U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-CM-23-003

# FY 2023 Third Quarter Performance Metric: Improving Simulations of Atmospheric River Impacts

Paul Ullrich
Alan Rhoades
Chaopeng Shen'
Yalan Song
Julia Szinai
David Yates
Yang Zhou

July 2023

## DISCLAIMER

# Contents



# Figures

# Tables

# 1.0 Product Definition

Atmospheric rivers (ARs) account for the majority of global water vapor transport from the equator to the poles (Zhu and Newell 1998). These vast plumes, often thousands of kilometers long, carry more water than the world's terrestrial rivers. Consequently, ARs are important sources of water to communities throughout the western United States. AR-driven precipitation is estimated to contribute to as much as 50% of western United States total annual water resources (Dettinger et al. 2011) and is largely responsible for the interannual variability of mountain snowpack, one of the West's largest natural reservoirs of water (Siirila-Woodburn et al. 2021). However, ARs can also be hazardous to life and infrastructure, contributing to more than 80% of flood-related damage in the same region (Corringham et al. 2019).

U.S. Department of Energy (USDOE) investments in modeling of the Earth system have greatly improved our capability to simulate and understand ARs and their impacts. Development of the Energy Exascale Earth System Model (E3SM), in conjunction with novel model evaluation and analysis methods, has led to improvements in simulated AR climatology and a capability to simulate and experiment with individual decision-relevant AR events in a global Earth-system model. Investments in land-surface and hydrologic modeling systems, including modern process-based models and models built upon cutting-edge machine learning (ML) approaches, have also allowed us to better capture total available water resources and estimate flood risk. Furthermore, support for the development of water management models has driven insights into the interplay between water availability and demand in the coupled human-Earth system. Using these tools, novel scientific research has led to a deeper understanding of AR impacts, the sensitivity of those impacts to changes in the large-scale environment, and potential vulnerability to AR hazards. These insights are important to inform practitioners and stakeholders about potential future socioeconomic and infrastructural risks from ARs.

Prior work has shown that in order to resolve landfalling AR-induced impacts on coastal communities, fine horizontal resolution (≤28 km) is needed in Earth system models (Rhoades et al., 2020a, 2020b, 2021, 2023). This is because finer horizontal resolutions allow Earth system model simulations to better resolve features and processes, such as coastal topography, air-sea contrast, rain-snow partitioning, and hydrological responses, which are responsible for shaping AR impacts (Demory et al. 2014). Finer-horizontal-resolution simulations are computationally affordable using regionally refined mesh capabilities of the Energy Exascale Earth System Model (RRM-E3SM). The RRM-E3SM has enabled simulations with local grid spacing as fine as 3km, ensuring land-atmosphere feedbacks related to rapidly varying topography are captured with high fidelity.

In the first 2023 quarterly report, the performance of E3SM in its low-resolution (LR) and high-resolution (HR) configurations was examined for ARs, with results demonstrating substantially improved performance for the newer HR configuration. In the second 2023 quarterly report, RRM-E3SM was shown to exhibit similarly improved performance to the HR configuration at a fraction of the computational cost. In contrast, this third 2023 quarterly report focuses on AR impacts, which are examined in two ways: first, a suite of impacts-relevant climatological metrics and diagnostics for ARs in the western United States are used to assess E3SM's ability to simulate AR climatology; and second, this report evaluates RRM-E3SM performance for modeling an AR event that produced widespread impacts throughout California – namely, the historic 1997 New Year's Flood. For reference, this report uses the

European Centre for Medium-Range Weather Forecasts (ECMWF) version 5 land analysis (ERA5-Land) at 9km horizontal grid spacing, U.S. Snow Telemetry (SNOTEL) data, and U.S. Geological Survey (USGS) stream gage data.

This report provides clear evidence of USDOE advancements in simulating AR impacts. At the climatological scale, E3SM performs well at capturing the fraction of precipitation, snowpack, and runoff from ARs, although at higher resolutions E3SM tends to produce more frequent and more intense storms than ERA5. These long-running simulations further include ARs that are similarly extreme to some of the most extreme ARs in the historical record. At the weather time scale, E3SM is demonstrably effective at simulating a high-impact historic AR event and, in conjunction with a suite of DOE-supported water management and hydrologic models, closely matched streamflow and reservoir inflow, and flagged areas of flood risk associated with the event. Meteorological inputs from RRM-E3SM even yield streamflow forecasts that better matched observations than a popular gridded meteorological product. All DOE-supported water models show clear improvement over presently employed process-based and machine learning-based alternatives.

# 2.0 Product Documentation

This report examines simulated AR impacts on both climatological and weather (i.e., event-focused) timescales. Notably, information derived from both analyses have utility to stakeholders for evaluating infrastructure vulnerabilities and for adaptation planning.

At climatological timescales, we analyze results from E3SMv2 in its low-resolution configuration (E3SM-LR, 110km grid spacing) and a regionally refined configuration (RRM-E3SM, 14km grid spacing over the contiguous U.S.) against observations from the western U.S. SNOTEL station network (Hufkens 2022) and reanalysis data from ERA5-Land (Muñoz 2019). Simulations cover the near-term historical period of 1986-2014. The E3SM simulations are run under Atmospheric Model Intercomparison Project (AMIP) protocols with prescribed sea surface temperatures and sea-ice extents, and fully coupled atmosphere (EAM) and land surface (ELM) models.

For the climatological study, we investigate to what degree our models capture AR impacts on water resource availability in the western U.S., including statistics of total precipitation and snow water equivalent. This study also explores ARs under the Ralph et al. 2019 AR category scale, which categorizes ARs into beneficial or hazardous to water resource management using integrated vapor transport and a USDOE-supported AR tracking algorithm, TempestExtremes (Ullrich et al. 2021). Finally, we examine the effect of ARs on runoff, snow cover fraction, and snow water equivalent: metrics relevant to both water resource management, specifically water supply reliability, and flood risk exposure.

At the weather timescale, we analyze results from E3SMv2 hindcast simulations of the 1997 New Year's Flood, which was driven by a strong, persistent AR. Confirming the ability of a model to simulate specific historical weather events is important to build confidence in the model's ability to represent more general events in a climatological context. Meteorological data from (1) RRM-E3SM, (2) E3SM at uniform coarse resolution and (3) a gridded meteorological product are subsequently fed into three impacts models: the Water Evaluation and Planning (WEAP) model, which simulates hydraulic connectivity between elements of the water management system, the Pennsylvania State University long short-term memory (LSTM) model, which is a pure data-driven ML model for streamflow, and a

cutting-edge, ML-based differentiable parameter learning (dPL) model. The LSTM and dPL models are subsequently used for simulating reservoir inflows during the flood event. These models are then evaluated against USGS streamflow observations during the flood event using conventional metrics such as Nash-Sutcliffe Efficiency and differences between peak flow.

# 3.0 Results

## 3.1 Climatological Representation of AR Impacts

We first compare E3SM-LR and RMM-E3SM simulations to observations from the western United States Snow Telemetry (SNOTEL) station network. Historical hydroclimatic simulation benchmarks include annual total precipitation and annual peak snow water equivalent (SWE), two key contributors to water resource availability and reliability in the West (Siirila-Woodburn et al. 2021). Figure 1 shows the distribution of values from E3SM-LR and RRM-E3SM against SNOTEL stations over 1986-2014 (evaluated at the nearest grid cell). E3SM-LR and RRM-E3SM systematically underrepresent both precipitation and SWE across SNOTEL station locations. However, RRM-E3SM matches more closely to observations than E3SM-LR, attributed to a better spatial representation of major mountain regions (mean elevation and spatial variability). A better representation of mountainous regions allows for storms to be more realistically orographically uplifted and the magnitude and phase (rain versus snow) of precipitation to build snowpack to be more in line with SNOTEL station observations.

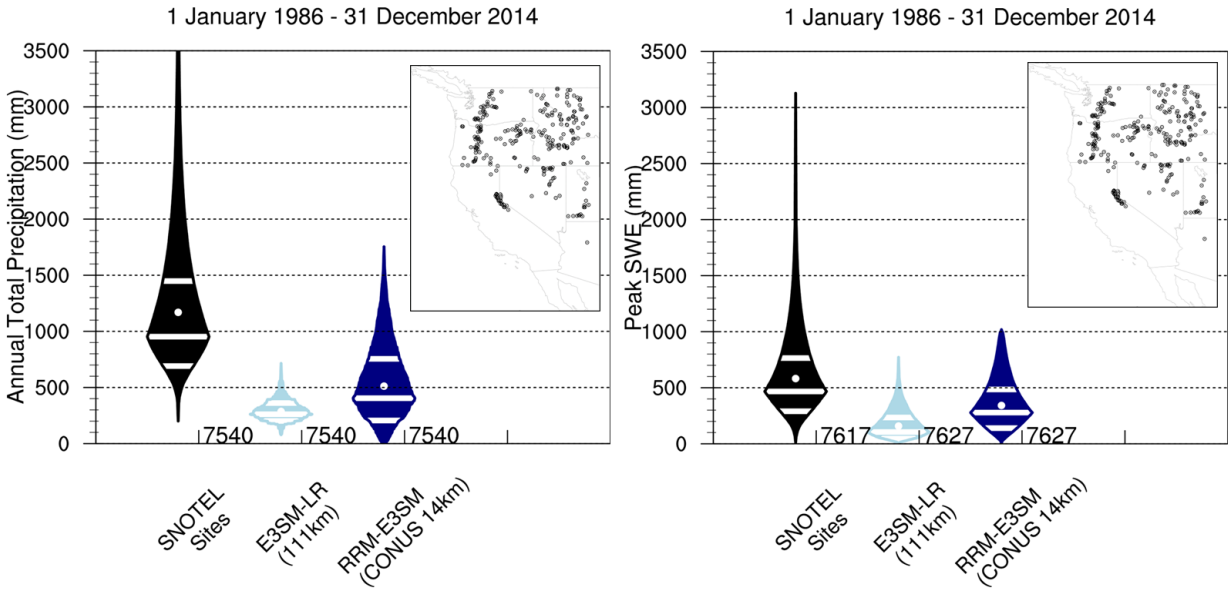


**Figure 1.** E3SM (111km) and RRM-E3SM (14km continental United States [CONUS]) historically simulated estimates of annual total precipitation and annual peak snow water equivalent (SWE). 260 (263) SNOTEL stations provide annual total precipitation and annual peak SWE observations across the western U.S. Snow Telemetry (SNOTEL) network for 1986-2014. Each violin plot represents, from the bottom up, the minimum, 25th percentile (white line), median (white line), average (white dot), 75th percentile (white line) and maximum values across the 29-years and 260-263 SNOTEL stations.

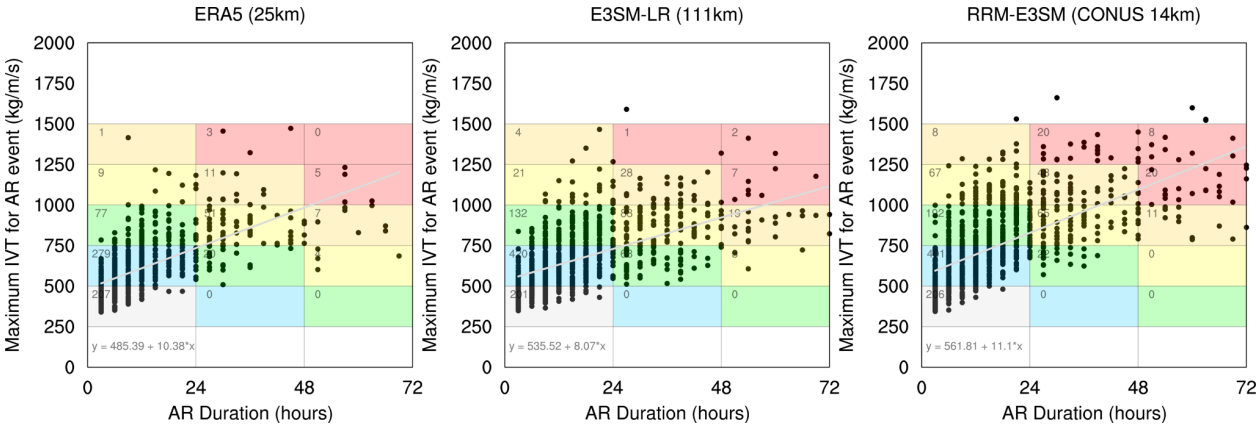


**Figure 2.** Western U.S. landfalling AR climatology for ERA5 (25km), E3SM (111km), and RRM-E3SM (14km CONUS). Each dot represents a single landfalling AR. Dots are overlain on the Ralph et al. 2019 AR category scale designation ranging from category 1 to 5. The category scale relates maximum integrated vapor transport over an AR landfall (y-axis) with the AR landfall duration (x-axis). Category 1-2 (blue/green boxes) are ARs that are mostly beneficial to water resources. Category 3 (yellow boxes) are ARs that are a balance between beneficial and hazardous. Category 4-5 (orange/red) are ARs that are primarily hazardous to water resources.

Figure 2 highlights the distribution of western United States landfalling ARs broken down by AR intensity and duration under the Ralph et al. 2019 AR category scale. The category scale was designed to better convey an AR's potential to either generate beneficial (category 1-3) or hazardous (category 3-5) outcomes to water resources. E3SM-LR produces more ARs that reach at least category 1 on landfall (783 events) than ERA5 (467 events), a result consistent with the second quarterly report. However, E3SM-LR's distribution of ARs among categories is quite similar to ERA5 when normalized by the total number of landfalling ARs to reach at least category 1: of this subset of ARs, E3SM-LR (ERA5) estimated that 78% (80%) of all landfalling ARs reached category 1-2, 14% (14%) reached category 3, and 8% (6%) reached category 4-5. RRM-E3SM produces even more ARs reaching at least category 1 on landfall (862 events). These events were also stronger, on average – among the subset of ARs reaching category 1, 72% reached category 1-2, 15% reached category 3, and 13% reached category 4-5. This indicates RRM-E3SM may overestimate flood risk from ARs and underestimate the frequency of beneficial ARs. This propensity for producing strong ARs appears to be, at least in part, attributable to higher tropospheric wind speeds and column-integrated moisture at high resolution in E3SM, as documented in earlier reports.

The increase in skill in precipitation and snowpack highlighted in Figure 1 is partly owed to better representations of landfalling AR-related land surface impacts. Figure 3 shows the snow-covered area (SCA) during AR landfalls, the AR-contributed changes to SWE (dSWE), and runoff efficiency (total annual runoff/total annual precipitation). Unlike E3SM-LR, RRM-E3SM largely captures the spatial variability of ERA5-Land. Without regional refinement, SCA during AR landfalls is generally non-existent in maritime mountains such as the Cascades and Sierra Nevada. Similarly, both the dSWE and runoff efficiency contributed by ARs is more muted and with less topographic imprint. For dSWE, RRM-E3SM shows that landfalling ARs contribute about 30-40% snow accumulation over the Sierra Nevada, which is consistent with ERA5-Land. Notably, RRM-E3SM also captures rain-on-snow effects from "warm" ARs (namely, negative dSWE and enhanced runoff), which are a driver of the most severe historic floods (Rhoades et al. 2023). On the other hand, while E3SM-LR roughly captures the AR

contribution over mountain regions in Washington, it fails to represent the AR contribution to dSWE over the Sierra Nevada. Although both E3SM-LR and RRM-E3SM underestimate the runoff efficiency compared to ERA5-Land, RRM-E3SM demonstrates great improvement in the representation of runoff efficiency by showing relatively higher runoff efficiency over mountainous regions, which agrees with ERA5-Land.

An even finer horizontal resolution than was employed in RRM-E3SM, i.e., O(10km) to O(1km), is likely needed to more precisely resolve the relationship between AR orographic precipitation and the build-up of mountain snowpack. At present, this is not feasible in RRM-E3SM for centennial-length climate projections. Therefore, to evaluate the necessary and sufficient model horizontal resolution needed to represent landfalling AR characteristics, rain-snow partitioning, and AR-associated impacts, we employ a weather-event-based or storyline approach.

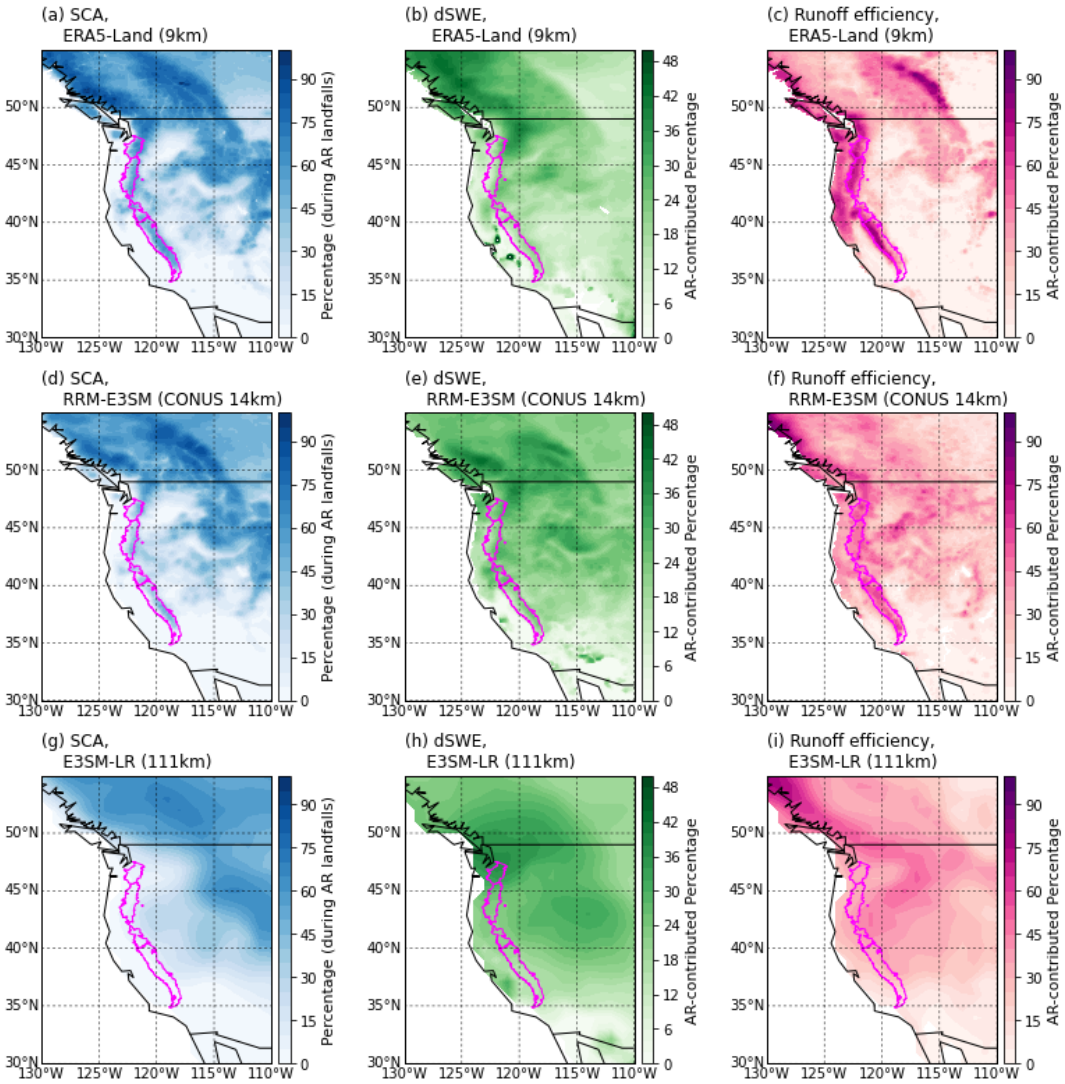


**Figure 3.** Climatological patterns of (left column) snow-covered area (SCA) during AR landfalls, (middle column) AR-contributed change of snow water equivalent (dSWE), and (right column) runoff efficiency during AR landfalls for (a-b) ERA5 Land (9km), (d-f) RRM-E3SM (14km CONUS), and (g-i) E3SM-LR (111km). The magenta outline denotes the Cascades and Sierra Nevada mountain ranges.

## 3.2 Simulating the 1997 New Year's Flood

Climatological metrics are important tools for ensuring models can capture the frequency and character of atmospheric rivers at a high level. However, practitioner and stakeholder decisions are often guided by historic events that had exceptional impact. Throughout the western United States, AR events are the leading cause of flood damage (Corringham et al. 2022). This suggests that our ability to quantify vulnerability to flooding depends on our capability to model flood impacts from the most extreme AR events. The most extreme ARs are inherently rare, with only a handful available in the historical record. Thus, it is impossible to assess model performance on these extremes using climatological metrics alone. To evaluate model performance on such high-impact extremes, we can instead simulate these events as they occurred historically and compare the simulated storm to observations from the time (a procedure commonly referred to as hindcasting).

The New Year's Flood that occurred in January 1997 is perhaps the most memorable AR-driven flooding event in recent history. As many of the tributaries of the Sacramento River flow westward from the Sierra Nevada Mountains onto the valley floor, the risk of flooding has long been known. Flood events in 1907 and 1955 gave rise to the development of an extensive levee system throughout the valley, with the hopes of minimizing property damage and eliminating loss of life. Despite these efforts, the New Year's flood still devastated various regions of California, with heavy rain and melting snow causing flooding along the Sacramento River and its tributaries, and widespread damage, evacuations, and disruptions to transportation. The 1997 New Year's Flood was particularly bad for the Sierra watersheds to the northwest of Sacramento, with more than 40 inches of precipitation reported in portions of the Feather River, and where below the Oroville Dam, thousands of people were evacuated due to the risk posed by the severe flooding and a possible dam failure (Hunrichs et al. 1998). Near Marysville, at the confluence of the Feather and Yuba rivers, three people died and roughly 1,000 homes were destroyed following a levee break on the eastern bank of the Feather River (Figure 4). Many communities were evacuated, flooding left houses and neighborhoods underwater, and loss of life occurred due to the disaster. This type of compound extreme event might become more commonplace in the future, warranting further investigation with modeling approaches that can help build infrastructure resiliency.

In this section we evaluate two approaches for modeling impacts associated with the New Year's Flood of 1997. In the first approach we use the WEAP coupled hydrologic and water systems model. In the second approach we show how machine learning methods can be used to improve our simulations of flooding during these events. A summary of the relevant simulation details is given in Table 1.

**Figure 4.** The town of Arboga in Yuba County is inundated following a levee break on the eastern bank of the Feather River. Photo by the California Department of Water Resources, public domain.

**Table 1**. Summary of 1997 New Year's Flood event and simulation configuration.

| | | | |
|---|---|---|---|
| Time period | 1996-1997 | | |
| Forcings | RRM-E3SM simulation | E3SM-LR simulation | gridMET |
| Area of focus | Sacramento/San Joaquin Basin<br>Feather/American River | | |
| Streamflow gage data and Reservoir inflow data | U.S. Geological Survey (USGS) GAGES-II data set<br>California Department of Water Resources (DWR) California Data Exchange Data Center (CDEC) reservoir data | | |
| Meteorological variables | Precipitation, maximum daily temperature, minimum daily temperature, near-surface wind speed, near-surface vapor pressure, day length | | |
| Hydrologic models | WEAP, LSTM, differentiable parameter-learning (dPL) model | | |
| Metrics | Nash-Sutcliffe Efficiency (NSE). Peak-flow bias. | | |

## 3.3 Investigating Streamflow and Flooding with the WEAP Model

WEAP is a class of models commonly referred to as "systems" or "water management models." Water management models differ from more process-oriented hydrologic models in that they include the

representation of human interventions into the hydrologic cycle. These interventions, prominent throughout California, include reservoirs, diversions, canal networks, levees, tunnels, and pumps, which are used to capture, store, and move water and protect vulnerable areas from flooding (Yates et al. 2005, 2021). WEAP operates on the basic principles of water balance and can represent complex river basin systems. Moreover, WEAP can simulate a broad range of natural and engineered components of these systems, including rainfall runoff, baseflow, and groundwater recharge from precipitation; flood extent; sectoral demand analyses; water conservation; water rights and allocation priorities; reservoir operations; hydropower generation; pollution tracking and water quality; vulnerability assessments; and ecosystem requirements.

The water systems of California are managed through a complex set of rules and governance structures to maximize potential water resource benefits while protecting vulnerable areas from flooding. The Oroville and New Bullards Bar Reservoirs are examples of this interplay, where operators seek to coordinate their management such that releases do not exceed downstream levee design capacities. In the case of major flooding throughout the region, controlled releases from Lake Oroville should be regulated in such a way that 1) the Feather River flows do not exceed a flow objective of 180,000 cubic feet per second (cfs) above the Yuba River confluence at Marysville and that 2) the combined flows in the Feather and Yuba Rivers do not exceed 300,000 cfs in the Feather River downstream from Marysville.

Figure 5 is a map of the Sacramento River Basin, and shows the key tributaries and locations identified for this study, while Figure 6 is a similar map schematic, but shows how the region is represented in the WEAP model. The WEAP model subdivides the watersheds into hydrologic response units (HRUs) or sub-catchments, land use, and 1000m elevation bands. The intersection of these sub-catchments, elevation bands, and land use results in more than 500 unique HRUs that are used to simulate the hydrologic response of the river basins to the meteorological forcing, including both to and below the key reservoirs. Table 2 and Figure 7 summarize and show the simulated inflows to the primary reservoirs and the flows of the Yolo Bypass, which are emphasized in this study. The Yolo Bypass is a key flood control feature in California's Sacramento Valley where, through a system of weirs, the bypass diverts floodwaters from the Sacramento River away from the state's capital city of Sacramento and other nearby riverside communities. The Yolo Bypass was instrumental in redirecting the January 1997 flood waters away from major population centers, thus minimizing the overall impacts of the flood event.

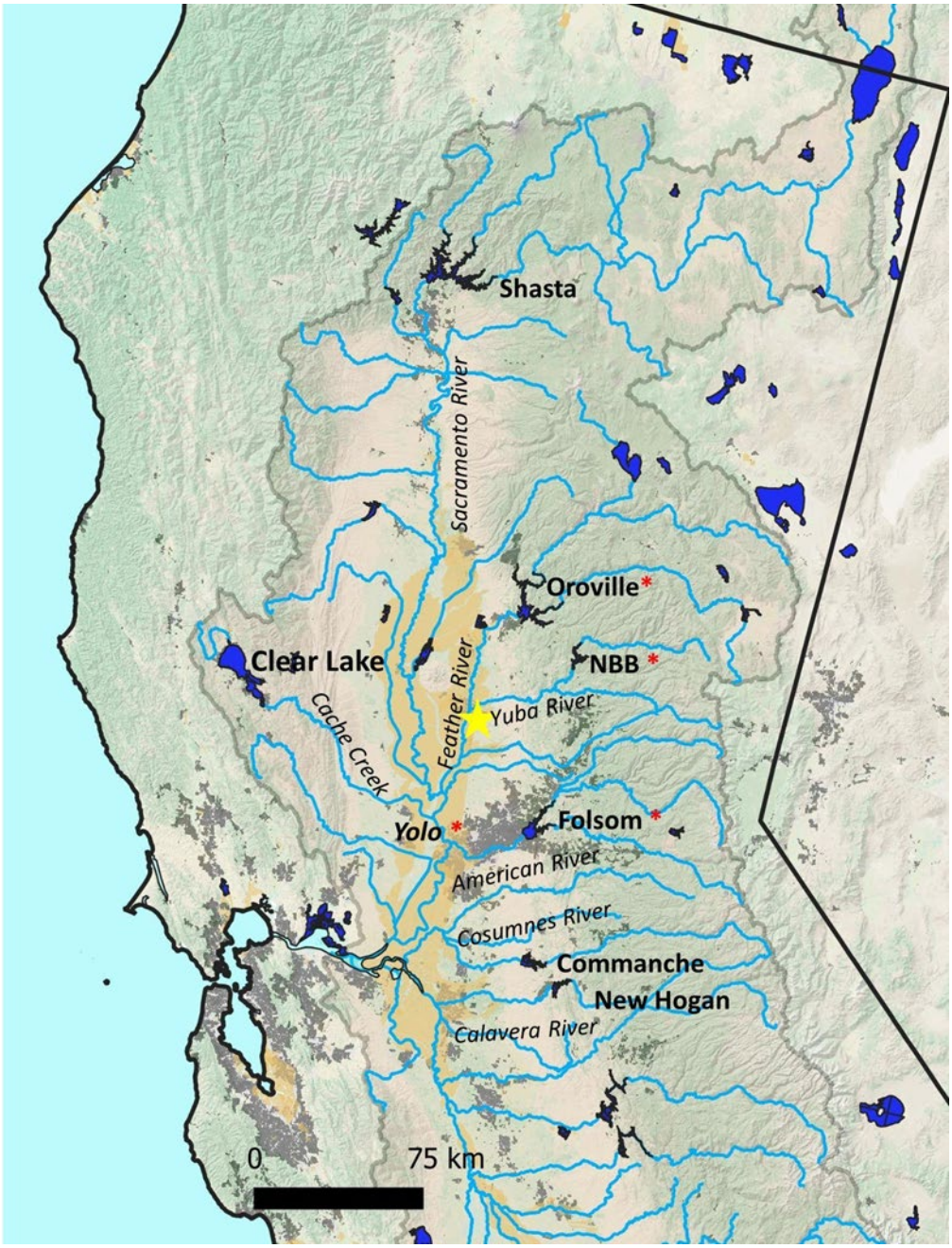


**Figure 5**. Map of the Sacramento River and key tributaries explored in the study. The tan shaded areas near the center of the map are places protected by the extensive levee system, while the gray shaded areas represent populated places. The yellow star marks the confluence of the Yuba and Feather Rivers, where severe levee breaches and flooding occurred, including loss of life. The red stars mark the Feather River inflows to Lake Oroville (ORO), the Yuba River inflows to New Bullards Bar Reservoir (NBB), the American River inflows to Folsom Lake (FOL), and the Yolo Bypass (YOL), which are gauging locations for which the performance of the WEAP model is evaluated for its ability to represent the January 1997 flood event.

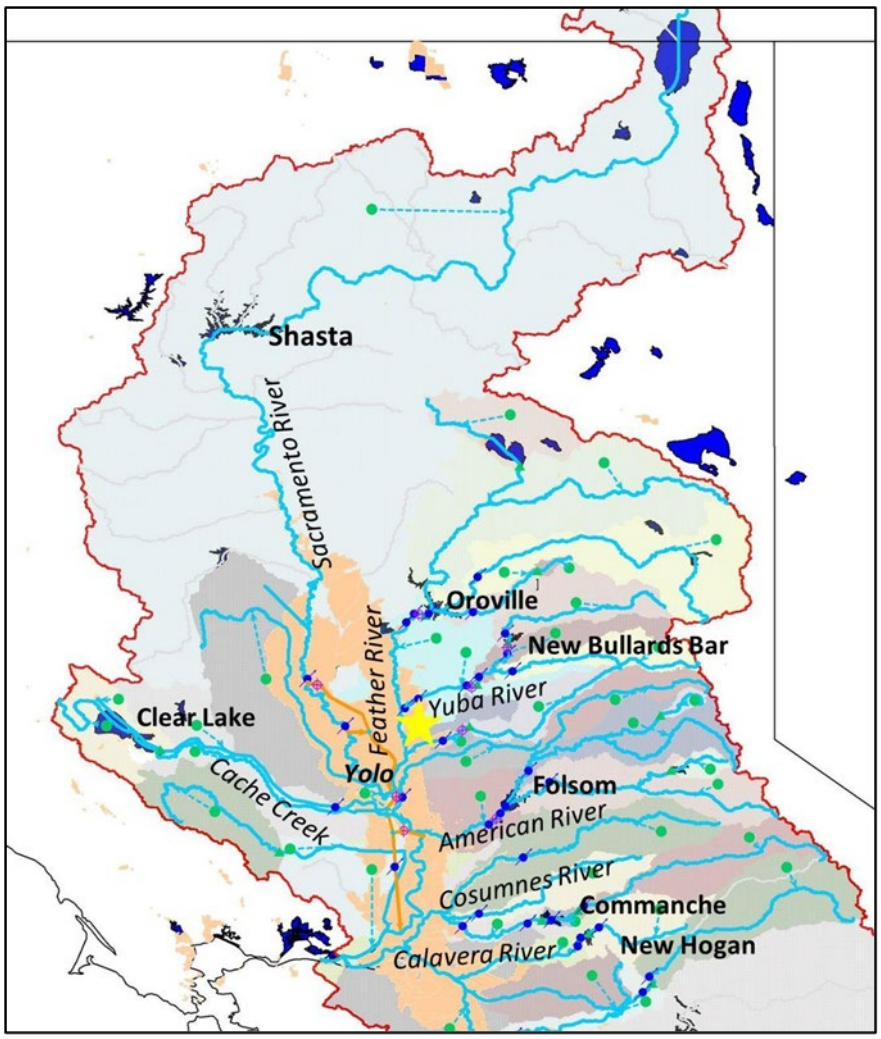


**Figure 6**. Engineering schematic showing some of the detailed representation of the water system as represented by the WEAP model for the same general domain as is shown in the shaded relief map of Figure 5.

The reference meteorological forcing for the WEAP model was the gridMET data sets (Abatzoglou 2013), which were used to force the model starting on 1 October 1995 and extending through 30 September 1997. This implies that two complete water years (i.e., 1996 and 1997) were simulated to evaluate the relatively brief, albeit catastrophic, New Year's flood of 1997. Crucial to understanding the hydrologic response, the model is "spun up" to capture the antecedent soil moisture and snowpack conditions throughout the Sacramento Basin, ahead of the actual flood event. The gridMET meteorological data includes daily estimates of precipitation (mm/day), surface air temperature (°C), relative humidity, and wind speed (m/s). The gridMET, the RRM-E3SM, and E3SM-LR data were mapped to the individual HRUs within the WEAP model, where the daily gridMET precipitation data were simply replaced by the RRM-E3SM and the E3SM-LR data for the flood event, which includes just the four days of 31-Dec-1996 and 1-Jan-1997 to 3-Jan-1997. We note that over the Yuba and Feather Rivers, the peak event precipitation totals for the three data sets-gridMET, RRM-E3SM, and E3SM-LR are generally lower than some of the localized peak estimates with total precipitation that exceeded

40 inches (Hunrichs et al. 1998), although the RRM-E3SM precipitation across the region was considerably greater than the other estimates.

Table 2 provides a summary of performance statistics for the WEAP-simulated flows based on the three different meteorological forcing data sets for four locations highlighted in Figure 5 in red stars, against gauge data from Department of Water Resources' California Data Exchange Center (CDEC). These performance metrics include the Nash-Sutcliffe Efficiency (NSE) and relative peak difference, $d_{max}$, (peak difference between simulation and observation divided by the peak observation, $Q_{max}$). NSE is a commonly used metric for evaluating streamflow, and is defined as:

$$NSE = 1 - \frac{\sum_t (Q_o^t - Q_m^t)^2}{\sum_t (Q_o^t - \overline{Q}_o)^2}$$

where the sums are taken over time, the subscript *m* denotes the modeled streamflow, the subscript *o* denotes the observed streamflow, and the overbar denotes the time mean. Because it is quadratic in the difference between model and observation, it puts more emphasis on differences during high flows.

**Table 2**. WEAP model performance against gauge data from California Data Exchange Center (CDEC) at four locations for the 1996 and 1997 water years for each of the meteorological forcings for the New Year's flood of 1997. The locations include the Feather River Inflows to Oroville Reservoir (ORO), the Yuba River inflows into New Bullards Bar (NBB), American River inflows into Folsom (FOL), and the flows in the Yolo Bypass (YOL), and the performance metrics include the Nash Sutcliffe Efficiency (NSE) coefficient, and the peak flow bias are shown (value of 0.0 would reflect no difference). The best scores under each metric are highlighted in blue.

| **Forcing** | **NSE** | | | | **1997 NYF Peak Flow Bias ($\|d_{max}\|/Q_{max}$)** | | | |
|---|---|---|---|---|---|---|---|---|
| | **ORO** | **NBB** | **FOL** | **YOL** | **ORO** | **NBB** | **FOL** | **YOL** |
| RRM-E3SM 0.0325 | **0.64** | **0.58** | **0.77** | **0.77** | **0.39** | **0.50** | **0.29** | **0.65** |
| E3SM-LR | 0.42 | 0.37 | 0.47 | 0.46 | 0.72 | 0.75 | 0.73 | 0.83 |
| gridMet | 0.51 | 0.42 | 0.64 | 0.69 | 0.66 | 0.63 | 0.56 | 0.72 |

The results suggest that the RRM-E3SM simulations at 0.0325° (3.5km) horizontal resolution are significantly superior to coarser E3SM-LR forcing and the more comparable-resolution gridMET forcing. The superior performance of RRM-E3SM compared to gridMET is likely not accidental and reflects the difficulty in producing gridded precipitation products by interpolating data between observing stations. As noted by Lundquist et al. (2019), at independent observation sites, models generally outperform gridded precipitation estimates by a factor of 2. Specific to ARs, errors in gridded precipitation can be

significant, with underpredicted storms leading to a 20% error in water-year total median California statewide snowfall (Lundquist et al. 2015).

Figure 7 shows the time series of simulated and observed flows into the three key reservoirs: the Feather River inflows to Oroville Reservoir (ORO), the Yuba River inflows into New Bullards Bar (NBB), and American River inflows into Folsom (FOL); and the flows in the Yolo Bypass (YOL). Generally, the WEAP-simulated flows are smaller than the observed flows into these reservoirs, with the flows that are forced with the RRM-E3SM data higher than the other flows, except for the flows into New Bullards Bar (NBB). The peak flows through the Yolo Bypass are similar to the observed flows, although the receding limbs of WEAP-simulated flows tend to be steeper than the observations, suggesting longer persistent flood inundation compared with what WEAP simulated.

Figure 8 shows the simulated flows along the Feather River, above and below Marysville at the confluence of the Yuba River (yellow star in the maps of Figure 5 and 6). The horizontal lines show the regulated flow thresholds that water managers try to stay below to minimize levee breaching and the generation of flood-inundated areas. Here, the superiority of the RRM-E3SM forcing at 0.0325º (3.5km) to drive the WEAP hydrologic model is apparent, as its peak flows are considerably higher than those flows that make use of the gridMET and the E3SM-LR forcing and the WEAP-simulated flows with RRM-E3SM forcing more closely matching the observed peak flows. This simulation highlights the immense risk posed by this flood event at this location, where a levee breach led to widespread flooding and loss of life and property (Figure 4). Figure 8 also includes a scenario that assumes there are no storage reservoirs and forces the WEAP model with the RRM-E3SM data, showing considerably higher flows that far exceeded the acceptable limits, demonstrating the benefits of flood storage.

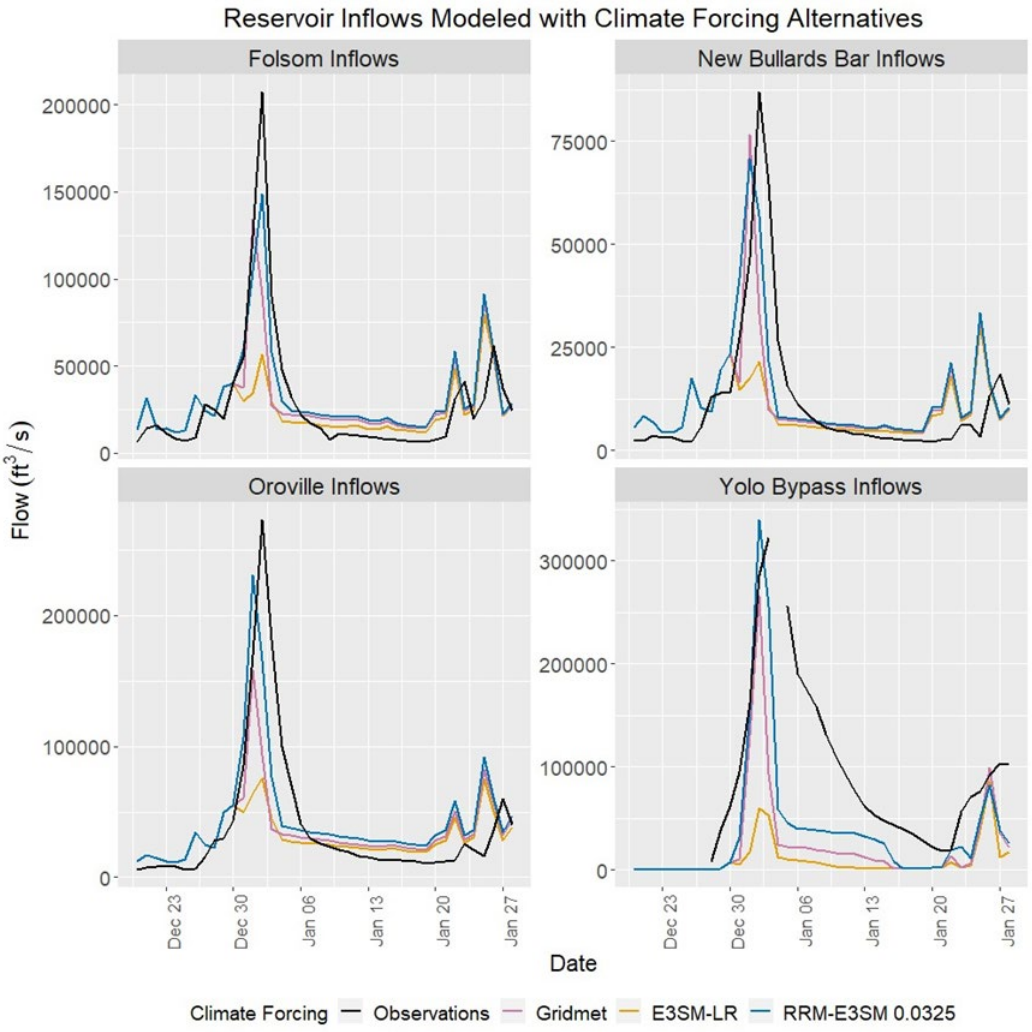


**Figure 7**. Inflows into the four select reservoirs and flood control locations including the American River inflows to Folsom Lake (FOL), the Yuba River inflows to New Bullards Bar Reservoir (NBB), the Feather River inflows to Lake Oroville (ORO), and the Yolo Bypass (YOL).

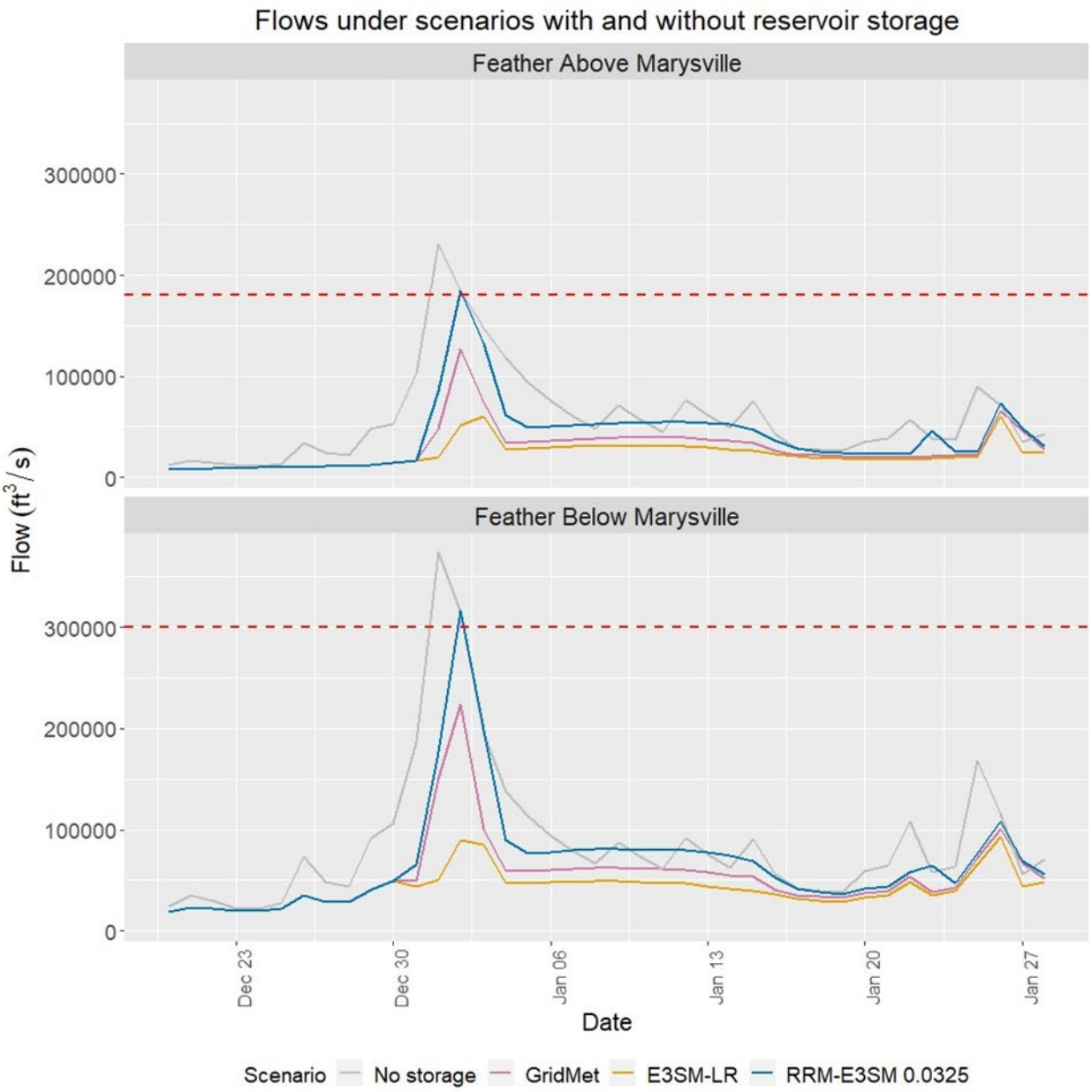


**Figure 8**. Flows under scenarios with and without reservoir storage at two key points: Feather River Above Marysville, and Below the Feather/Yuba Rivers Confluence at Marysville. For the three forcing scenarios (gridMET, E3SM-LR, and RRM-E3SM). The "No Storage" scenario assumes no reservoir storage throughout the basin and uses the RRM-E3SM forcing.

## 3.4 Modeling Streamflow with ML Methods

The ML framework employed in this report was developed with funding from the USDOE and consists of two distinct models: The first model is a purely data-driven model called long short-term memory (LSTM), which is a widely used architecture in hydrology. It takes as inputs atmospheric forcing time series data and static basin attributes, including physiographic attributes and anthropogenic influences (Ouyang et al. 2021). The output of this model is a daily streamflow time series. The second

hydrologic model employed in this study is a type of physics-informed ML model referred to as a "differentiable" parameter learning model (dPL). This model employs structural constraints, e.g., mass balance and energy equations, in conjunction with a data-driven architecture. With dPL models, connected neural networks can be trained together with the physical model priors. One neural network is trained to serve physical parameters, e.g., storage capacity of the unsaturated zone, for all training sites, given dynamical forcings and static attributes as inputs. This neural network is connected to a simple conceptual hydrologic model HBV (Hydrologiska Byråns Vattenbalansavdelning) (Bergström 1992, 1976), which provides outputs such as streamflow, baseflow, recharge, evapotranspiration, and snow water equivalent (Tsai et al. 2021, Feng et al. 2022a, 2022b, Shen et al. 2023). Such dPL models extrapolate better outside of the training regime (Feng et al. 2023) and are expected to perform better for extremes.

The LSTM and dPL models were both trained using data from 777 USGS stations in California. The training data consisted of daily high-spatial resolution (~4-km, 1/24th degree) surface meteorological data from gridMET, static attributes used in Ouyang et al. (2021), and streamflow observations from USGS, spanning the period from 1991 to 2019. The performance of both LSTM and dPL models was evaluated for the 1997 water year, using the Nash-Sutcliffe model efficiency coefficient (NSE) and relative peak difference, dmax (peak difference between simulation and observation divided by the peak observation, Qmax). The disparity in peak values provides an indication of the models' predictive error in estimating streamflow during the 1997 New Year's flood event, which produced peak flow for that water year. Three sets of forcing data were applied: gridMET, E3SM-LR, and RRM-E3SM. For both E3SM data sets, precipitation and temperature data from December 31, 1996 to January 3, 1997 supplanted gridMET data during the flood event. From the upstream Sacramento region, 122 USGS gauge stations associated with 14 reservoirs (Table 3) were selected for evaluation. During the evaluation of the models, we substituted the precipitation and temperature data for the corresponding gridMET data, spanning a 4-day period, to compare the streamflow predictions during the 1997 New Year's flood event, using different forcing data sources.

The comparison led to three conclusions: First, the hydrologic models forced with high-resolution precipitation and temperature data from RRM-E3SM better captured the 1997 New Year's flood for both medium-sized (Table 4 and Figure 9) and small basins (Table 5). More specifically, these simulations produced a noticeably better median peak error metric, indicating that the high-resolution RRM-E3SM simulation served its purpose in better representing the impact of extreme events. Second, the dPL model, which combines deep learning with a physical hydrological model, can noticeably outperform the LSTM (data-driven) model in medium-sized basins where the drainage area is larger than 1000 $km^2$ (Table 4 and Figure 9). The difference in performance is larger within the larger basins (Table 4) than with the smaller basins, presumably because of the difficulty of river routing. This result suggests the physical process representation in the dPL model produces an advantage in representing extreme events. A purely data-driven model like LSTM tends to underpredict the magnitude of events that are outside of the range of the training data. However, the constraints on the neural networks inside the dPL model circumvented this problem. Third, as shown in Table 6, for the 40 basins inside the study domain that overlap with the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) data set (Newman et al. 2015), the ML models yielded much higher performance compared to Sacramento Soil Moisture Accounting (SAC-SMA), a conventional operational hydrologic model used by many Regional River Forecasting Centers, when employing the Daymet climate data set (Thornton et al. 2016).

**Table 3**. Selected reservoirs in the upstream Sacramento region.

| River | Reservoir | Station Name |
|---|---|---|
| Feather River | Lake Oroville | ORO |
| Yuba River | Englebright | ENG |
| Yuba River | New Bullards Bar | BUL |
| Bear River | Camp Far West | CFW |
| MF American River | Folsom | FOL |
| Mokulmne River | Camanche Reservoir | CMN |
| Mokulmne River | Pardee Reservoir | PAR |
| Stanislaus River | New Melones | NML |
| Calavera River | New Hogan Lake | NHG |
| SF Feather River | Little Grass Valley | LGV |
| HW NF Feather River | Lake Almanor | ALM |
| HW SF American River | Union Valley | UNV |
| SF Bear River | French Meadows | FDM |
| Cache Crk | Clear Lake | CLA |

**Table 4**. Model performance in 36 reservoir medium-sized basins (>1000 km$^2$) upstream of Sacramento. The best-performing combination of model and forcing data is highlighted in blue.

| **Model** | **Forcing** | **median (NSE)** | **1997 NYF Basin Median peak error ($\|d_{max}\|/Q_{max}$)** |
|---|---|---|---|
| LSTM | RRM-E3SM 0.0325 degree | 0.721 | 38.0% |
| LSTM | E3SM-LR simulation (ne30) regridded to 0.0325 degree | 0.521 | 60.7% |
| LSTM | gridMET | 0.618 | 37.0% |
| **dPL** | **RRM-E3SM 0.0325 degree** | **0.772** | **20.1%** |
| dPL | E3SM-LR simulation (ne30) regridded to 0.0325 degree | 0.556 | 59.1% |
| dPL | gridMET | 0.741 | 27.9% |

**Table 5**. Model performance in 86 small basins upstream of the reservoirs in the study domain (<=1000 km$^2$) upstream of Sacramento.

| Model | Forcing | median (NSE) | 1997 NYF Basin Median Peak Error ($\|d_{max}\|/Q_{max}$) |
|---|---|---|---|
| LSTM | RRM-E3SM 0.0325 degree | 0.666 | 42.4% |
| LSTM | E3SM-LR simulation (ne30) regridded to 0.0325 degree | 0.305 | 71.1% |
| LSTM | gridMET | 0.583 | 42.6% |
| **dPL** | **RRM-E3SM 0.0325 degree** | **0.666** | **40.5%** |
| dPL | E3SM-LR simulation (ne30) regridded to 0.0325 degree | 0.321 | 77.4% |
| dPL | gridMET | 0.607 | 49.3% |

**Table 6**. Model performance in 40 CAMELS in California.

| Model | Forcing | median (NSE) | 1997 NYF Basin Median ($\|d_{max}\|/Q_{max}$) |
|---|---|---|---|
| LSTM | Daymet | 0.756 | 23.7% |
| dPL | Daymet | 0.780 | 25.8% |
| SAC-SMA | Daymet | 0.677 | 48.8% |

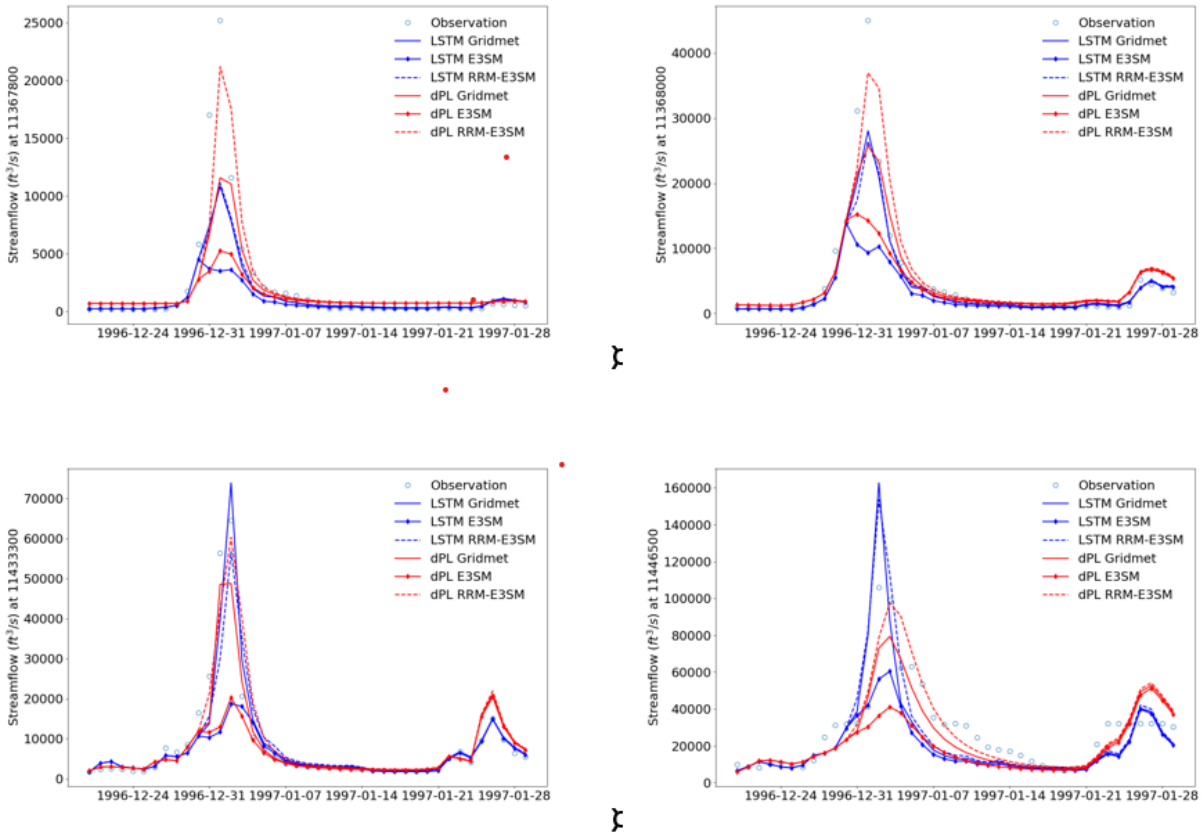


**Figure 9**. Time series of streamflow from observations and models using different forcing data at four medium-sized basins upstream of Sacramento.

# 4.0 Summary

The results highlighted in this report provide clear evidence that USDOE funding has led to improvements in simulating AR impacts. On climatological timescales, it has been shown that E3SM-simulated ARs drive a clear response in water resource-relevant variables (e.g., precipitation, snowpack, and runoff) that is consistent with observations. E3SM-LR and RRM-E3SM are further able to simulate the most extreme AR events, assessed using the Ralph et al. (2019) AR category scale. Further, with more accurate representation of topography, changes in mountain snowpack derived from AR precipitation and associated runoff responses are simulated more accurately by RRM-E3SM than E3SM-LR. To assess the capacity of the model to perform weather event- or "storyline"-type simulations of historic high-impact AR events, E3SM was used in both its low-resolution and RRM configurations to simulate the 1997 New Year's Flood. Results from E3SM-LR and RRM-E3SM were then fed into a water management model, WEAP, and a pair of data-driven hydrologic models. Simulated streamflow and reservoir inputs matched much closer to observations when these models were fed data from RRM-E3SM, suggesting the input meteorology from RRM-E3SM was much closer to reality than either E3SM-LR or gridMET, a commonly employed gridded meteorological data product. This result is consistent with Lundquist et al. (2019), who noted that atmospheric models frequently outperform gridded precipitation products, particularly in the extreme. Further, with the RRM-E3SM input, WEAP was able to flag a point along the Feather River where a historic levee breach during the 1997 New Year's

Flood had produced widespread inundation. Simulated streamflow with the ML models also showed significant improvement over a commonly employed process-based model, and among the two ML models, the differentiable parameter learning (dPL) model developed in part from USDOE investments, demonstrated further improvement over the LSTM.

ARs are generally thought of as producing two major impacts: first, they replenish reservoirs and snowpack, being responsible for 50% of the total water resource needs of the western U.S.; second, when too much precipitation falls, they can cause widespread flooding that can lead to loss of life and infrastructure. This report substantiates and documents the major accomplishments made possible through USDOE investments for enhancing the ability to capture both forms of impacts, and for producing salient information for stakeholders and practitioners to prepare for these events.

# 5.0 References

Abatzoglou, JT. 2013. "Development of gridded surface meteorological data for ecological applications and modelling." *International Journal of Climatology* 33(1): 121–131, https://doi.org/10.1002/joc.3413

Bergström, S. 1992. The HBV model – its structure and applications (No. RH No. 4), SMHI Reports. Swedish Meteorological and Hydrological Institute (SMHI). Norrköping, Sweden.

Bergström, S. 1976. Development and application of a conceptual runoff model for Scandinavian catchments (Ph.D. thesis). Swedish Meteorological and Hydrological Institute (SMHI). Norköping, Sweden.

Corringham, TW, FM Ralph, A Gershunov, DR Cayan, and CA Talbot. 2019. "Atmospheric rivers drive flood damages in the western United States." *Science Advances* 5(12), https://doi.org/10.1126/sciadv.aax4631

Demory, ME, PL Vidale, MJ Roberts, P Berrisford, J Strachan, R Schiemann, and MS Mizielinski. 2014. "The role of horizontal resolution in simulating drivers of the global hydrological cycle." *Climate Dynamics* 42: 2201–2225, https://doi.org/10.1007/s00382-013-1924-4

Dettinger, MD, FM Ralph, T Das, PJ Neiman, and DR Cayan. 2011. "Atmospheric rivers, floods and the water resources of California." *Water* 3(2): 445–478, https://doi.org/10.3390/w3020445

Feng, D, H Beck, K Lawson, and C Shen. 2022a. "The suitability of differentiable, learnable hydrologic models for ungauged regions and climate change impact assessment." *Hydrology and Earth System Sciences Discussions* 1–28, https://doi.org/10.5194/hess-2022-245

Feng, D, J Liu, K Lawson, and C Shen. 2022b. "Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy." *Water Resource Research* 58(10), e2022WR032404, https://doi.org/10.1029/2022WR032404

Hufkens, K. 2022. "SNOTELR: A toolbox to facilitate easy SNOTEL data exploration and downloads." in R. Zenodo, https://doi.org/10.5281/zenodo.7012728

Hunrichs, RA, DA Pratt, and RW Meyer. 1998. Magnitude and frequency of the floods of January 1997 in Northern and Central California, preliminary determinations. Open File Report 98-626, US Geological Survey.

Lundquist, JD, M Hughes, B Henn, ED Gutmann, B Livneh, J Dozier, and P Neiman. 2015. "High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada, California." *Journal of Hydrometeorology* 16(4): 1773–1792, https://doi.org/10.1175/JHM-D-15-0019.1

Lundquist, JD, M Hughes, ED Gutmann, and S Kapnick. 2019. "Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks." *Bulletin of the American Meteorological Society* 100(12): 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1

Muñoz Sabater, J. 2019. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.e2161bac (accessed 12-07-2023).

Newman, AJ, MP Clark, K Sampson, A Wood, LE Hay, A Bock, RJ Viger, D Blodgett, L Brekke, JR Arnold, T Hopson, and Q Duan. 2015. "Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance." *Hydrologic Earth System Science* 19(1): 209–223, https://doi.org/10.5194/hess-19-209-2015

Ouyang, W, K Lawson, D Feng, L Ye, C Zhang, and C Shen. 2021. "Continental-scale streamflow modeling of basins with reservoirs: Towards a coherent deep-learning-based strategy." *Journal of Hydrology* 599: 126455, https://doi.org/10.1016/j.jhydrol.2021.126455

Ralph, FM, JJ Rutz, JM Cordeira, M Dettinger, M Anderson, D Reynolds, LJ Schick, and C Smallcomb. 2019. "A scale to characterize the strength and impacts of atmospheric rivers." *Bulletin of the American Meteorological Society* 100(2): 269–289, https://doi.org/10.1175/BAMS-D-18-0023.1

Rhoades, AM, AD Jones, TA O'Brien, JP O'Brien, PA Ullrich, and CM Zarzycki. 2020. "Influences of North Pacific Ocean domain extent on the western US winter hydroclimatology in variable-resolution CESM." *Journal of Geophysical Research – Atmospheres* 125(14): e2019JD031977, https://doi.org/10.1029/2019JD031977

Rhoades, AM, AD Jones, A Srivastava, H Huang, TA O'Brien, CM Patricola, PA Ullrich, M Wehner, and Y Zhou. 2020. "The shifting scales of western US landfalling atmospheric rivers under climate change." *Geophysical Research Letters* 47(17): e2020GL089096, https://doi.org/10.1029/2020GL089096

Rhoades, AM, MD Risser, DA Stone, MF Wehner, and AD Jones. 2021. "Implications of warming on western United States landfalling atmospheric rivers and their flood damages." *Weather and Climate Extremes* 32: 100326, https://doi.org/10.1016/j.wace.2021.100326

Rhoades, AM, CM Zarzycki, HAI Díaz, M Ombadi, U Pasquier, A Srivastava, BJ Hatchett, E Dennis, A Heggli, RR McCrary, and SA McGinnis. 2023. "Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model." Preprint, https://doi.org/10.22541/essoar.168319830.09780775/v1

Shen, C, AP Appling, P Gentine, T Bandai, H Gupta, A Tartakovsky, M Baity-Jesi, F Fenicia, D Kifer, L Li, X Liu, W Ren, Y Zheng, CJ Harman, M Clark, M Farthing, D Feng, P Kumar, D Aboelyazeed, F Rahmani, HE Beck, T Bindas, D Dwivedi, K Fang, M Höge, C Rackauckas, T Roy, C Xu, and K Lawson. 2023. Differentiable modeling to unify machine learning and physical models and advance geosciences. https://doi.org/10.48550/arXiv.2301.04027

Siirila-Woodburn, ER, AM Rhoades, BJ Hatchett, LS Huning, J Szinai, C Tague, PS Nico, DR Feldman, AD Jones, WD Collins, and L Kaatz. 2021. "A low-to-no snow future and its impacts on water resources in the western United States." *Nature Reviews Earth & Environment* 2(11): 800–819, https://doi.org/10.1038/s43017-021-00219-y

Thornton, PE, MM Thornton, BW Mayer, Y Wei, R Devarakonda, RS Vose, and RB Cook. 2016. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. Oak Ridge National Laboratory Distributed Active Archive Center. https://doi.,org/10.3334/ORNLDAAC/1328

Tsai, W-P, D Feng, M Pan, H Beck, K Lawson, Y Yang, J Liu, and C Shen. 2021. "From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling." *Nature Communications* 12: 5988, https://doi.org/10.1038/s41467-021-26107-z

Ullrich, PA, CM Zarzycki, EE McClenny, MC Pinheiro, AM Stansfield, and KA Reed. 2021. "TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets." *Geoscientific Model Development* 14(8): 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021

Yates, D, J Sieber, D Purkey, and A Huber-Lee. 2005. "WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model." *Water International* 30(4): 487–500, https://doi.org/10.1080/02508060508691893

Yates, D, VK Mehta, A Huber-Lee, A McCluskey, and D Purkey. 2021. "Exploring the Water-Energy Nexus in California via an Integrative Modeling Approach." *Journal of Water Resources Planning and Management* 147(12): 04021084, https://doi.org/10.1061/(ASCE)WR.1943-5452.000143

Zhu, Y, and RE Newell. 1998. "A proposed algorithm for moisture fluxes from atmospheric rivers." *Monthly Weather Review* 126(3): 725–735, https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2

U.S. DEPARTMENT OF
ENERGY

Office of Science