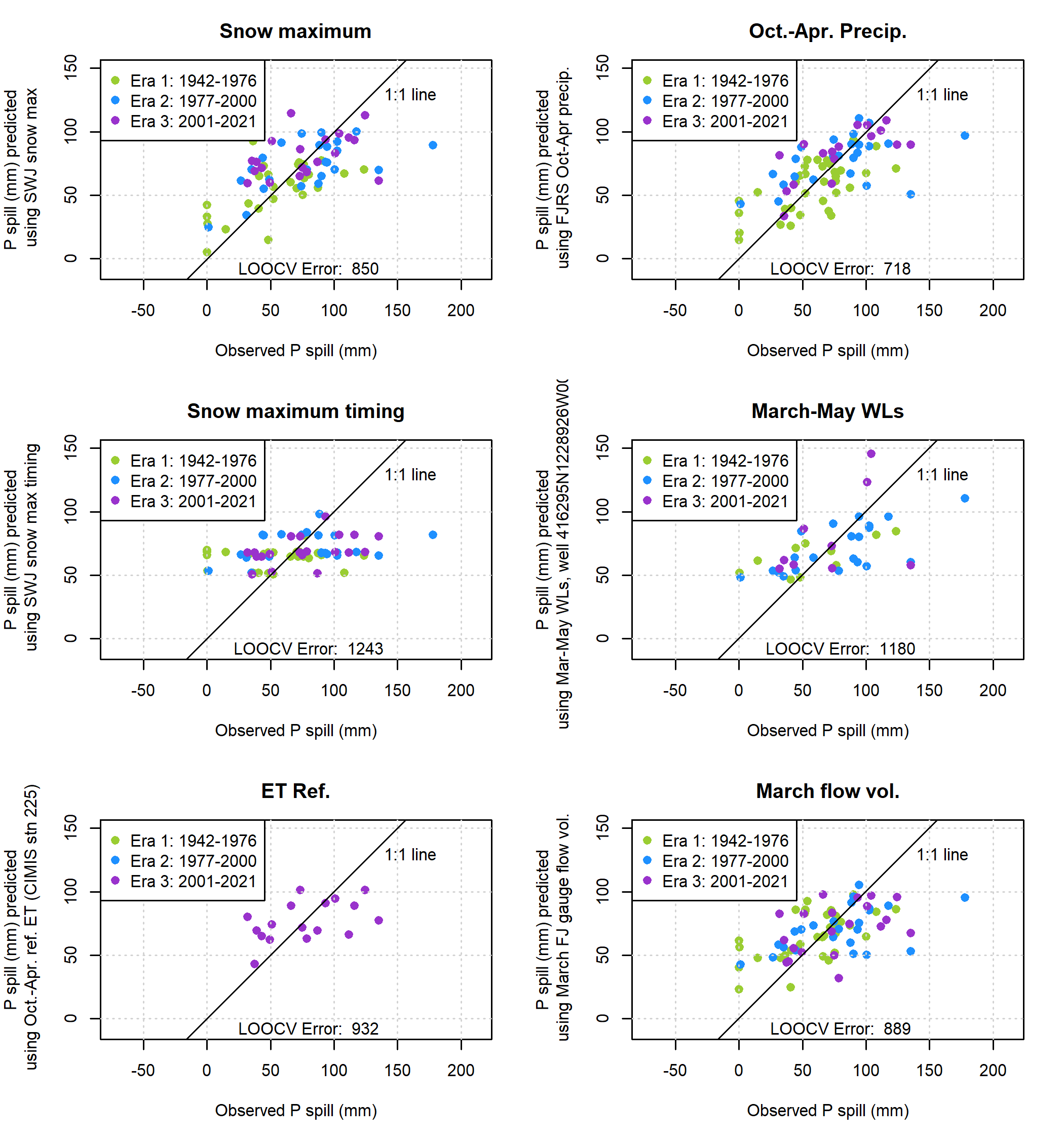
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor 1 | Predictor 2 | Predictor 3 | n | Log Like. | AIC | LOOCV | R squared |
| SWJ\_max\_wc\_mm | USC00043182\_oct\_apr\_mm | SWJ\_jday\_of\_max | 67 | -119 | 247 | 2.3 | 0.63 |
| SWJ\_max\_wc\_mm | USC00043182\_oct\_apr\_mm | 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 |

## Model selection criteria - *Pspill*

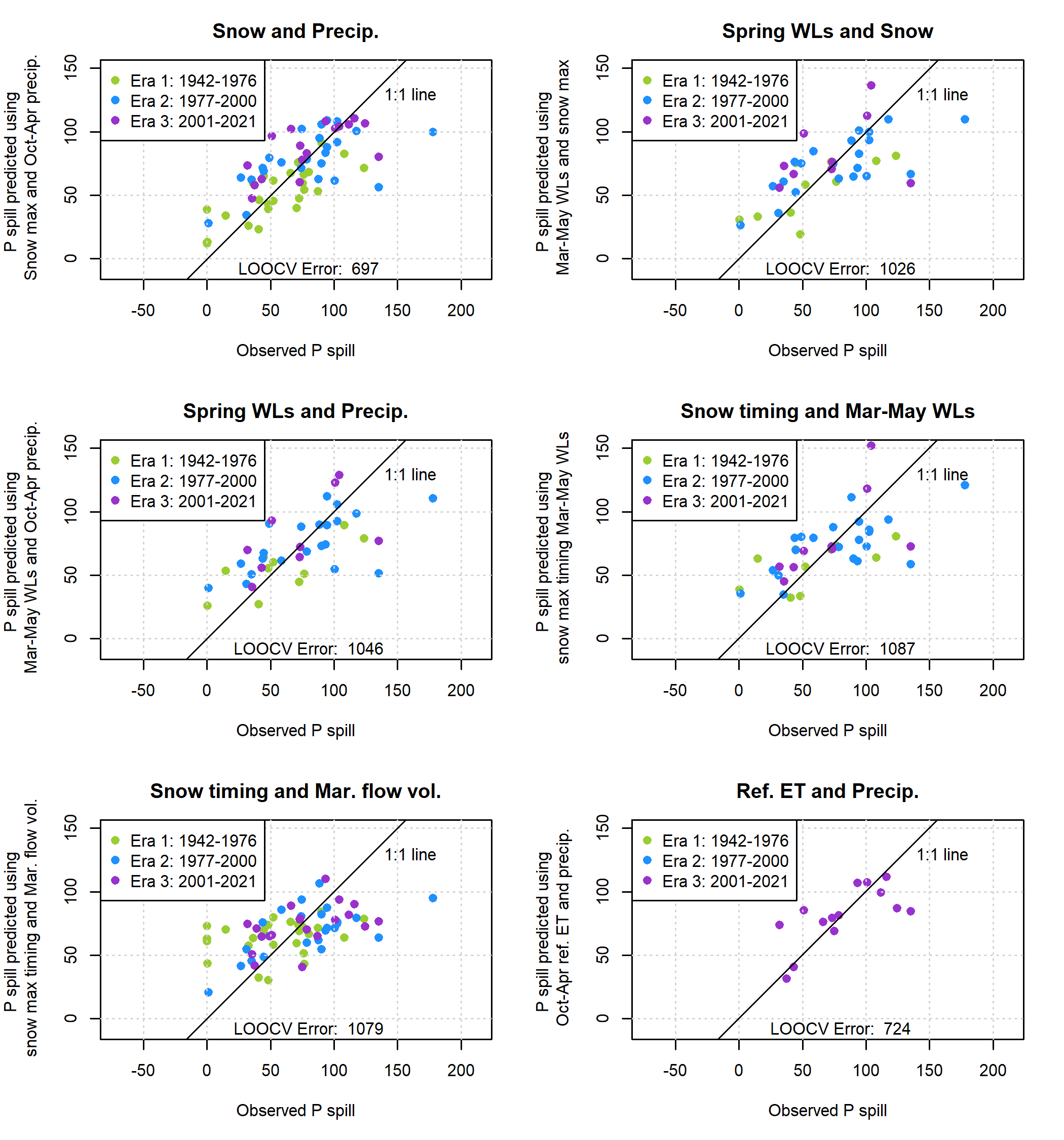
Diagnostics used to select the predictive models for *Pspill* are shown below and discussed in Results. Predictors are abbreviated in tables and described briefly in Table A4; for more information on potential predictors see Section 2.3.

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

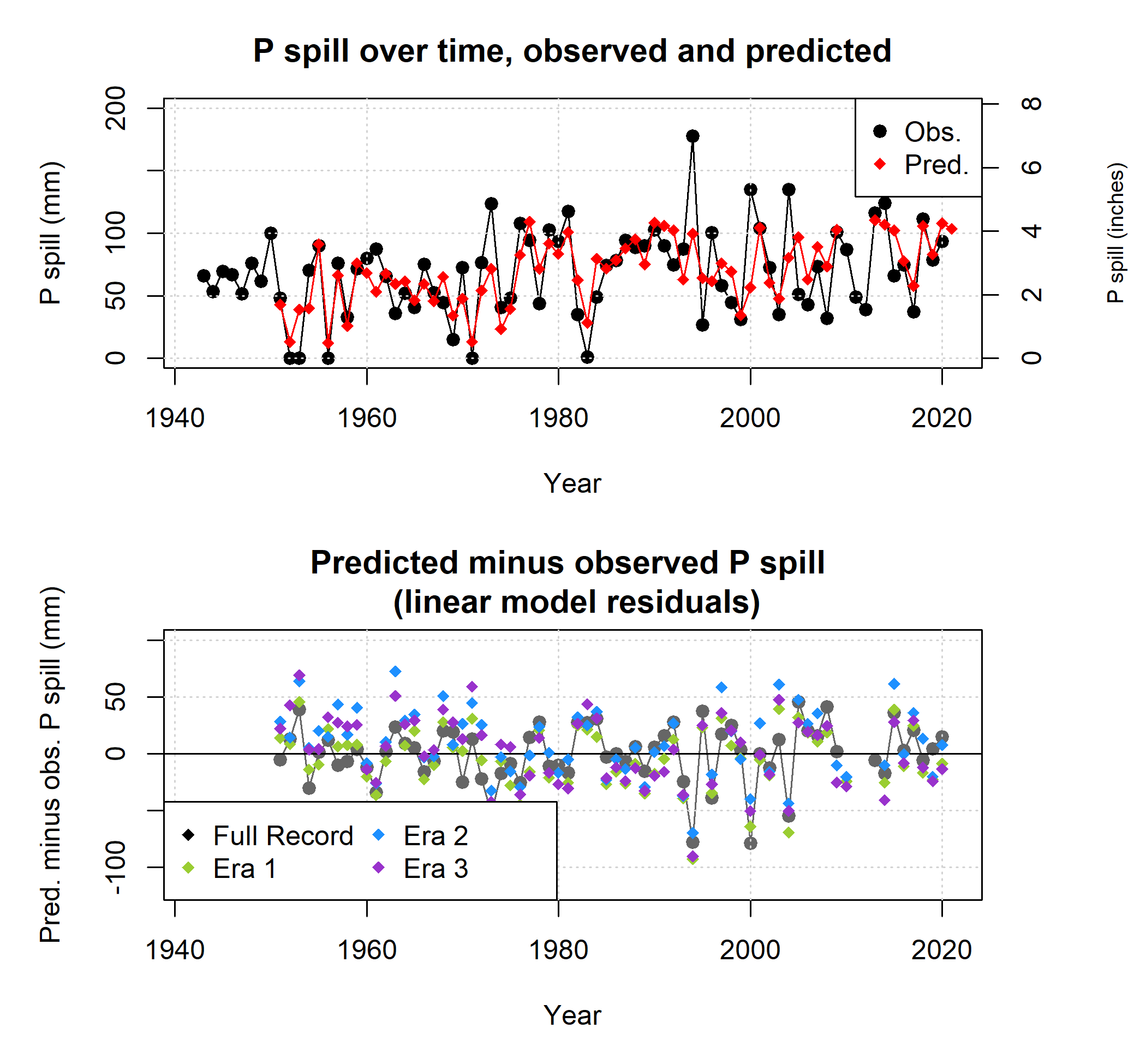
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Predictor ID | Predictor Descrip. | n | Log Likelihood | AIC | LOOCV | R squared |
| SWJ\_max\_wc\_mm | Snow maximum | 70 | -333 | 673 | 850 | 0.38 |
| USC00043182\_oct\_apr\_mm | Oct.-Apr. Precip. | 75 | -351 | 708 | 718 | 0.43 |
| 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 |

## Diagnostic plots for selected models

Standard diagnostic plots for the selected predictive models for *Vmin* and *Pspill*. For *Vmin*, these diagnostic plots highlight outlier record 42, which corresponds to the year 1983, when an early September storm followed a wet year. For *Pspill*, three lesser outliers are highlighted, corresponding to water years 1994, 2000 and 2004, in which the three highest values of *Pspill* 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.

Residuals vs Fitted

Residuals

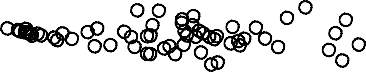
4

8

2

6

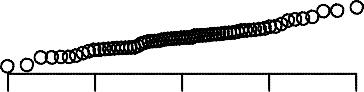
Normal Q−Q



42

12

61



42

12

61

0 2 4 6

−2

−2

−2 −1 0 1 2

Fitted values

Standardized residuals

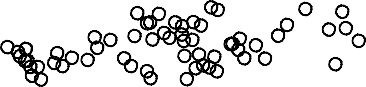
Theoretical Quantiles

Scale−Location

Standardized residuals

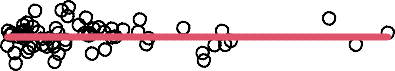
Residuals vs Leverage

1



42

61 12



Cook's dista28nce

42

22

1.5

6

0.5

2

0 2 4 6

0.0

−2

0.00 0.05 0.10 0.15

Fitted values

Standardized residuals

Leverage

**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.

*.* This manuscript emerged from dissertation work funded by Siskiyou County SGMA planning grants, with funding from California water bonds.

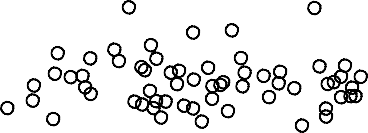
Residuals vs Fitted

Residuals

50

2

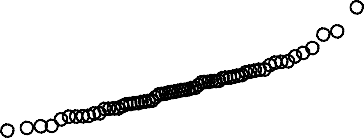
Normal Q−Q



59

53

63



53 59

63

20 40 60 80 100

−50

−2 0

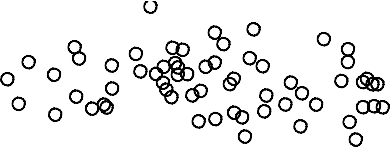
−2 −1 0 1 2

Fitted values

Standardized residuals

Theoretical Quantiles

Scale−Location



59

53

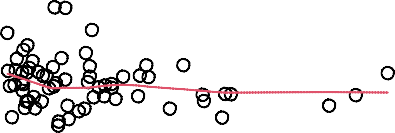
63

Standardized residuals

2

Residuals vs Leverage

0.5



Cook's distance

5593

63

1.0

20 40 60 80 100

0.0

0.00 0.05 0.10 0.15

−2 0

Fitted values

Standardized residuals

Leverage

**Figure A10.** Diagnostic plots for the selected predictive models of P spill.