

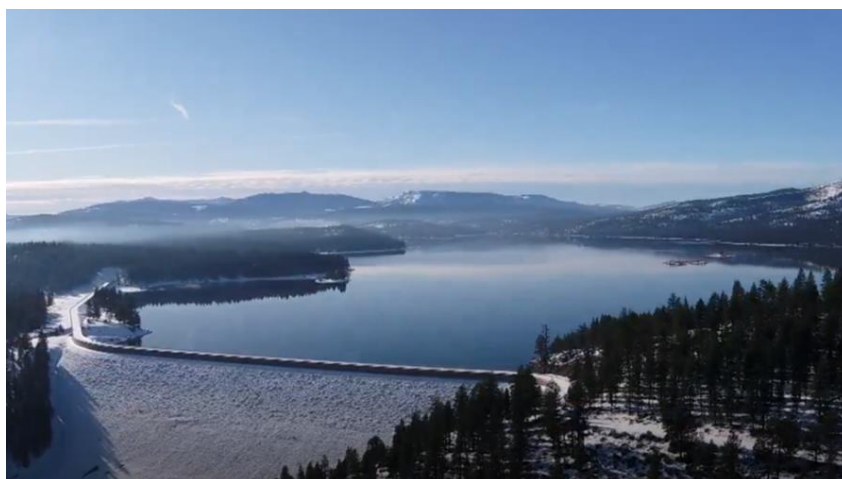


Atmospheric Rivers, Reservoir Storage, and Streamflow Forecast Skill: Assessing the Potential for Forecast-Informed Reservoir Operations in the Truckee River Basin

Christine M. Albano
Seshadri Rajagopal
Michael Imgarten

July 2023

Publication No. 41295



Stampede Reservoir
Photo Credit: Jeff Deitch, SkyVision

Prepared by
Division of Hydrologic Sciences, Desert Research Institute

Prepared for
US Army Engineer Research and Development
Coastal and Hydraulics Laboratory
3909 Halls Ferry Road
Vicksburg, MS 39180

THIS PAGE INTENTIONALLY LEFT BLANK

Atmospheric Rivers, Reservoir Storage, and Streamflow Forecast Skill: Assessing the Potential for Forecast-Informed Reservoir Operations in the Truckee River Basin

Christine M. Albano
Seshadri Rajagopal
Michael Imgarten

July 2023

Publication No. 41295

Prepared by
Division of Hydrologic Sciences, Desert Research Institute

Prepared for
US Army Engineer Research and Development
Coastal and Hydraulics Laboratory
3909 Halls Ferry Road
Vicksburg, MS 39180

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this project was to provide a preliminary understanding of challenges and opportunities for forecast-informed reservoir operations (FIRO) along the Truckee River by 1) quantifying relationships between atmospheric river water vapor transport (AR IVT) above the Truckee River basin and observed reservoir storage changes and 2) assessing streamflow forecast skill. Reservoir storage changes linearly with AR IVT, with some storms classified as “Extreme”, according to the Atmospheric River (AR) Scale, resulting in storage changes of $>1/3$ total capacity at three of six reservoirs during 1980-2017. Of the annual peak flows that occurred during ARs, all occurred during storms classified on the AR scale as “Moderate” or greater. Based on available (2015-2020) deterministic and ensemble streamflow forecast data, considerable improvement in accuracy is observed starting at a 7-day lead. During high-flow events, ensemble forecasts exhibit a bias toward underprediction and under-dispersion. For these events deterministic forecasts show superior accuracy relative to the ensemble median but, on average, still under-forecast flows by 15 percent and 35 percent, at 1- and 4-day leads, respectively. Results presented here are a first step in assessing AR-reservoir relationships and forecast skill for the Truckee River Basin. Understanding the relative contributions of storm characteristics and antecedent conditions to streamflow responses and to uncertainties and biases of hydrologic forecasts are important next steps.

ACKNOWLEDGEMENTS

The authors wish to thank Maria Vasquez for assistance with report production and review. This work was conducted for the Coastal and Hydraulics Laboratory (CHL), U.S. Army Engineer Research and Development Center (ERDC). The authors wish to thank, Michael Dettinger (Center for Western Weather and Water Extremes – Scripps Institution of Oceanography) and Rosemary Carroll (Desert Research Institute; DRI) for their technical assistance and contributions to this project. In addition, the authors thank Ian Floyd (ERDC) and Kumud Acharya (DRI) for their input and assistance with project design. Authors also wish to thank Sean McKenna, Dennis Hallema, and Guo Yu (DRI) and Brett Whitin (NOAA California Nevada River Forecast Center) for reviewing and providing feedback that helped to improve the contents of this report. Funding for Christine Albano and Seshadri Rajagopal was provided by the U.S. Army Corps of Engineers under grant agreement number W912HZ1920011. Funding for Michael Imgarten was provided by the California-Nevada River Forecast Center.

CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF FIGURES	vi
LIST OF TABLES.....	vii
LIST OF ACRONYMS	viii
INTRODUCTION	1
METHODS	2
Study Area	2
Methods.....	3
Reservoir Storage and Atmospheric River Relationships.....	3
Assessment of Wintertime Streamflow Forecast Skill	3
RESULTS AND DISCUSSION.....	6
Reservoir Storage and Atmospheric River Relationships.....	6
Assessment of Winter Streamflow Forecast Skill	8
CONCLUSIONS.....	18
RECOMMENDATIONS.....	18
REFERENCES	19

LIST OF FIGURES

1. Atmospheric Rivers and Truckee River floods.....	1
2. Truckee River Basin Study Area	2
3. Linear relationships between storm total IVT and changes in reservoir storage in six reservoirs along the Truckee River for the 27 largest IVT storms between 1980 and 2017.. ..	6
4. Precipitation in relation to AR storm-total IVT (a) and Precipitation and storm duration in relation to AR storm-total IVT (b).....	7
5. Change in snow water equivalent (SWE) 3 days prior to flood peaks for each of four flood-producing AR events.....	9
6. Average and range of R^2 values between forecast and observed 3-day flow volumes as a function of forecast lead time for the ensemble median and deterministic.	10
7. Scatter plots of observed and forecasted 3-day flow volumes and best-fit linear regression lines for top 10% AR days for the deterministic and ensemble median forecasts at a 1-day lead time.....	10
8. Rank histograms for top 10% flow and other AR days at a 1-day lead time for select sites..	12
9. Rank histograms for top 10% flow and other AR days at longer five- and ten- day lead times for select sites.	13
10. Forecast hit rates at varying lead times.....	14
11. Forecast critical success index (CSI) at varying lead times.....	15
12. Frequency bias as a function of lead time.....	15
13. Time series comparisons of observed and predicted 1-, 3-, and 5- day flow accumulations at corresponding lead times at the Truckee River at Farad gage during winter 2017, when multiple large AR events occurred in the Truckee River Basin..	17

LIST OF TABLES

1. List of California-Nevada River Forecast Center (CNRFC) points assessed	4
2. Annual peak flood volumes for four AR-induced flooding events and antecedent reservoir conditions.....	8
3. R^2 between forecast and observed 3-day flow volumes at 1- and 4-day lead times for the ensemble median and deterministic forecasts at select sites and for subsets of AR days with 90 th percentile or greater flow magnitudes, other AR days, and non-AR days.	11
4. Tabulation of forecast skill metrics at 1- and 4-day lead times for flows greater than the 1980-2019 90 th percentile.	14
5. Tabulation of forecast skill metrics at 1- and 4-day lead times for flows less than the 1980-2019 50 th percentile.	16

LIST OF ACRONYMS

AR	Atmospheric River
CHL	Coastal and Hydraulics Laboratory
CNRFC	California-Nevada River Forecast Center
ERDC	Engineer Research and Development Center
FIRO	Forecast-informed reservoir operations
GEFS	Global Ensemble Forecast System
HEFS	Hydrologic Ensemble Forecast System
IVT	Integrated water vapor transport
NOAA	National Oceanic and Atmospheric Administration
NWIS	National Water Information System
RFC	River Forecast Center
SNOTEL	SNOpack TELemetry
STIVT	Storm-total integrated water vapor transport
SWE	Snow-water equivalent
USACE	U.S. Army Corps of Engineers
USBR	U.S. Bureau of Reclamation
USGS	U.S. Geological Survey

INTRODUCTION

Reservoir operations in the Truckee River basin were designed based on the historical flow regime to achieve a balance between maximizing water supply and delivery for a diversity of water uses, while also minimizing flood risks. Over the past few decades, and especially the past few years, current operational rules as codified in the USACE Water Control Manual have proven constraining under extreme precipitation variability and warmer temperatures, which have shifted runoff to earlier in the season when reservoirs cannot accommodate or capture these flows because storage space is allocated to minimize flood risks (US. Department of Interior, 2019). Moreover, climate change projections consistently indicate warming temperatures that will significantly affect rain-snow ratios, likely resulting in a transition to more consistently earlier runoff regimes that current infrastructure and operational rules were not designed for. Earlier capture of streamflow in reservoirs has the potential to increase water supplies (Sterle *et al.*, 2020), but this may come at the cost of increased flood risks. A forecast-informed reservoir operations strategy (FIRO) has the potential to mitigate against these risks, but its effectiveness is, in part, contingent on the ability to reliably forecast flood events, in particular. In the Truckee River basin, both water supply and floods are strongly linked to wintertime atmospheric rivers (Figure 1) but despite improvements in forecasting these events at long lead times, the timing and amount of spillover precipitation onto the lee side of the Sierra Nevada remains a key uncertainty. In addition, storm runoff volumes in the Truckee basin are highly sensitive to rain-snow elevation, which is also difficult to forecast. Finally, antecedent snowpack and soil conditions have the potential to modulate runoff volumes but factors controlling the strength of these modulations are incompletely understood and monitored.

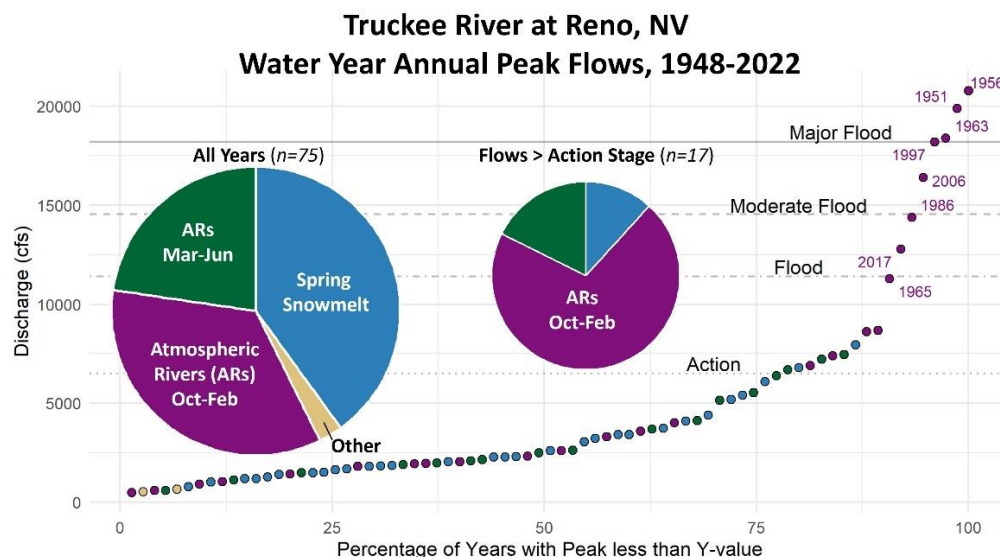


Figure 1. Atmospheric Rivers and Truckee River floods. The majority of annual peak flows during 1948-2022 occurred during atmospheric river events. All flood flows occurred during atmospheric rivers in the Oct-Feb time period, as did the majority of flows above Action stage. Peak flow values are based on USGS data. AR occurrences prior to 1980 are based on Gershunov *et al.*, 2017. [Flood categories are based on stage-discharge relationships used by National Weather Service](#), current as of May 2023.

There are two objectives of this study. The first is to assess linkages between ARs and reservoir storage to better understand the conditions under which AR-driven flooding occurs in the Truckee River basin. The second is to conduct an assessment of streamflow forecast skill in the Truckee River over a 5-year period (2015-2019) to provide a preliminary understanding of potential forecast-related challenges and opportunities for FIRO.

METHODS

STUDY AREA

The Truckee River originates at 2700 m elevation on the eastern slope of the Sierra Nevada Mountain range and flows 195 km northeastward through the cities of Reno and Sparks to its eventual terminus at Pyramid Lake (Figure 2). A large proportion of streamflow originates in the winter from snowpack and precipitation in the upper 25 percent of the basin in California, while the majority of water demand occurs within the lower part of the basin in the state of Nevada (U.S. Bureau of Reclamation (USBR), 2016). The Truckee River is a highly managed system that includes three dammed natural lakes that serve as water supply reservoirs (Lake Tahoe, Donner Lake, and Independence Lake) and four additional flood control reservoirs (Stampede, Boca, Prosser Creek, and Martis Creek) with fixed-date operations in accordance with U.S. Army Corps of Engineers flood control criteria. Diverse uses of Truckee River water supplies include municipal, industrial, agricultural, and environmental instream flows.

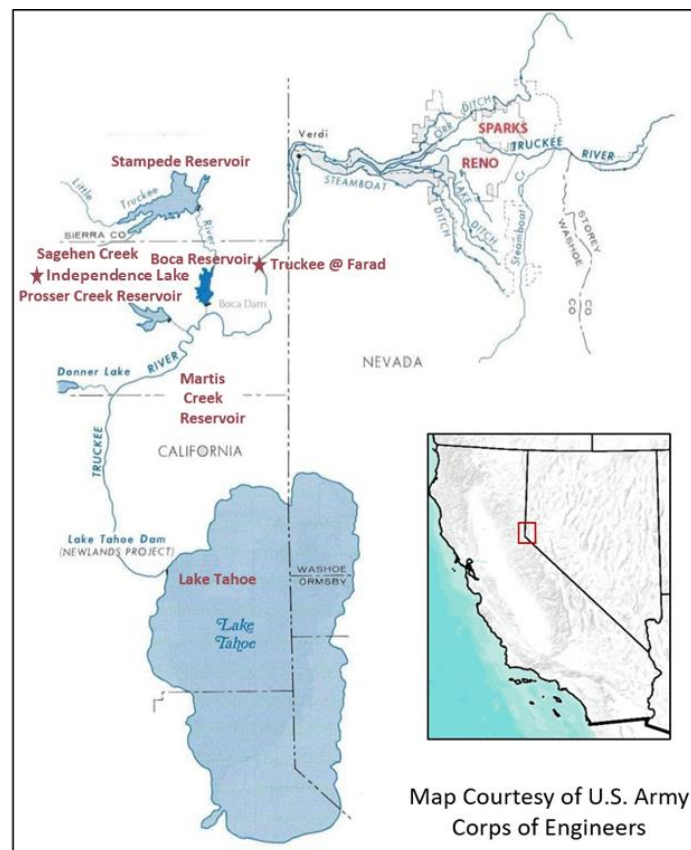


Figure 2. Truckee River Basin Study Area

METHODS

Reservoir Storage and Atmospheric River Relationships

To assess relationships between reservoir storage and ARs, a chronology of multi-day AR events from 1980 – 2017 at the 39.5 N and 120 W latitude and longitude grid cell (roughly centered on Verdi, NV) were identified using the Rutz AR detection algorithm (Rutz *et al.*, 2015) on the Modern-Era Retrospective Analysis for Research and Applications, Version 2 dataset (MERRA-2; Gelaro & Coauthors, 2017) and methods described in Albano *et al.*, 2020. A total of 185 events were identified. A subset of 27 events were selected based on a storm total integrated water vapor transport (IVT) threshold of $IVT > 4 \times 10^7$ kg/m, which equates to a ‘moderate’ AR event with an average IVT of ~500 kg/m/s over a duration of 24 hours (Ralph *et al.*, 2019). Storm total IVT is used because it provides a measure of both the magnitude and duration of the AR event (Ralph *et al.*, 2019), it tends to scale linearly with precipitation (Rutz *et al.*, 2014), and because IVT can be more predictably forecasted than precipitation at week or longer lead times (Lavers *et al.*, 2016). The difference between reservoir storage at the start and end of each event was calculated based on USGS streamflow gage data to assess change in reservoir storage associated with each event and these changes were compared to storm total IVT to determine the relationship. Causes for differences in flood amounts were further investigated using long-term SNOpack TELEmetry (SNOTEL) weather station data.

Assessment of Wintertime Streamflow Forecast Skill

To assess streamflow forecast skill, readily available forecast data were compared to observed flows for winter months (Oct-Apr) for water years 2015-2020. Forecast data from the National Weather Service California-Nevada River Forecast Center (CNRFC) were compiled for the 12 sites shown in Table 1. Two types of forecast data were used, including 1) the 6-hour deterministic forecast, which predicts streamflow out to 5 day lead times (only available for a subset of sites – see Table 1), and 2) the short-range hourly Hydrologic Ensemble Forecast System (HEFS; <https://cnrfc.noaa.gov/documentation/hefsAtCnrfc.pdf>) streamflow predictions from the California-Nevada River Forecast Center (CNRFC; <https://www.cnrfc.noaa.gov/ensembleHourlyProductCSV.php>; accessed May 2020), which include hourly streamflow forecasts with lead times up to 30 days. Both sets of forecasts are produced using a coupling of the Sacramento Soil Moisture Accounting Model (SAC-SMA; Burnash *et al.*, 1973) and the Snow Accumulation and Ablation Model (SNOW-17; Anderson 1973) model and are initiated with the same initial model states. The ensemble forecasts, HEFS, provide an uncertainty range based on 60 or more individual forecasts with differing future conditions. At the longest lead times, the ensemble forecasts are based on historical climatology. At shorter lead times, they integrate and weight information from other RFC forecasts and forecast products such as the Global Ensemble Forecast System (GEFS). More specifically, uncertainty and skill from the deterministic RFC quantitative precipitation forecasts are used for 1- to 3-day lead times and the GEFS mean precipitation uncertainty is used thereafter. Forecast data were aggregated from hourly/six-hourly time steps to daily averages. Prior to doing so, the time conversion from GMT to Pacific Time Zone was applied to ensure that forecast data coincided as closely as possible with daily streamflow and reservoir storage change observations.

Table 1. List of California-Nevada River Forecast Center (CNRFC) points assessed (see Figure 2). Highlighted rows indicate focal points for which results are presented in this report. Sites indicated by * are those for which deterministic forecasts were available.

CNRFC Code	CNRFC Site Name	CNRFC Data Type
BCAC1F	Little Truckee - Boca Dam, Natural Flow	Reservoir Inflow
DNRC1I	Donner Lake	Reservoir Inflow
*FARC1F	Truckee – Farad, Natural Flow	Forecast Point
ILAC1I	Independence Lake	Reservoir Inflow
MTSC1I	Martis Creek	Reservoir Inflow
*PSRC1	Prosser Creek - Prosser Creek Dam	Reservoir Inflow
SCRN2	Steamboat Creek - Steamboat	Other Point
SGNC1	Sagehen Creek - Truckee	Other Point
*STPC1F	Little Truckee - Stampede Dam, Natural Flow	Reservoir Inflow
TAHC1	Lake Tahoe	Reservoir Inflow
TRCC1	Truckee near Truckee	Forecast Point
TRRN2	Truckee at Reno	Forecast Point

Observational data of reservoir storage change and outflow were obtained for eight sites from the CNRFC. These are the same data that are used to calibrate the streamflow forecast model and originate from the Truckee River Operating Agreement information system monthly Water Master Reports (<https://www.troa.net/reports/mwmr/>; viewed May 15, 2020), which incorporate modeled estimates of reservoir evaporation. Reservoir inflows were calculated as change in storage plus outflow based on equations found here:

<http://www.troa.net/tis/details.html#160040> (viewed May 15, 2020). Daily flow data for the other four sites: Steamboat Creek (10349300), Sagehen Creek (10343500), Truckee near Truckee (10338000), and Truckee at Reno (10348000) were obtained from the USGS NWIS system.

Several metrics of forecast skill were calculated based on comparisons of the forecast and observed daily and aggregated multi-day (e.g., 3-day volumes, centered on the forecast lead time) flow data. Because FIRO viability may rely most heavily on forecast skill during AR days when the highest flows occur, data were subset into 3 groups in order to assess differences in forecast skill among non-AR days, AR days when daily flows were in the top 10 percentile based on the 1980-2019 time period, and other AR days when flows were not in the top 10 percent. For each of the three subsets, metrics of forecast accuracy and multiplicative bias (see below) were calculated for both the deterministic and the ensemble median forecasts at multiple lead times. In addition, a rank histogram analysis was completed for the three subsets to qualitatively assess the reliability of the forecast ensemble at varying lead times. To further explore forecast skill for storm and non-storm flows, hit rates, false alarm rates, the Critical Stress Index and frequency bias were calculated for 3-day flows greater than the 1980-2019 90th percentile (i.e., top 10% flow events) and less than the 50th percentile (i.e., non-storm flows). All metrics were calculated using the Verification package in R (Gilleland, 2015) and are described below. Additional information on these and other potential metrics is available at

<https://www.cawcr.gov.au/projects/verification/> (viewed October 29, 2020)

Coefficient of determination (R^2) between observed and forecasted flow - the proportion of variance in observed 3-day flow volumes that is explained by forecasted flow. This was calculated for the three subsets of days - > top 10 percent Flow AR days, other AR days, non-AR days.

Multiplicative bias (M.Bias) between observed and forecasted 3-day flow volumes – the ratio of average forecast magnitude to average observed magnitude. M.Bias > 1 indicates forecasted flow magnitudes are greater than observed, M.Bias=1 indicates no bias, M.Bias < 1 indicates forecasted flow magnitudes are less than observed. This was calculated for the three subsets of days - top 10 percent Flow AR days, other AR days, non-AR days.

Rank histograms of observed 3-day flows relative to ensemble members – visualizes the frequencies with which observed 3-day flows fall at a given rank relative to a 59-member forecast ensemble. A uniform distribution of frequencies (flat histogram) indicates good forecast reliability. Higher frequencies at low ranks (right skew) indicate an overprediction bias, while higher frequencies at high ranks (left skew) indicate an underprediction bias. A U-shaped distribution indicates that the forecast ensemble is under-dispersed, while a hump-shaped distribution indicates overdispersion. Rank histograms were developed for top 10 percent Flow AR and Other AR days.

Hit rate for flows > 90th percentile and < 50th percentile - the proportion of 3-day flow occurrences greater than the 90th percentile or less than the 50th percentile that were correctly forecasted to be greater than the 90th percentile. Ranges from 0 (no skill) to 1 (perfect skill).

False alarm rate for flows > 90th percentile and < 50th percentile - the proportion of 3-day flow occurrences less than the 90th percentile that were incorrectly forecasted to be greater than the 90th percentile. Ranges from 0 (perfect skill) to 1 (no skill).

Critical Success Index (CSI) for flows > 90th percentile and < 50th percentile – the proportion of 3-day flow occurrences greater than the 90th percentile or less than the 50th percentile that were correctly forecasted (this metric is similar to the *hit rate* but includes a penalty based on the *false alarm rate*). Ranges from 0 (no skill) to 1 (perfect skill).

Frequency Bias (F.Bias) for flows > 90th percentile and < 50th percentile – the ratio of the frequency of forecasted flows greater than the 90th percentile or less than the 50th percentile to the observed frequency. F.Bias > 1 indicates over-forecast, F.Bias=1 indicates no bias, F.Bias < 1 indicates under-forecast.

For the sake of efficiency, results are presented here for only a subset of sites (See Table 2), which were selected to represent a range of contributing area sizes and elevation distributions (Figure 2), including a large flow volume, high elevation site (Tahoe), a small flow volume high elevation site (Sagehen Creek), two flood control sites (Prosser and Stampede), and a downstream ‘integrator’ site (Truckee River at Farad) where flows are an important determinant of operational decisions and water rights distributions for the entire Truckee and Carson River systems.

RESULTS AND DISCUSSION

Reservoir Storage and Atmospheric River Relationships

The relationships between AR storm total IVT and reservoir storage changes are generally linear, with each reservoir having a different IVT threshold at which storage increases by 1/3 capacity (Figure 3). Three of six reservoirs analyzed exceeded this threshold at least once, including Donner, Prosser, and Boca. In these cases, the thresholds based on the linear regression were 11, 12.5, and 19×10^7 kg/m, respectively. Several outlier AR events, for which little storage change occurred despite large IVT amounts appeared to be due to relatively lower precipitation efficiencies (i.e., Precipitation: IVT; Figure 4 – orange shaded box), which could be driven by a number of different factors related to rain shadowing, the vertical distributions of wind and water vapor within the AR, or the approach angle of the AR, which affects the amount of uplift and consequently precipitation production (Dettinger *et al.*, 2019). These outliers also occurred during the early season (October 2016 and November 1996) when soils tend to be drier, which can also reduce runoff efficiencies.

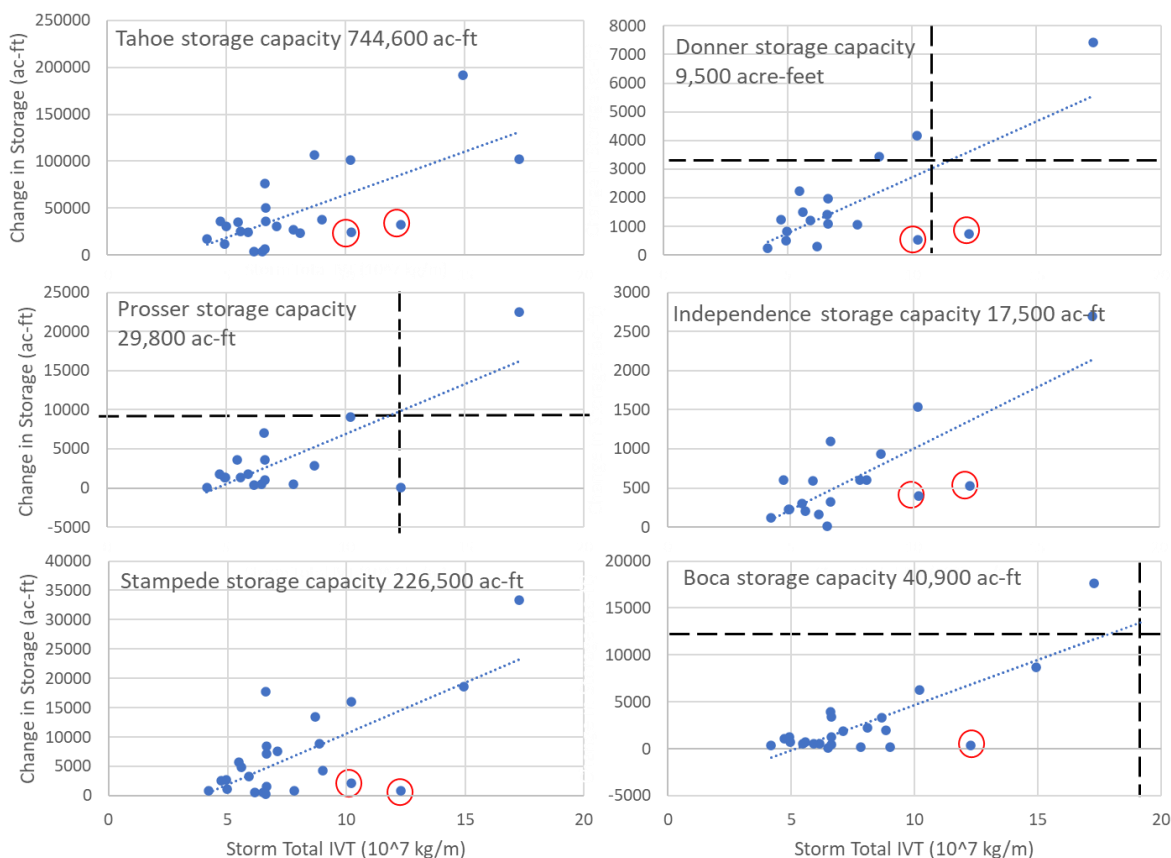


Figure 3. Linear relationships between storm total IVT and changes in reservoir storage in six reservoirs along the Truckee River for the 27 largest IVT storms between 1980 and 2017. Circled points indicate outlier storms with smaller changes in storage than expected given IVT amounts. Vertical lines indicate the IVT threshold at which change in storage is greater than or equal to 1/3 of the storage capacity. Note that this threshold is not reached in Independence, Stampede, and Tahoe. See Figure 2 for locations.

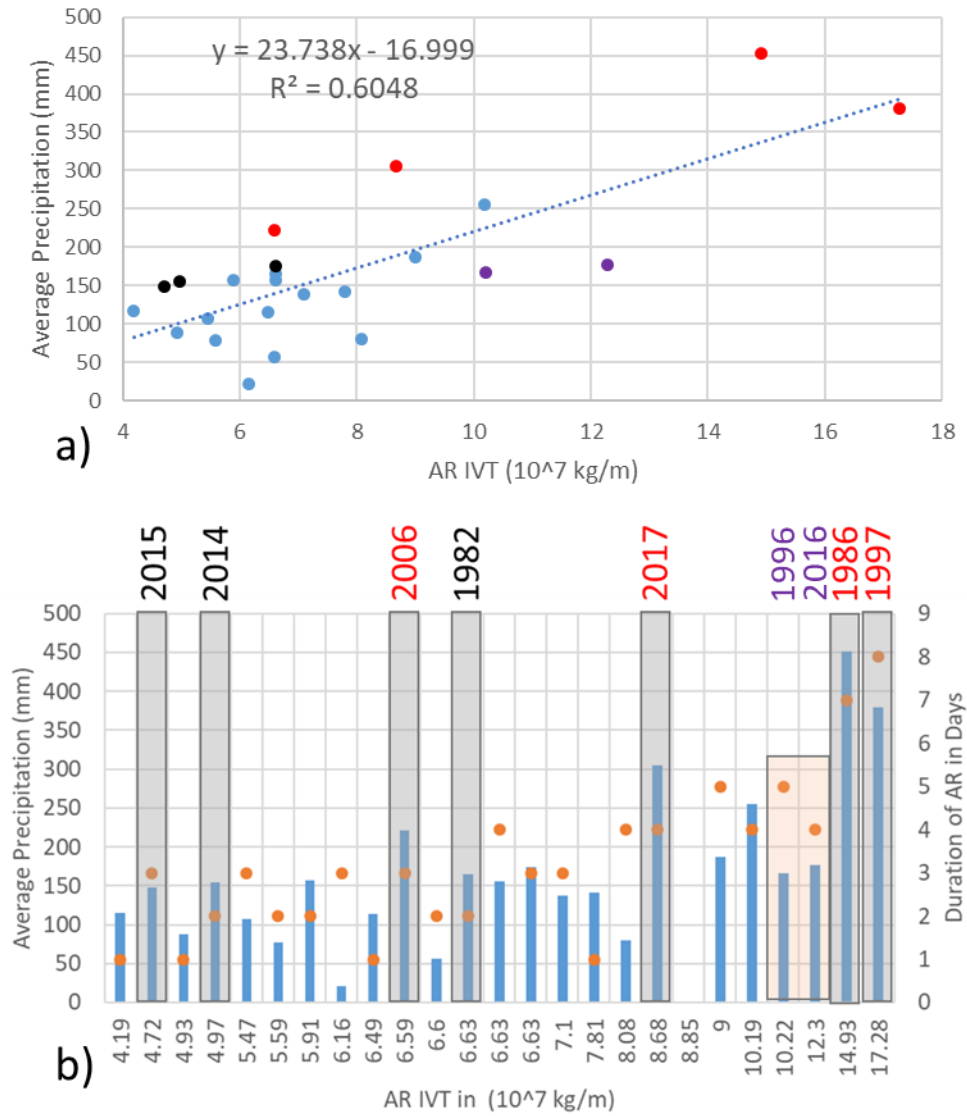


Figure 4. Precipitation in relation to AR storm-total IVT (a) and Precipitation (blue bars) and storm duration (orange dots) in relation to AR storm-total IVT (b). Low precipitation efficiency outlier storms highlighted in Figure 3 are indicated by purple points (a) and text (b). Black points (a) and text and gray shading (b) indicates water years in which annual peak flows resulted from AR storms. Red points (a) and text (b) indicates those AR storms that resulted in exceedance of flood stage at the Truckee at Reno gage.

Of the 27 storms, four of these resulted in annual peak flows at the Truckee at Reno stream gage that were near flood stage, including storms occurring on 2/17/1986, 1/2/1997, 12/31/2005, and 1/9/2017 (Table 2, Figure 4). Other storms which resulted in annual peak flows that did not cause flooding occurred in 1982, 2014, and 2015 (Figure 4). Overall, no storms with less than 6.59×10^7 kg/m storm total IVT caused an exceedance in flood stage at the Truckee at Reno gage. The linear least squares regression R^2 between storm total IVT and precipitation was 0.6, indicating a strong linear relationship (Figure 4a), and storms that produced annual peak flows tended to have larger precipitation amounts than expected based on this relationship.

Table 2. Annual peak flood volumes for four AR-induced flooding events and antecedent reservoir conditions.

Event Date	Peak Flood Volume - Truckee @	Reservoir % of Maximum Capacity 1- week prior to AR Event					
	Reno (cfs)	Tahoe	Donner	Prosser	Independence	Stampede	Boca
1/9/2017	12,800	36	57	30	80	50	36
2/17/1986	14,400	76				54	57
12/31/2005	16,400	49	55	40	90	80	60
1/2/1997	18,200	90	44	32	84	90	47

Assessment of one-week prior reservoir conditions for the four flood-stage storms indicate greater flood volumes are associated with greater antecedent reservoir storage volumes (as % of maximum capacity) in Tahoe and Stampede – the largest reservoirs, but no relation is observed for other reservoirs (Table 2). Assessment of snowpack conditions three days prior to each event also indicate relations with flood volumes, with decreases in snow-water equivalent (SWE) observed at some or most SNOTEL stations prior to the two largest flow events (12/31/2005 and 1/2/1997), indicating snowmelt contributions to runoff in these cases and mostly increases in SWE observed for the other two smaller flood events (Figure 5).

Assessment of Winter Streamflow Forecast Skill

Based on the coefficient of determination (R^2 values) between observed and forecasted 3-day flows (Figure 6, Table 3), forecast skill clearly improves as a function of shorter lead times. This is especially the case for the top 10 percent (i.e., > 90th percentile) flow AR days where there is rapid improvement from virtually no skill at an 8-day lead time to an among-site average of 60 percent of the variation in the observed flows explained by the ensemble median forecast flow at a 1-day lead time. For the top 10 percent flow AR days, the deterministic forecast is consistently more skillful, explaining an among-site average of 75 percent variation between forecast and observed at a 1-day lead time. The poorer skill of the ensemble forecasts relative to the deterministic forecasts is due in part to the blending and weighting of short-range forecasts from the RFC and GEFS with climatology. Blending with climatology effectively and systematically dilutes the forecast magnitudes of large storms, particularly at longer lead times. Non-AR and Other AR days consistently have higher skill relative to the top 10 percent flow AR days and the deterministic and ensemble median forecast skills are more similar to each other in these cases.

The multiplicative bias of the forecasts indicates underprediction of flow volumes during top 10 percent flow AR days (Table 3, Figure 7), with 3-day flow volumes averaging 30 percent greater than the ensemble median and 15 percent greater than the deterministic at a 1-day lead time. Forecasts are less biased on Non-AR and Other AR days and have a greater tendency toward overprediction than underprediction (Table 3). These biases are further illustrated by rank histograms, which provide an indication of how the magnitudes of observed flows compare to individual ensemble members. For the top 10 percent flow AR days at 1-day lead time, the

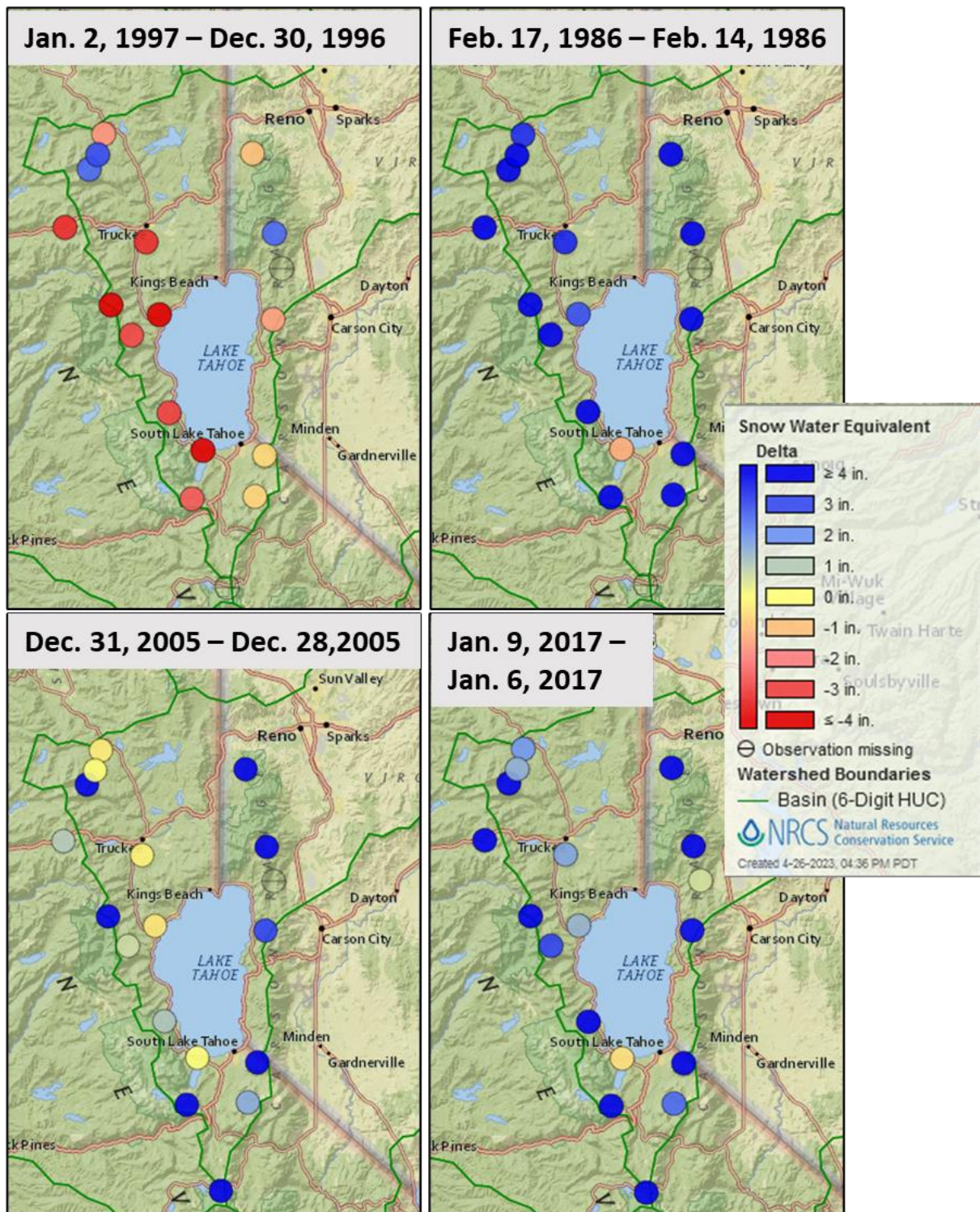


Figure 5. Change in snow water equivalent (SWE) 3 days prior to flood peaks for each of four flood-producing AR events. Warm colored dots indicate decreases in SWE during the storm event and blue dots indicate increases.

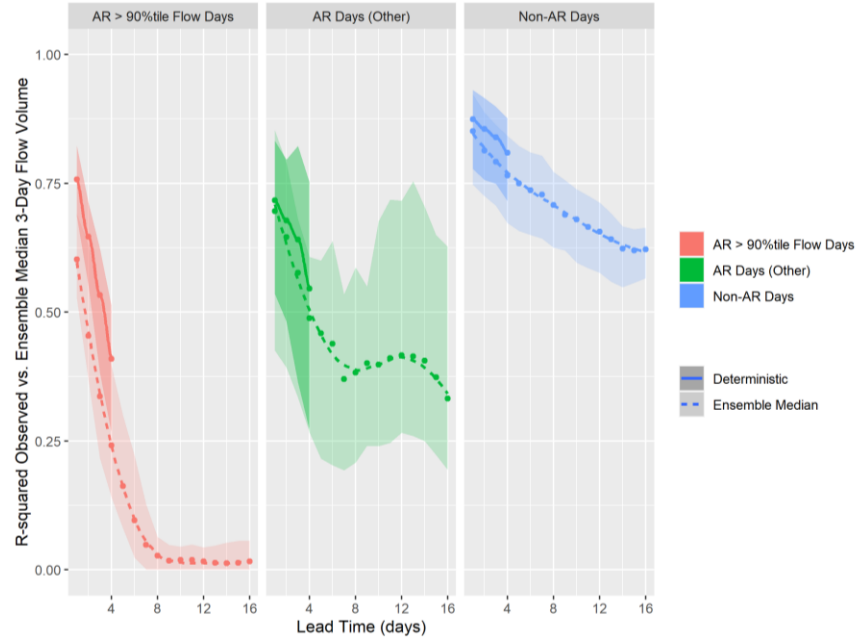


Figure 6. Average and range of R^2 values between forecast and observed 3-day flow volumes as a function of forecast lead time for the ensemble median (5-sites) and deterministic (3-sites).

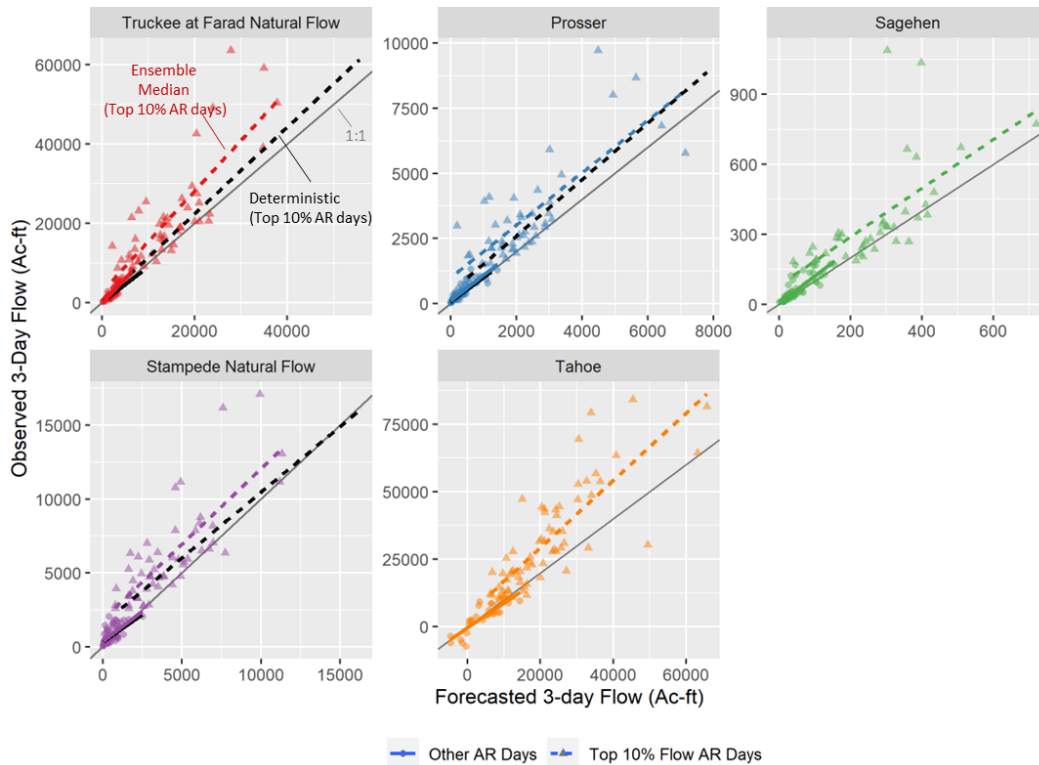


Figure 7. Scatter plots of observed and forecasted 3-day flow volumes and best-fit linear regression lines for top 10% AR days for the deterministic (black dashed line) and ensemble median (colored dashed line) forecasts at a 1-day lead time. The gray line indicates the 1:1 line indicating perfect correspondence between forecasted and observed. In all cases, both best fit lines fall above the 1:1 line, indicating an underprediction bias. The deterministic best-fit line tends to fall closer to the 1:1 line, indicating less bias and higher accuracy.

Table 3. R^2 between forecast and observed 3-day flow volumes at 1- and 4-day lead times for the ensemble median and deterministic forecasts at select sites and for subsets of AR days with 90th percentile or greater flow magnitudes, other AR days, and non-AR days. Green to red color ramps indicate high to low skill, respectively. The orange to blue color ramp indicates bias toward under- and over-forecasting flow magnitudes, respectively.

Site		Average Daily Flow (cfs)		R2						Multiplicative Bias					
				>90%tile AR Days		Other AR Days		Non-AR Days		>90%tile AR Days		Other AR Days		Non-AR Days	
				Deterministic						Deterministic					
1- Day		4- Day		1- Day		4- Day		1- Day		4- Day		1- Day		4- Day	
Truckee-Farad Prosser Res. Stampede Res.	506	0.82	0.45	0.78	0.75	0.93	0.88	0.89	0.65	1.14	1.12	1.06	1.02		
	76	0.69	0.27	0.83	0.61	0.91	0.84	0.80	0.61	0.96	0.92	1.00	0.93		
	138	0.77	0.51	0.53	0.27	0.78	0.72	0.85	0.66	1.10	1.08	1.05	1.02		
	Ensemble Median						Ensemble Median								
	1 Day		4- Day		1- Day		4- Day		1- Day		4- Day		1- Day		4- Day
Truckee-Farad Prosser Res. Stampede Res. Sagehen Creek Lake Tahoe	506	0.62	0.20	0.85	0.58	0.92	0.84	0.68	0.47	1.12	1.09	1.03	0.94		
	76	0.55	0.14	0.78	0.52	0.90	0.81	0.70	0.46	0.97	0.90	0.96	0.85		
	138	0.63	0.30	0.43	0.26	0.77	0.67	0.75	0.54	0.97	0.99	0.99	0.98		
	9.5	0.69	0.40	0.70	0.61	0.75	0.73	0.70	0.50	1.10	1.08	1.01	0.94		
	156	0.53	0.17	0.71	0.47	0.93	0.77	0.70	0.42			1.07	0.96		

histograms are left skewed (Figure 8), meaning that the observed flows frequently rank higher than most ensemble members, which indicates underprediction bias. This bias increases at longer lead times (Figure 9) when the forecast is more heavily influenced by climatology. In the case of Other AR days (and non-AR days – not shown), the U-shape of the rank histograms is due to the observed flows more frequently ranking among the highest or lowest ensemble members, indicating that the ensemble is under-dispersed. For some sites (e.g., Sagehen, Tahoe, Truckee-Farad), this under-dispersion tendency lessens at longer lead time.

Binary metrics of forecast skill for the top 10 percent flow AR days indicate that about 90 percent of the times a top 10 percent flow occurs, it is forecasted to be in the top 10 percent (hit rate) at a 1-day lead time in the case of the deterministic forecasts. In the case of the ensemble median, the hit rate is closer to 80 percent, on average across sites (Table 4, Figure 10). For longer lead times forecasted in the ensemble, hit rates start to improve at 8- to 11-day lead times, depending on the site, but are still unreliable with some sites only reaching a 50 percent hit rate at 4-6 days lead times (Figure 10). Hit rates tend to be lower in the upper basin sites

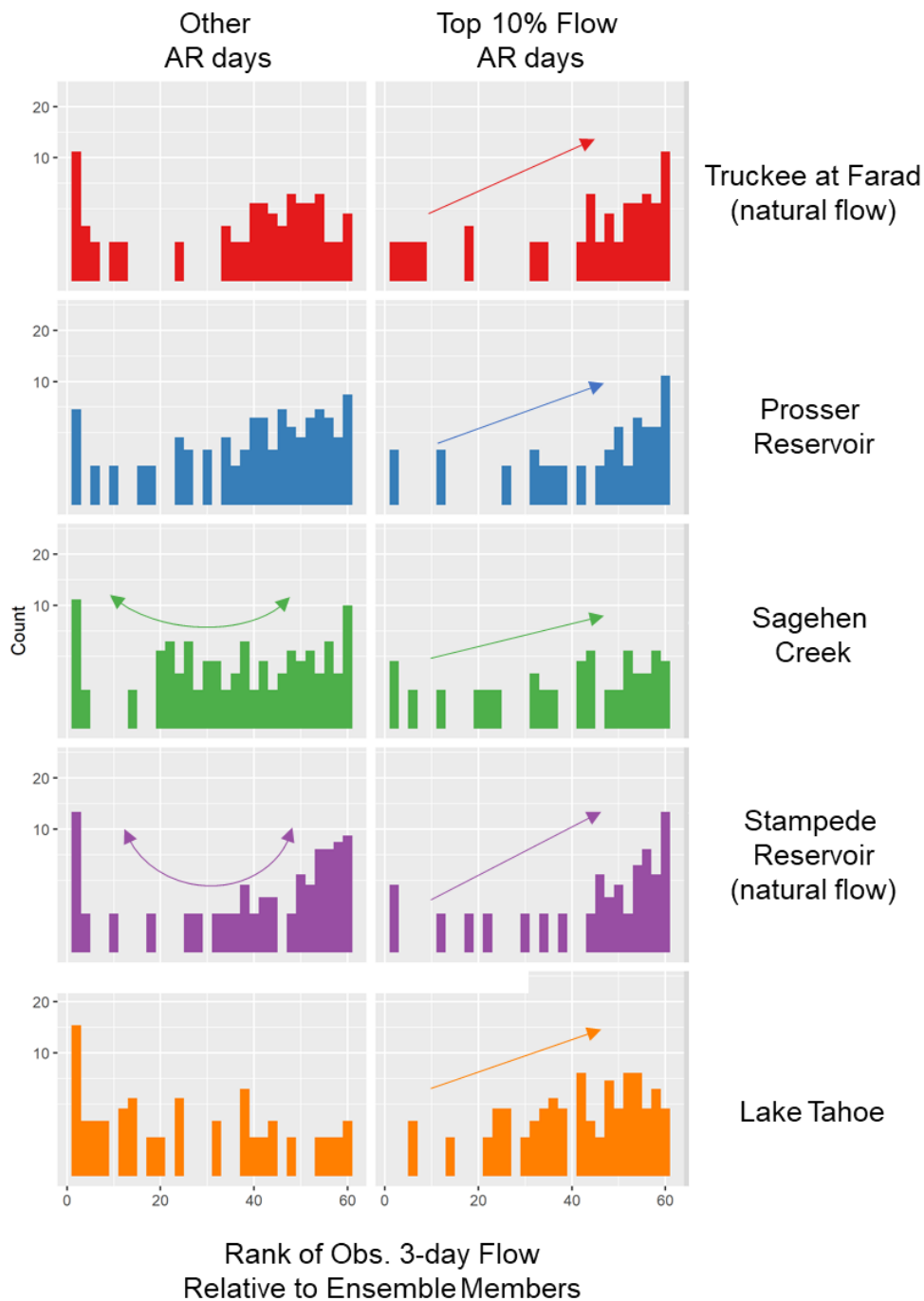


Figure 8. Rank histograms for top 10% flow and other AR days at a 1-day lead time for select sites. The bar heights indicate the frequency with which the observed 3-day flow falls at a given rank relative to the 59 ensemble members. A flat histogram would indicate perfect forecast reliability. The upward slope of the histograms from left to right for most sites on the top 10% flow AR days indicates that the observed flow tends to be higher than most ensemble members (i.e., the ensemble is biased toward underprediction). The U-shape of the histograms for some sites indicate under-dispersion - meaning that the observed flows frequently fall outside of the ensemble range. This is more commonly the case for the lower flow AR days.

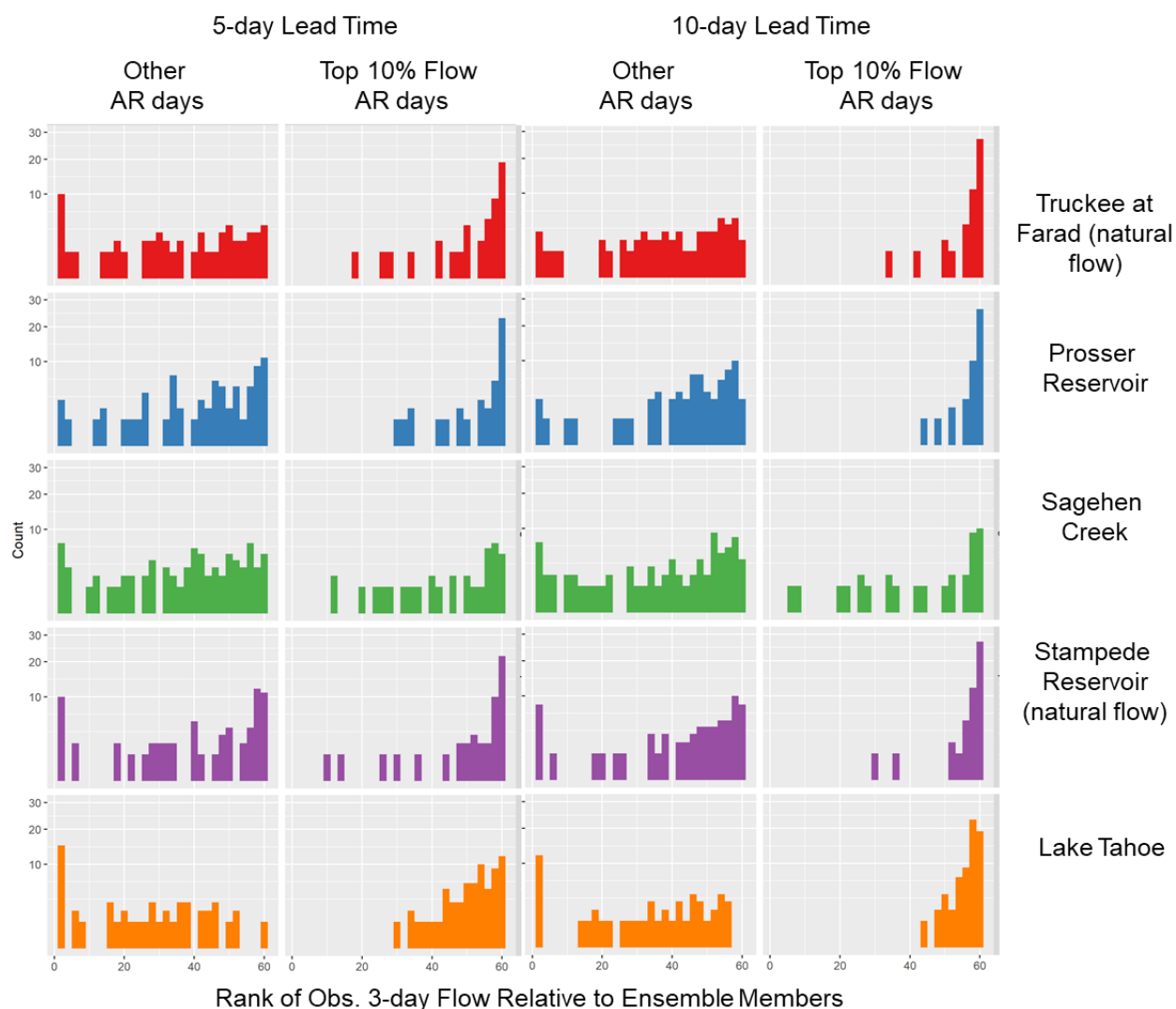


Figure 9. Rank histograms for top 10% flow and other AR days at longer five- and ten- day lead times for select sites. Relative to Figure 8 (1-day lead time), the underprediction bias increases (higher bar heights on the right) at longer lead times, especially for top 10% flow days but the under-dispersion of Other AR days tends to decrease (becomes less U-shaped). See Figure 8 for additional explanation.

(Tahoe, Prosser) and higher in the lower basin sites (Sagehen, Stampede, Truckee-Farad). False alarm rates are low across the board, indicating that when a top 10 percent flow is forecasted, it rarely turns out not to be one. The CSI integrates both hit rates and false alarm rates and given the uniformly low false alarm rates it mirrors the hit rate with improved –but still relatively poor – forecast skill observed at lead times of 8-11 days (Figure 11). Finally, the frequency bias – the ratio of the frequencies of top 10 percent flows observed to those predicted indicate the superior forecast skill of the deterministic model relative to the ensemble median at short lead times, with underprediction bias of the ensemble median increasing somewhat linearly at increasing lead times for most sites (Figure 12).

Table 4. Tabulation of forecast skill metrics at 1- and 4-day lead times for flows greater than the 1980-2019 90th percentile. Green to red color ramps indicate high to low skill, respectively. The green to blue color ramp indicates bias toward under- and over-forecasting the frequency of 90th percentile events, respectively.

	Hit Rate (Q> 90 percentile)		False Alarm Rate (Q> 90 percentile)		Critical Success Index (Q> 90 percentile)		Frequency Bias (Q> 90 percentile)	
	Deterministic							
Site	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day
Truckee at Farad	0.90	0.80	0.03	0.03	0.79	0.69	1.03	0.95
Prosser Reservoir	0.87	0.76	0.01	0.02	0.81	0.67	0.95	0.90
Stampede Reservoir	0.90	0.81	0.03	0.03	0.79	0.68	1.04	0.98
	Ensemble Median							
	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day
Truckee at Farad	0.81	0.69	0.02	0.02	0.74	0.62	0.91	0.80
Prosser Reservoir	0.82	0.55	0.01	0.01	0.76	0.52	0.90	0.59
Stampede Reservoir	0.88	0.76	0.01	0.02	0.83	0.67	0.93	0.91
Sagehen Creek	0.85	0.72	0.03	0.03	0.74	0.63	1.00	0.87
Lake Tahoe	0.65	0.39	0.03	0.02	0.56	0.35	0.81	0.48

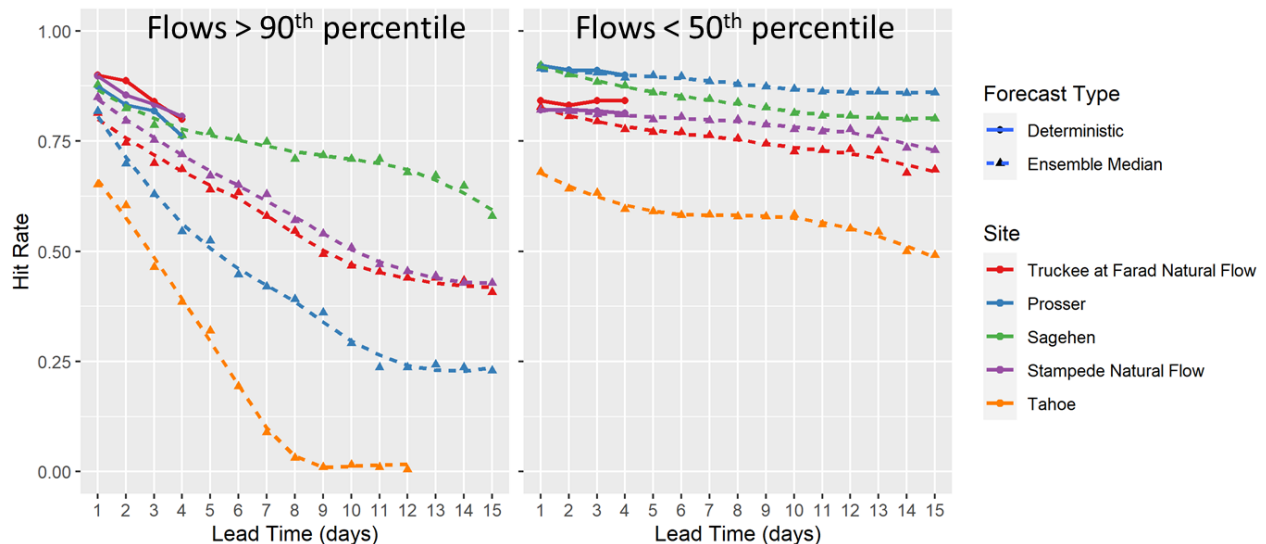


Figure 10. Forecast hit rates at varying lead times. Hit rates for > 90th percentile flows are low at long lead times but clearly improve starting at 8- to 11- day lead times, with the deterministic forecast exhibiting higher skill than the ensemble median. In contrast, hit rates for < 50th percentile flows are consistently high, change very gradually with lead time, and the deterministic and ensemble median have similar skill.

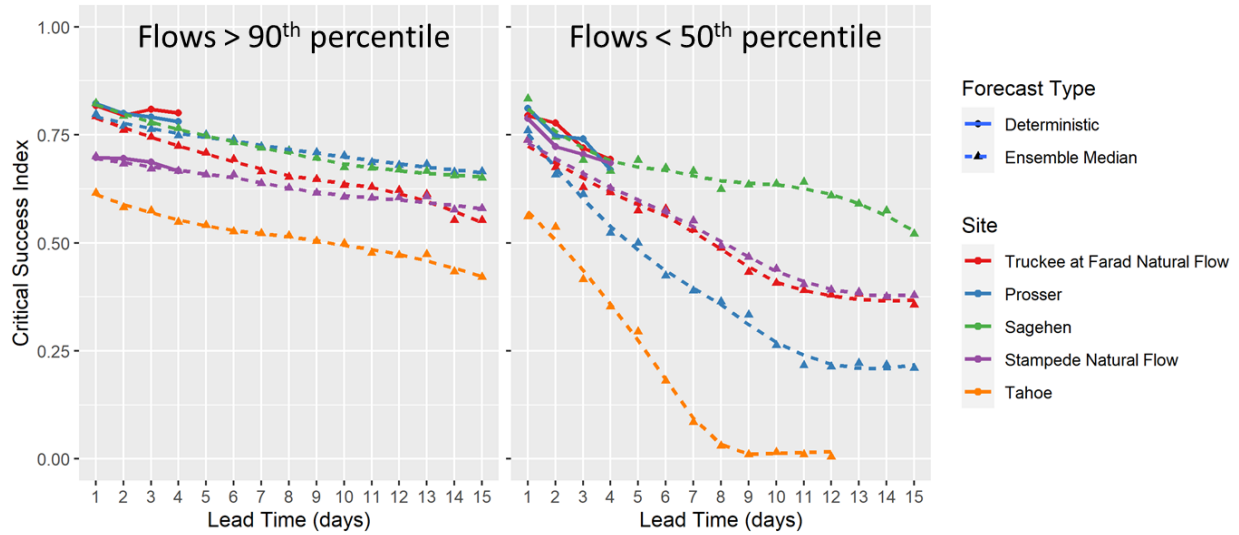


Figure 11. Forecast critical success index (CSI) at varying lead times. CSI values follow similar patterns to the hit rates shown in Figure 10 given that false alarm rates are consistently low and thus have little influence on the CSI.

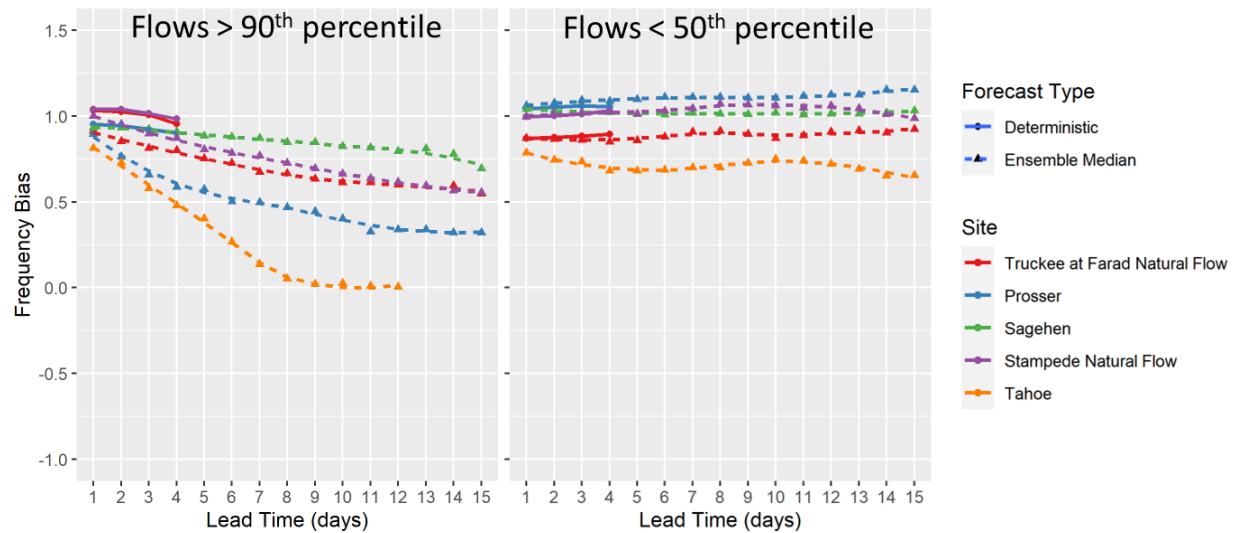


Figure 12. Frequency bias as a function of lead time. Frequency bias for $> 90^{\text{th}}$ percentile flows is low (close to 1) with a tendency toward underprediction by the ensemble median and overprediction by the deterministic forecasts at short lead times. At longer lead times, frequency bias of the ensemble median increases. For $< 50^{\text{th}}$ percentile flows, frequency bias stays relatively small and consistent from short to long lead times and over vs. underprediction bias tends to vary by site rather than by forecast type.

Binary metrics of forecast skill for low flow conditions (< 50th percentile; Table 5) show several patterns that differ from the top 10 percent flows. First, hit rates are consistently higher and do not vary as much with lead time (Figure 10), suggesting a greater ability to predict low-flow events at long lead times relative to high flows. Second, at short lead times, the deterministic and ensemble median show relatively similar hit rates and CSI scores, which points to the ensemble median having less (underprediction) bias at low flows relative to high flows, as indicated by other results in this study. Third, the frequency biases for < 50th percentile flows stay relatively small and consistent from short to long lead times and tendencies toward over vs. underprediction bias tends to vary by site (Figure 12) rather than by forecast type, as is the case for the top 10 percent flows.

The forecast metric results described above provide an abstract view of forecast skill over the 5-year study period. Several large AR storms during the winter of 2017 provided an opportunity to take a closer look at how well the forecast ensembles captured the associated large flow events and the circumstances under which forecast skill varies. Figure 13 enables comparisons of observed, deterministic forecast, and ensemble forecast median and range for 1-, 3- and 5-day lead times for the Truckee River at Farad -the most downstream site in this study. These plots show that most flood peaks were captured in the ensemble range at 1- to 5-day lead times. Observed flows most often fall outside the ensemble range on the rising and falling limbs of the hydrograph and during non-AR periods when the ensemble range is narrow (i.e., due to lack of precipitation and absence of hydrologic uncertainty in the forecast).

Table 5. Tabulation of forecast skill metrics at 1- and 4-day lead times for flows less than the 1980-2019 50th percentile. Green to red color ramps indicate high to low skill, respectively. The green to blue color ramp indicates bias toward under- and over-forecasting the frequency of < 50th percentile events, respectively.

	Hit Rate (Q< 50 percentile)		False Alarm Rate (Q<50 percentile)		Critical Stress Index (Q<50 percentile)		Frequency Bias (Q<50 percentile)	
	Deterministic							
Site	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day
Truckee at Farad	0.84	0.84	0.01	0.02	0.82	0.80	0.87	0.89
Prosser Reservoir	0.92	0.90	0.12	0.15	0.82	0.78	1.04	1.05
Stampede Reservoir	0.82	0.81	0.07	0.09	0.70	0.67	1.00	1.03
	Ensemble Median							
	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day	1-Day	4-Day
Truckee at Farad	0.83	0.78	0.02	0.04	0.79	0.72	0.87	0.85
Prosser Reservoir	0.92	0.89	0.13	0.18	0.80	0.75	1.06	1.09
Stampede Reservoir	0.82	0.81	0.07	0.08	0.70	0.67	0.99	1.03
Sagehen Creek	0.92	0.88	0.11	0.14	0.82	0.76	1.04	1.02
Lake Tahoe	0.68	0.60	0.09	0.08	0.62	0.55	0.79	0.68

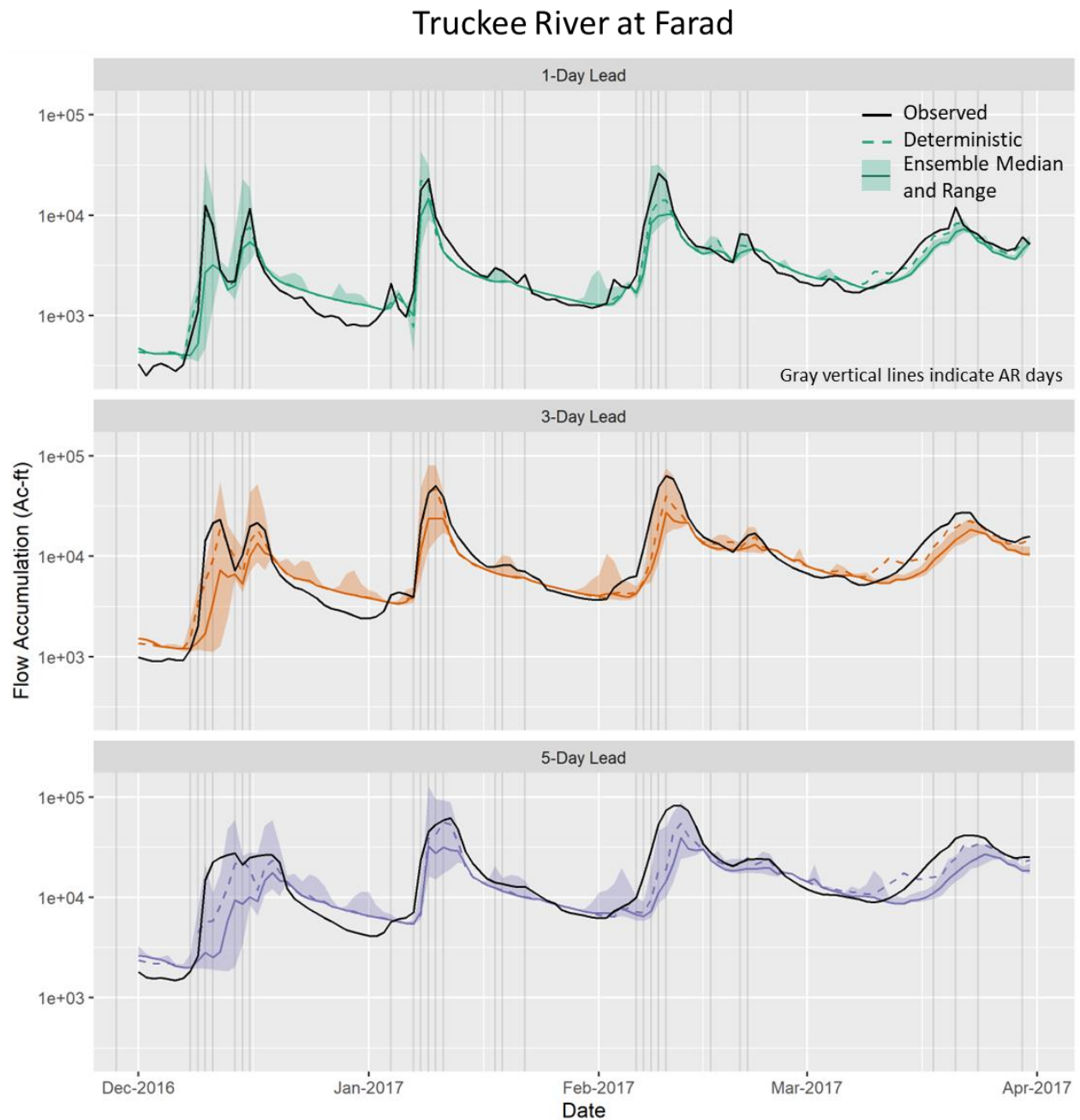


Figure 13. Time series comparisons of observed (black line) and predicted 1-, 3-, and 5- day flow accumulations at corresponding lead times at the Truckee River at Farad gage during winter 2017, when multiple large AR events occurred in the Truckee River Basin. Most observed flood peaks were captured within the ensemble range but observed flows were sometimes outside the ensemble within the rising and falling limbs of the hydrographs.

CONCLUSIONS

Quantifying forecast skill is a key component to assessing the viability of FIRO. In this study, a preliminary analysis of streamflow forecast skill is conducted using 5 years of readily available deterministic and ensemble forecast data. Taken together, the results described above indicate that AR-driven streamflow and flood potential are clearly detectable in the forecast ensembles at 7-day plus lead times but there is an overall tendency for the ensembles to under-forecast flow volumes, even at short lead times, for high flow events that occur during ARs. Under high flow conditions, deterministic forecasts show superior accuracy and less multiplicative bias relative to the ensemble median, but still under-forecast the highest flow events by 15 percent, on average, at a 1-day lead time and 35 percent, on average, at a 4-day lead time (Table 2). Under low-flow conditions, forecast skill is more similar between the deterministic and ensemble median and forecasts tend to have higher accuracy and less underprediction bias. Importantly, low flow conditions tended to have higher skill at longer lead times for most of the metrics assessed, indicating that dry periods are better forecasted at long lead times than large storm events. The ability to forecast prolonged dry periods has the potential to be as valuable for FIRO as forecasting large flow events, depending on the FIRO strategy employed, as accurate forecasting of these periods can indicate potential to retain water supplies within reservoirs (Jasperse *et al.*, 2020).

Differences in data, analytical approaches, and time periods analyzed and the fact that FIRO assessments are still relatively rare make it challenging to make direct comparisons of the results presented above to other studies. The recently completed Lake Mendocino FIRO Final Viability Assessment (Jasperse *et al.*, 2020; Weihs *et al.*, 2019) reports deterministic forecast skill for 24-hour inflows to Lake Mendocino as ranging from R^2 of 0.9 at a one-day lead time to over 0.5 at a five day lead time, but their analysis is ‘dominated by many days with no rain and stable inflows’. These results are on-par with the non-AR day results reported here but how results for higher flow conditions from this study would compare is unknown. The Lake Mendocino Final Viability Assessment also assessed forecast ensemble reliability for inflows, albeit with a slightly different approach afforded by the use of a multi-decadal timeseries of ensemble hindcasts, which were not available at the time the Truckee Basin study described here was conducted. From a qualitative standpoint, the Lake Mendocino study highlighted issues such as underprediction biases and under-dispersion of the ensemble, which are both issues identified in the forecast reliability assessment conducted here. However, the Lake Mendocino study indicates under-dispersion of the ensemble in association with the highest flow volumes and underprediction bias for lower flow volumes which is the opposite of what was observed here (i.e., in the Truckee, high flows are underpredicted and under-dispersion of the ensemble occurs at low flows). The reasons for these differences in unknown but merit further study.

RECOMMENDATIONS

The results presented here represent a first step in assessing forecast skill for the Truckee River Basin in relation to FIRO and many avenues of research still need to be pursued. Important next steps include 1) evaluation of deterministic and ensemble quantitative precipitation forecasts to understand the relative role these play in streamflow forecast skill and uncertainty,

2) expanding analyses of forecast skill to include more years, ideally using multi-decadal hindcast data, 3) understanding the relative contributions of precipitation forecast uncertainty, storm characteristics, and antecedent snow and soils conditions to uncertainties and biases of hydrologic forecasts, 4) contextualizing analyses and results in terms relevant to the FIRO strategies being considered and to the physical and operational constraints of the Truckee River system, and 5) assessing similarities and differences between forecast skill in the Truckee River basin and other places where FIRO studies have previously been conducted. The code and workflows developed as part of this project provide the opportunity to readily address many of these tasks.

REFERENCES

- Albano, C. M., Dettinger, M. D., McCarthy, M. I., Schaller, K. D., Welborn, T. L., & Cox, D. A. (2016). Application of an extreme winter storm scenario to identify vulnerabilities, mitigation options, and science needs in the Sierra Nevada mountains, USA. *Natural Hazards*, 80(2), 879–900. <https://doi.org/10.1007/s11069-015-2003-4>
- Albano, C. M., Dettinger, M. D., & Harpold, A. A. (2020). Patterns and drivers of atmospheric river precipitation and hydrologic impacts across the Western US. *Journal of Hydrometeorology*, 21, 143–159.
- Anderson, E. A. (1973). National Weather Service River Forecast System -- Snow Accumulation and Ablation Model. *NOAA Technical Memorandum NWS HYDRO-17*, US Dept. of Commerce, Silver Spring, MD, 217p.
- Burnash, R. J., Ferral, R. L., & McGuire, R.A. (1973). A generalized streamflow simulation system conceptual modeling for digital computers, U.S. Department of Commerce National Weather Service and State of California Department of Water Resources.
- Dettinger, M., Lavers, D., Rutz, J., Viale, M., Wade, A., & White, A. B. (2019). Effects of Atmospheric Rivers. In F.M. Ralph, M. D. Dettinger, D. E. Waliser, & J. J. Rutz (Eds.), *Atmospheric Rivers* (p. 40). Springer International.
- Gelaro, R., & Coauthors. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30, 5419–5454.
- Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A. & Rutz, J.J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America, *Geophys. Res. Lett.*, 44, doi:10.1002/2017GL074175.
- Gilleland, E. (2015). Package “Verification”: Weather Forecast Verification Utilities v. 1.42.
- Jasperse, J., Ralph, F., Anderson, M., Brekke, L., Malasavage, N., Dettinger, M., *et al.* (2020). Lake Mendocino Forecast Informed Reservoir Operations Final Viability Assessment. UC San Diego.
- Lavers, D. A., Waliser, D. E., Ralph, F. M., & Dettinger, M. D. (2016). Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophysical Research Letters*, 43, 2275–2282. <https://doi.org/10.1002/2016GL067765>

- US. Dept. of Interior, B. of R. (2019). Water SMART - Water Management Options Pilot for the Truckee Basin. Letter Proposal July 2019.
- Ralph, F Martin, Dettinger, M., Cordeira, J. M., Rutz, J. J., Schick, L., Anderson, M., *et al.* (2019). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bull. Amer. Meteor. Soc.*, 100, 269–879. <https://doi.org/10.1175/BAMS-D-18-0023.1>
- Rutz, J. J., & Steenburgh, W. J. (2012). Quantifying the role of atmospheric rivers in the interior western United States. *Atmospheric Science Letters*, 13(4), 257–261. <https://doi.org/10.1002/asl.392>
- Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Monthly Weather Review*, 142, 905–921. <https://doi.org/10.1175/MWR-D-13-00168.1>
- Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2015). The inland penetration of atmospheric rivers over western North America: A lagrangian analysis. *Monthly Weather Review*, 143(5), 1924–1944. <https://doi.org/10.1175/MWR-D-14-00288.1>
- Sterle, K., Jose, L., Coors, S., Singletary, L., Pohll, G., & Rajagopal, S. (2020). Collaboratively Modeling Reservoir Reoperation to Adapt to Earlier Snowmelt Runoff. *Journal of Water Resources Planning and Management*, 146(1). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001136](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001136)
- U.S. Bureau of Reclamation (USBOR). (2016). Chapter 9: Truckee River Basin. -Reclamation Climate Change and Water 2016.
- Weihs, R. R., Reynolds, D. W., Hartman, R., Scott, L., Kozlowski, D., & Ralph, F. M. (2019). Assessing Quantitative Precipitation and Inflow Forecast Skill for Potential Forecast-Informed Reservoir Operations for Lake Mendocino. ERDC/LAB TR-0X-X

STANDING DISTRIBUTION LIST

Ian Floyd
 US Army Engineer Research and
 Development
 Coastal and Hydraulics Laboratory
 3909 Halls Ferry Road
 Vicksburg, MS 39180
Ian.E.Floyd@erdc.dren.mil
Travis.A.Dahl@erdc.dren.mil
Elissa.M.Yeates@erdc.dren.mil
Cary.A.Talbot@erdc.dren.mil
joseph.c.forbis@usace.army.mil

Nevada State Library and Archives
 State Publications
 100 North Stewart Street
 Carson City, NV 89701-4285
NSLstatepubs@admin.nv.gov

Archives Getchell Library
 University of Nevada, Reno
 1664 N. Virginia St.
 Reno, NV 89557
cklenke@unr.edu

Document Section, Library
 University of Nevada, Las Vegas
 4505 Maryland Parkway
 Las Vegas, NV 89154
sue.wainscott@unlv.edu

†Library
 Southern Nevada Science Center
 Desert Research Institute
 755 E. Flamingo Road
 Las Vegas, NV 89119-7363

***All on distribution list receive one PDF copy,
 unless otherwise noted.***

† 2 copies