Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model

Alan M. Rhoades[1], Colin M. Zarzycki[2], Héctor Alejandro Inda Díaz[1], Mohammed Ombadi[1], Ulysse Pasquier[1], Abhishekh Srivastava[3], Benjamin J Hatchett[4], Eli Dennis[5], Anne Heggli[4], Rachel Rose McCrary[6], Seth A. McGinnis[6], Stefan R. Rahimi-Esfarjani[7], Emily A Slinskey[8], Paul Ullrich[9], Michael F Wehner[10], and Andrew D Jones[11]

[1]Lawrence Berkeley National Laboratory
[2]Pennsylvania State University
[3]University of California, Davis
[4]Desert Research Institute
[5]Unknown
[6]National Center for Atmospheric Research (UCAR)
[7]University of California Los Angeles
[8]University of California, Los Angeles
[9]University of California Davis
[10]Lawrence Berkeley National Laboratory (DOE)
[11]Lawrence Berkeley Laboratory

May 4, 2023

Abstract

The 1997 New Year's flood event was the most costly in California's history. This compound extreme event was driven by a category 5 atmospheric river that led to widespread snowmelt. Extreme precipitation, snowmelt, and saturated soils produced heavy runoff causing widespread inundation in the Sacramento Valley. This study recreates the 1997 flood using the Regionally Refined Mesh capabilities of the Energy Exascale Earth System Model (RRM-E3SM) under prescribed ocean conditions. Understanding the processes causing extreme events inform practical efforts to anticipate and prepare for such events in the future, and also provides a rich context to evaluate model skill in representing extremes. Three California-focused RRM grids, with horizontal resolution refinement of 14km down to 3.5km, and six forecast lead times, 28 December 1996 at 00Z through 30 December 1996 at 12Z, are assessed for their ability to recreate the 1997 flood. Planetary to synoptic scale atmospheric circulations and integrated vapor transport are weakly influenced by horizontal resolution refinement over California. Topography and mesoscale circulations, such as the Sierra barrier jet, are prominently influenced by horizontal resolution. The finest resolution RRM-E3SM simulation best represents storm total precipitation and storm duration snowpack changes. Traditional time-series and causal analysis frameworks are used to examine runoff sensitivities state-wide and above major reservoirs. These frameworks show that horizontal resolution plays a more prominent role in shaping reservoir inflows, namely the magnitude and time-series shape, than forecast lead time, 2-to-4 days prior to the 1997 flood onset.

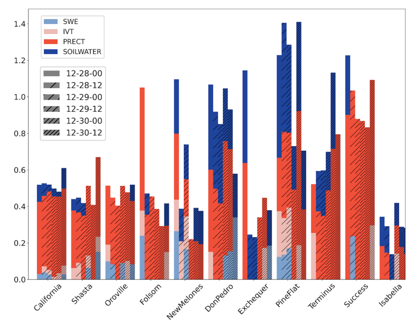

Figure 16. Same as Figure 12, however, each stacked bar chart represents one of the six forecasts produced by RRM-E3SM (3.5km) and conveys the strength of causal influence of four hydrometeorological variables, integrated vapor transport (IVT), total precipitation (PRECT), snow water equivalent (SWE), and 10 cm soil moisture (SOILWATER), on total runoff (overland flow, interflow, and baseflow). The forecast initialization date is indicated by different styles of hatching.

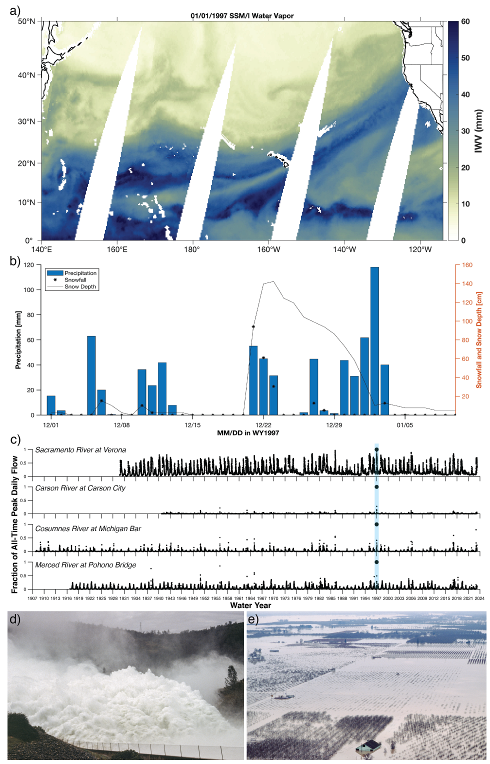

Figure 1. a) Special Sensor Microwave Imager (SSM/I) integrated water vapor on 1 January 1997. b) Tahoe City precipitation, snowfall, and snow depth from 1 December 1996 to 10 January 1997. c) Examples of all-time peak daily flows set during the event on major river systems in California and Nevada. d) Reservoir releases from Lake Oroville approached 4,530 cubic meters per second (160,000 cubic feet per second). (e) Flooding inundated the Sacramento Valley of California following heavy rainfall and snowmelt. Images d) and e) courtesy of the California Department of Water Resources.

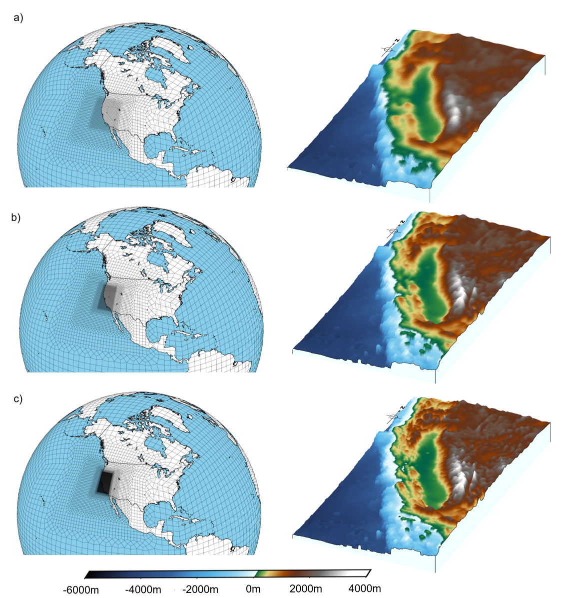

Figure 2. The Regionally Refined Mesh enabled Energy Exascale Earth System Model (RRM-E3SM) cases used to recreate the 1997 flood at horizontal resolutions of a) 0.125° (∼14km) b) 0.063° (∼7km) and c) 0.031° (∼3.5km) focused over California. Each RRM-E3SM case's topography is provided to the right of the grid refinement map. Note that ocean bathymetry is not represented in the RRM-E3SM simulations, but is included here for illustrative purposes.

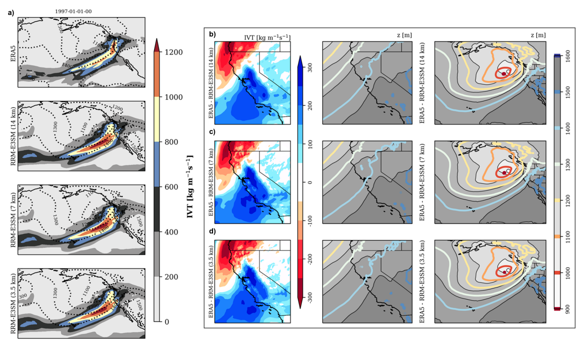

Figure 3. a) Forecast ensemble average integrated vapor transport (IVT) with 850mb geopotential height (dashed; units in meters) fields for ERA5 and each RRM-E3SM case. b) Difference in IVT between ERA5 and RRM-E3SM (14km), RRM-E3SM (7km) and RRM-E3SM (3.5km) (top, middle, and bottom rows, respectively), when the AR makes landfall in California on 1 January 1997. c-d) 850 mb geopotential height for ERA5 (gray-to-white contours) and RRM-E3SM (colored contours) over California (c) and the Northeastern Pacific (d), also at the time of AR landfall.

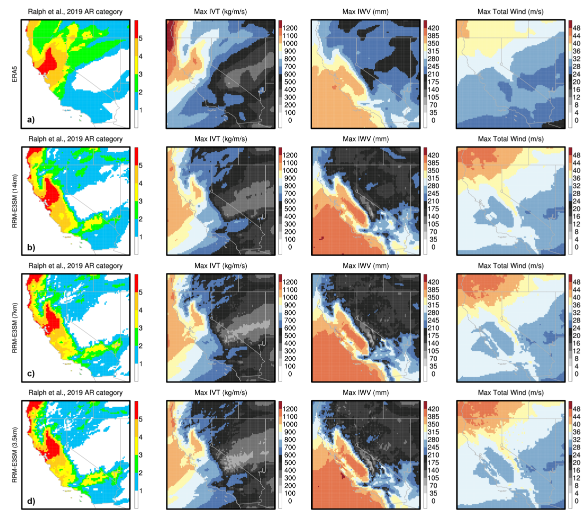

Figure 4. AR characteristics for the forecast ensemble average between the period of 31 December 1996 up to 4 January 1997. Characteristics include the Ralph et al. (2019) category scale (left column), maximum integrated vapor transport (IVT, second column), maximum integrated water vapor (IWV, third column), and maximum integrated total wind (right column) for a) ERA5 b) RRM-E3SM (14km) c) RRM-E3SM (7km) and d) RRM-E3SM (3.5km).

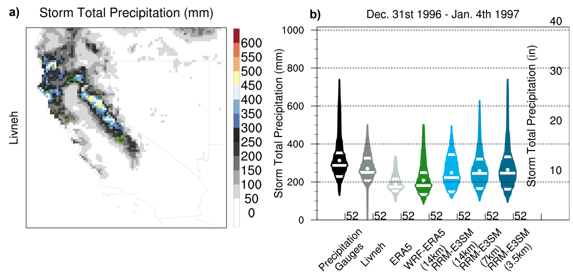

Figure 5. a) Storm total precipitation (31 December 1996 to 4 January 1997) from the Livneh product. Green dots highlight the locations of the 52 precipitation gauges used by NOAA to produce the 1997 flood event storm summary (`https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). b) Violin plots of reanalysis and model estimate storm total precipitation derived from the nearest grid cell to the 52 stations shown in a). The mean is shown with a white dot, and white lines indicate the 25th, median, and 75th percentiles. The shape of each violin reflects the probability density function of the data.

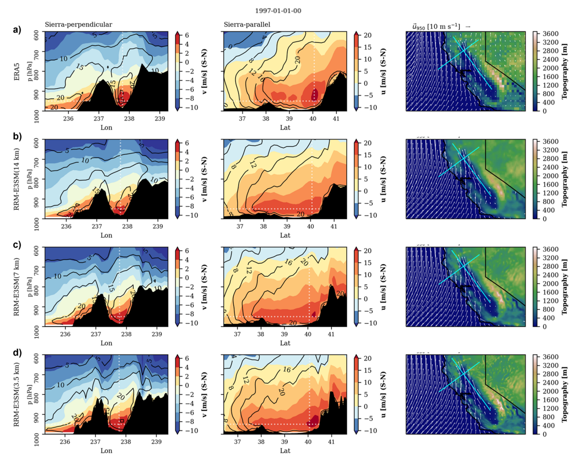

Figure 6. Sierra-perpendicular and Sierra-parallel cross sections of meridional (v) and zonal (u) winds at the start of the 1997 flood event AR landfall (1 January 1997) for ERA5 and the six-forecast ensemble average estimates provided by RRM-E3SM. The longitudinal and latitudinal cross-section transect lines are shown on the right-most column sub-panel figures overlaid on California. In the case of Sierra-perpendicular (Sierra-parallel), positive values mean that winds are blowing from South to North (West to East).

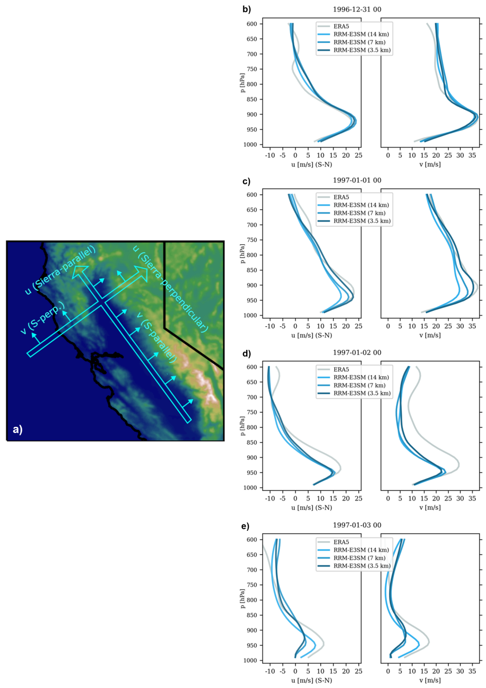

Figure 7. Sierra-parallel and Sierra-perpendicular vertical profiles of zonal (u) and meridional (v) wind speeds at the latitudinal location of the jet maxima with altitude for ERA5 and the six-forecast ensemble average RRM-E3SM simulations. a) shows the latitudinal and longitudinal transects and positive wind direction from the Sierra perspective. b-e) shows the vertical wind profiles at the intersection of the transects for the duration of the 1997 flood (31 December 1996 through 3 January 1997).

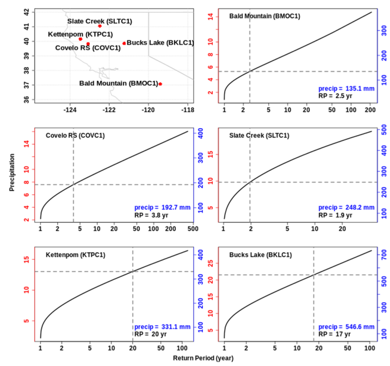

Figure 8. Return periods of the 4-day precipitation totals (Rx4day; 31 December 1996 through 3 January 1997) estimated using a non-stationary GEV framework on the Livneh product. To estimate the return period, the annual maxima of the Rx4day are interpolated to the precipitation gauge locations using first-order conservative remapping. The five stations shown (out of 52 total) are selected to indicate the minimum, 25th, 50th, 75th, and maximum Rx4day across the gauge locations. The left (right) y-axis provides Rx4day in English (metric) units. The horizontal and vertical dashed lines show the Rx4day and the corresponding return period in the Livneh product, as do the annotations in the bottom right. The x-axis (return period) is plotted on the log scale.

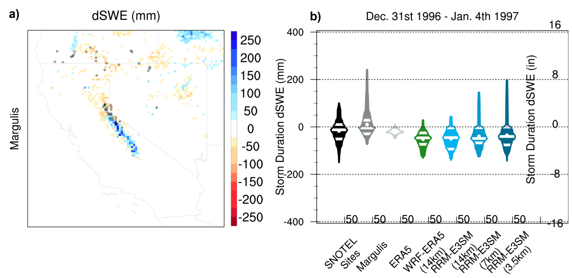

Figure 9. a) Storm duration change in snow water equivalent, dSWE, (31 December 1996 through 4 January 1997) from the Margulis product. Black dots highlight the locations of the 50 SNOTEL stations within the vicinity of the 1997 flood. b) Violin plots of reanalysis and model estimate storm duration dSWE derived from the nearest grid cell to the 50 stations shown in a). The mean is shown with a white dot, and white lines indicate the 25th, median, and 75th percentiles. The shape of each violin reflects the probability density function of the data.

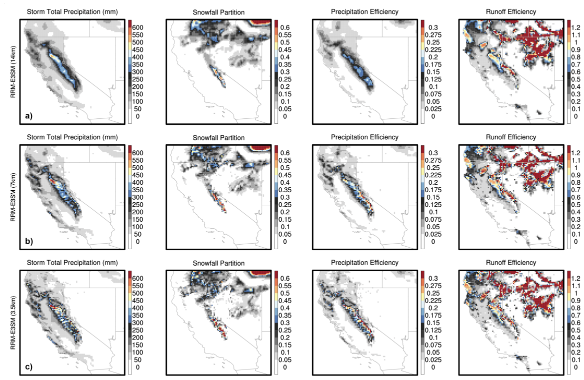

Figure 10. Forecast ensemble average precipitation characteristics, including storm total precipitation, snowfall partition, precipitation efficiency, and runoff efficiency for a) RRM-E3SM (14km) b) RRM-E3SM (7km) and c) RRM-E3SM (3.5km) over the overlapping forecast period of 31 December 1996 to 4 January 1997.

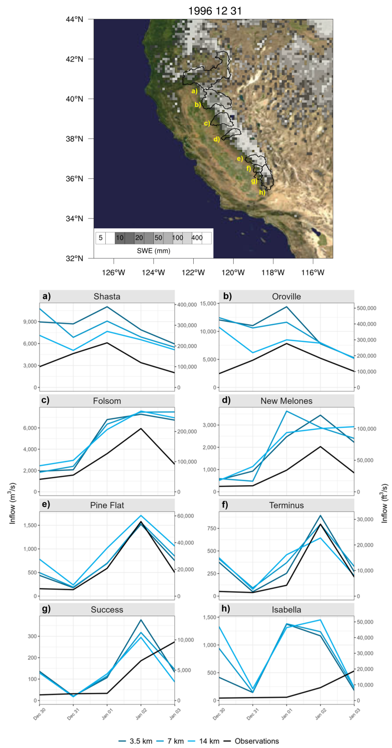

Figure 11. Forecast ensemble average reservoir inflow rates from each of the RRM-E3SM simulations across eight major reservoirs in California. The top figure shows the location of the eight reservoirs and the areal extent of the watersheds that feed into them (black outlines) overlaid onto Margulis product estimates of snow water equivalent, SWE, at the start of the 1997 flood. The black lines in the sub-panel plots represent measured inflows into each reservoir.

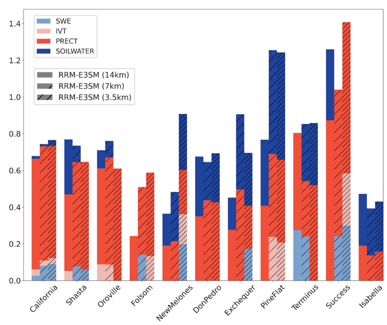

Figure 12. Causal inference estimates for the magnitude of the impact of hydrometeorological variables on total runoff (overland flow, interflow, and baseflow). The four variables include integrated vapor transport (IVT), total precipitation (PRECT), snow water equivalent (SWE), and 10 cm soil moisture content (SOILWATER). The magnitude of the influence of each variable on total runoff (overland flow, interflow and baseflow) is represented by an individual component of a stacked bar chart. Each component has a range between 0 and 1. RRM-E3SM cases (designated by hatching) are stacked next to each other for each region assessed including California (Hydrologic Unit Code 18) and the headwater regions of the 10 major reservoirs in California (ordered by latitude from northernmost to southernmost).

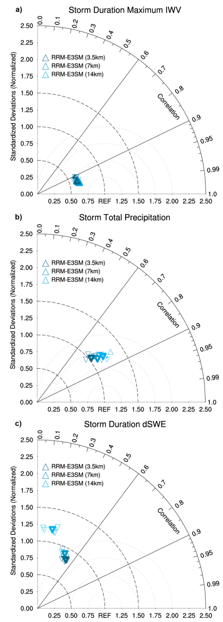

Figure 13. Taylor diagrams representing all grid cells within the hydrologic unit code (HUC-2) California Region, region 18 in Seaber et al. (1987), for the forecast period of 31 December 1996 up to 4 January 1997. a) Storm duration maximum integrated water vapor (IWV) compared to ERA5; b) storm total precipitation compared to the Livneh product; and c) storm duration change in snow water equivalent, dSWE, compared to the Margulis product. Each triangle represents one of the six RRM-E3SM forecasts initialized from 28 December 1996 at 00Z to 30 December 1996 at 12Z. Bold triangles represent the forecast ensemble average. Upward (downward) triangle orientation represents a positive (negative) bias compared to each reference dataset. Black radial lines provide general guidance for groupings of Pearson pattern correlation. The black and gray dashed azimuthal lines centered around REF indicate the root mean squared error and standard deviations from the reference dataset.

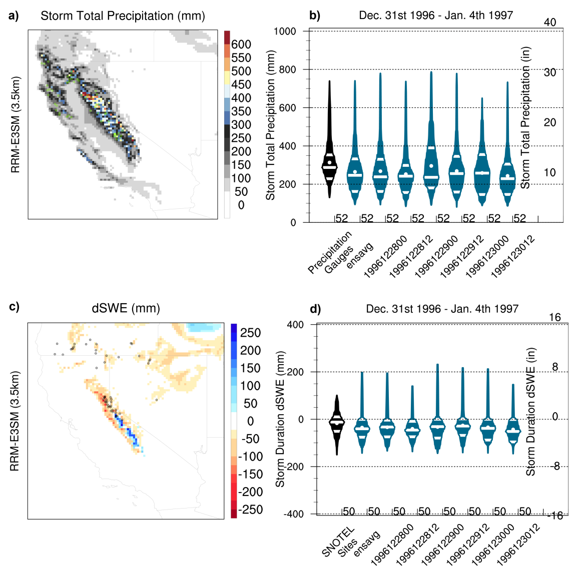

Figure 14. Same as Figures 5 and 9, but the violin plots now compare the initialization dates for each of the six RRM-E3SM (3.5km) forecasts. Panels a) and b) show storm total precipitation and panels c) and d) storm duration change in snow water equivalent (dSWE). The six-forecast ensemble average (ensavg) is also shown in black.

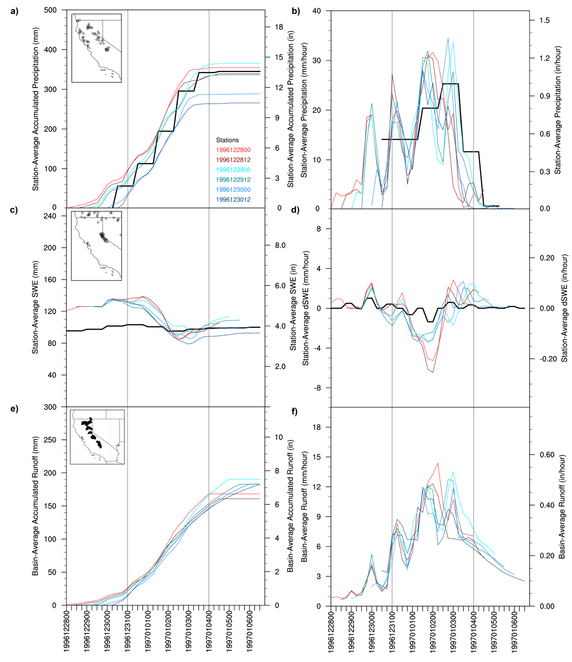

Figure 15. Time series for precipitation, snow water equivalent, and runoff simulated by RRM-E3SM (3.5km) across forecast lead time evaluated at station locations and in regions identified in the upper left maps. The left-column sub-panel plots represent cumulative totals and the right-column sub-panel plots represent hourly rates. Black lines represent station observations. Vertical gray lines indicate the period during which the 1997 flood occurred.

Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model

Alan M. Rhoades[1], Colin M. Zarzycki[2], Héctor A. Inda-Diaz[1], Mohammed Ombadi[1,3], Ulysse Pasquier[1], Abhishekh Srivastava[4], Benjamin J. Hatchett[5], Eli Dennis[6], Anne Heggli[5], Rachel McCrary[7], Seth McGinnis[7], Stefan Rahimi-Esfarjani[6], Emily Slinskey[6], Paul A. Ullrich[1,4,8], Michael Wehner[9], and Andrew D. Jones[1,10]

[1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[2]Department of Meteorology and Atmospheric Science, Penn State University, State College, PA, USA
[3]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA
[4]Department of Land, Air, and Water Resources, University of California, Davis, CA, USA
[5]Desert Research Institute, Reno, NV, USA
[6]Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA
[7]National Center for Atmospheric Research, Boulder, CO, USA
[8]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA
[9]Computational Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[10]Energy and Resources Group, University of California, Berkeley, Berkeley, CA, USA

Key Points:

- Energy Exascale Earth System Model forecasts at 3.5km grid spacing skillfully recreate the hydrometeorology of California's 1997 flood
- Horizontal resolution alters the representation of key flood drivers such as the Sierra barrier jet, precipitation extremes, and snowmelt
- Forecast lead time 2-to-4 days prior to the onset of the 1997 flood minimally influences forecast precipitation and snowmelt skill

Corresponding author: Alan M. Rhoades, `arhoades@lbl.gov`

Abstract

The 1997 New Year's flood event was the most costly in California's history. This compound extreme event was driven by a category 5 atmospheric river that led to widespread snowmelt. Extreme precipitation, snowmelt, and saturated soils produced heavy runoff causing widespread inundation in the Sacramento Valley. This study recreates the 1997 flood using the Regionally Refined Mesh capabilities of the Energy Exascale Earth System Model (RRM-E3SM) under prescribed ocean conditions. Understanding the processes causing extreme events inform practical efforts to anticipate and prepare for such events in the future, and also provides a rich context to evaluate model skill in representing extremes. Three California-focused RRM grids, with horizontal resolution refinement of 14km down to 3.5km, and six forecast lead times, 28 December 1996 at 00Z through 30 December 1996 at 12Z, are assessed for their ability to recreate the 1997 flood. Planetary to synoptic scale atmospheric circulations and integrated vapor transport are weakly influenced by horizontal resolution refinement over California. Topography and mesoscale circulations, such as the Sierra barrier jet, are prominently influenced by horizontal resolution. The finest resolution RRM-E3SM simulation best represents storm total precipitation and storm duration snowpack changes. Traditional time-series and causal analysis frameworks are used to examine runoff sensitivities state-wide and above major reservoirs. These frameworks show that horizontal resolution plays a more prominent role in shaping reservoir inflows, namely the magnitude and time-series shape, than forecast lead time, 2-to-4 days prior to the 1997 flood onset.

Plain Language Summary

The 1997 California New Year's flood event caused over a billion dollars in damages. This storm became a central part in guiding efforts to reduce flood risks. Earth system models are increasingly asked to recreate extreme weather events. However, the ability of Earth system models to recreate such events requires rigorous testing. Testing ensures that models provide value in anticipating and planning for future flood events. This is particularly important given the changing climate. We evaluated the Department of Energy's flagship Earth system model, the Energy Exascale Earth System Model, in its ability to recreate the weather and flood characteristics of the 1997 flood. The model resolution, important for resolving mountain terrain and storm interactions, and forecast lead time, important for storm progression accuracy, are assessed. The multi-forecast average from the highest-resolution model best recreates the observed precipitation, snowpack changes, and flood characteristics. Our findings provide confidence that the highest resolution model could be used to study how a 1997-like flood event would be altered in a warmer world.

Introduction

California is especially susceptible to major cool season flood events (Kattelmann, 1997). Atmospheric rivers (ARs) are largely responsible, accounting for 84% of flood damages in the western United States (Corringham et al., 2019). The most notable California flood event, measured by its intensity, duration, and inundation area, occurred in 1861/1862 (Porter et al., 2011; Huang & Swain, 2022). It was thought to be AR-driven and inundated portions of both the Sacramento and San Joaquin valleys and portions of the present-day metropolitan area of Los Angeles. Because of its impact, this event has emerged as an important "design storm" for California water managers and led to the development of the colloquially termed "ARkStorm", which combines aspects of AR-induced flood events that occurred in 1969 and 1986. The 1861/1862 flood event happened during a time in California's history when the population density and built infrastructure was at a much smaller scale than today. Since the 1860s, urbanization has resulted in the loss of floodplains in many communities that are vulnerable to flooding despite significant investments in constructing flood control infrastructure (Whipple et al., 2017; Whipple & Viers, 2019). In many low-lying regions throughout the Central Valley, aging levee systems and subsidence continue to expose populations and industries to flood impacts (Hanak & Lund, 2012). Sequences of heavy precipitation-producing storms, many of which were ARs, during the winters of 2017 and 2023 highlight the present susceptibility of California to major riverine flooding. Climate change may further exacerbate impacts felt by these storms (Gershunov et al., 2019; Rhoades et al., 2021; Corringham et al., 2022; Huang & Swain, 2022), particularly in the most underserved communities (Wing et al., 2022), highlighting the need for detailed analyses aimed at understanding how these storms drive compound extremes under historical and future climate conditions.

The most costly flood event ($1.6 billion) in California history was the New Year's flood event of 1997, hereafter "1997 flood" (Lott et al., 1997). Major flood losses occurred throughout the western United States, including losses of $500 million in Nevada and $125 million in Washington. The combination of flood area and severity across the western United States ranks the 1997 flood as the #2 superflood between 1950 and 2010 (Tarouilly et al., 2021). At least half a million people were displaced by the flooding and the majority of California counties (43/58) were declared disaster zones (Lott et al., 1997).

The 1997 flood was primarily made up of three storms that occurred between 25 December 1996 and 2 January 1997 with inundation afterward (Galewsky & Sobel, 2005). Antecedent conditions played an important role in driving up the economic cost of this event; earlier storms throughout late November and December of 1996 built an abundant snowpack and elevated soil moisture content throughout the Central Valley and the Sierra Nevada (Figure 1). Between 30 December 1996 and 3 January 1997 storms produced more than 750 mm of precipitation in certain regions of northern California (e.g., 840 mm, or 33 in, at Bucks Lake in Plumas County, California; (Figure 1; `https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). Heavy rainfall with snow above 3,000 m elevation commenced on 30 December 1996; the Central Sierra Snow Lab (CSSL; located at 2,100 m) reported 137 mm of rainfall on 30-31 December 1996 (Osterhuber & Schwartz, 2021). On New Year's Day of 1997, an extreme AR event made landfall (Figure 1). Maximum temperatures at 2,100 m elevation hit 7°C and reached 3°C at 2,900 m on 1 January 1997 when 120 mm of rain fell at the CSSL (Osterhuber & Schwartz, 2021; Heggli et al., 2022). Prior to the onset of rainfall on 29 December 1996, snow densities were ready to produce terrestrial water input (32%), rising to 35% on 30 December 1996 (Heggli et al., 2022). The CSSL lost 100 mm of snow water equivalent (SWE) between 30 December 1996 and 1 January 1997 ultimately contributing to the development of a warm-snow drought water year (Hatchett & McEvoy, 2018). When combined with saturated soils and sufficiently ripe snowpack to melt and convey water to the land surface, the extreme multi-day precipitation caused major rivers to reach flood stage, with several setting all-time peak flows (Figure 1; `https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). As a result of the December-January storms, this two-month period set the record for the wettest since records began in 1920, measured via California's 8-station index, with a total of ∼1,200 mm of precipitation. However, despite the wet start, the remainder of the water year was drier than normal leading to below-normal snowpack and reservoir levels at the end of the required flood pool period in April. The 1997 flood event thus represents an object lesson both for the study of extreme precipitation and runoff but also for reservoir and flood management in a highly variable climate.

A growing area of climate research is focused on understanding cascading, compound, and/or sequential hydrometeorological extreme events (Fish et al., 2019; AghaKouchak et al., 2020; Raymond et al., 2020). Simultaneously, the climate research community has sought to provide more credible and salient decision-relevant information to practitioners and management communities through iterative, co-produced research (Lemos et al., 2018; Jagannathan et al., 2021; Siirila-Woodburn et al., 2021). Examining historically significant, decision-relevant extreme events, through high-resolution climate model "storyline" recreations can be both be useful for water resource managers (Shepherd, 2019; Gutowski et al., 2020; Bukovsky et al., 2023) and have also been frequently used in event attribution studies (Wehner et al., 2019). Storylines are physically based model recreations of impactful weather events, often chosen through iterative discussions between scientists and stakeholders, that are then simulated under plausible past and future climate scenarios. However, it is important to note that while such studies can provide information on the local dynamic and thermodynamic effects of climate change on extreme events, they do not provide information about the influence of large-scale circulation changes on the return probability of such events.

Storyline event recreations also have practical model development implications. Climate models are mostly optimized around mean state performance for different hydrometeorological performance metrics (Fasullo, 2020), rather than extremes. This is especially true from the perspective of land-atmosphere interactions that drive compound extremes (La Follette et al., 2021). Storyline approaches can also help to convey information on model uncertainty, namely the role of structural and scenario uncertainty (Lehner et al., 2020), in a more understandable and decision-relevant way. Therefore, the recreation of the 1997 flood is a useful exercise in understanding the nature of extreme events and determining whether our cutting-edge modeling approaches are fit for purpose in simulating them. An additional benefit of storyline approaches is that the climate models used and the resultant climate research conducted becomes tailored toward greater practitioner relevance over time (Lemos et al., 2012).

In this study, we recreate the 1997 flood using the U.S. Department of Energy's flagship climate model, the Energy Exascale Earth System Model, and its regionally refined mesh capabilities (RRM-E3SM). We chose the 1997 flood because it is the flood of record most recently experienced by current water managers, was relatively well-monitored by a network of meteorological and hydrologic measurements, and occurred during a period in which atmospheric reanalysis products have higher skill (Uppala et al., 2005; Hersbach et al., 2020). This event also allows us to assess the relative contributions of E3SM horizontal resolution and forecast initialization time in shaping the fidelity of the flood event recreation. We pay particular attention to the interactions across the submodels of E3SM (e.g., atmospheric and land-surface) and their representation of key hydrometeorological variables before/during/after the event. This is the first time RRM-E3SM has been systematically used, across resolution and forecast lead time, to generate a storyline recreation of a western United States hydrometeorological extreme. Our scientific questions include:

(1) To what degree does horizontal model resolution influence land-atmosphere interactions and hydrometeorological impacts associated with the 1997 flood?
(2) What is the forecast lead time that best balances the short-term antecedent preconditioning of soils and snowpack and post-storm impacts when recreating the 1997 flood?
(3) Is RRM-E3SM fit-for-purpose in representing a compound extreme event such as the 1997 flood?

The manuscript is organized as follows. We first highlight details about our RRM-E3SM experimental setup. We then discuss the various *in-situ*, reanalysis, regional climate model, and gridded climate products used to assess and juxtapose RRM-E3SM skill in recreating the 1997 flood. We then discuss our results and how they fit within the broader literature. Finally, we summarize our major findings and provide suggestions for future research.

Methods

Energy Exascale Earth System Model (E3SM) version 2

The Energy Exascale Earth System Model version 2 (E3SMv2; Golaz et al., 2022) used for this analysis allows for regionally refined mesh (RRM-E3SM) simulations over a targeted region of interest. Recent studies find that RRM-E3SM performs comparably to uniform 0.25° (∼25km) horizontal resolution simulations for water cycle-related processes and provides several improvements to uniform 1.00° (∼111km) horizontal resolution simulations (Tang et al., 2019, 2022). These improvements are particularly important in regions of complex terrain such as the California Sierra Nevada. A detailed description of E3SMv2's atmospheric dynamical core, physics and dynamics, horizontal grids, vertical discretization, radiation, tracer transport schemes, and subgrid-scale parameterization choices (e.g., cloud microphysics scheme) can be found in Golaz et al. (2022). More specific findings related to RRM-E3SM are described in Tang et al. (2022), while Harrop et al. (2022) provides additional details on water cycle process fidelity in both the atmosphere and land-surface in E3SM at uniform horizontal resolutions of 1.00° versus 0.25° over the United States.

The RRM-E3SM meshes were produced using TempestRemap (Ullrich & Taylor, 2015; Ullrich et al., 2016); the topography was generated with the NCAR_Topo tool (Lauritzen et al., 2015) and smoothed for model stability purposes using the framework discussed in Zarzycki et al. (2015) and a coefficient of 3e^{-16} (c in Equation 1 of Zarzycki et al., 2015). The refinement regions and topographic representation in the simulations over California for the three RRM-E3SM cases are shown in Figure 2. Hereafter, RRM-E3SM simulations with a maximum refinement resolution over California at 14km, 7km, and 3.5km will be referred to as, RRM-E3SM (14km), RRM-E3SM (7km), and RRM-E3SM (3.5km), respectively. In all simulations, the E3SM default setting of 72 vertical levels is used. As found in other variable-resolution and regionally refined mesh Earth system model analyses over the last decade, horizontal resolution influences the simulation fidelity of synoptic-to-mesoscale trajectory of storm tracks and eddies (Rauscher et al., 2013; Rauscher & Ringler, 2014; Sakaguchi et al., 2016; Liu et al., 2023). Resolution also influences the representation of topography, which in turn affects how coastal landfalling storms are orographically uplifted, the rain-snow partitioning of the storm's precipitation, and the build-up and evolution of mountain snowpack throughout the cool-season (Rhoades et al., 2016; Huang et al., 2016; Wu et al., 2017; Rhoades, Ullrich, & Zarzycki, 2018; Rhoades, Ullrich, Zarzycki, Johansen, et al., 2018; Xu et al., 2018; Rhoades, Jones, O'Brien, et al., 2020; Rhoades, Jones, Srivastava, et al., 2020; Bambach et al., 2021; Xu et al., 2021; Maina et al., 2022). Similarly, land-surface cover and soil heterogeneity increase at finer resolutions, which can alter the surface-through-subsurface water and energy balance interactions of the hydrologic cycle (e.g., soil moisture).

Betacast

The 1997 flood event forecast ensemble was produced for six different 8-day periods starting on 28 December 1996 at 00Z through 30 December 1996 at 12Z, initialized at 12-hour increments between those dates, using the "Betacast" framework described in Zarzycki et al. (2014) and the Atmosphere Model Intercomparison Project (AMIP) protocols (Gates et al., 1999). The land surface conditions are spun-up for five years prior to the first forecast, with a standalone simulation of the E3SM Land Surface Model (ELM) forced by the 6-hourly atmospheric data from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5; Copernicus Climate Change Service Climate Data Store (CDS), 2017). This ensures that antecedent land surface conditions (namely soil moisture content and mountain snowpack) are consistent with the actual 1997 flood event conditions on the day each RRM-E3SM forecast is started. Subsequent forecast cycles use the 12-hour land forecast from the previous cycle for initialization. This approach gives nearly identical results to spinning up each forecast cycle's land surface independently (not shown).

The atmospheric initial state is generated using high-order remap algorithms to take data from the ERA5 reanalyses and map them onto the corresponding RRM-E3SM grid. The pressure field is adjusted based on the technique in Trenberth et al. (1993) to account for differences in ERA5 and RRM-E3SM orography that may result in geostrophic imbalances. Observed ocean surface conditions (i.e., sea surface temperatures and sea ice extent) are also prescribed by interpolating NOAA Optimum Interpolation (OI) data (Reynolds et al., 2007) to the model grid. After initialization from ERA5, the RRM-E3SM forecasts are "free-running": the atmosphere and land surface models are fully coupled and allowed to freely solve the governing equations that drive these systems.

All RRM-E3SM simulations utilize the hydrostatic dynamical core in E3SM. Notably, the effective resolution is 4-5x the actual grid spacing (Ullrich, 2014; Klaver et al., 2020). Further, it has been shown that non-hydrostatic dynamical cores minimally influence midlatitude wintertime precipitation (slight drying) from resolutions of 36-to-4km, even in idealized mountain environments (Yang et al., 2017; Liu et al., 2022). With each 2x refinement in horizontal resolution, the RRM-E3SM dynamics and physics timestep and second-order viscosity diffusion strength at the model top were halved. For RRM-E3SM (14km), the atmospheric dynamics and physics timesteps and diffusion strength were 40 and 600 seconds and $4e^{-4}$, for RRM-E3SM (7km) they were 20 and 300 seconds and $2e^{-4}$, and for RRM-E3SM (3.5km) they were 10 and 150 seconds and $1e^{-4}$, respectively. The only additional differences across cases were the macrophysics-microphysics subgrid-scale parameterization substeps, set to 6 in RRM-E3SM (14km) and RRM-E3SM (7km) and 3 in RRM-E3SM (3.5km).

Atmospheric River Detection and Categorization

We used TempestExtremes (TE; namely the SpineARs and StitchBlobs algorithms) to detect the primary AR that made landfall during the 1997 flood on 1 January 1997 (Ullrich & Zarzycki, 2017; Zarzycki & Ullrich, 2017). TE is a "relative threshold" based AR detector (ARDT), meaning that it is minimally sensitive to fixed thresholding issues (i.e., an AR event only exists beyond ~250 kg/m/s), which may have important implications for assessing future AR characteristic changes (O'Brien et al., 2022). Our parameter settings for TE and the extensions made to TE to estimate AR landfalling characteristics, such as the AR category scale (Ralph et al., 2019), are important for estimating water resource impacts (e.g., AR-induced flood damages in Corringham et al., 2022) as discussed in more detail in Rhoades, Jones, O'Brien, et al. (2020), Rhoades, Jones, Srivastava, et al. (2020) and Rhoades et al. (2021). Although it is advantageous to use several ARDTs for climatology-based analyses of ARs (O'Brien et al., 2022), particularly when assessing climate change-related impacts, we use only TE because the primary AR during the 1997 flood was a category 5 event and recent findings in Zhou et al. (2021) have shown that ARDTs largely agree when identifying characteristics of category 4-5 AR events.

Validation

To evaluate the hydrometeorological forecast skill of RRM-E3SM in recreating the 1997 flood, we use a mixture of *in-situ* observations, reanalysis, gridded climate products, and more conventional regional climate modeling strategies. We obtained *in-situ* observations from 50 sites in the SNOw TELemetry (SNOTEL) network (`https://www.nrcs.usda.gov/wps/portal/wcc/home/snowClimateMonitoring/snowpack/snowpackMaps`) and 52 precipitation gauge sites from the California Data Exchange Center (CDEC) that are used in the National Oceanic and Atmospheric Administration (NOAA) storm summary (`https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). We obtained daily reservoir inflow observations from the US Army Corps of Engineers Water Control Data System (`https://www.spk-wc.usace.army.mil/plots/california`

`.html`), retrieving inflow information for the 1997 Water Year from the Shasta, Oroville, Folsom, New Melones, Pine Flat, Terminus, Success, and Isabella Reservoirs.

We used reanalysis and gridded climate products to evaluate storm-total precipitation and pre-and post-event changes in snow water equivalent (SWE). Storm-total precipitation is evaluated against Pierce et al. (2021) which is an updated version of the Livneh product (Livneh et al., 2015), hereafter Livneh, and against the ERA5 reanalysis product, due to its use in providing initial conditions for the RRM-E3SM simulations. According to Pierce et al. (2021), the updated Livneh product better preserves extreme event precipitation totals by more systematically accounting for daily time adjustments in precipitation gauge data (i.e., rounding-related issues related to the time of day the station observation is taken). We also conducted a preliminary analysis comparing Livneh with other widely used gridded climate products, Newman et al. (2015) (Newman) and Daly et al. (2008) (Parameter-elevation Regressions on Independent Slopes Model, PRISM) as shown in Figure S1. Compared with the 52 precipitation gauge measurements, we found that Livneh was either a better estimate (compared with Newman) or was indistinguishable (compared with PRISM) in its representation of the 4-day precipitation totals produced during the 1997 flood. In order to estimate the return periods of the 4-day precipitation totals during the 1997 flood, we applied a non-stationary generalized extreme value (NS-GEV) analysis to the annual maximum of 4-day precipitation totals (Rx4day) in the Livneh product interpolated to the 52 gauge locations using the first-order conservative remapping (P. W. Jones, 1999). In the NS-GEV framework, we first apply the Mann–Kendall (MK) trend test (Mann, 1945) to the Rx4day data at each gauge location to determine if the data has a significant trend at the 5% level. If the Rx4day data at a location has a significant trend, we fit time as a covariate in the location or/and scale parameters of the GEV distribution fitted to the Rx4day data at that gauge location. The complete procedure is outlined in Srivastava et al. (2021).

We assess pre- and post-event changes in SWE against the Fang et al. (2022) western United States-wide snow reanalysis product (hereafter Margulis due to it being an updated version of Margulis et al., 2016). The Margulis reanalysis product has shown skill in estimating peak SWE in the California Sierra Nevada when compared with airborne LiDAR SWE measurements (e.g., 1 April mean SWE depth differences of -0.15 to +0.05 m across 2015-2021), which have essentially become the snow community standard for spatially complete estimates of snow depth and SWE in recent years (Painter et al., 2016; Stillinger et al., 2023). We also compare and contrast RRM-E3SM skill with a set of simulations produced with a more traditional and widely-used dynamical downscaling approach. These simulations were produced using the Weather Research and Forecasting (WRF) model run at 14km resolution over California that is bounded laterally and at the model top with ERA5 (A. D. Jones et al., 2022). All gridded data that is intercompared has been regridded from its native grid resolution to a regular latitude-longitude grid resolution of 14 km using bilinear interpolation provided by the Earth System Modeling Framework (ESMF) Offline Regridding Weight Generator (The NCAR Command Language (Version 6.6.2), 2022).

Causal Inference

The complexity of Earth system interactions within the RRM-E3SM simulations and the large number of grid cells within the spatial domain of analysis makes it difficult to unambiguously disentangle the impact of resolution and forecast lead time on processes and interactions between hydrometeorological variables. Thus, in the present study, we use causal inference to gain insights into the interactions between atmospheric and land-surface variables on one hand, and total runoff on the other. To the best of our knowledge, this is the first application of this framework for this style of problem. Causal inference allows us to move beyond canonical correlation analysis while reducing the dimensionality of analysis to investigate interactions in the model. The goal of causal inference methods is to determine causal relationships between hydrometeorological variables by using concepts of statistical conditional independence on time series data. These methods are gaining popularity in the Earth and environmental sciences community (Sugihara et al., 2012; Runge et al., 2019; Ombadi et al., 2020; Runge, 2023) and offer a unique perspective to evaluate relationships.

We use the Peter-Clark (PC) algorithm (Spirtes & Glymour, 1991), a causal inference method that utilizes graph theory and graphical rules to recover causal relations from time series data. The PC algorithm starts with a fully connected graph where all variables are causally related to each other, then iteratively and systematically removes causal relations using conditional independence tests. One of the main advantages of the PC algorithm is its ability to reduce the number of variables in the conditioning set, thereby mitigating the "curse of dimensionality". We chose to use the PC algorithm because it provides good performance in hydrometeorological systems, especially in controlling the number of falsely detected causal links (Ombadi et al., 2020). For our conditional independence tests, we used information-theoretic conditional independence instead of partial correlation due to its ability to detect nonlinear relationships (Ombadi et al., 2021). Our causal analysis considers contemporaneous causality between the time series of the five key hydrometeorological variables evaluated in this study (i.e., integrated vapor transport [IVT], precipitation, SWE, 10 cm soil moisture content, and total runoff volume) for all grid cells within a specific spatial domain (e.g., California-wide or the mountainous headwaters of a surface reservoir). Causality was assessed at a statistical significance level of 0.05.

Results and Discussion

Murphy (1993) provides terminology to discuss forecast verification qualities that both forecasters and users of forecasts find important. In this study, we will evaluate RRM-E3SM's representation of the California New Year's flood event of 1997 according to forecast quality (forecast correspondence to observations) and forecast value (forecast utility to decision makers). We use the effects of horizontal resolution and forecast lead time to assess forecast quality and value via measures of bias (the difference between forecast and observation), association (linear correlation between forecast and observation), sharpness (forecast capability in representing extremes), and through measures of value (e.g., reservoir inflow volumes).

Resolution influence on atmospheric process representation of the 1997 flood

We first compare the influence of regional grid refinement over California by evaluating how the representation of the large-scale atmospheric circulations that shaped the landfalling AR on New Year's Day of 1997 differ according to the resolution of the regional refinement domain. Figure 3 compares the large-scale IVT fields and circulation patterns of ERA5 and the three grid refinement resolutions at the start of the major AR landfall on 1 January 1997. The RRM-E3SM values are six-member forecast averages. The RRM-E3SM simulations forecast the low-pressure center near the Pacific Northwest coastline further southwest than it is in ERA5 on this date (Supplemental Figure S2). The simulations generally agree across resolutions on the spatial distribution of AR categories from the California Bay Area up through the Sacramento Valley (Figure 4 and

Figure S3). Agreement is also found with ERA5 in the northern portions of California, particularly with regard to category 5 AR conditions (Figure S4); however, all RRM-E3SM simulations systematically produce AR categories that are too high in southern California. This appears to be due to a disagreement in the AR width and/or the centroid of the AR landfall location with ERA5, which occurs further South (as indicated by positive IVT anomaly from central to southern California in Figure 3) and due to uniformly higher wind speeds (Figure S4). Notably, ERA5 may under-represent AR activity in southern California compared to other reanalyses (Collow et al., 2022).

Although IVT is important from a forecasting perspective, particularly since it allows for longer forecast lead times than precipitation (Lavers et al., 2016), IVT is simply one metric indicating the potential for precipitation to occur, and its orientation with respect to terrain can suppress or enhance precipitation (Ricciotti & Cordeira, 2022). Therefore, we also evaluate how the precipitation potential across RRM-E3SM simulations is realized in the 1997 flood, particularly its association and sharpness. The forecast ensemble average storm total precipitation amounts are shown in Figure 5. This figure compares simulated precipitation values with reanalysis and gridded climate products as well as a conventionally used regional climate model (WRF, forced by ERA5) at the grid cells nearest to the 52 precipitation gauges used in NOAA's storm summary of the 1997 flood. Refinement from 14km to 3.5km in RRM-E3SM has an appreciable effect on the statistical distribution of storm total precipitation, including the mean, median, and maximum. RRM-E3SM (3.5km) matches the distribution of storm total precipitation at the 52 precipitation gauge sites better than other datasets, including the Livneh product. RRM-E3SM (3.5km) agreement (r=0.73) in storm total precipitation holds across individual precipitation gauge sites as well (Figure S5), particularly precipitation gauges in the northern Sierra Nevada, which have the highest precipitation totals (e.g., Buck's Lake and La Porte). Note that the WRF simulations were conducted at 14km resolution and do not represent an even comparison with RRM-E3SM (7km) or RRM-E3SM (3.5km). The superior skill of models, relative to statistical interpolation and extrapolation techniques utilized in gridded climate products, in representing mountain precipitation processes have been noted before (J. Lundquist et al., 2019).

In contrast to landfalling AR characteristics, we found storm total precipitation to be resolution-dependent. We hypothesize that this is likely a result of more realistic topographic representations of California's Coast Ranges and Sierra Nevada. In addition, we hypothesize that important mesoscale circulation features known to influence the spatiotemporal characteristics of precipitation in northern California are better resolved. One such feature is the Sierra Barrier Jet (SBJ), a classic terrain-parallel low-level jet. The SBJ results from the blocking, slowing, and subsequent counter-clockwise turning of low-level winds as they interact with the Sierra Nevada in a stable or moist-neutral environment. The SBJ has a typical core of peak winds at $\sim$500m to 1km ($\sim$950-900 hPa) above the Central Valley with wind speeds $\geq$15 m/s (Neiman et al., 2010, 2013). The location and strength of the SBJ play an important role in driving California's precipitation maxima during AR events (Neiman et al., 2013). This precipitation maximum usually occurs northwest and upstream of the Sierra Nevada crest, typically around the Buck's Lake precipitation gauge (39.85°N, 121.24°W) in the headwaters of the Oroville Dam. To examine RRM-E3SM skill in representing the SBJ, we compare winds using analogous cross-sections and transect lines outlined in Hughes et al. (2012) that dissect the typical locations of the SBJ in California.

Figure 6 shows cross-sections of zonal and meridional winds for ERA5 and the RRM-E3SM simulations at the start of the AR landfall on 1 January 1997. Similarly to previous findings, wind speeds are generally stronger in RRM-E3SM cases compared with ERA5. However, the altitude, latitudinal, and longitudinal locations of the wind speed maximum do generally agree with ERA5. RRM-E3SM simulates the SBJ and locates its core between 950-900 hPa at around 40°N, 122°W. Resolution plays an important role in better resolving the location of the wind speed maximum both with altitude and latitudinally. Similarly, RRM-E3SM (3.5km) shows higher wind speeds from 1000-900 hPa and more orographic uplift potential along the windward sides of both the Coast Ranges and the Sierra Nevada. This favors more orographic precipitation, as is shown in Figure 5.

To assess RRM-E3SM skill in representing the entire lifecycle of the SBJ, we now show vertical profiles of both meridional and zonal winds, from both a Sierra-parallel and Sierra-perpendicular perspective, compared with ERA5 (Figure 7). Prior to the onset of the flood event, on 31 December 1996, the RRM-E3SM simulations show the jet beginning to form at the right altitude relative to ERA5, but slightly stronger. On the first day of the flood event (1 January 1997), RRM-E3SM (3.5km) best represents the altitude location ($\sim$950-1000 hPa) and strength (20-25 m/s) of the SBJ. The jet altitude and latitudinal location and strength match with the findings of Neiman et al. (2013)

for other couplets of AR-SBJ events identified using a combination of *in-situ* measurements including vertical wind profilers and reanalysis products. The RRM-E3SM results also corroborate the conclusion made by Hughes et al. (2012) that approximately a six-kilometer horizontal resolution is needed to properly represent the SBJ in model simulations. However, regardless of RRM-E3SM resolution, the SBJ becomes both weaker and/or lower in altitude relative to ERA5 on 3-4 January 1997.

Resolution influence on land-surface process representation of the 1997 flood

Although the 1997 flood was one of the most costly and damaging floods in northern California history, a non-stationary return period analysis of the Livneh product at the 52 gauge sites indicates that it was, at most, a 1-in-20-year event at a few gauge locations, based on 4-day precipitation total estimates over the 105-year record covering 1915-2019 (Figure 8). At 50% of gauge locations, the return period of the event was less than 6 years. This implies that the flooding was notable due to it being a compound extreme shaped by not only the precipitation provided by the sequence of storms, culminating in a category 5 AR landfall on 1 January 1997 but also antecedent land surface conditions that were primed for snowmelt and runoff generation. The importance of antecedent conditions and land surface feedbacks was shown by Ivancic and Shaw (2015) where only 36% of the 99th percentile discharge events occurred due to a 99th percentile precipitation event when evaluated CONUS-wide between 1950-2000.

To evaluate the role that antecedent and land surface conditions played in shaping the flood event, we now assess the change in snow water equivalent, or dSWE, for the category 5 AR storm duration (Figure 9). Analogously to the storm total precipitation analysis, we show storm duration dSWE across 50 SNOTEL sites throughout northern California, southern Oregon, and Nevada compared to the Margulis product. Model resolution also plays an important role in the distributions of both positive and negative dSWE in the California Sierra Nevada. This is likely due to the influence of topographic resolution on the simulated freezing level and the rain-snow partitioning of the AR event, which in turn influences the land surface representation of the accumulation and ablation of the mountain snowpack at mid-to-high elevations. The 50 SNOTEL sites indicate that more negative dSWE occurred over the duration of the 1997 flood (-152 mm / -6 in). However, at higher elevations, positive dSWE also occurred (+102 mm /

+4 in). In comparison, the Margulis product indicates that more positive dSWE occurred (up to +254 mm, or +10 inches, in certain locations). Although a general negative dSWE skew in the statistical distribution is shown for RRM-E3SM, with every 2x refinement in resolution over California the simulations more closely approximate the statistical distribution from the 50 SNOTEL location observations.

Figure 10 shows the effects of resolution on the spatial representation of precipitation and runoff characteristics. The differences across each RRM-E3SM case are explicitly shown in Figure S6. Storm total precipitation is enhanced at finer horizontal resolutions, particularly along the Coast Range and crest of the Sierra Nevada, upwards of 250 mm in RRM-E3SM (3.5km) relative to RRM-E3SM (14km). However, a general dry (wet) bias across RRM-E3SM simulations is seen in northwestern California's Klamath Mountains (Sierra Nevada) when compared with the Livneh product (Figure S7). Notably, the Livneh product had a general dry bias compared with precipitation gauge measurements (Figure 5 and S5). This indicates that Sierra Nevada crest precipitation overestimates in RRM-E3SM may not be as severe as is shown in Figure S7, corroborates the findings of J. Lundquist et al. (2019), and would support the claims made about the underrepresentation of gridded climate products' AR-related precipitation in J. D. Lundquist et al. (2015).

Model resolution also plays a key role in shaping both the rain-snow partitioning of precipitation and the efficiency at which water vapor becomes precipitation (Figure 10 and S6). Snowfall is enhanced by upwards of 20% in high-elevation regions of the California Sierra Nevada, particularly in the headwaters of the American River through the Kern River watersheds. Similarly, the precipitation efficiency (the amount of precipitation per unit of integrated water vapor) is enhanced by upwards of 20% throughout the Klamath Mountains, Coastal Ranges, and the Sierra Nevada in RRM-E3SM (3.5km). The combination of enhanced and more efficient precipitation and alterations to rain-snow partitioning changes the signature of runoff efficiency (the total runoff amount per total precipitation amount). Runoff efficiency is generally enhanced by upwards of 60% at low- to mid-elevations in northern California in RRM-E3SM (3.5km) compared to RRM-E3SM (14km), whereas in the high-elevation southern Sierra Nevada, a decrease is simulated. The enhanced runoff efficiency in RRM-E3SM (3.5km) is likely associated with more precipitation that is falling on wetter soils and, importantly, more snowmelt (as seen with more grid cells with runoff efficiencies at or exceeding 1). Conversely, runoff efficiencies decline in RRM-E3SM (3.5km) where snowfall is enhanced, which agrees with SNOTEL sites that indicate that positive dSWE changes occurred during the 1997 flood (Figure 9).

Even without a calibrated hydrologic model, comparing simulated inflow to observed inflow provides context for how well the model captures the key hydrologic-focused land-atmosphere interactions. This is because, in order to properly estimate reservoir inflows in the context of the 1997 flood, it is necessary that the model properly forecast the AR translational speed, plume intensity, and landfall location; the antecedent land surface conditions (e.g., snowpack and soil moisture); and the land-atmosphere interactions during and after the storm. Furthermore, model evaluation should also be done in decision-relevant regions (e.g., watersheds) instead of arbitrary latitude-longitude boxes. Therefore, to evaluate the value of the RRM-E3SM forecasts, we investigate reservoir inflows from the headwaters of eight major reservoirs, which represent a third (13.3 million-acre feet) of California's surface reservoir storage (Figure 11). Reservoir inflows are computed as basin averages of total runoff provided by the land-surface model in RRM-E3SM. In the headwaters of the two largest reservoirs (Lakes Shasta and Oroville), all simulations overestimate inflows, and resolution systematically increases the volume of water flowing through the system. This may be due to several factors, including a lack of parameter calibration in the land surface model (e.g., soil characteristics) and/or antecedent soil moisture being too high. Unfortunately, we could not find estimates of soil moisture content, from either *in-situ* or remote sensing sources, and were unable to evaluate soil moisture as we did precipitation and snowpack. We were also unable to find piezometer data recording groundwater height changes.

Although the magnitude of reservoir inflows is biased even in RRM-E3SM (3.5km), the shape of the reservoir inflow time series improves at finer resolutions in both Shasta and Oroville, with a more distinct peak inflow on 1 January 1997. This resolution dependence also holds for two other key northern California reservoirs (e.g., Folsom and New Melones). Unlike the results for Shasta and Oroville, the antecedent conditions (i.e., reservoir inflows at the beginning of 30 December 1996) in Folsom and New Melones Reservoirs seem to play a lesser role in model performance, with model drift in reservoir inflow estimates starting to occur one to two days after the forecasts have begun. Moving further south along the western slopes of the Sierra Nevada to Pine Flat and Terminus, RRM-E3SM (3.5km) matches reservoir inflows remarkably well, regardless of antecedent condition issues. Finally, RRM-E3SM simulations in the headwaters of Success and Isabella reservoirs match neither the amplitude nor shape of reservoir inflows, particularly Isabella. The lack of match between simulated and observed inflows is likely influenced by infrastructure and/or management decisions made above the reservoirs in these headwater regions, especially since RRM-E3SM simulations do not account for these factors.

To better contextualize RRM-E3SM runoff forecasts across resolution, we employ the PC causal inference algorithm with conditional mutual information test (Spirtes & Glymour, 1991; Ombadi et al., 2020). The influential strength of four hydrometeorological variables (i.e., IVT, precipitation, SWE, and 10 cm soil moisture content) on total runoff (overland flow, interflow, and baseflow) across California and within its 10 major reservoir headwater regions is shown in Figure 12 and Figure S8. The higher the stacked bar, the more variance is explained in total runoff. Each of the four hydrometeorological variables contributes a value ranging between zero and one, with a maximum possible total of four across variables. Across California, our causal analysis framework agrees with our prior suggestions that resolution plays an important role in amplifying the strength that both soil moisture content and SWE play in total runoff magnitude. With that said, atmospheric conditions (IVT and precipitation) heavily influence the total runoff signal across California comprising 84-94% of the total variance explained by the four chosen hydrometeorological variables (Figure S9). However, this causal relationship does change considerably from one reservoir headwater region to another (particularly in the central to southern Sierra Nevada).

Through this causal inference framework, we can also see that in certain reservoir headwater regions, resolution plays a systematic role in either adding more interactions between total runoff (more components contributing to each stacked bar) and all of the hydrometeorological variables (e.g., New Melones) or simplifying interactions to a single (e.g., Oroville) or fewer hydrometeorological variable(s) (e.g., Shasta). In other headwater regions, there is an insensitivity to resolution (e.g., Don Pedro and Isabella). In New Melones Lake, where runoff interaction diversity increases the most, IVT and SWE play no role in shaping runoff in RRM-E3SM (14km) and RRM-E3SM (7km), with a nearly a 50/50 split between precipitation and soil moisture, whereas RRM-E3SM (3.5km) shows a more equal interaction between all four hydrometeorological variables and runoff. Conversely, in Lakes Shasta and Oroville, three hydrometeorological variables play a key role in runoff forecasts in RRM-E3SM (14km) and RRM-E3SM (7km), yet precipitation becomes the dominant variable of influence in RRM-E3SM (3.5km), 91% and 100%, respectively (Figure S9). Finally, both Lake Don Pedro and Isabella Lake have an insensitivity to resolution where precipitation and soil moisture content play comparable roles in shaping total runoff across RRM-E3SM simulations.

Forecast lead time influence on atmospheric and land-surface process representation of the 1997 flood

To summarize the resolution dependence of RRM-E3SM simulations found thus far, we use Taylor diagrams (Figure 13) to show that although large-scale meteorology is relatively insensitive to finer horizontal resolutions (14km to 3.5km), even for landfalling AR characteristics (Figure 4), storm characteristics (e.g., storm total precipitation) and land-atmosphere interactions (e.g., storm duration dSWE) are sensitive to resolution. Dispersion in model results associated with forecast lead time is also shown. This will be the focus for the rest of our analysis, but to decrease the dimensionality of our analysis we focus on the best-performing simulation, RRM-E3SM (3.5km).

In RRM-E3SM (3.5km) both storm total precipitation and storm duration dSWE are weakly and not systematically sensitive to forecast lead time (Figure 14). The highest storm total precipitation and positive storm duration dSWE occurred in the forecast that was initialized on 1996-12-29 at 00Z. This finding is counter to our original hypothesis that forecast skill should increase as forecast lead time gets closer to 31 December 1996. This assumption was made because the 30 December 1996 at 12Z forecast has the least amount of time to drift from the conditions provided by ERA5 which could influence, for example, the AR intensity, landfall location, and translational speed.

Although forecast lead time does not appear to have a significant influence on storm total precipitation and storm duration dSWE over the period of 31 December 1996 to 4 January 1997, these metrics may mask temporal dependencies. To determine whether there are important diurnal and/or day-to-day differences across forecast lead times, Figure 15 shows both 6-hourly rates and cumulative 6-hourly totals for precipitation, dSWE, and runoff. The cumulative total precipitation estimated at the 52 precipitation gauge stations is well bracketed by the six RRM-E3SM (3.5km) forecasts. Hourly rates in precipitation show that precipitation diverges most across the six forecasts on 3 January

1997 (or four to six days post initialization of the forecast). From the perspective of dSWE, evaluated across the 50 SNOTEL sites, the six forecasts generally have similar tendencies throughout the flood period, but also disagree most on 3-4 January 1997. Negative dSWE values, an indication of the magnitude of snow ablation caused by the AR, were highest on 3 January 1997 in both observations and forecasts. The forecast spread on 3 January 1997 was -2 mm/hour to -7 mm/hour, which was generally stronger than was observed at SNOTEL sites. Undoubtedly, the spread in precipitation and SWE across forecasts from 3-4 January 1997 influenced runoff rates and totals in the reservoir headwater regions.

Finally, we evaluate how RRM-E3SM (3.5km) forecast lead time influences the causal strength and relationship between runoff and the four key hydrometeorological variables (i.e., IVT, precipitation, SWE, and 10 cm soil moisture content) over the period of 31 December 1996 to 4 January 1997. Interestingly, California-wide causal strength of the hydrometeorological variables on runoff generally is maintained across the six forecast lead times. Atmospheric conditions (IVT and precipitation) dominate the runoff signal (74-87% range across forecasts for the total variance explained for the four hydrometeorological variables chosen). The dominance of atmospheric conditions on runoff across forecasts holds in the headwaters of both Lakes Shasta and Oroville. However, akin to the resolution-focused results, antecedent conditions and land surface feedbacks play a larger role in shaping runoff in the reservoir headwater regions of the central to southern Sierra Nevada. For example, in the central and southern Sierra Nevada (New Melones Lake, Lake Don Pedro, and Isabella Lake) the role of antecedent and land surface conditions represents 46-51%, 40-51%, and 30-51%, respectively, on the causal relationship with runoff. Again, these percentages represent the range across forecasts for the total variance explained for just the four hydrometeorological variables chosen. The comparative randomness of forecast lead time relative to resolution on the causal strength and relationship of hydrometeorological variables on total runoff is likely due to the difficulty of exactly recreating the category 5 AR event life cycle. ARs have complex spatiotemporal structures that are hard to predict at watershed scales, particularly the AR landfall location latitude; the sweeping comma-shaped nature, topographic orthogonality, and translational speed of the AR plume at landfall; and the precise precipitation magnitude and rain-snow partitioning over the storm duration. This combined with biases in the forecast land-surface initial conditions, most of which are not truly constrained by *in-situ* observations (e.g., soil moisture probe data and groundwater table levels), could help to explain the randomness of forecast lead time on total runoff at individual reservoir regions.

Summary and Conclusions

We used a storyline approach to recreate California's flood of record, the New Year's flood of 1997, using a regionally refined Earth system modeling approach, RRM-E3SM. This is the first time RRM-E3SM has been used to systematically evaluate a key western United States hydrometeorological extreme event. We assessed how both forecast lead time and model horizontal resolution focused over California influenced forecast skill in recreating the flood event. Across several formal measures of forecast quality and value, RRM-E3SM (3.5km) had the highest skill in recreating the 1997 flood compared with lower-resolution versions of E3SM validated against *in-situ*, reanalysis, and gridded climate products.

RRM-E3SM's ability to simulate the North Pacific large-scale circulation patterns and IVT fields and landfalling AR characteristics prior to and during the 1997 flood were minimally influenced by the refinement of horizontal resolution over California. RRM-E3SM simulations largely agreed with ERA5 in the northern portions of California, particularly for extreme AR conditions. However, all RRM-E3SM simulations systematically produce excessively high AR categories in southern California; this is due to elevated amounts of water vapor in southern California and winds that are systematically higher than ERA5 throughout California. Regional refinement resolution in E3SM is important to the representation of storm total precipitation and storm duration changes in snow water equivalent. We find that RRM-E3SM (3.5km) best represents the statistical distributions of storm total precipitation at 52 precipitation gauge sites, with particular improvement in the precipitation maxima. We attribute this to a better representation of both California's mountainous topography as well as important mesoscale circulations in driving precipitation location and magnitude, notably the Sierra barrier jet. Enhanced snowfall at higher elevations and snowpack ablation at low-to-mid elevations are also better represented in RRM-E3SM (3.5km), as shown by comparison to 50 snow pillow sites and a gridded climate product.

Reservoir inflows represent the integrated watershed response resulting from interactions between atmospheric processes with topography. These interactions drive the simulated precipitation patterns and subsequently interact with land surface processes such as snowpack accumulation and melt, soil moisture content, and surface-through-subsurface flow. Simulated inflows exhibit mixed forecast skill across RRM-E3SM simulations. In general, reservoir inflow time series magnitude and, in some cases, shape were off across RRM-E3SM simulations. This is partly due to the integrated surface-through-subsurface hydrology being simulated with uncalibrated (or "out-of-the-box") parameter settings. Using these parameter values shows how E3SM's default settings, often optimized for mean state skill, represent extreme runoff. Notably, although uncalibrated, RRM-E3SM (3.5km) more consistently matched the time series shape of reservoir inflows across five of the eight major reservoirs in California. Future work will leverage the skillfully-resolved atmospheric fields, particularly in RRM-E3SM (3.5km), to run offline integrated hydrologic models (Maina et al., 2022) to assess partitioning between overland flow and groundwater recharge and/or water infrastructure models (Yates et al., 2022) to assess flood inundation potential associated with management decisions.

In addition to not accounting for water management infrastructure in E3SM, there were difficulties in validating certain aspects of the 1997 flood. Specifically, although the antecedent conditions (e.g., soil moisture content and groundwater table levels) provided by the "Betacast" offline five-year ELM spinup procedure driven by ERA5 meteorology undoubtedly shaped reservoir inflow estimates, more observationally-constrained initial conditions for the simulations were not available. Soil moisture content data (both *in-situ* and remote sensing-based estimates) were impossible to find at sub-monthly timescales prior to the year 2000 and, in particular, in mountains from missing data gaps, partly due to the effects of complex terrain and cloudy days on satellite retrievals. Similarly, observational estimates of groundwater table depths (e.g., piezometers and/or satellite-based estimates) were not publicly available.

Forecast lead time resulted in a random effect on the hydrometeorological representation of the 1997 flood. We speculate this is because the forecast lead times chosen (2-to-4 days prior to the 1997 flood onset) were comfortably within the forecast predictability of large-scale synoptic events like ARs (Haiden et al., 2021) and results were therefore dependent on more chaotic spinup processes, mesoscale processes with the main precipitation shield, and small-scale interactions of flow with orography. Although examining the sub-seasonal-to-seasonal forecast skill of E3SM is beyond the scope of this study, L'Heureux et al. (2021) has shown that precipitation forecast skill across seven Earth system model forecasts for California begins to sharply drop with lead times of 8-14 days. Alternatively, to isolate why 2-to-4 day forecast lead time had a relatively random effect on storm total precipitation RRM-E3SM can be run similarly to a weather forecast model, where data produced outside of the regionally refined domain is swapped with reanalysis data (Kruse et al., 2022; Zhang et al., 2022), to better constrain the lateral boundary conditions and, ultimately, the lifecycle of the AR propagation and landfall. Alternatively, the use of perturbed physics ensembles may help to further constrain which subgrid-scale parameterization most influenced drift in AR propagation and landfall and hydrometeorological characteristics of the RRM-E3SM forecasts (Mulholland et al., 2017). Last, given the noted uncertainties in land surface initial conditions, an AR-induced flood event that overlaps with recent high-resolution satellite-based estimates (Vergopolan et al., 2022) could be performed with RRM-E3SM to better isolate the role of antecedent conditions (e.g., soil moisture content) on flood event characteristics (e.g., reservoir inflows). Practically, the lack of hydrometeorological sensitivity with forecast lead time between two to four days prior to the onset of the flood event implies that if a flood manager is interested in event evolution at a specific point an ensemble forecast approach is necessary (e.g., simulations spanning multiple lead times and/or perturbed physics).

Overall, RRM-E3SM (3.5km) forecast ensemble average skill in recreating the 1997 flood gives confidence in its utility to aid flood resiliency planning. To further the utility of these storyline simulations, in future work, we will investigate flood characteristics if a 1997-like flood event were to have happened without anthropogenic climate change or were to happen again at different global warming levels. We hope that these storyline recreations of the 1997 flood event in past and future climates can supplement ongoing efforts in water resource agency flood resiliency planning efforts related to extreme events, especially those involving compounding and/or cascading processes.

Acknowledgments

This study was primarily funded by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy Regional and Global Model Analysis (RGMA) Program. Authors Inda-Diaz, Ombadi, Pasquier, Rhoades, Srivastava, Ullrich, and Wehner were funded by "the Calibrated and Systematic Characterization, Attribution and Detection of Extremes (CASCADE)" Science Focus Area (award no. DE-AC02-05CH11231). Authors Dennis, Jones, McCrary, McGinnis, Rahimi-Esfarjani, Rhoades, Slinskey, Srivastava, Ullrich, and Zarzycki were funded by the "An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US" project (award no. DE-SC0016605). Authors Hatchett and Heggli received support from the Nevada Department of Transportation under agreement P296-22-803. We would also like to acknowledge Smitha Buddhavarapu and Kripa Jagannathan for the considerable time and energy they provided in facilitating scientist-stakeholder discussions in the HyperFACETS project. The facilitated discussions played a key role in helping us to choose the 1997 flood event as a featured storyline within the HyperFACETS project. We also appreciate their constructive suggestions on our manuscript draft.

Open Research

Analysis and model simulations were performed using the National Energy Research Scientific Computing Center (NERSC), specifically Cori-Haswell and Cori-KNL supercomputing facilities. ERA5 is publicly available at the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) at `https://cds.climate.copernicus.eu/#!/search?text=ERA5`. The SSM/I data used in Figure 1 are produced by Remote Sensing Systems. Data are available at `www.remss.com/missions/ssmi`. The Betacast source code is available at `https://github.com/zarzycki/betacast`. The RRM-CESM simulations generated for this study will be made accessible via a NERSC Science Gateway at the time of publication.

References

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., … Sadegh, M. (2020). Climate Extremes and Compound Hazards in a Warming World. *Annual Review of Earth and Planetary Sciences*, *48*(1), 519-548. doi: 10.1146/annurev-earth-071719-055228

Bambach, N. E., Rhoades, A. M., Hatchett, B. J., Jones, A. D., Ullrich, P. A., & Zarzycki, C. M. (2021). Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile. *International Journal of Climatology*, 1-29. doi: https://doi.org/10.1002/joc.7379

Bukovsky, M. S., Gutowski, W., Mearns, L. O., Paquin, D., & Pryor, S. C. (2023). Climate storylines. *Bulletin of the American Meteorological Society*, *104*(1), E96 - E98. doi: https://doi.org/10.1175/BAMS-D-22-0224.1

Collow, A. B. M., Shields, C. A., Guan, B., Kim, S., Lora, J. M., McClenny, E. E., ... Wehner, M. (2022). An Overview of ARTMIP's Tier 2 Reanalysis Intercomparison: Uncertainty in the Detection of Atmospheric Rivers and Their Associated Precipitation. *Journal of Geophysical Research: Atmospheres*, *127*(8), e2021JD036155. doi: https://doi.org/10.1029/2021JD036155

Copernicus Climate Change Service Climate Data Store (CDS). (2017). *Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.* Retrieved 2022-10-13, from `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview`

Corringham, T. W., McCarthy, J., Shulgina, T., Gershunov, A., Cayan, D. R., & Ralph, F. M. (2022). Climate change contributions to future atmospheric river flood damages in the western United States. *Scientific reports*, *12*(1), 1–9. doi: https://doi.org/10.1038/s41598-022-15474-2

Corringham, T. W., Ralph, F. M., Gershunov, A., Cayan, D. R., & Talbot, C. A. (2019). Atmospheric rivers drive flood damages in the western United States. *Science Advances*, *5*(12). doi: 10.1126/sciadv.aax4631

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

Fang, Y., Liu, Y., & Margulis, S. A. (2022). A western United States snow reanalysis dataset over the Landsat era from water years 1985 to 2021. *Scientific Data*, *9*(1), 1–17. doi: https://doi.org/10.1038/s41597-022-01768-7

Fasullo, J. T. (2020). Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geoscientific Model Development*, *13*(8), 3627–3642. doi: 10.5194/gmd-13-3627-2020

Fish, M. A., Wilson, A. M., & Ralph, F. M. (2019, 10). Atmospheric River Families:

Definition and Associated Synoptic Conditions. *Journal of Hydrometeorology*, *20*(10), 2091-2108. doi: 10.1175/JHM-D-18-0217.1

Galewsky, J., & Sobel, A. (2005). Moist dynamics and orographic precipitation in northern and central california during the new year's flood of 1997. *Monthly Weather Review*, *133*(6), 1594 - 1612. doi: 10.1175/MWR2943.1

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., ... Williams, D. N. (1999). An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *80*(1), 29–56. doi: 10.1175/1520-0477(1999)080⟨0029:AOOTRO⟩2.0.CO;2

Gershunov, A., Shulgina, T., Clemesha, R. E., Guirguis, K., Pierce, D. W., Dettinger, M. D., ... others (2019). Precipitation regime change in Western North America: the role of Atmospheric Rivers. *Scientific reports*, *9*(1), 1–11. doi: https://doi.org/10.1038/s41598-019-46169-w

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., ... Bader, D. C. (2022). The DOE E3SM Model Version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, *n/a*(n/a), e2022MS003156. doi: https://doi.org/10.1029/2022MS003156

Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., ... Zarzycki, C. (2020). The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society*, *101*(5), E664 - E683. doi: 10.1175/BAMS-D-19-0113.1

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., & Richardson, D. (2021, 01/2021). Evaluation of ECMWF forecasts, including the 2020 upgrade. *ECMWF Technical Memoranda*(880). Retrieved from `https://www.ecmwf.int/node/19879` doi: 10.21957/6njp8byz4

Hanak, E., & Lund, J. R. (2012). Adapting California's water management to climate change. *Climatic change*, *111*, 17–44. doi: https://doi.org/10.1007/s10584-011-0241-3

Harrop, B., Balaguru, K., Golaz, J.-C., Leung, L., Mahajan, S., Rhoades, A., ... Mametjanov, A. (2022). Evaluating the water cycle over CONUS using multiple metrics for the Energy Exascale Earth System Model version 1 (E3SMv1) across resolutions. *Journal of Advances in Modeling Earth Systems*(In Revision).

Hatchett, B. J., & McEvoy, D. J. (2018). Exploring the origins of snow drought in the northern sierra nevada, california. *Earth Interactions*, *22*(2), 1 - 13. doi: https://doi.org/10.1175/EI-D-17-0027.1

Heggli, A., Hatchett, B., Schwartz, A., Bardsley, T., & Hand, E. (2022). Toward snowpack runoff decision support. *iScience*, *25*(5), 104240. doi: https://doi.org/10.1016/j.isci.2022.104240

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Huang, X., Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2016). An evaluation of the variable-resolution CESM for modeling California's climate. *Journal of Advances in Modeling Earth Systems*, *8*(1), 345–369. doi: 10.1002/2015MS000559

Huang, X., & Swain, D. L. (2022). Climate change is increasing the risk of a california megaflood. *Science Advances*, *8*(32), eabq0995. doi: 10.1126/sciadv.abq0995

Hughes, M., Neiman, P. J., Sukovich, E., & Ralph, M. (2012). Representation of the Sierra Barrier Jet in 11 years of a high-resolution dynamical reanalysis downscaling compared with long-term wind profiler observations. *Journal of Geophysical Research: Atmospheres*, *117*(D18). doi: https://doi.org/10.1029/2012JD017869

Ivancic, T. J., & Shaw, S. B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, *133*, 681–693. doi: https://doi.org/10.1007/s10584-015-1476-1

Jagannathan, K., Jones, A. D., & Ray, I. (2021). The making of a metric: Co-producing decision-relevant climate science. *Bulletin of the American Meteorological Society*, *102*(8), E1579 - E1590. doi: 10.1175/BAMS-D-19-0296.1

Jones, A. D., Rastogi, D., Vahmani, P., Stansfield, A., Reed, K., Thurber, T., … Rice, J. S. (2022). IM3/HyperFACETS Thermodynamic Global Warming

(TGW) Simulation Datasets (v1.0.0) [Data set]. *MSD-LIVE Data Repository*. doi: https://doi.org/10.57931/1885756

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*(9), 2204 - 2210. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/127/9/1520-0493_1999_127_2204_fasocr_2.0.co_2.xml` doi: https://doi.org/10.1175/1520-0493(1999)127⟨2204:FASOCR⟩2.0.CO;2

Kattelmann, R. (1997). Flooding from rain-on-snow events in the Sierra Nevada. *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, *239*, 59–66.

Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, *21*(4), e952. doi: https://doi.org/10.1002/asl.952

Kruse, C. G., Bacmeister, J. T., Zarzycki, C. M., Larson, V. E., & Thayer-Calder, K. (2022). Do Nudging Tendencies Depend on the Nudging Timescale Chosen in Atmospheric Models? *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS003024. doi: https://doi.org/10.1029/2022MS003024

La Follette, P. T., Teuling, A. J., Addor, N., Clark, M., Jansen, K., & Melsen, L. A. (2021). Numerical daemons of hydrological models are summoned by extreme precipitation. *Hydrology and Earth System Sciences*, *25*(10), 5425–5446. Retrieved from `https://hess.copernicus.org/articles/25/5425/2021/` doi: 10.5194/hess-25-5425-2021

Lauritzen, P. H., Bacmeister, J. T., Callaghan, P. F., & Taylor, M. A. (2015, dec). NCAR_topo (v1.0): NCAR global model topography generation software for unstructured grids. *Geoscientific Model Development*, *8*(12), 3975–3986. Retrieved from `https://doi.org/10.5194%2Fgmd-8-3975-2015` doi: 10.5194/gmd-8-3975-2015

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*(5), 2275-2282. doi: https://doi.org/10.1002/2016GL067765

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., . . .

Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. Retrieved from https://esd.copernicus.org/articles/11/491/2020/ doi: 10.5194/esd-11-491-2020

Lemos, M. C., Arnott, J. C., Ardoin, N. M., Baja, K., Bednarek, A. T., Dewulf, A., ... others (2018). To co-produce or not to co-produce. *Nature Sustainability*, *1*(12), 722–724. doi: https://doi.org/10.1038/s41893-018-0191-0

Lemos, M. C., Kirchhoff, C. J., & Ramprasad, V. (2012). Narrowing the climate information usability gap. *Nature climate change*, *2*(11), 789–794. doi: https://doi.org/10.1038/nclimate1614

Liu, W., Ullrich, P. A., Guba, O., Caldwell, P. M., & Keen, N. D. (2022). An Assessment of Nonhydrostatic and Hydrostatic Dynamical Cores at Seasonal Time Scales in the Energy Exascale Earth System Model (E3SM). *Journal of Advances in Modeling Earth Systems*, *14*(2), e2021MS002805. doi: https://doi.org/10.1029/2021MS002805

Liu, W., Ullrich, P. A., Li, J., Zarzycki, C., Caldwell, P. M., Leung, L. R., & Qian, Y. (2023). The June 2012 North American Derecho: A Testbed for Evaluating Regional and Global Climate Modeling Systems at Cloud-Resolving Scales. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003358. doi: https://doi.org/10.1029/2022MS003358

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, *2*(150042). doi: doi:10.1038/sdata.2015.42

Lott, N., Sittel, M. C., & Ross, D. (1997). The winter of '96-'97 : west coast flooding. *National Climatic Data Center technical report ; 97-01*, 1 - 22. doi: https://repository.library.noaa.gov/view/noaa/13812

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (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. doi: https://doi.org/10.1175/BAMS-D-19-0001.1

Lundquist, J. D., Hughes, M., Henn, B., Gutmann, E. D., Livneh, B., Dozier, J., & Neiman, P. (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. doi: https://doi.org/10.1175/JHM-D-15-0019.1

L'Heureux, M. L., Tippett, M. K., & Becker, E. J. (2021). Sources of subseasonal skill and predictability in wintertime california precipitation forecasts. *Weather and Forecasting*, *36*(5), 1815 - 1826. doi: https://doi.org/10.1175/WAF-D-21-0061.1

Maina, F. Z., Rhoades, A., Siirila-Woodburn, E. R., & Dennedy-Frank, P.-J. (2022). Projecting end-of-century climate extremes and their impacts on the hydrology of a representative California watershed. *Hydrology and Earth System Sciences*, *26*(13), 3589–3609. doi: 10.5194/hess-26-3589-2022

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245–259. doi: http://www.jstor.org/stable/1907187

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A landsat-era sierra nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, *17*(4), 1203 - 1221. doi: https://doi.org/10.1175/JHM-D-15-0177.1

Mulholland, D. P., Haines, K., Sparrow, S. N., & Wallom, D. (2017). Climate model forecast biases assessed with a perturbed physics ensemble. *Climate Dynamics*, *49*, 1729–1746. doi: https://doi.org/10.1007/s00382-016-3407-x

Murphy, A. H. (1993). What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, *8*(2), 281 - 293. doi: 10.1175/1520-0434(1993)008⟨0281:WIAGFA⟩2.0.CO;2

Neiman, P. J., Hughes, M., Moore, B. J., Ralph, F. M., & Sukovich, E. M. (2013). Sierra barrier jets, atmospheric rivers, and precipitation characteristics in northern california: A composite perspective based on a network of wind profilers. *Monthly Weather Review*, *141*(12), 4211 - 4233. doi: https://doi.org/10.1175/MWR-D-13-00112.1

Neiman, P. J., Sukovich, E. M., Ralph, F. M., & Hughes, M. (2010). A seven-year wind profiler–based climatology of the windward barrier jet along california's northern sierra nevada. *Monthly Weather Review*, *138*(4), 1206 - 1233. doi: https://doi.org/10.1175/2009MWR3170.1

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., . . . Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous united states. *Journal of Hydrometeorology*, *16*(6), 2481 - 2500. doi: https://doi.org/10.1175/JHM-D-15-0026.1

Ombadi, M., Nguyen, P., Sorooshian, S., & lin Hsu, K. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7). Retrieved from `https://doi.org/10.1029/2020wr027251` doi: 10.1029/2020wr027251

Ombadi, M., Nguyen, P., Sorooshian, S., & lin Hsu, K. (2021). How much information on precipitation is contained in satellite infrared imagery? *Atmospheric Research*, *256*, 105578. Retrieved from `https://doi.org/10.1016/j.atmosres.2021.105578` doi: 10.1016/j.atmosres.2021.105578

Osterhuber, R., & Schwartz, A. (2021). Snowpack, precipitation, and temperature measurements at the Central Sierra Snow Laboratory for water years 1971 to 2019. *Dryad, Dataset*. doi: https://doi.org/10.6078/D1941T

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L.-R., ... Zhou, Y. (2022). Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD036013. doi: https://doi.org/10.1029/2021JD036013

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., ... Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139-152. doi: https://doi.org/10.1016/j.rse.2016.06.018

Pierce, D. W., Su, L., Cayan, D. R., Risser, M. D., Livneh, B., & Lettenmaier, D. P. (2021). An extreme-preserving long-term gridded daily precipitation dataset for the conterminous united states. *Journal of Hydrometeorology*, *22*(7), 1883 - 1895. Retrieved from `https://journals.ametsoc.org/view/journals/hydr/22/7/JHM-D-20-0212.1.xml` doi: 10.1175/JHM-D-20-0212.1

Porter, K., Wein, A., Alpers, C. N., Baez, A., Barnard, P. L., Carter, J., ... others (2011). *Overview of the ARkStorm scenario* (Tech. Rep.). US Geological Survey. doi: https://pubs.er.usgs.gov/publication/ofr20101312

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., ... Smallcomb, C. (2019). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269-289. doi: 10.1175/BAMS-D-18-0023.1

Rauscher, S. A., & Ringler, T. D. (2014). Impact of Variable-Resolution Meshes on Midlatitude Baroclinic Eddies Using CAM-MPAS-A. *Monthly Weather Review*, *142*(11), 4256–4268. doi: 10.1175/MWR-D-13-00366.1

Rauscher, S. A., Ringler, T. D., Skamarock, W. C., & Mirin, A. A. (2013). Exploring a Global Multiresolution Modeling Approach Using Aquaplanet Simulations. *Journal of Climate*, *26*(8), 2432-2452. doi: 10.1175/JCLI-D-12-00154.1

Raymond, C., Horton, R. M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., … others (2020). Understanding and managing connected extreme events. *Nature climate change*, *10*(7), 611–621. doi: https://doi.org/10.1038/s41558-020-0790-4

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. doi: 10.1175/2007JCLI1824.1

Rhoades, A. M., Huang, X., Ullrich, P. A., & Zarzycki, C. M. (2016). Characterizing Sierra Nevada Snowpack Using Variable-Resolution CESM. *Journal of Applied Meteorology and Climatology*, *55*(1), 173–196. doi: 10.1175/JAMC-D-15-0156.1

Rhoades, A. M., Jones, A. D., O'Brien, T. A., O'Brien, J. P., Ullrich, P. A., & Zarzycki, C. M. (2020). Influences of North Pacific Ocean Domain Extent on the Western U.S. Winter Hydroclimatology in Variable-Resolution CESM. *Journal of Geophysical Research: Atmospheres*, *125*(14), e2019JD031977. doi: 10.1029/2019JD031977

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., … Zhou, Y. (2020). The Shifting Scales of Western U.S. Landfalling Atmospheric Rivers Under Climate Change. *Geophysical Research Letters*, *47*(17), e2020GL089096. doi: https://doi.org/10.1029/2020GL089096

Rhoades, A. M., Risser, M. D., Stone, D. A., Wehner, M. F., & Jones, A. D. (2021). Implications of warming on western United States landfalling atmospheric rivers and their flood damages. *Weather and Climate Extremes*, *32*, 100326. doi: https://doi.org/10.1016/j.wace.2021.100326

Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2018). Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM. *Climate Dynamics*, *50*(1), 261–288. doi: 10.1007/s00382-017-3606-0

Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A., Morrison, H., ... Collins, W. D. (2018). Sensitivity of Mountain Hydroclimate Simulations in Variable-Resolution CESM to Microphysics and Horizontal Resolution. *Journal of Advances in Modeling Earth Systems*, *10*(6), 1357–1380. doi: 10.1029/2018MS001326

Ricciotti, J. A., & Cordeira, J. M. (2022). Summarizing relationships among landfalling atmospheric rivers, integrated water vapor transport, and california watershed precipitation 1982–2019. *Journal of Hydrometeorology*, *23*(9), 1439 - 1454. doi: https://doi.org/10.1175/JHM-D-21-0119.1

Runge, J. (2023). Modern causal inference approaches to investigate biodiversity-ecosystem functioning relationships. *nature communications*, *14*(1), 1917. doi: https://doi.org/10.1038/s41467-023-37546-1

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., ... others (2019). Inferring causation from time series in Earth system sciences. *Nature communications*, *10*(1), 2553. doi: https://doi.org/10.1038/s41467-019-10105-3

Sakaguchi, K., Lu, J., Leung, L. R., Zhao, C., Li, Y., & Hagos, S. (2016). Sources and pathways of the upscale effects on the Southern Hemisphere jet in MPAS-CAM4 variable-resolution simulations. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1786-1805. doi: 10.1002/2016MS000743

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic unit maps: US Geological Survey water supply paper 2294.* U.S. Geological Survey. doi: https://pubs.usgs.gov/wsp/wsp2294/

Shepherd, T. G. (2019). Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2225), 20190013. doi: 10.1098/rspa.2019.0013

Siirila-Woodburn, E., Rhoades, A., Hatchett, B., Huning, L., Szinai, J., Tague, C., ... Kaatz, L. (2021). Evidence of a low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth and Environment*, *2*, 800–819. doi: https://doi.org/10.1038/s43017-021-00219-y

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, *9*(1), 62–72. Retrieved from https://doi.org/10.1177/089443939100900106 doi: 10.1177/089443939100900106

Srivastava, A. K., Grotjahn, R., Ullrich, P. A., & Sadegh, M. (2021). Pooling data improves multimodel idf estimates over median-based idf estimates: Analysis over the susquehanna and florida. *Journal of Hydrometeorology*, *22*(4), 971 - 995. Retrieved from https://journals.ametsoc.org/view/journals/hydr/22/4/JHM-D-20-0180.1.xml doi: https://doi.org/10.1175/JHM-D-20-0180.1

Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., & Bair, E. H. (2023). Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets. *The Cryosphere*, *17*(2), 567–590. doi: 10.5194/tc-17-567-2023

Sugihara, G., May, R., Ye, H., hao Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting Causality in Complex Ecosystems. *Science*, *338*(6106), 496-500. doi: 10.1126/science.1227079

Tang, Q., Golaz, J.-C., Van Roekel, L. P., Taylor, M. A., Lin, W., Hillman, B. R., ... Bader, D. C. (2022). The Fully Coupled Regionally Refined Model of E3SM Version 2: Overview of the Atmosphere, Land, and River. *Geoscientific Model Development Discussions*, *2022*, 1–64. Retrieved from https://gmd.copernicus.org/preprints/gmd-2022-262/ doi: 10.5194/gmd-2022-262

Tang, Q., Klein, S. A., Xie, S., Lin, W., Golaz, J.-C., Roesler, E. L., ... Zheng, X. (2019). Regionally refined test bed in e3sm atmosphere model version 1 (eamv1) and applications for high-resolution modeling. *Geoscientific Model Development*, *12*(7), 2679–2706. Retrieved from https://gmd.copernicus.org/articles/12/2679/2019/ doi: 10.5194/gmd-12-2679-2019

Tarouilly, E., Li, D., & Lettenmaier, D. P. (2021). Western U.S. Superfloods in the Recent Instrumental Record. *Water Resources Research*, *57*(9), e2020WR029287. doi: https://doi.org/10.1029/2020WR029287

The NCAR Command Language (Version 6.6.2). (2022). *Boulder, Colorado: UCAR/NCAR/CISL/TDD.* (http://dx.doi.org/10.5065/D6WD3XH5)

Trenberth, E., Berry, C., & Buja, E. (1993). *Vertical interpolation and truncation of* *model-coordinate data.* doi: 10.5065/D6HX19NH

Ullrich, P. A. (2014). Understanding the treatment of waves in atmospheric models. Part 1: The shortest resolved waves of the 1D linearized shallow-water equations. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1426-1440. doi: https://doi.org/10.1002/qj.2226

Ullrich, P. A., Devendran, D., & Johansen, H. (2016). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part ii. *Monthly Weather Review*, *144*(4), 1529–1549.

Ullrich, P. A., & Taylor, M. A. (2015). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part i. *Monthly Weather Review*, *143*(6), 2419–2440.

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090. doi: 10.5194/gmd-10-1069-2017

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., ... Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961-3012. doi: https://doi.org/10.1256/qj.04.176

Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., ... Wood, E. F. (2022). High-Resolution Soil Moisture Data Reveal Complex Multi-Scale Spatial Variability Across the United States. *Geophysical Research Letters*, *49*(15), e2022GL098586. doi: https://doi.org/10.1029/2022GL098586

Wehner, M. F., Zarzycki, C. M., & Patricola, C. M. (2019). Estimating the human influence on tropical cyclone intensity as the climate changes. In J. Collins & K. J. Walsh (Eds.), *Hurricane risk* (pp. 235–260). Springer.

Whipple, A. A., & Viers, J. H. (2019). Coupling landscapes and river flows to restore highly modified rivers. *Water Resources Research*, *55*(6), 4512-4532. doi: https://doi.org/10.1029/2018WR022783

Whipple, A. A., Viers, J. H., & Dahlke, H. E. (2017). Flood regime typology for floodplain ecosystem management as applied to the unregulated Cosumnes River of California, United States. *Ecohydrology*, *10*(5), e1817. doi: https://doi.org/10.1002/eco.1817

Wing, O. E., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M.,

. . . Kousky, C. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, *12*(2), 156–162. doi: https://doi.org/10.1038/s41558-021-01265-6

Wu, C., Liu, X., Lin, Z., Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., . . . Rahimi-Esfarjani, S. R. (2017). Exploring a Variable-Resolution Approach for Simulating Regional Climate in the Rocky Mountain Region Using the VR-CESM. *Journal of Geophysical Research: Atmospheres*, *122*(20), 10,939–10,965. doi: 10.1002/2017JD027008

Xu, Z., Di Vittorio, A., Zhang, J., Rhoades, A., Xin, X., Xu, H., & Xiao, C. (2021). Evaluating Variable-Resolution CESM Over China and Western United States for Use in Water-Energy Nexus and Impacts Modeling. *Journal of Geophysical Research: Atmospheres*, *126*(15), e2020JD034361. doi: https://doi.org/10.1029/2020JD034361

Xu, Z., Rhoades, A. M., Johansen, H., Ullrich, P. A., & Collins, W. D. (2018). An Intercomparison of GCM and RCM Dynamical Downscaling for Characterizing the Hydroclimatology of California and Nevada. *Journal of Hydrometeorology*, *19*(9), 1485–1506. doi: 10.1175/JHM-D-17-0181.1

Yang, Q., Leung, L. R., Lu, J., Lin, Y.-L., Hagos, S., Sakaguchi, K., & Gao, Y. (2017). Exploring the effects of a nonhydrostatic dynamical core in high-resolution aquaplanet simulations. *Journal of Geophysical Research: Atmospheres*, *122*(6), 3245-3265. doi: https://doi.org/10.1002/2016JD025287

Yates, D., Szinai, J., & Jones, A. D. (2022). Modeling the Water Systems of the Western US to Support Climate-Resilient Electricity System Planning. *Authorea Preprints*. doi: 10.1002/essoar.10512623.1

Zarzycki, C. M., Jablonowski, C., & Taylor, M. A. (2014). Using Variable-Resolution Meshes to Model Tropical Cyclones in the Community Atmosphere Model. *Monthly Weather Review*, *142*(3), 1221–1239. doi: 10.1175/MWR-D-13-00179.1

Zarzycki, C. M., Jablonowski, C., Thatcher, D. R., & Taylor, M. A. (2015). Effects of localized grid refinement on the general circulation and climatology in the community atmosphere model. *Journal of Climate*, *28*(7), 2777 - 2803. doi: https://doi.org/10.1175/JCLI-D-14-00599.1

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, *44*(2), 1141-1149. doi: 10.1002/2016GL071606

Zhang, S., Zhang, K., Wan, H., & Sun, J. (2022). Further improvement and evaluation of nudging in the E3SM Atmosphere Model version 1 (EAMv1): simulations of the mean climate, weather events, and anthropogenic aerosol effects. *Geoscientific Model Development*, *15*(17), 6787–6816. doi: 10.5194/gmd-15-6787-2022

Zhou, Y., O'Brien, T. A., Ullrich, P. A., Collins, W. D., Patricola, C. M., & Rhoades, A. M. (2021). Uncertainties in Atmospheric River Lifecycles by Detection Algorithms: Climatology and Variability. *Journal of Geophysical Research: Atmospheres*, *126*(8), e2020JD033711. doi: https://doi.org/10.1029/2020JD033711

Figure 1.

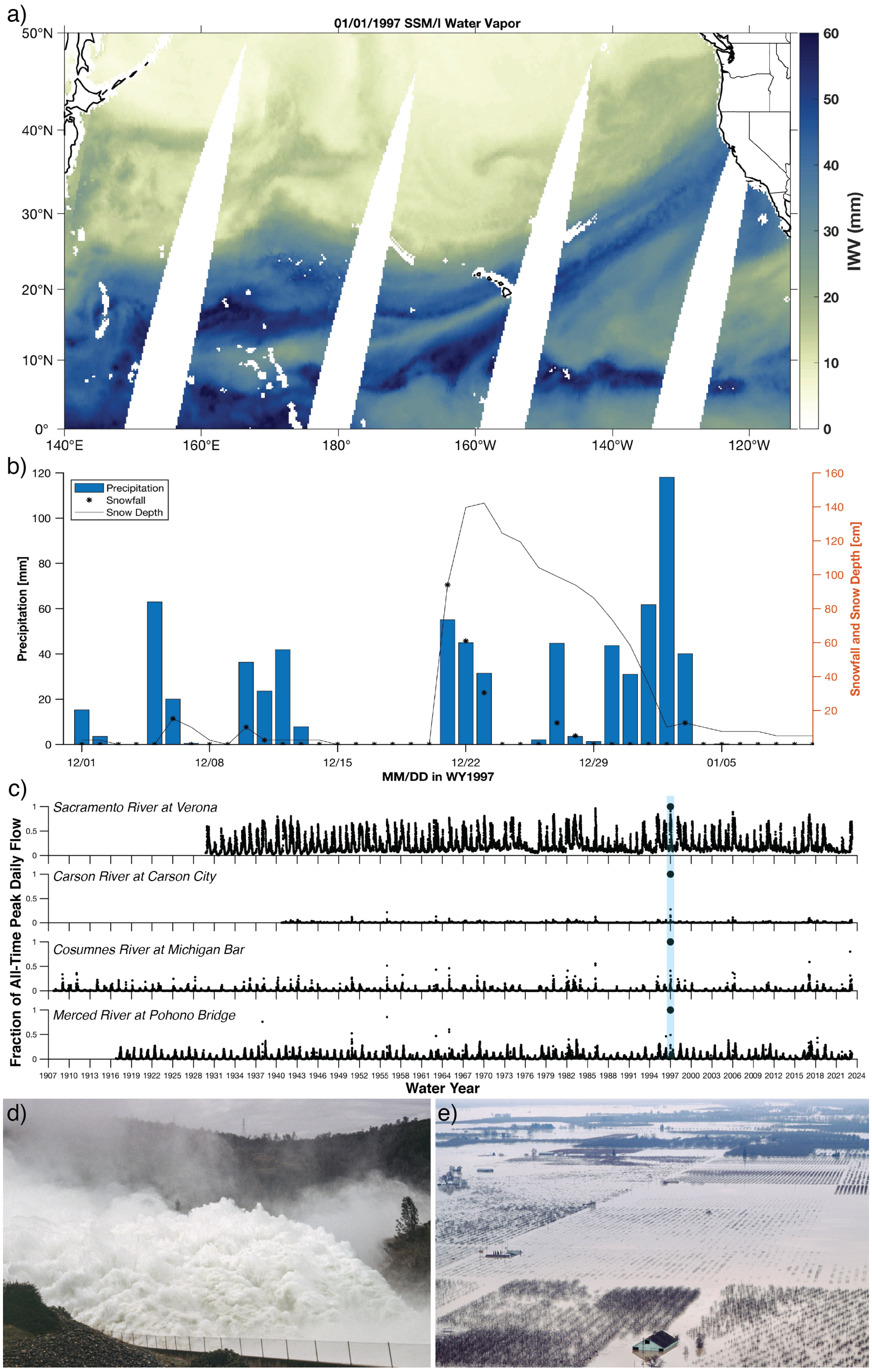

Figure 1. a) Special Sensor Microwave Imager (SSM/I) integrated water vapor on 1 January 1997. b) Tahoe City precipitation, snowfall, and snow depth from 1 December 1996 to 10 January 1997. c) Examples of all-time peak daily flows set during the event on major river systems in California and Nevada. d) Reservoir releases from Lake Oroville approached 4,530 cubic meters per second (160,000 cubic feet per second). (e) Flooding inundated the Sacramento Valley of California following heavy rainfall and snowmelt. Images d) and e) courtesy of the California Department of Water Resources.

Figure 2.

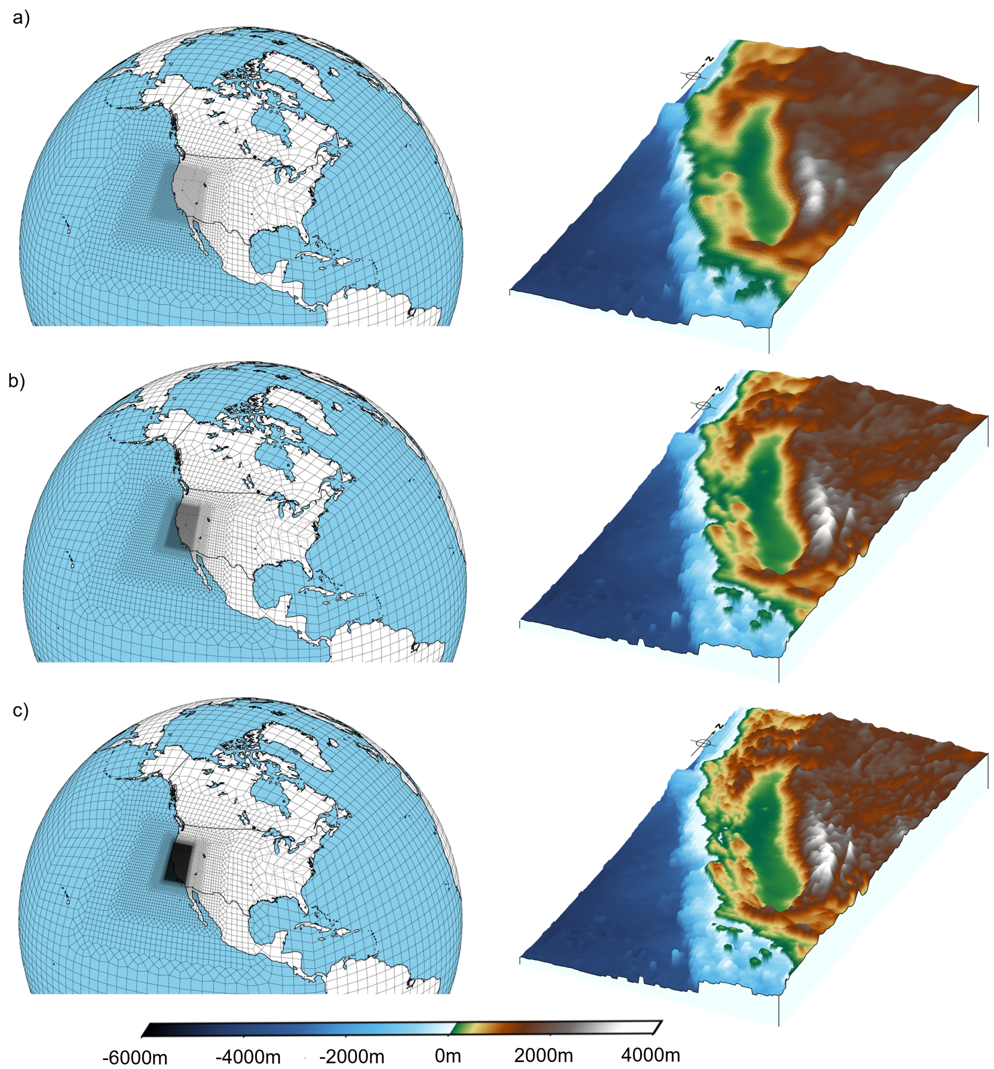

Figure 2. The Regionally Refined Mesh enabled Energy Exascale Earth System Model (RRM-E3SM) cases used to recreate the 1997 flood at horizontal resolutions of a) 0.125° (∼14km) b) 0.063° (∼7km) and c) 0.031° (∼3.5km) focused over California. Each RRM-E3SM case's topography is provided to the right of the grid refinement map. Note that ocean bathymetry is not represented in the RRM-E3SM simulations, but is included here for illustrative purposes.

Figure 3.

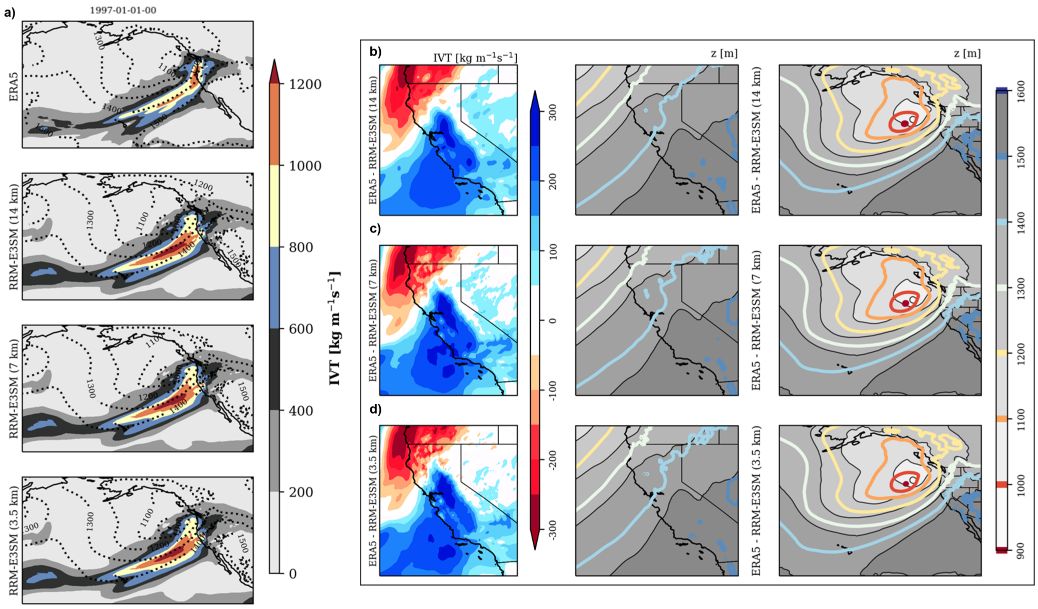

Figure 3. a) Forecast ensemble average integrated vapor transport (IVT) with 850mb geopotential height (dashed; units in meters) fields for ERA5 and each RRM-E3SM case. b) Difference in IVT between ERA5 and RRM-E3SM (14km), RRM-E3SM (7km) and RRM-E3SM (3.5km) (top, middle, and bottom rows, respectively), when the AR makes landfall in California on 1 January 1997. c-d) 850 mb geopotential height for ERA5 (gray-to-white contours) and RRM-E3SM (colored contours) over California (c) and the Northeastern Pacific (d), also at the time of AR landfall.

Figure 4.

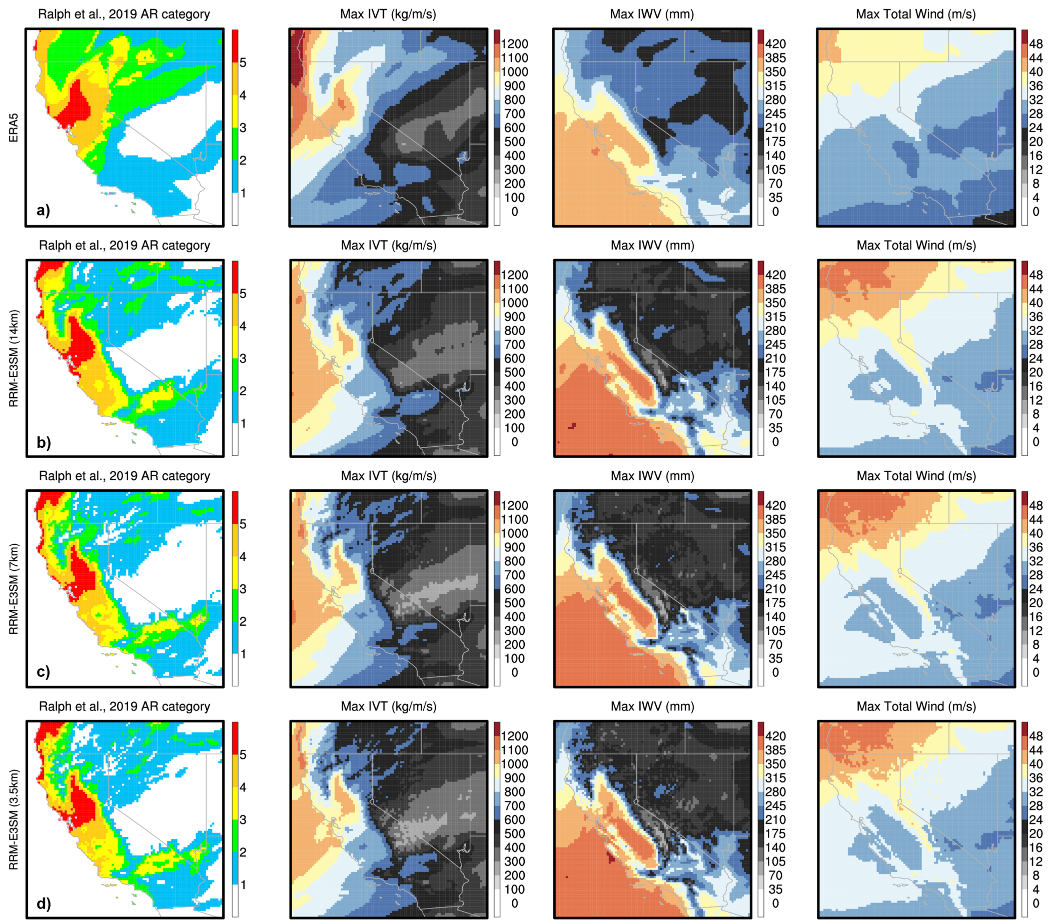

Figure 4. AR characteristics for the forecast ensemble average between the period of 31 December 1996 up to 4 January 1997. Characteristics include the Ralph et al. (2019) category scale (left column), maximum integrated vapor transport (IVT, second column), maximum integrated water vapor (IWV, third column), and maximum integrated total wind (right column) for a) ERA5 b) RRM-E3SM (14km) c) RRM-E3SM (7km) and d) RRM-E3SM (3.5km).

Figure 5.

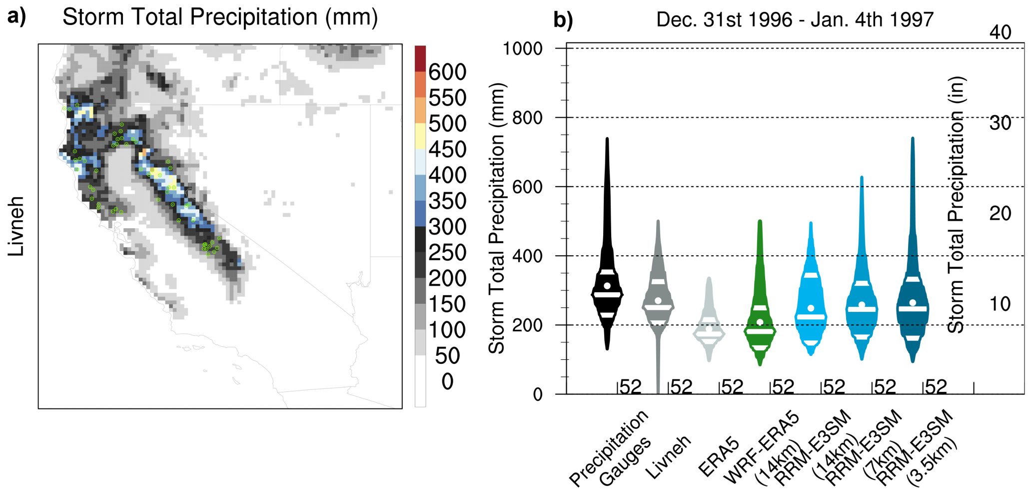

Figure 5. a) Storm total precipitation (31 December 1996 to 4 January 1997) from the Livneh product. Green dots highlight the locations of the 52 precipitation gauges used by NOAA to produce the 1997 flood event storm summary (`https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). b) Violin plots of reanalysis and model estimate storm total precipitation derived from the nearest grid cell to the 52 stations shown in a). The mean is shown with a white dot, and white lines indicate the 25th, median, and 75th percentiles. The shape of each violin reflects the probability density function of the data.

Figure 6.

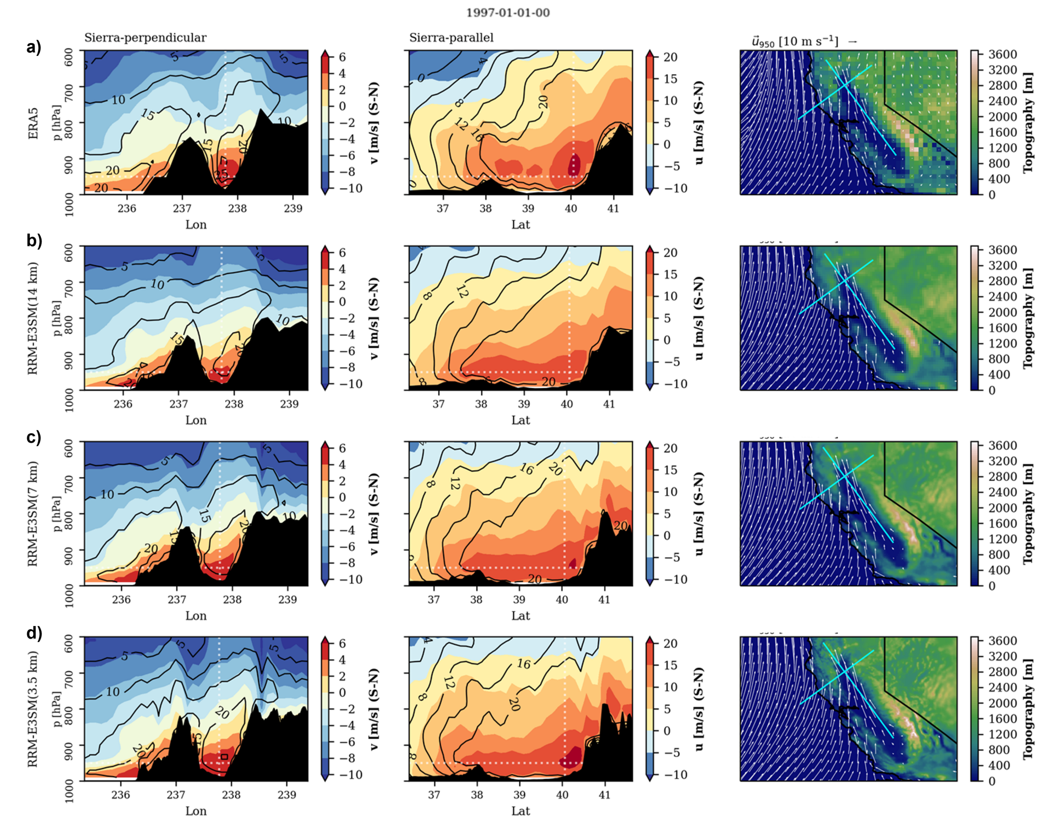

Figure 6. Sierra-perpendicular and Sierra-parallel cross sections of meridional (v) and zonal (u) winds at the start of the 1997 flood event AR landfall (1 January 1997) for ERA5 and the six-forecast ensemble average estimates provided by RRM-E3SM. The longitudinal and latitudinal cross-section transect lines are shown on the right-most column sub-panel figures overlaid on California. In the case of Sierra-perpendicular (Sierra-parallel), positive values mean that winds are blowing from South to North (West to East).

Figure 7.

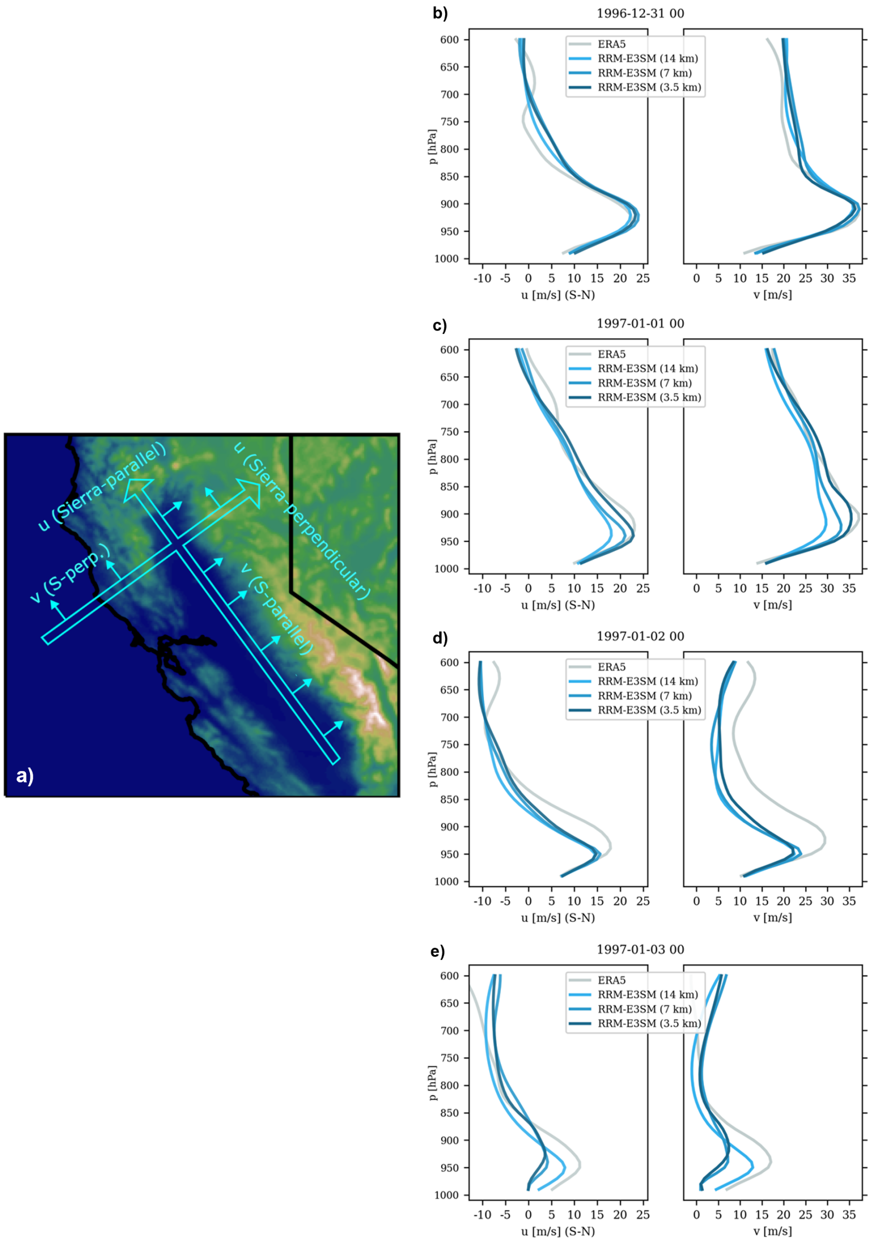

Figure 7. Sierra-parallel and Sierra-perpendicular vertical profiles of zonal (u) and meridional (v) wind speeds at the latitudinal location of the jet maxima with altitude for ERA5 and the six-forecast ensemble average RRM-E3SM simulations. a) shows the latitudinal and longitudinal transects and positive wind direction from the Sierra perspective. b-e) shows the vertical wind profiles at the intersection of the transects for the duration of the 1997 flood (31 December 1996 through 3 January 1997).

Figure 8.

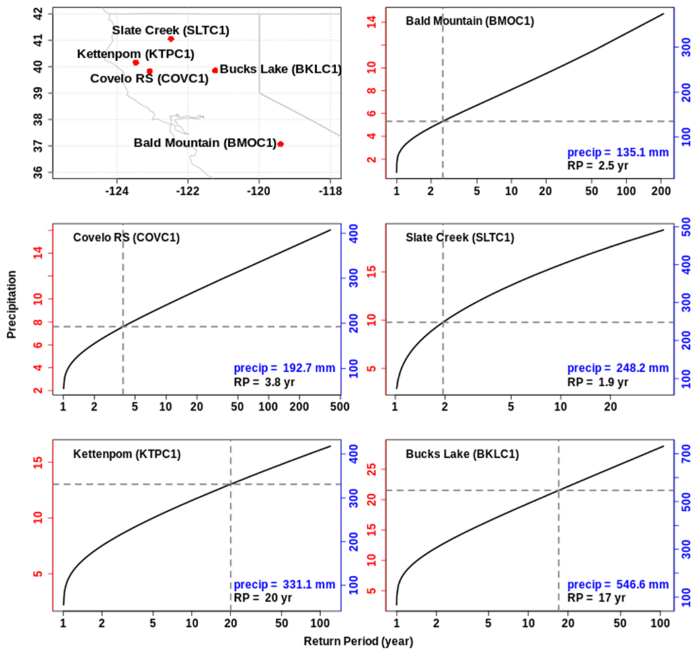

Figure 8. Return periods of the 4-day precipitation totals (Rx4day; 31 December 1996 through 3 January 1997) estimated using a non-stationary GEV framework on the Livneh product. To estimate the return period, the annual maxima of the Rx4day are interpolated to the precipitation gauge locations using first-order conservative remapping. The five stations shown (out of 52 total) are selected to indicate the minimum, 25th, 50th, 75th, and maximum Rx4day across the gauge locations. The left (right) y-axis provides Rx4day in English (metric) units. The horizontal and vertical dashed lines show the Rx4day and the corresponding return period in the Livneh product, as do the annotations in the bottom right. The x-axis (return period) is plotted on the log scale.

Figure 9.

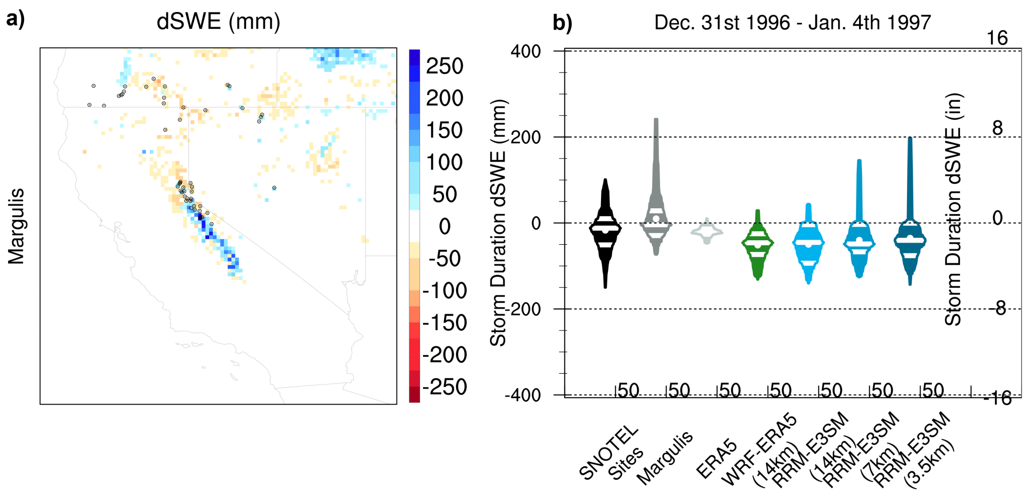

Figure 9. a) Storm duration change in snow water equivalent, dSWE, (31 December 1996 through 4 January 1997) from the Margulis product. Black dots highlight the locations of the 50 SNOTEL stations within the vicinity of the 1997 flood. b) Violin plots of reanalysis and model estimate storm duration dSWE derived from the nearest grid cell to the 50 stations shown in a). The mean is shown with a white dot, and white lines indicate the 25th, median, and 75th percentiles. The shape of each violin reflects the probability density function of the data.

Figure 10.

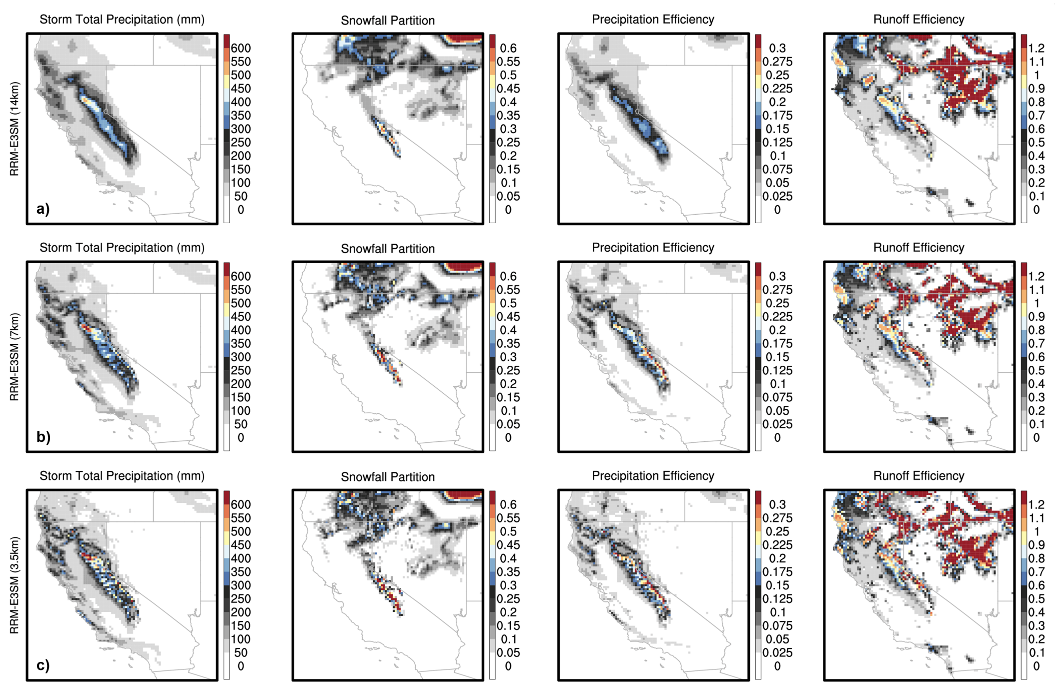

Figure 10. Forecast ensemble average precipitation characteristics, including storm total precipitation, snowfall partition, precipitation efficiency, and runoff efficiency for a) RRM-E3SM (14km) b) RRM-E3SM (7km) and c) RRM-E3SM (3.5km) over the overlapping forecast period of 31 December 1996 to 4 January 1997.

Figure 11.

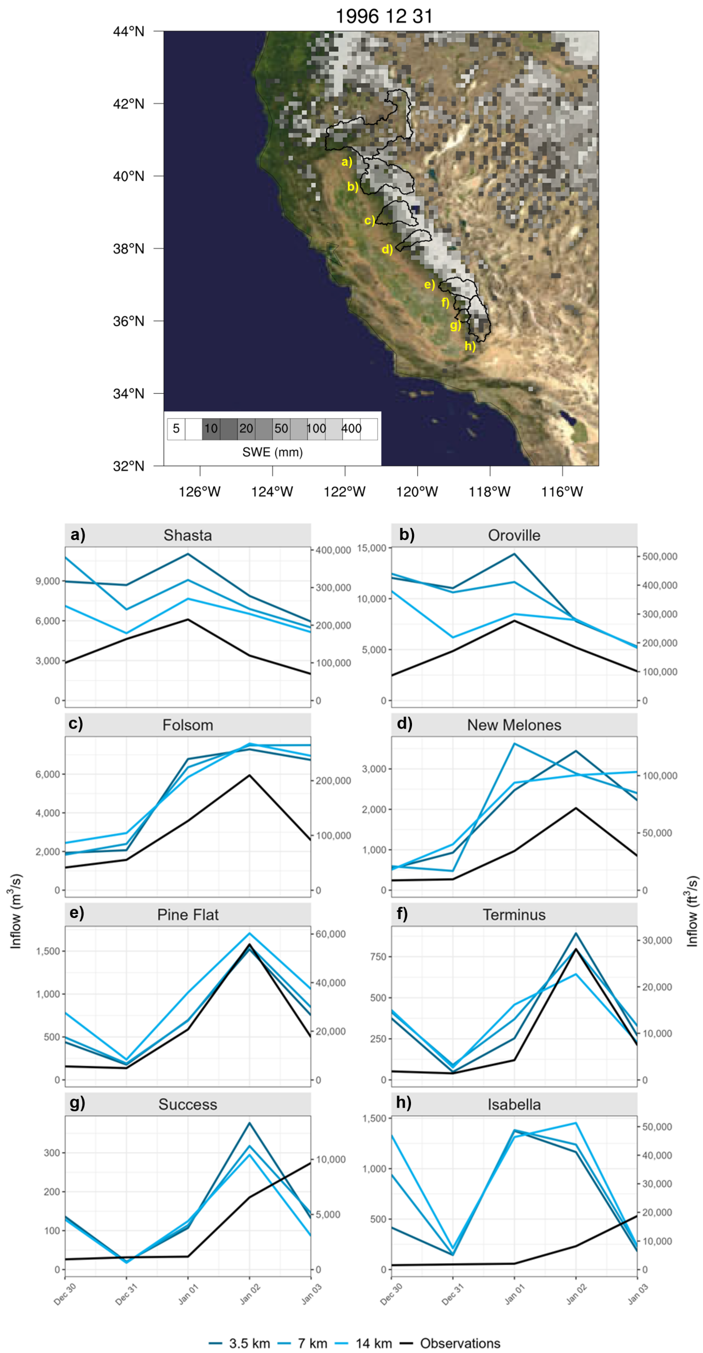

Figure 11. Forecast ensemble average reservoir inflow rates from each of the RRM-E3SM simulations across eight major reservoirs in California. The top figure shows the location of the eight reservoirs and the areal extent of the watersheds that feed into them (black outlines) overlaid onto Margulis product estimates of snow water equivalent, SWE, at the start of the 1997 flood. The black lines in the sub-panel plots represent measured inflows into each reservoir.

Figure 12.

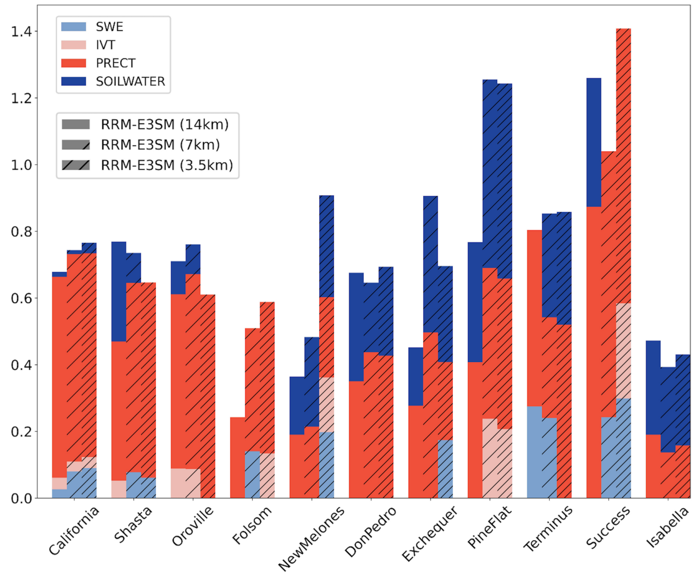

Figure 12. Causal inference estimates for the magnitude of the impact of hydrometeorological variables on total runoff (overland flow, interflow, and baseflow). The four variables include integrated vapor transport (IVT), total precipitation (PRECT), snow water equivalent (SWE), and 10 cm soil moisture content (SOILWATER). The magnitude of the influence of each variable on total runoff (overland flow, interflow and baseflow) is represented by an individual component of a stacked bar chart. Each component has a range between 0 and 1. RRM-E3SM cases (designated by hatching) are stacked next to each other for each region assessed including California (Hydrologic Unit Code 18) and the headwater regions of the 10 major reservoirs in California (ordered by latitude from northernmost to southernmost).

Figure 13.

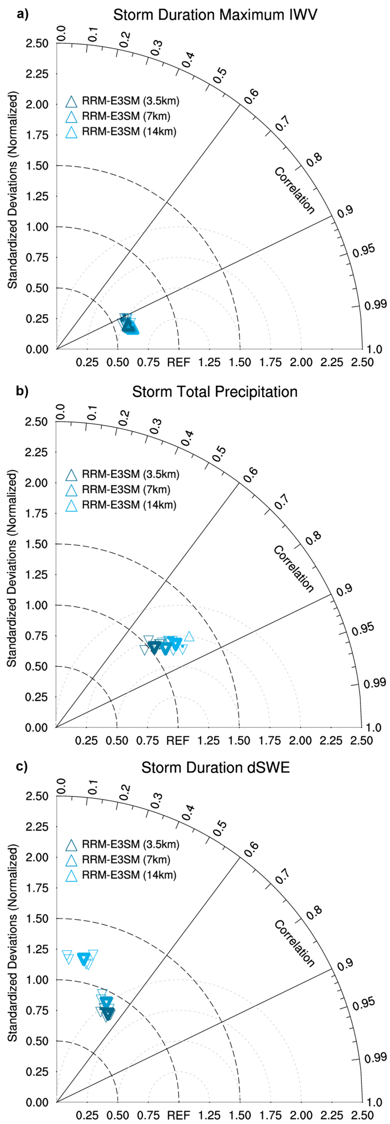

Figure 13. Taylor diagrams representing all grid cells within the hydrologic unit code (HUC-2) California Region, region 18 in Seaber et al. (1987), for the forecast period of 31 December 1996 up to 4 January 1997. a) Storm duration maximum integrated water vapor (IWV) compared to ERA5; b) storm total precipitation compared to the Livneh product; and c) storm duration change in snow water equivalent, dSWE, compared to the Margulis product. Each triangle represents one of the six RRM-E3SM forecasts initialized from 28 December 1996 at 00Z to 30 December 1996 at 12Z. Bold triangles represent the forecast ensemble average. Upward (downward) triangle orientation represents a positive (negative) bias compared to each reference dataset. Black radial lines provide general guidance for groupings of Pearson pattern correlation. The black and gray dashed azimuthal lines centered around REF indicate the root mean squared error and standard deviations from the reference dataset.

Figure 14.

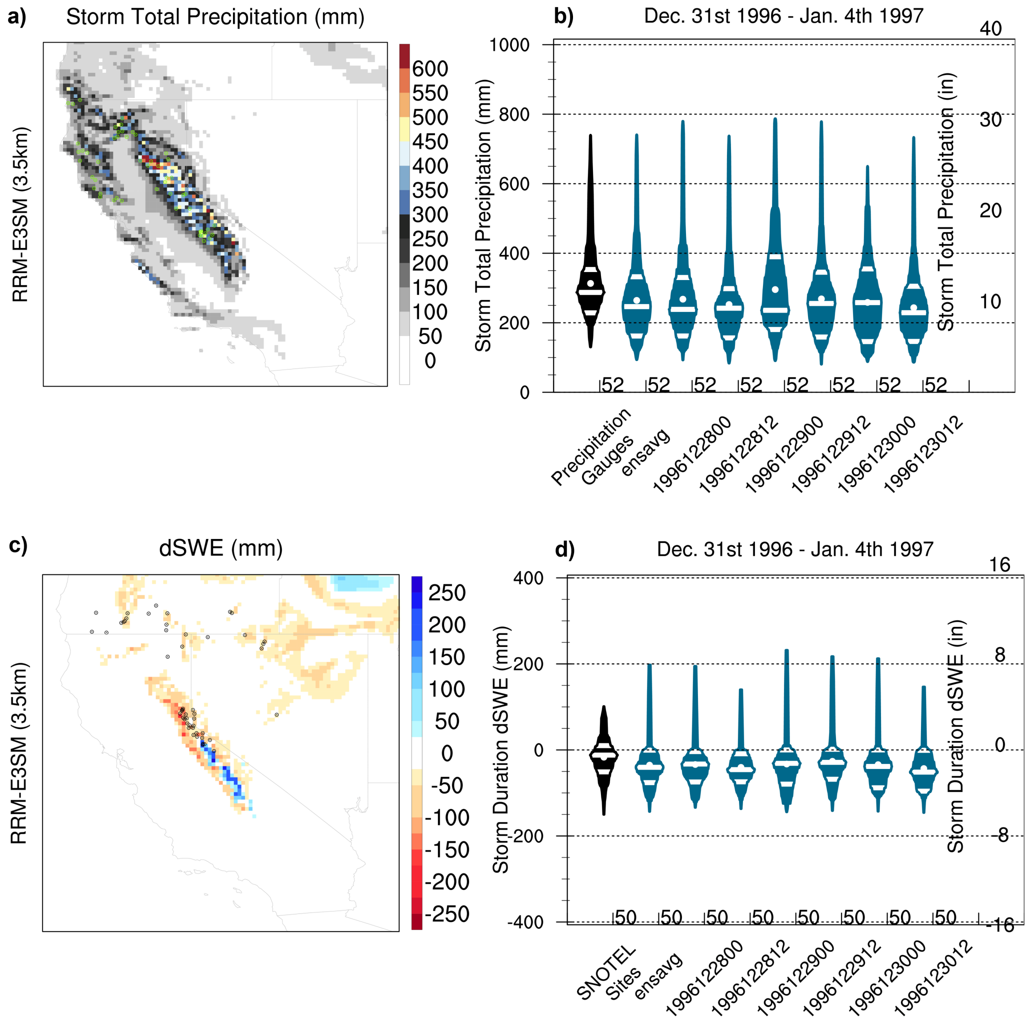

Figure 14. Same as Figures 5 and 9, but the violin plots now compare the initialization dates for each of the six RRM-E3SM (3.5km) forecasts. Panels a) and b) show storm total precipitation and panels c) and d) storm duration change in snow water equivalent (dSWE). The six-forecast ensemble average (ensavg) is also shown in black.

Figure 15.

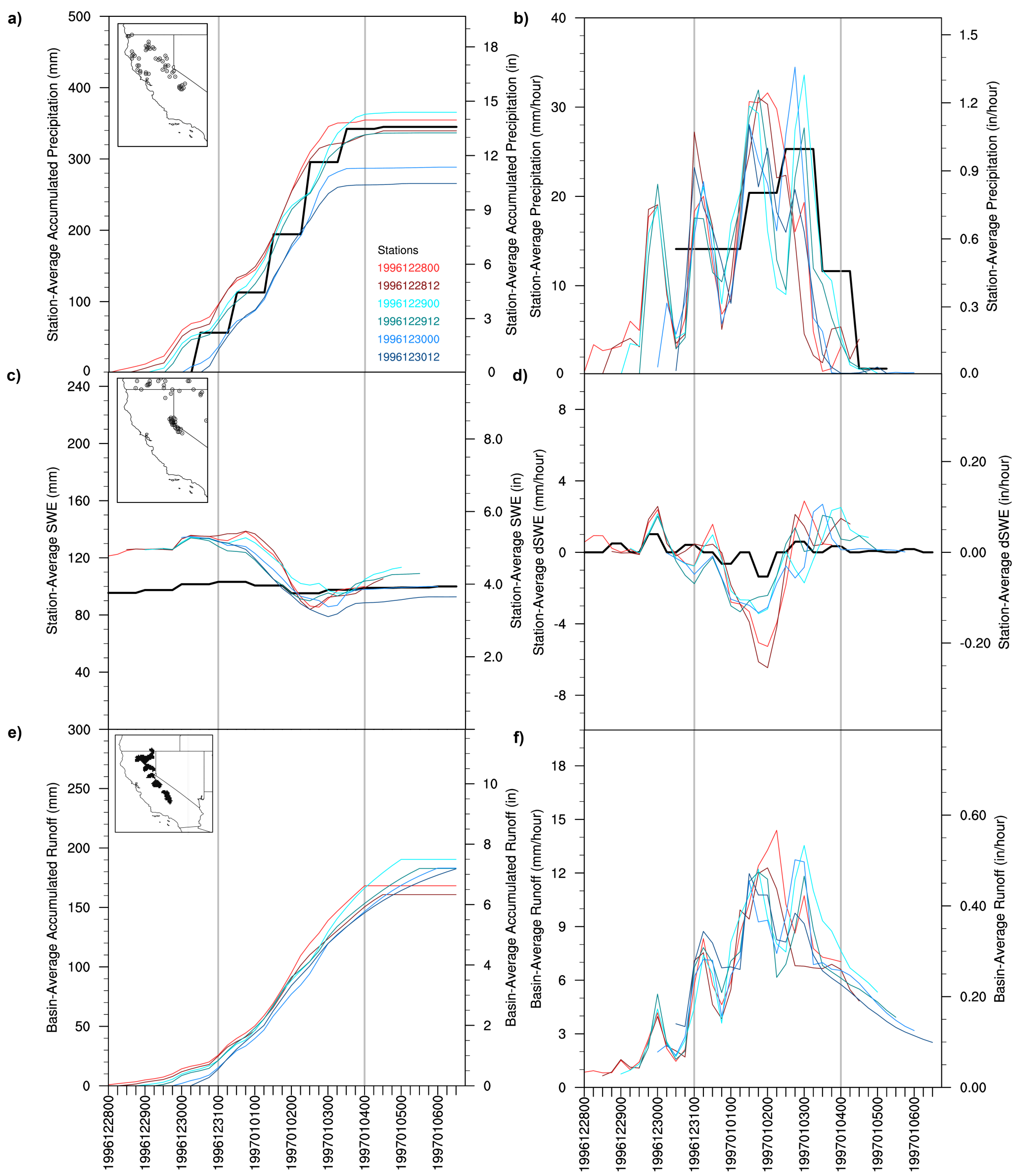

Figure 15. Time series for precipitation, snow water equivalent, and runoff simulated by RRM-E3SM (3.5km) across forecast lead time evaluated at station locations and in regions identified in the upper left maps. The left-column sub-panel plots represent cumulative totals and the right-column sub-panel plots represent hourly rates. Black lines represent station observations. Vertical gray lines indicate the period during which the 1997 flood occurred.

Figure 16.

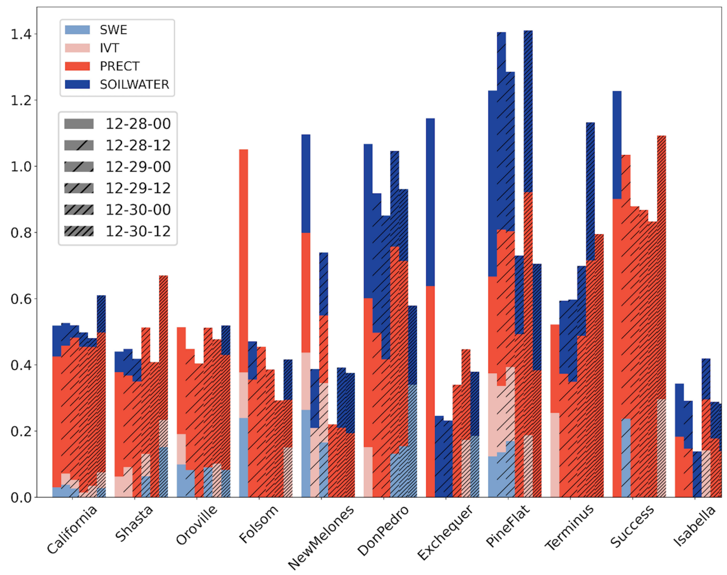

Figure 16. Same as Figure 12, however, each stacked bar chart represents one of the six forecasts produced by RRM-E3SM (3.5km) and conveys the strength of causal influence of four hydrometeorological variables, integrated vapor transport (IVT), total precipitation (PRECT), snow water equivalent (SWE), and 10 cm soil moisture (SOILWATER), on total runoff (overland flow, interflow, and baseflow). The forecast initialization date is indicated by different styles of hatching.

Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model

Alan M. Rhoades[1], Colin M. Zarzycki[2], Héctor A. Inda-Diaz[1], Mohammed Ombadi[1,3], Ulysse Pasquier[1], Abhishekh Srivastava[4], Benjamin J. Hatchett[5], Eli Dennis[6], Anne Heggli[5], Rachel McCrary[7], Seth McGinnis[7], Stefan Rahimi-Esfarjani[6], Emily Slinskey[6], Paul A. Ullrich[1,4,8], Michael Wehner[9], and Andrew D. Jones[1,10]

[1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[2]Department of Meteorology and Atmospheric Science, Penn State University, State College, PA, USA
[3]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA
[4]Department of Land, Air, and Water Resources, University of California, Davis, CA, USA
[5]Desert Research Institute, Reno, NV, USA
[6]Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA
[7]National Center for Atmospheric Research, Boulder, CO, USA
[8]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA
[9]Computational Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[10]Energy and Resources Group, University of California, Berkeley, Berkeley, CA, USA

Key Points:

- Energy Exascale Earth System Model forecasts at 3.5km grid spacing skillfully recreate the hydrometeorology of California's 1997 flood
- Horizontal resolution alters the representation of key flood drivers such as the Sierra barrier jet, precipitation extremes, and snowmelt
- Forecast lead time 2-to-4 days prior to the onset of the 1997 flood minimally influences forecast precipitation and snowmelt skill

Corresponding author: Alan M. Rhoades, `arhoades@lbl.gov`

Abstract

The 1997 New Year's flood event was the most costly in California's history. This compound extreme event was driven by a category 5 atmospheric river that led to widespread snowmelt. Extreme precipitation, snowmelt, and saturated soils produced heavy runoff causing widespread inundation in the Sacramento Valley. This study recreates the 1997 flood using the Regionally Refined Mesh capabilities of the Energy Exascale Earth System Model (RRM-E3SM) under prescribed ocean conditions. Understanding the processes causing extreme events inform practical efforts to anticipate and prepare for such events in the future, and also provides a rich context to evaluate model skill in representing extremes. Three California-focused RRM grids, with horizontal resolution refinement of 14km down to 3.5km, and six forecast lead times, 28 December 1996 at 00Z through 30 December 1996 at 12Z, are assessed for their ability to recreate the 1997 flood. Planetary to synoptic scale atmospheric circulations and integrated vapor transport are weakly influenced by horizontal resolution refinement over California. Topography and mesoscale circulations, such as the Sierra barrier jet, are prominently influenced by horizontal resolution. The finest resolution RRM-E3SM simulation best represents storm total precipitation and storm duration snowpack changes. Traditional time-series and causal analysis frameworks are used to examine runoff sensitivities state-wide and above major reservoirs. These frameworks show that horizontal resolution plays a more prominent role in shaping reservoir inflows, namely the magnitude and time-series shape, than forecast lead time, 2-to-4 days prior to the 1997 flood onset.

Plain Language Summary

The 1997 California New Year's flood event caused over a billion dollars in damages. This storm became a central part in guiding efforts to reduce flood risks. Earth system models are increasingly asked to recreate extreme weather events. However, the ability of Earth system models to recreate such events requires rigorous testing. Testing ensures that models provide value in anticipating and planning for future flood events. This is particularly important given the changing climate. We evaluated the Department of Energy's flagship Earth system model, the Energy Exascale Earth System Model, in its ability to recreate the weather and flood characteristics of the 1997 flood. The model resolution, important for resolving mountain terrain and storm interactions, and forecast lead time, important for storm progression accuracy, are assessed. The multi-forecast

average from the highest-resolution model best recreates the observed precipitation, snowpack changes, and flood characteristics. Our findings provide confidence that the highest resolution model could be used to study how a 1997-like flood event would be altered in a warmer world.

Introduction

California is especially susceptible to major cool season flood events (Kattelmann, 1997). Atmospheric rivers (ARs) are largely responsible, accounting for 84% of flood damages in the western United States (Corringham et al., 2019). The most notable California flood event, measured by its intensity, duration, and inundation area, occurred in 1861/1862 (Porter et al., 2011; Huang & Swain, 2022). It was thought to be AR-driven and inundated portions of both the Sacramento and San Joaquin valleys and portions of the present-day metropolitan area of Los Angeles. Because of its impact, this event has emerged as an important "design storm" for California water managers and led to the development of the colloquially termed "ARkStorm", which combines aspects of AR-induced flood events that occurred in 1969 and 1986. The 1861/1862 flood event happened during a time in California's history when the population density and built infrastructure was at a much smaller scale than today. Since the 1860s, urbanization has resulted in the loss of floodplains in many communities that are vulnerable to flooding despite significant investments in constructing flood control infrastructure (Whipple et al., 2017; Whipple & Viers, 2019). In many low-lying regions throughout the Central Valley, aging levee systems and subsidence continue to expose populations and industries to flood impacts (Hanak & Lund, 2012). Sequences of heavy precipitation-producing storms, many of which were ARs, during the winters of 2017 and 2023 highlight the present susceptibility of California to major riverine flooding. Climate change may further exacerbate impacts felt by these storms (Gershunov et al., 2019; Rhoades et al., 2021; Corringham et al., 2022; Huang & Swain, 2022), particularly in the most underserved communities (Wing et al., 2022), highlighting the need for detailed analyses aimed at understanding how these storms drive compound extremes under historical and future climate conditions.

The most costly flood event ($1.6 billion) in California history was the New Year's flood event of 1997, hereafter "1997 flood" (Lott et al., 1997). Major flood losses occurred throughout the western United States, including losses of $500 million in Nevada and $125 million in Washington. The combination of flood area and severity across the western United States ranks the 1997 flood as the #2 superflood between 1950 and 2010 (Tarouilly et al., 2021). At least half a million people were displaced by the flooding and the majority of California counties (43/58) were declared disaster zones (Lott et al., 1997).

The 1997 flood was primarily made up of three storms that occurred between 25 December 1996 and 2 January 1997 with inundation afterward (Galewsky & Sobel, 2005). Antecedent conditions played an important role in driving up the economic cost of this event; earlier storms throughout late November and December of 1996 built an abundant snowpack and elevated soil moisture content throughout the Central Valley and the Sierra Nevada (Figure 1). Between 30 December 1996 and 3 January 1997 storms produced more than 750 mm of precipitation in certain regions of northern California (e.g., 840 mm, or 33 in, at Bucks Lake in Plumas County, California; (Figure 1; `https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). Heavy rainfall with snow above 3,000 m elevation commenced on 30 December 1996; the Central Sierra Snow Lab (CSSL; located at 2,100 m) reported 137 mm of rainfall on 30-31 December 1996 (Osterhuber & Schwartz, 2021). On New Year's Day of 1997, an extreme AR event made landfall (Figure 1). Maximum temperatures at 2,100 m elevation hit 7°C and reached 3°C at 2,900 m on 1 January 1997 when 120 mm of rain fell at the CSSL (Osterhuber & Schwartz, 2021; Heggli et al., 2022). Prior to the onset of rainfall on 29 December 1996, snow densities were ready to produce terrestrial water input (32%), rising to 35% on 30 December 1996 (Heggli et al., 2022). The CSSL lost 100 mm of snow water equivalent (SWE) between 30 December 1996 and 1 January 1997 ultimately contributing to the development of a warm-snow drought water year (Hatchett & McEvoy, 2018). When combined with saturated soils and sufficiently ripe snowpack to melt and convey water to the land surface, the extreme multi-day precipitation caused major rivers to reach flood stage, with several setting all-time peak flows (Figure 1; `https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). As a result of the December-January storms, this two-month period set the record for the wettest since records began in 1920, measured via California's 8-station index, with a total of ~1,200 mm of precipitation. However, despite the wet start, the remainder of the water year was drier than normal leading to below-normal snowpack and reservoir levels at the end of the required flood pool period in April. The 1997 flood event thus represents an object lesson both for the study of extreme precipitation and runoff but also for reservoir and flood management in a highly variable climate.

A growing area of climate research is focused on understanding cascading, compound, and/or sequential hydrometeorological extreme events (Fish et al., 2019; AghaKouchak et al., 2020; Raymond et al., 2020). Simultaneously, the climate research community has sought to provide more credible and salient decision-relevant information to practitioners and management communities through iterative, co-produced research (Lemos et al., 2018; Jagannathan et al., 2021; Siirila-Woodburn et al., 2021). Examining historically significant, decision-relevant extreme events, through high-resolution climate model "storyline" recreations can be both be useful for water resource managers (Shepherd, 2019; Gutowski et al., 2020; Bukovsky et al., 2023) and have also been frequently used in event attribution studies (Wehner et al., 2019). Storylines are physically based model recreations of impactful weather events, often chosen through iterative discussions between scientists and stakeholders, that are then simulated under plausible past and future climate scenarios. However, it is important to note that while such studies can provide information on the local dynamic and thermodynamic effects of climate change on extreme events, they do not provide information about the influence of large-scale circulation changes on the return probability of such events.

Storyline event recreations also have practical model development implications. Climate models are mostly optimized around mean state performance for different hydrometeorological performance metrics (Fasullo, 2020), rather than extremes. This is especially true from the perspective of land-atmosphere interactions that drive compound extremes (La Follette et al., 2021). Storyline approaches can also help to convey information on model uncertainty, namely the role of structural and scenario uncertainty (Lehner et al., 2020), in a more understandable and decision-relevant way. Therefore, the recreation of the 1997 flood is a useful exercise in understanding the nature of extreme events and determining whether our cutting-edge modeling approaches are fit for purpose in simulating them. An additional benefit of storyline approaches is that the climate models used and the resultant climate research conducted becomes tailored toward greater practitioner relevance over time (Lemos et al., 2012).

In this study, we recreate the 1997 flood using the U.S. Department of Energy's flagship climate model, the Energy Exascale Earth System Model, and its regionally refined mesh capabilities (RRM-E3SM). We chose the 1997 flood because it is the flood of record most recently experienced by current water managers, was relatively well-monitored by a network of meteorological and hydrologic measurements, and occurred during a period in which atmospheric reanalysis products have higher skill (Uppala et al., 2005; Hersbach et al., 2020). This event also allows us to assess the relative contributions of E3SM horizontal resolution and forecast initialization time in shaping the fidelity of the flood event recreation. We pay particular attention to the interactions across the submodels of E3SM (e.g., atmospheric and land-surface) and their representation of key hydrometeorological variables before/during/after the event. This is the first time RRM-E3SM has been systematically used, across resolution and forecast lead time, to generate a storyline recreation of a western United States hydrometeorological extreme. Our scientific questions include:

(1) To what degree does horizontal model resolution influence land-atmosphere interactions and hydrometeorological impacts associated with the 1997 flood?
(2) What is the forecast lead time that best balances the short-term antecedent preconditioning of soils and snowpack and post-storm impacts when recreating the 1997 flood?
(3) Is RRM-E3SM fit-for-purpose in representing a compound extreme event such as the 1997 flood?

The manuscript is organized as follows. We first highlight details about our RRM-E3SM experimental setup. We then discuss the various *in-situ*, reanalysis, regional climate model, and gridded climate products used to assess and juxtapose RRM-E3SM skill in recreating the 1997 flood. We then discuss our results and how they fit within the broader literature. Finally, we summarize our major findings and provide suggestions for future research.

Methods

Energy Exascale Earth System Model (E3SM) version 2

The Energy Exascale Earth System Model version 2 (E3SMv2; Golaz et al., 2022) used for this analysis allows for regionally refined mesh (RRM-E3SM) simulations over a targeted region of interest. Recent studies find that RRM-E3SM performs comparably to uniform 0.25° (∼25km) horizontal resolution simulations for water cycle-related processes and provides several improvements to uniform 1.00° (∼111km) horizontal resolution simulations (Tang et al., 2019, 2022). These improvements are particularly important in regions of complex terrain such as the California Sierra Nevada. A detailed description of E3SMv2's atmospheric dynamical core, physics and dynamics, horizontal grids, vertical discretization, radiation, tracer transport schemes, and subgrid-scale parameterization choices (e.g., cloud microphysics scheme) can be found in Golaz et al. (2022). More specific findings related to RRM-E3SM are described in Tang et al. (2022), while Harrop et al. (2022) provides additional details on water cycle process fidelity in both the atmosphere and land-surface in E3SM at uniform horizontal resolutions of 1.00° versus 0.25° over the United States.

The RRM-E3SM meshes were produced using TempestRemap (Ullrich & Taylor, 2015; Ullrich et al., 2016); the topography was generated with the NCAR_Topo tool (Lauritzen et al., 2015) and smoothed for model stability purposes using the framework discussed in Zarzycki et al. (2015) and a coefficient of $3e^{-16}$ (c in Equation 1 of Zarzycki et al., 2015). The refinement regions and topographic representation in the simulations over California for the three RRM-E3SM cases are shown in Figure 2. Hereafter, RRM-E3SM simulations with a maximum refinement resolution over California at 14km, 7km, and 3.5km will be referred to as, RRM-E3SM (14km), RRM-E3SM (7km), and RRM-E3SM (3.5km), respectively. In all simulations, the E3SM default setting of 72 vertical levels is used. As found in other variable-resolution and regionally refined mesh Earth system model analyses over the last decade, horizontal resolution influences the simulation fidelity of synoptic-to-mesoscale trajectory of storm tracks and eddies (Rauscher et al., 2013; Rauscher & Ringler, 2014; Sakaguchi et al., 2016; Liu et al., 2023). Resolution also influences the representation of topography, which in turn affects how coastal landfalling storms are orographically uplifted, the rain-snow partitioning of the storm's precipitation, and the build-up and evolution of mountain snowpack throughout the cool-season (Rhoades et al., 2016; Huang et al., 2016; Wu et al., 2017; Rhoades, Ullrich, & Zarzycki, 2018; Rhoades, Ullrich, Zarzycki, Johansen, et al., 2018; Xu et al., 2018; Rhoades, Jones, O'Brien, et al., 2020; Rhoades, Jones, Srivastava, et al., 2020; Bambach et al., 2021; Xu et al., 2021; Maina et al., 2022). Similarly, land-surface cover and soil heterogeneity increase at finer resolutions, which can alter the surface-through-subsurface water and energy balance interactions of the hydrologic cycle (e.g., soil moisture).

Betacast

The 1997 flood event forecast ensemble was produced for six different 8-day periods starting on 28 December 1996 at 00Z through 30 December 1996 at 12Z, initialized at 12-hour increments between those dates, using the "Betacast" framework described in Zarzycki et al. (2014) and the Atmosphere Model Intercomparison Project (AMIP) protocols (Gates et al., 1999). The land surface conditions are spun-up for five years prior to the first forecast, with a standalone simulation of the E3SM Land Surface Model (ELM) forced by the 6-hourly atmospheric data from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5; Copernicus Climate Change Service Climate Data Store (CDS), 2017). This ensures that antecedent land surface conditions (namely soil moisture content and mountain snowpack) are consistent with the actual 1997 flood event conditions on the day each RRM-E3SM forecast is started. Subsequent forecast cycles use the 12-hour land forecast from the previous cycle for initialization. This approach gives nearly identical results to spinning up each forecast cycle's land surface independently (not shown).

The atmospheric initial state is generated using high-order remap algorithms to take data from the ERA5 reanalyses and map them onto the corresponding RRM-E3SM grid. The pressure field is adjusted based on the technique in Trenberth et al. (1993) to account for differences in ERA5 and RRM-E3SM orography that may result in geostrophic imbalances. Observed ocean surface conditions (i.e., sea surface temperatures and sea ice extent) are also prescribed by interpolating NOAA Optimum Interpolation (OI) data (Reynolds et al., 2007) to the model grid. After initialization from ERA5, the RRM-E3SM forecasts are "free-running": the atmosphere and land surface models are fully coupled and allowed to freely solve the governing equations that drive these systems.

All RRM-E3SM simulations utilize the hydrostatic dynamical core in E3SM. Notably, the effective resolution is 4-5x the actual grid spacing (Ullrich, 2014; Klaver et al., 2020). Further, it has been shown that non-hydrostatic dynamical cores minimally influence midlatitude wintertime precipitation (slight drying) from resolutions of 36-to-4km, even in idealized mountain environments (Yang et al., 2017; Liu et al., 2022). With each 2x refinement in horizontal resolution, the RRM-E3SM dynamics and physics timestep and second-order viscosity diffusion strength at the model top were halved. For RRM-E3SM (14km), the atmospheric dynamics and physics timesteps and diffusion strength were 40 and 600 seconds and $4e^{-4}$, for RRM-E3SM (7km) they were 20 and 300 seconds and $2e^{-4}$, and for RRM-E3SM (3.5km) they were 10 and 150 seconds and $1e^{-4}$, respectively. The only additional differences across cases were the macrophysics-microphysics subgrid-scale parameterization substeps, set to 6 in RRM-E3SM (14km) and RRM-E3SM (7km) and 3 in RRM-E3SM (3.5km).

Atmospheric River Detection and Categorization

We used TempestExtremes (TE; namely the SpineARs and StitchBlobs algorithms) to detect the primary AR that made landfall during the 1997 flood on 1 January 1997 (Ullrich & Zarzycki, 2017; Zarzycki & Ullrich, 2017). TE is a "relative threshold" based AR detector (ARDT), meaning that it is minimally sensitive to fixed thresholding issues (i.e., an AR event only exists beyond ~250 kg/m/s), which may have important implications for assessing future AR characteristic changes (O'Brien et al., 2022). Our parameter settings for TE and the extensions made to TE to estimate AR landfalling characteristics, such as the AR category scale (Ralph et al., 2019), are important for estimating water resource impacts (e.g., AR-induced flood damages in Corringham et al., 2022) as discussed in more detail in Rhoades, Jones, O'Brien, et al. (2020), Rhoades, Jones, Srivastava, et al. (2020) and Rhoades et al. (2021). Although it is advantageous to use several ARDTs for climatology-based analyses of ARs (O'Brien et al., 2022), particularly when assessing climate change-related impacts, we use only TE because the primary AR during the 1997 flood was a category 5 event and recent findings in Zhou et al. (2021) have shown that ARDTs largely agree when identifying characteristics of category 4-5 AR events.

Validation

To evaluate the hydrometeorological forecast skill of RRM-E3SM in recreating the 1997 flood, we use a mixture of *in-situ* observations, reanalysis, gridded climate products, and more conventional regional climate modeling strategies. We obtained *in-situ* observations from 50 sites in the SNOw TELemetry (SNOTEL) network (`https://www.nrcs.usda.gov/wps/portal/wcc/home/snowClimateMonitoring/snowpack/snowpackMaps`) and 52 precipitation gauge sites from the California Data Exchange Center (CDEC) that are used in the National Oceanic and Atmospheric Administration (NOAA) storm summary (`https://www.cnrfc.noaa.gov/storm_summaries/ol.php?storm=jan1997`). We obtained daily reservoir inflow observations from the US Army Corps of Engineers Water Control Data System (`https://www.spk-wc.usace.army.mil/plots/california`

`.html`), retrieving inflow information for the 1997 Water Year from the Shasta, Oroville, Folsom, New Melones, Pine Flat, Terminus, Success, and Isabella Reservoirs.

We used reanalysis and gridded climate products to evaluate storm-total precipitation and pre-and post-event changes in snow water equivalent (SWE). Storm-total precipitation is evaluated against Pierce et al. (2021) which is an updated version of the Livneh product (Livneh et al., 2015), hereafter Livneh, and against the ERA5 reanalysis product, due to its use in providing initial conditions for the RRM-E3SM simulations. According to Pierce et al. (2021), the updated Livneh product better preserves extreme event precipitation totals by more systematically accounting for daily time adjustments in precipitation gauge data (i.e., rounding-related issues related to the time of day the station observation is taken). We also conducted a preliminary analysis comparing Livneh with other widely used gridded climate products, Newman et al. (2015) (Newman) and Daly et al. (2008) (Parameter-elevation Regressions on Independent Slopes Model, PRISM) as shown in Figure S1. Compared with the 52 precipitation gauge measurements, we found that Livneh was either a better estimate (compared with Newman) or was indistinguishable (compared with PRISM) in its representation of the 4-day precipitation totals produced during the 1997 flood. In order to estimate the return periods of the 4-day precipitation totals during the 1997 flood, we applied a non-stationary generalized extreme value (NS-GEV) analysis to the annual maximum of 4-day precipitation totals (Rx4day) in the Livneh product interpolated to the 52 gauge locations using the first-order conservative remapping (P. W. Jones, 1999). In the NS-GEV framework, we first apply the Mann–Kendall (MK) trend test (Mann, 1945) to the Rx4day data at each gauge location to determine if the data has a significant trend at the 5% level. If the Rx4day data at a location has a significant trend, we fit time as a covariate in the location or/and scale parameters of the GEV distribution fitted to the Rx4day data at that gauge location. The complete procedure is outlined in Srivastava et al. (2021).

We assess pre- and post-event changes in SWE against the Fang et al. (2022) western United States-wide snow reanalysis product (hereafter Margulis due to it being an updated version of Margulis et al., 2016). The Margulis reanalysis product has shown skill in estimating peak SWE in the California Sierra Nevada when compared with airborne LiDAR SWE measurements (e.g., 1 April mean SWE depth differences of -0.15 to +0.05 m across 2015-2021), which have essentially become the snow community standard for spatially complete estimates of snow depth and SWE in recent years (Painter et al., 2016; Stillinger et al., 2023). We also compare and contrast RRM-E3SM skill with a set of simulations produced with a more traditional and widely-used dynamical downscaling approach. These simulations were produced using the Weather Research and Forecasting (WRF) model run at 14km resolution over California that is bounded laterally and at the model top with ERA5 (A. D. Jones et al., 2022). All gridded data that is intercompared has been regridded from its native grid resolution to a regular latitude-longitude grid resolution of 14 km using bilinear interpolation provided by the Earth System Modeling Framework (ESMF) Offline Regridding Weight Generator (The NCAR Command Language (Version 6.6.2), 2022).

Causal Inference

The complexity of Earth system interactions within the RRM-E3SM simulations and the large number of grid cells within the spatial domain of analysis makes it difficult to unambiguously disentangle the impact of resolution and forecast lead time on processes and interactions between hydrometeorological variables. Thus, in the present study, we use causal inference to gain insights into the interactions between atmospheric and land-surface variables on one hand, and total runoff on the other. To the best of our knowledge, this is the first application of this framework for this style of problem. Causal inference allows us to move beyond canonical correlation analysis while reducing the dimensionality of analysis to investigate interactions in the model. The goal of causal inference methods is to determine causal relationships between hydrometeorological variables by using concepts of statistical conditional independence on time series data. These methods are gaining popularity in the Earth and environmental sciences community (Sugihara et al., 2012; Runge et al., 2019; Ombadi et al., 2020; Runge, 2023) and offer a unique perspective to evaluate relationships.

We use the Peter-Clark (PC) algorithm (Spirtes & Glymour, 1991), a causal inference method that utilizes graph theory and graphical rules to recover causal relations from time series data. The PC algorithm starts with a fully connected graph where all variables are causally related to each other, then iteratively and systematically removes causal relations using conditional independence tests. One of the main advantages of the PC algorithm is its ability to reduce the number of variables in the conditioning set, thereby mitigating the “curse of dimensionality”. We chose to use the PC algorithm because it provides good performance in hydrometeorological systems, especially in controlling the number of falsely detected causal links (Ombadi et al., 2020). For our conditional independence tests, we used information-theoretic conditional independence instead of partial correlation due to its ability to detect nonlinear relationships (Ombadi et al., 2021). Our causal analysis considers contemporaneous causality between the time series of the five key hydrometeorological variables evaluated in this study (i.e., integrated vapor transport [IVT], precipitation, SWE, 10 cm soil moisture content, and total runoff volume) for all grid cells within a specific spatial domain (e.g., California-wide or the mountainous headwaters of a surface reservoir). Causality was assessed at a statistical significance level of 0.05.

Results and Discussion

Murphy (1993) provides terminology to discuss forecast verification qualities that both forecasters and users of forecasts find important. In this study, we will evaluate RRM-E3SM's representation of the California New Year's flood event of 1997 according to forecast quality (forecast correspondence to observations) and forecast value (forecast utility to decision makers). We use the effects of horizontal resolution and forecast lead time to assess forecast quality and value via measures of bias (the difference between forecast and observation), association (linear correlation between forecast and observation), sharpness (forecast capability in representing extremes), and through measures of value (e.g., reservoir inflow volumes).

Resolution influence on atmospheric process representation of the 1997 flood

We first compare the influence of regional grid refinement over California by evaluating how the representation of the large-scale atmospheric circulations that shaped the landfalling AR on New Year's Day of 1997 differ according to the resolution of the regional refinement domain. Figure 3 compares the large-scale IVT fields and circulation patterns of ERA5 and the three grid refinement resolutions at the start of the major AR landfall on 1 January 1997. The RRM-E3SM values are six-member forecast averages. The RRM-E3SM simulations forecast the low-pressure center near the Pacific Northwest coastline further southwest than it is in ERA5 on this date (Supplemental Figure S2). The simulations generally agree across resolutions on the spatial distribution of AR categories from the California Bay Area up through the Sacramento Valley (Figure 4 and

Figure S3). Agreement is also found with ERA5 in the northern portions of California, particularly with regard to category 5 AR conditions (Figure S4); however, all RRM-E3SM simulations systematically produce AR categories that are too high in southern California. This appears to be due to a disagreement in the AR width and/or the centroid of the AR landfall location with ERA5, which occurs further South (as indicated by positive IVT anomaly from central to southern California in Figure 3) and due to uniformly higher wind speeds (Figure S4). Notably, ERA5 may under-represent AR activity in southern California compared to other reanalyses (Collow et al., 2022).

Although IVT is important from a forecasting perspective, particularly since it allows for longer forecast lead times than precipitation (Lavers et al., 2016), IVT is simply one metric indicating the potential for precipitation to occur, and its orientation with respect to terrain can suppress or enhance precipitation (Ricciotti & Cordeira, 2022). Therefore, we also evaluate how the precipitation potential across RRM-E3SM simulations is realized in the 1997 flood, particularly its association and sharpness. The forecast ensemble average storm total precipitation amounts are shown in Figure 5. This figure compares simulated precipitation values with reanalysis and gridded climate products as well as a conventionally used regional climate model (WRF, forced by ERA5) at the grid cells nearest to the 52 precipitation gauges used in NOAA's storm summary of the 1997 flood. Refinement from 14km to 3.5km in RRM-E3SM has an appreciable effect on the statistical distribution of storm total precipitation, including the mean, median, and maximum. RRM-E3SM (3.5km) matches the distribution of storm total precipitation at the 52 precipitation gauge sites better than other datasets, including the Livneh product. RRM-E3SM (3.5km) agreement (r=0.73) in storm total precipitation holds across individual precipitation gauge sites as well (Figure S5), particularly precipitation gauges in the northern Sierra Nevada, which have the highest precipitation totals (e.g., Buck's Lake and La Porte). Note that the WRF simulations were conducted at 14km resolution and do not represent an even comparison with RRM-E3SM (7km) or RRM-E3SM (3.5km). The superior skill of models, relative to statistical interpolation and extrapolation techniques utilized in gridded climate products, in representing mountain precipitation processes have been noted before (J. Lundquist et al., 2019).

In contrast to landfalling AR characteristics, we found storm total precipitation to be resolution-dependent. We hypothesize that this is likely a result of more realistic topographic representations of California's Coast Ranges and Sierra Nevada. In addition, we hypothesize that important mesoscale circulation features known to influence the spatiotemporal characteristics of precipitation in northern California are better resolved. One such feature is the Sierra Barrier Jet (SBJ), a classic terrain-parallel low-level jet. The SBJ results from the blocking, slowing, and subsequent counter-clockwise turning of low-level winds as they interact with the Sierra Nevada in a stable or moist-neutral environment. The SBJ has a typical core of peak winds at ∼500m to 1km (∼950-900 hPa) above the Central Valley with wind speeds ≥15 m/s (Neiman et al., 2010, 2013). The location and strength of the SBJ play an important role in driving California's precipitation maxima during AR events (Neiman et al., 2013). This precipitation maximum usually occurs northwest and upstream of the Sierra Nevada crest, typically around the Buck's Lake precipitation gauge (39.85°N, 121.24°W) in the headwaters of the Oroville Dam. To examine RRM-E3SM skill in representing the SBJ, we compare winds using analogous cross-sections and transect lines outlined in Hughes et al. (2012) that dissect the typical locations of the SBJ in California.

Figure 6 shows cross-sections of zonal and meridional winds for ERA5 and the RRM-E3SM simulations at the start of the AR landfall on 1 January 1997. Similarly to previous findings, wind speeds are generally stronger in RRM-E3SM cases compared with ERA5. However, the altitude, latitudinal, and longitudinal locations of the wind speed maximum do generally agree with ERA5. RRM-E3SM simulates the SBJ and locates its core between 950-900 hPa at around 40°N, 122°W. Resolution plays an important role in better resolving the location of the wind speed maximum both with altitude and latitudinally. Similarly, RRM-E3SM (3.5km) shows higher wind speeds from 1000-900 hPa and more orographic uplift potential along the windward sides of both the Coast Ranges and the Sierra Nevada. This favors more orographic precipitation, as is shown in Figure 5.

To assess RRM-E3SM skill in representing the entire lifecycle of the SBJ, we now show vertical profiles of both meridional and zonal winds, from both a Sierra-parallel and Sierra-perpendicular perspective, compared with ERA5 (Figure 7). Prior to the onset of the flood event, on 31 December 1996, the RRM-E3SM simulations show the jet beginning to form at the right altitude relative to ERA5, but slightly stronger. On the first day of the flood event (1 January 1997), RRM-E3SM (3.5km) best represents the altitude location (∼950-1000 hPa) and strength (20-25 m/s) of the SBJ. The jet altitude and latitudinal location and strength match with the findings of Neiman et al. (2013)

for other couplets of AR-SBJ events identified using a combination of *in-situ* measurements including vertical wind profilers and reanalysis products. The RRM-E3SM results also corroborate the conclusion made by Hughes et al. (2012) that approximately a six-kilometer horizontal resolution is needed to properly represent the SBJ in model simulations. However, regardless of RRM-E3SM resolution, the SBJ becomes both weaker and/or lower in altitude relative to ERA5 on 3-4 January 1997.

Resolution influence on land-surface process representation of the 1997 flood

Although the 1997 flood was one of the most costly and damaging floods in northern California history, a non-stationary return period analysis of the Livneh product at the 52 gauge sites indicates that it was, at most, a 1-in-20-year event at a few gauge locations, based on 4-day precipitation total estimates over the 105-year record covering 1915-2019 (Figure 8). At 50% of gauge locations, the return period of the event was less than 6 years. This implies that the flooding was notable due to it being a compound extreme shaped by not only the precipitation provided by the sequence of storms, culminating in a category 5 AR landfall on 1 January 1997 but also antecedent land surface conditions that were primed for snowmelt and runoff generation. The importance of antecedent conditions and land surface feedbacks was shown by Ivancic and Shaw (2015) where only 36% of the 99th percentile discharge events occurred due to a 99th percentile precipitation event when evaluated CONUS-wide between 1950-2000.

To evaluate the role that antecedent and land surface conditions played in shaping the flood event, we now assess the change in snow water equivalent, or dSWE, for the category 5 AR storm duration (Figure 9). Analogously to the storm total precipitation analysis, we show storm duration dSWE across 50 SNOTEL sites throughout northern California, southern Oregon, and Nevada compared to the Margulis product. Model resolution also plays an important role in the distributions of both positive and negative dSWE in the California Sierra Nevada. This is likely due to the influence of topographic resolution on the simulated freezing level and the rain-snow partitioning of the AR event, which in turn influences the land surface representation of the accumulation and ablation of the mountain snowpack at mid-to-high elevations. The 50 SNOTEL sites indicate that more negative dSWE occurred over the duration of the 1997 flood (-152 mm / -6 in). However, at higher elevations, positive dSWE also occurred (+102 mm /

+4 in). In comparison, the Margulis product indicates that more positive dSWE occurred (up to +254 mm, or +10 inches, in certain locations). Although a general negative dSWE skew in the statistical distribution is shown for RRM-E3SM, with every 2x refinement in resolution over California the simulations more closely approximate the statistical distribution from the 50 SNOTEL location observations.

Figure 10 shows the effects of resolution on the spatial representation of precipitation and runoff characteristics. The differences across each RRM-E3SM case are explicitly shown in Figure S6. Storm total precipitation is enhanced at finer horizontal resolutions, particularly along the Coast Range and crest of the Sierra Nevada, upwards of 250 mm in RRM-E3SM (3.5km) relative to RRM-E3SM (14km). However, a general dry (wet) bias across RRM-E3SM simulations is seen in northwestern California's Klamath Mountains (Sierra Nevada) when compared with the Livneh product (Figure S7). Notably, the Livneh product had a general dry bias compared with precipitation gauge measurements (Figure 5 and S5). This indicates that Sierra Nevada crest precipitation overestimates in RRM-E3SM may not be as severe as is shown in Figure S7, corroborates the findings of J. Lundquist et al. (2019), and would support the claims made about the underrepresentation of gridded climate products' AR-related precipitation in J. D. Lundquist et al. (2015).

Model resolution also plays a key role in shaping both the rain-snow partitioning of precipitation and the efficiency at which water vapor becomes precipitation (Figure 10 and S6). Snowfall is enhanced by upwards of 20% in high-elevation regions of the California Sierra Nevada, particularly in the headwaters of the American River through the Kern River watersheds. Similarly, the precipitation efficiency (the amount of precipitation per unit of integrated water vapor) is enhanced by upwards of 20% throughout the Klamath Mountains, Coastal Ranges, and the Sierra Nevada in RRM-E3SM (3.5km). The combination of enhanced and more efficient precipitation and alterations to rain-snow partitioning changes the signature of runoff efficiency (the total runoff amount per total precipitation amount). Runoff efficiency is generally enhanced by upwards of 60% at low- to mid-elevations in northern California in RRM-E3SM (3.5km) compared to RRM-E3SM (14km), whereas in the high-elevation southern Sierra Nevada, a decrease is simulated. The enhanced runoff efficiency in RRM-E3SM (3.5km) is likely associated with more precipitation that is falling on wetter soils and, importantly, more snowmelt (as seen with more grid cells with runoff efficiencies at or exceeding 1). Conversely, runoff efficiencies decline in RRM-E3SM (3.5km) where snowfall is enhanced, which agrees with SNOTEL sites that indicate that positive dSWE changes occurred during the 1997 flood (Figure 9).

Even without a calibrated hydrologic model, comparing simulated inflow to observed inflow provides context for how well the model captures the key hydrologic-focused land-atmosphere interactions. This is because, in order to properly estimate reservoir inflows in the context of the 1997 flood, it is necessary that the model properly forecast the AR translational speed, plume intensity, and landfall location; the antecedent land surface conditions (e.g., snowpack and soil moisture); and the land-atmosphere interactions during and after the storm. Furthermore, model evaluation should also be done in decision-relevant regions (e.g., watersheds) instead of arbitrary latitude-longitude boxes. Therefore, to evaluate the value of the RRM-E3SM forecasts, we investigate reservoir inflows from the headwaters of eight major reservoirs, which represent a third (13.3 million-acre feet) of California's surface reservoir storage (Figure 11). Reservoir inflows are computed as basin averages of total runoff provided by the land-surface model in RRM-E3SM. In the headwaters of the two largest reservoirs (Lakes Shasta and Oroville), all simulations overestimate inflows, and resolution systematically increases the volume of water flowing through the system. This may be due to several factors, including a lack of parameter calibration in the land surface model (e.g., soil characteristics) and/or antecedent soil moisture being too high. Unfortunately, we could not find estimates of soil moisture content, from either *in-situ* or remote sensing sources, and were unable to evaluate soil moisture as we did precipitation and snowpack. We were also unable to find piezometer data recording groundwater height changes.

Although the magnitude of reservoir inflows is biased even in RRM-E3SM (3.5km), the shape of the reservoir inflow time series improves at finer resolutions in both Shasta and Oroville, with a more distinct peak inflow on 1 January 1997. This resolution dependence also holds for two other key northern California reservoirs (e.g., Folsom and New Melones). Unlike the results for Shasta and Oroville, the antecedent conditions (i.e., reservoir inflows at the beginning of 30 December 1996) in Folsom and New Melones Reservoirs seem to play a lesser role in model performance, with model drift in reservoir inflow estimates starting to occur one to two days after the forecasts have begun. Moving further south along the western slopes of the Sierra Nevada to Pine Flat and Terminus, RRM-E3SM (3.5km) matches reservoir inflows remarkably well, regardless of antecedent condition issues. Finally, RRM-E3SM simulations in the headwaters of Success and Isabella reservoirs match neither the amplitude nor shape of reservoir inflows, particularly Isabella. The lack of match between simulated and observed inflows is likely influenced by infrastructure and/or management decisions made above the reservoirs in these headwater regions, especially since RRM-E3SM simulations do not account for these factors.

To better contextualize RRM-E3SM runoff forecasts across resolution, we employ the PC causal inference algorithm with conditional mutual information test (Spirtes & Glymour, 1991; Ombadi et al., 2020). The influential strength of four hydrometeorological variables (i.e., IVT, precipitation, SWE, and 10 cm soil moisture content) on total runoff (overland flow, interflow, and baseflow) across California and within its 10 major reservoir headwater regions is shown in Figure 12 and Figure S8. The higher the stacked bar, the more variance is explained in total runoff. Each of the four hydrometeorological variables contributes a value ranging between zero and one, with a maximum possible total of four across variables. Across California, our causal analysis framework agrees with our prior suggestions that resolution plays an important role in amplifying the strength that both soil moisture content and SWE play in total runoff magnitude. With that said, atmospheric conditions (IVT and precipitation) heavily influence the total runoff signal across California comprising 84-94% of the total variance explained by the four chosen hydrometeorological variables (Figure S9). However, this causal relationship does change considerably from one reservoir headwater region to another (particularly in the central to southern Sierra Nevada).

Through this causal inference framework, we can also see that in certain reservoir headwater regions, resolution plays a systematic role in either adding more interactions between total runoff (more components contributing to each stacked bar) and all of the hydrometeorological variables (e.g., New Melones) or simplifying interactions to a single (e.g., Oroville) or fewer hydrometeorological variable(s) (e.g., Shasta). In other headwater regions, there is an insensitivity to resolution (e.g., Don Pedro and Isabella). In New Melones Lake, where runoff interaction diversity increases the most, IVT and SWE play no role in shaping runoff in RRM-E3SM (14km) and RRM-E3SM (7km), with a nearly a 50/50 split between precipitation and soil moisture, whereas RRM-E3SM (3.5km) shows a more equal interaction between all four hydrometeorological variables and runoff. Conversely, in Lakes Shasta and Oroville, three hydrometeorological variables play a key role in runoff forecasts in RRM-E3SM (14km) and RRM-E3SM (7km), yet precipitation becomes the dominant variable of influence in RRM-E3SM (3.5km), 91% and 100%, respectively (Figure S9). Finally, both Lake Don Pedro and Isabella Lake have an insensitivity to resolution where precipitation and soil moisture content play comparable roles in shaping total runoff across RRM-E3SM simulations.

Forecast lead time influence on atmospheric and land-surface process representation of the 1997 flood

To summarize the resolution dependence of RRM-E3SM simulations found thus far, we use Taylor diagrams (Figure 13) to show that although large-scale meteorology is relatively insensitive to finer horizontal resolutions (14km to 3.5km), even for land-falling AR characteristics (Figure 4), storm characteristics (e.g., storm total precipitation) and land-atmosphere interactions (e.g., storm duration dSWE) are sensitive to resolution. Dispersion in model results associated with forecast lead time is also shown. This will be the focus for the rest of our analysis, but to decrease the dimensionality of our analysis we focus on the best-performing simulation, RRM-E3SM (3.5km).

In RRM-E3SM (3.5km) both storm total precipitation and storm duration dSWE are weakly and not systematically sensitive to forecast lead time (Figure 14). The highest storm total precipitation and positive storm duration dSWE occurred in the forecast that was initialized on 1996-12-29 at 00Z. This finding is counter to our original hypothesis that forecast skill should increase as forecast lead time gets closer to 31 December 1996. This assumption was made because the 30 December 1996 at 12Z forecast has the least amount of time to drift from the conditions provided by ERA5 which could influence, for example, the AR intensity, landfall location, and translational speed.

Although forecast lead time does not appear to have a significant influence on storm total precipitation and storm duration dSWE over the period of 31 December 1996 to 4 January 1997, these metrics may mask temporal dependencies. To determine whether there are important diurnal and/or day-to-day differences across forecast lead times, Figure 15 shows both 6-hourly rates and cumulative 6-hourly totals for precipitation, dSWE, and runoff. The cumulative total precipitation estimated at the 52 precipitation gauge stations is well bracketed by the six RRM-E3SM (3.5km) forecasts. Hourly rates in precipitation show that precipitation diverges most across the six forecasts on 3 January

1997 (or four to six days post initialization of the forecast). From the perspective of dSWE, evaluated across the 50 SNOTEL sites, the six forecasts generally have similar tendencies throughout the flood period, but also disagree most on 3-4 January 1997. Negative dSWE values, an indication of the magnitude of snow ablation caused by the AR, were highest on 3 January 1997 in both observations and forecasts. The forecast spread on 3 January 1997 was -2 mm/hour to -7 mm/hour, which was generally stronger than was observed at SNOTEL sites. Undoubtedly, the spread in precipitation and SWE across forecasts from 3-4 January 1997 influenced runoff rates and totals in the reservoir headwater regions.

Finally, we evaluate how RRM-E3SM (3.5km) forecast lead time influences the causal strength and relationship between runoff and the four key hydrometeorological variables (i.e., IVT, precipitation, SWE, and 10 cm soil moisture content) over the period of 31 December 1996 to 4 January 1997. Interestingly, California-wide causal strength of the hydrometeorological variables on runoff generally is maintained across the six forecast lead times. Atmospheric conditions (IVT and precipitation) dominate the runoff signal (74-87% range across forecasts for the total variance explained for the four hydrometeorological variables chosen). The dominance of atmospheric conditions on runoff across forecasts holds in the headwaters of both Lakes Shasta and Oroville. However, akin to the resolution-focused results, antecedent conditions and land surface feedbacks play a larger role in shaping runoff in the reservoir headwater regions of the central to southern Sierra Nevada. For example, in the central and southern Sierra Nevada (New Melones Lake, Lake Don Pedro, and Isabella Lake) the role of antecedent and land surface conditions represents 46-51%, 40-51%, and 30-51%, respectively, on the causal relationship with runoff. Again, these percentages represent the range across forecasts for the total variance explained for just the four hydrometeorological variables chosen. The comparative randomness of forecast lead time relative to resolution on the causal strength and relationship of hydrometeorological variables on total runoff is likely due to the difficulty of exactly recreating the category 5 AR event life cycle. ARs have complex spatiotemporal structures that are hard to predict at watershed scales, particularly the AR landfall location latitude; the sweeping comma-shaped nature, topographic orthogonality, and translational speed of the AR plume at landfall; and the precise precipitation magnitude and rain-snow partitioning over the storm duration. This combined with biases in the forecast land-surface initial conditions, most of which are not truly constrained by *in-situ* observations (e.g., soil moisture probe data and groundwater table levels), could help to explain the randomness of forecast lead time on total runoff at individual reservoir regions.

Summary and Conclusions

We used a storyline approach to recreate California's flood of record, the New Year's flood of 1997, using a regionally refined Earth system modeling approach, RRM-E3SM. This is the first time RRM-E3SM has been used to systematically evaluate a key western United States hydrometeorological extreme event. We assessed how both forecast lead time and model horizontal resolution focused over California influenced forecast skill in recreating the flood event. Across several formal measures of forecast quality and value, RRM-E3SM (3.5km) had the highest skill in recreating the 1997 flood compared with lower-resolution versions of E3SM validated against *in-situ*, reanalysis, and gridded climate products.

RRM-E3SM's ability to simulate the North Pacific large-scale circulation patterns and IVT fields and landfalling AR characteristics prior to and during the 1997 flood were minimally influenced by the refinement of horizontal resolution over California. RRM-E3SM simulations largely agreed with ERA5 in the northern portions of California, particularly for extreme AR conditions. However, all RRM-E3SM simulations systematically produce excessively high AR categories in southern California; this is due to elevated amounts of water vapor in southern California and winds that are systematically higher than ERA5 throughout California. Regional refinement resolution in E3SM is important to the representation of storm total precipitation and storm duration changes in snow water equivalent. We find that RRM-E3SM (3.5km) best represents the statistical distributions of storm total precipitation at 52 precipitation gauge sites, with particular improvement in the precipitation maxima. We attribute this to a better representation of both California's mountainous topography as well as important mesoscale circulations in driving precipitation location and magnitude, notably the Sierra barrier jet. Enhanced snowfall at higher elevations and snowpack ablation at low-to-mid elevations are also better represented in RRM-E3SM (3.5km), as shown by comparison to 50 snow pillow sites and a gridded climate product.

Reservoir inflows represent the integrated watershed response resulting from interactions between atmospheric processes with topography. These interactions drive the simulated precipitation patterns and subsequently interact with land surface processes such as snowpack accumulation and melt, soil moisture content, and surface-through-subsurface flow. Simulated inflows exhibit mixed forecast skill across RRM-E3SM simulations. In general, reservoir inflow time series magnitude and, in some cases, shape were off across RRM-E3SM simulations. This is partly due to the integrated surface-through-subsurface hydrology being simulated with uncalibrated (or "out-of-the-box") parameter settings. Using these parameter values shows how E3SM's default settings, often optimized for mean state skill, represent extreme runoff. Notably, although uncalibrated, RRM-E3SM (3.5km) more consistently matched the time series shape of reservoir inflows across five of the eight major reservoirs in California. Future work will leverage the skillfully-resolved atmospheric fields, particularly in RRM-E3SM (3.5km), to run offline integrated hydrologic models (Maina et al., 2022) to assess partitioning between overland flow and groundwater recharge and/or water infrastructure models (Yates et al., 2022) to assess flood inundation potential associated with management decisions.

In addition to not accounting for water management infrastructure in E3SM, there were difficulties in validating certain aspects of the 1997 flood. Specifically, although the antecedent conditions (e.g., soil moisture content and groundwater table levels) provided by the "Betacast" offline five-year ELM spinup procedure driven by ERA5 meteorology undoubtedly shaped reservoir inflow estimates, more observationally-constrained initial conditions for the simulations were not available. Soil moisture content data (both *in-situ* and remote sensing-based estimates) were impossible to find at sub-monthly timescales prior to the year 2000 and, in particular, in mountains from missing data gaps, partly due to the effects of complex terrain and cloudy days on satellite retrievals. Similarly, observational estimates of groundwater table depths (e.g., piezometers and/or satellite-based estimates) were not publicly available.

Forecast lead time resulted in a random effect on the hydrometeorological representation of the 1997 flood. We speculate this is because the forecast lead times chosen (2-to-4 days prior to the 1997 flood onset) were comfortably within the forecast predictability of large-scale synoptic events like ARs (Haiden et al., 2021) and results were therefore dependent on more chaotic spinup processes, mesoscale processes with the main precipitation shield, and small-scale interactions of flow with orography. Although examining the sub-seasonal-to-seasonal forecast skill of E3SM is beyond the scope of this study, L'Heureux et al. (2021) has shown that precipitation forecast skill across seven Earth system model forecasts for California begins to sharply drop with lead times of 8-14 days. Alternatively, to isolate why 2-to-4 day forecast lead time had a relatively random effect on storm total precipitation RRM-E3SM can be run similarly to a weather forecast model, where data produced outside of the regionally refined domain is swapped with reanalysis data (Kruse et al., 2022; Zhang et al., 2022), to better constrain the lateral boundary conditions and, ultimately, the lifecycle of the AR propagation and landfall. Alternatively, the use of perturbed physics ensembles may help to further constrain which subgrid-scale parameterization most influenced drift in AR propagation and landfall and hydrometeorological characteristics of the RRM-E3SM forecasts (Mulholland et al., 2017). Last, given the noted uncertainties in land surface initial conditions, an AR-induced flood event that overlaps with recent high-resolution satellite-based estimates (Vergopolan et al., 2022) could be performed with RRM-E3SM to better isolate the role of antecedent conditions (e.g., soil moisture content) on flood event characteristics (e.g., reservoir inflows). Practically, the lack of hydrometeorological sensitivity with forecast lead time between two to four days prior to the onset of the flood event implies that if a flood manager is interested in event evolution at a specific point an ensemble forecast approach is necessary (e.g., simulations spanning multiple lead times and/or perturbed physics).

Overall, RRM-E3SM (3.5km) forecast ensemble average skill in recreating the 1997 flood gives confidence in its utility to aid flood resiliency planning. To further the utility of these storyline simulations, in future work, we will investigate flood characteristics if a 1997-like flood event were to have happened without anthropogenic climate change or were to happen again at different global warming levels. We hope that these storyline recreations of the 1997 flood event in past and future climates can supplement ongoing efforts in water resource agency flood resiliency planning efforts related to extreme events, especially those involving compounding and/or cascading processes.

Acknowledgments

This study was primarily funded by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy Regional and Global Model Analysis (RGMA) Program. Authors Inda-Diaz, Ombadi, Pasquier, Rhoades, Srivastava, Ullrich, and Wehner were funded by "the Calibrated and Systematic Characterization, Attribution and Detection of Extremes (CASCADE)" Science Focus Area (award no. DE-AC02-05CH11231). Authors Dennis, Jones, McCrary, McGinnis, Rahimi-Esfarjani, Rhoades, Slinskey, Srivastava, Ullrich, and Zarzycki were funded by the "An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US" project (award no. DE-SC0016605). Authors Hatchett and Heggli received support from the Nevada Department of Transportation under agreement P296-22-803. We would also like to acknowledge Smitha Buddhavarapu and Kripa Jagannathan for the considerable time and energy they provided in facilitating scientist-stakeholder discussions in the HyperFACETS project. The facilitated discussions played a key role in helping us to choose the 1997 flood event as a featured storyline within the HyperFACETS project. We also appreciate their constructive suggestions on our manuscript draft.

Open Research

Analysis and model simulations were performed using the National Energy Research Scientific Computing Center (NERSC), specifically Cori-Haswell and Cori-KNL supercomputing facilities. ERA5 is publicly available at the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) at `https://cds.climate.copernicus.eu/#!/search?text=ERA5`. The SSM/I data used in Figure 1 are produced by Remote Sensing Systems. Data are available at `www.remss.com/missions/ssmi`. The Betacast source code is available at `https://github.com/zarzycki/betacast`. The RRM-CESM simulations generated for this study will be made accessible via a NERSC Science Gateway at the time of publication.

References

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., ... Sadegh, M. (2020). Climate Extremes and Compound Hazards in a Warming World. *Annual Review of Earth and Planetary Sciences*, *48*(1), 519-548. doi: 10.1146/annurev-earth-071719-055228

Bambach, N. E., Rhoades, A. M., Hatchett, B. J., Jones, A. D., Ullrich, P. A., & Zarzycki, C. M. (2021). Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile. *International Journal of Climatology*, 1-29. doi: https://doi.org/10.1002/joc.7379

Bukovsky, M. S., Gutowski, W., Mearns, L. O., Paquin, D., & Pryor, S. C. (2023). Climate storylines. *Bulletin of the American Meteorological Society*, *104*(1), E96 - E98. doi: https://doi.org/10.1175/BAMS-D-22-0224.1

Collow, A. B. M., Shields, C. A., Guan, B., Kim, S., Lora, J. M., McClenny, E. E., ... Wehner, M. (2022). An Overview of ARTMIP's Tier 2 Reanalysis Intercomparison: Uncertainty in the Detection of Atmospheric Rivers and Their Associated Precipitation. *Journal of Geophysical Research: Atmospheres*, *127*(8), e2021JD036155. doi: https://doi.org/10.1029/2021JD036155

Copernicus Climate Change Service Climate Data Store (CDS). (2017). *Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.* Retrieved 2022-10-13, from `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview`

Corringham, T. W., McCarthy, J., Shulgina, T., Gershunov, A., Cayan, D. R., & Ralph, F. M. (2022). Climate change contributions to future atmospheric river flood damages in the western United States. *Scientific reports*, *12*(1), 1–9. doi: https://doi.org/10.1038/s41598-022-15474-2

Corringham, T. W., Ralph, F. M., Gershunov, A., Cayan, D. R., & Talbot, C. A. (2019). Atmospheric rivers drive flood damages in the western United States. *Science Advances*, *5*(12). doi: 10.1126/sciadv.aax4631

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

Fang, Y., Liu, Y., & Margulis, S. A. (2022). A western United States snow reanalysis dataset over the Landsat era from water years 1985 to 2021. *Scientific Data*, *9*(1), 1–17. doi: https://doi.org/10.1038/s41597-022-01768-7

Fasullo, J. T. (2020). Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geoscientific Model Development*, *13*(8), 3627–3642. doi: 10.5194/gmd-13-3627-2020

Fish, M. A., Wilson, A. M., & Ralph, F. M. (2019, 10). Atmospheric River Families:

Definition and Associated Synoptic Conditions. *Journal of Hydrometeorology*, *20*(10), 2091-2108. doi: 10.1175/JHM-D-18-0217.1

Galewsky, J., & Sobel, A. (2005). Moist dynamics and orographic precipitation in northern and central california during the new year's flood of 1997. *Monthly Weather Review*, *133*(6), 1594 - 1612. doi: 10.1175/MWR2943.1

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., ... Williams, D. N. (1999). An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *80*(1), 29–56. doi: 10.1175/1520-0477(1999)080⟨0029:AOOTRO⟩2.0.CO;2

Gershunov, A., Shulgina, T., Clemesha, R. E., Guirguis, K., Pierce, D. W., Dettinger, M. D., ... others (2019). Precipitation regime change in Western North America: the role of Atmospheric Rivers. *Scientific reports*, *9*(1), 1–11. doi: https://doi.org/10.1038/s41598-019-46169-w

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., ... Bader, D. C. (2022). The DOE E3SM Model Version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, *n/a*(n/a), e2022MS003156. doi: https://doi.org/10.1029/2022MS003156

Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., ... Zarzycki, C. (2020). The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society*, *101*(5), E664 - E683. doi: 10.1175/BAMS-D-19-0113.1

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., & Richardson, D. (2021, 01/2021). Evaluation of ECMWF forecasts, including the 2020 upgrade. *ECMWF Technical Memoranda*(880). Retrieved from `https://www.ecmwf.int/node/19879` doi: 10.21957/6njp8byz4

Hanak, E., & Lund, J. R. (2012). Adapting California's water management to climate change. *Climatic change*, *111*, 17–44. doi: https://doi.org/10.1007/s10584-011-0241-3

Harrop, B., Balaguru, K., Golaz, J.-C., Leung, L., Mahajan, S., Rhoades, A., ... Mametjanov, A. (2022). Evaluating the water cycle over CONUS using multiple metrics for the Energy Exascale Earth System Model version 1 (E3SMv1) across resolutions. *Journal of Advances in Modeling Earth Systems*(In Revision).

Hatchett, B. J., & McEvoy, D. J. (2018). Exploring the origins of snow drought in the northern sierra nevada, california. *Earth Interactions*, *22*(2), 1 - 13. doi: https://doi.org/10.1175/EI-D-17-0027.1

Heggli, A., Hatchett, B., Schwartz, A., Bardsley, T., & Hand, E. (2022). Toward snowpack runoff decision support. *iScience*, *25*(5), 104240. doi: https://doi.org/10.1016/j.isci.2022.104240

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Huang, X., Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2016). An evaluation of the variable-resolution CESM for modeling California's climate. *Journal of Advances in Modeling Earth Systems*, *8*(1), 345–369. doi: 10.1002/2015MS000559

Huang, X., & Swain, D. L. (2022). Climate change is increasing the risk of a california megaflood. *Science Advances*, *8*(32), eabq0995. doi: 10.1126/sciadv.abq0995

Hughes, M., Neiman, P. J., Sukovich, E., & Ralph, M. (2012). Representation of the Sierra Barrier Jet in 11 years of a high-resolution dynamical reanalysis downscaling compared with long-term wind profiler observations. *Journal of Geophysical Research: Atmospheres*, *117*(D18). doi: https://doi.org/10.1029/2012JD017869

Ivancic, T. J., & Shaw, S. B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, *133*, 681–693. doi: https://doi.org/10.1007/s10584-015-1476-1

Jagannathan, K., Jones, A. D., & Ray, I. (2021). The making of a metric: Co-producing decision-relevant climate science. *Bulletin of the American Meteorological Society*, *102*(8), E1579 - E1590. doi: 10.1175/BAMS-D-19-0296.1

Jones, A. D., Rastogi, D., Vahmani, P., Stansfield, A., Reed, K., Thurber, T., … Rice, J. S. (2022). IM3/HyperFACETS Thermodynamic Global Warming

(TGW) Simulation Datasets (v1.0.0) [Data set]. *MSD-LIVE Data Repository*. doi: https://doi.org/10.57931/1885756

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*(9), 2204 - 2210. Retrieved from https://journals.ametsoc.org/view/journals/mwre/127/9/1520-0493_1999_127_2204_fasocr_2.0.co_2.xml doi: https://doi.org/10.1175/1520-0493(1999)127⟨2204:FASOCR⟩2.0.CO;2

Kattelmann, R. (1997). Flooding from rain-on-snow events in the Sierra Nevada. *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, *239*, 59–66.

Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, *21*(4), e952. doi: https://doi.org/10.1002/asl.952

Kruse, C. G., Bacmeister, J. T., Zarzycki, C. M., Larson, V. E., & Thayer-Calder, K. (2022). Do Nudging Tendencies Depend on the Nudging Timescale Chosen in Atmospheric Models? *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS003024. doi: https://doi.org/10.1029/2022MS003024

La Follette, P. T., Teuling, A. J., Addor, N., Clark, M., Jansen, K., & Melsen, L. A. (2021). Numerical daemons of hydrological models are summoned by extreme precipitation. *Hydrology and Earth System Sciences*, *25*(10), 5425–5446. Retrieved from https://hess.copernicus.org/articles/25/5425/2021/ doi: 10.5194/hess-25-5425-2021

Lauritzen, P. H., Bacmeister, J. T., Callaghan, P. F., & Taylor, M. A. (2015, dec). NCAR_topo (v1.0): NCAR global model topography generation software for unstructured grids. *Geoscientific Model Development*, *8*(12), 3975–3986. Retrieved from https://doi.org/10.5194%2Fgmd-8-3975-2015 doi: 10.5194/gmd-8-3975-2015

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*(5), 2275-2282. doi: https://doi.org/10.1002/2016GL067765

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., . . .

Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. Retrieved from https://esd.copernicus.org/articles/11/491/2020/ doi: 10.5194/esd-11-491-2020

Lemos, M. C., Arnott, J. C., Ardoin, N. M., Baja, K., Bednarek, A. T., Dewulf, A., ... others (2018). To co-produce or not to co-produce. *Nature Sustainability*, *1*(12), 722–724. doi: https://doi.org/10.1038/s41893-018-0191-0

Lemos, M. C., Kirchhoff, C. J., & Ramprasad, V. (2012). Narrowing the climate information usability gap. *Nature climate change*, *2*(11), 789–794. doi: https://doi.org/10.1038/nclimate1614

Liu, W., Ullrich, P. A., Guba, O., Caldwell, P. M., & Keen, N. D. (2022). An Assessment of Nonhydrostatic and Hydrostatic Dynamical Cores at Seasonal Time Scales in the Energy Exascale Earth System Model (E3SM). *Journal of Advances in Modeling Earth Systems*, *14*(2), e2021MS002805. doi: https://doi.org/10.1029/2021MS002805

Liu, W., Ullrich, P. A., Li, J., Zarzycki, C., Caldwell, P. M., Leung, L. R., & Qian, Y. (2023). The June 2012 North American Derecho: A Testbed for Evaluating Regional and Global Climate Modeling Systems at Cloud-Resolving Scales. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003358. doi: https://doi.org/10.1029/2022MS003358

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, *2*(150042). doi: doi:10.1038/sdata.2015.42

Lott, N., Sittel, M. C., & Ross, D. (1997). The winter of '96-'97 : west coast flooding. *National Climatic Data Center technical report ; 97-01*, 1 - 22. doi: https://repository.library.noaa.gov/view/noaa/13812

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (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. doi: https://doi.org/10.1175/BAMS-D-19-0001.1

Lundquist, J. D., Hughes, M., Henn, B., Gutmann, E. D., Livneh, B., Dozier, J., & Neiman, P. (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. doi: https://doi.org/10.1175/JHM-D-15-0019.1

L'Heureux, M. L., Tippett, M. K., & Becker, E. J. (2021). Sources of subseasonal skill and predictability in wintertime california precipitation forecasts. *Weather and Forecasting*, *36*(5), 1815 - 1826. doi: https://doi.org/10.1175/WAF-D-21-0061.1

Maina, F. Z., Rhoades, A., Siirila-Woodburn, E. R., & Dennedy-Frank, P.-J. (2022). Projecting end-of-century climate extremes and their impacts on the hydrology of a representative California watershed. *Hydrology and Earth System Sciences*, *26*(13), 3589–3609. doi: 10.5194/hess-26-3589-2022

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245–259. doi: http://www.jstor.org/stable/1907187

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A landsat-era sierra nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, *17*(4), 1203 - 1221. doi: https://doi.org/10.1175/JHM-D-15-0177.1

Mulholland, D. P., Haines, K., Sparrow, S. N., & Wallom, D. (2017). Climate model forecast biases assessed with a perturbed physics ensemble. *Climate Dynamics*, *49*, 1729–1746. doi: https://doi.org/10.1007/s00382-016-3407-x

Murphy, A. H. (1993). What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, *8*(2), 281 - 293. doi: 10.1175/1520-0434(1993)008⟨0281:WIAGFA⟩2.0.CO;2

Neiman, P. J., Hughes, M., Moore, B. J., Ralph, F. M., & Sukovich, E. M. (2013). Sierra barrier jets, atmospheric rivers, and precipitation characteristics in northern california: A composite perspective based on a network of wind profilers. *Monthly Weather Review*, *141*(12), 4211 - 4233. doi: https://doi.org/10.1175/MWR-D-13-00112.1

Neiman, P. J., Sukovich, E. M., Ralph, F. M., & Hughes, M. (2010). A seven-year wind profiler–based climatology of the windward barrier jet along california's northern sierra nevada. *Monthly Weather Review*, *138*(4), 1206 - 1233. doi: https://doi.org/10.1175/2009MWR3170.1

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., ... Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous united states. *Journal of Hydrometeorology*, *16*(6), 2481 - 2500. doi: https://doi.org/10.1175/JHM-D-15-0026.1

Ombadi, M., Nguyen, P., Sorooshian, S., & lin Hsu, K. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7). Retrieved from `https://doi.org/10.1029/2020wr027251` doi: 10.1029/2020wr027251

Ombadi, M., Nguyen, P., Sorooshian, S., & lin Hsu, K. (2021). How much information on precipitation is contained in satellite infrared imagery? *Atmospheric Research*, *256*, 105578. Retrieved from `https://doi.org/10.1016/j.atmosres.2021.105578` doi: 10.1016/j.atmosres.2021.105578

Osterhuber, R., & Schwartz, A. (2021). Snowpack, precipitation, and temperature measurements at the Central Sierra Snow Laboratory for water years 1971 to 2019. *Dryad, Dataset*. doi: https://doi.org/10.6078/D1941T

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L.-R., ... Zhou, Y. (2022). Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD036013. doi: https://doi.org/10.1029/2021JD036013

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., ... Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139-152. doi: https://doi.org/10.1016/j.rse.2016.06.018

Pierce, D. W., Su, L., Cayan, D. R., Risser, M. D., Livneh, B., & Lettenmaier, D. P. (2021). An extreme-preserving long-term gridded daily precipitation dataset for the conterminous united states. *Journal of Hydrometeorology*, *22*(7), 1883 - 1895. Retrieved from `https://journals.ametsoc.org/view/journals/hydr/22/7/JHM-D-20-0212.1.xml` doi: 10.1175/JHM-D-20-0212.1

Porter, K., Wein, A., Alpers, C. N., Baez, A., Barnard, P. L., Carter, J., ... others (2011). *Overview of the ARkStorm scenario* (Tech. Rep.). US Geological Survey. doi: https://pubs.er.usgs.gov/publication/ofr20101312

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., ... Smallcomb, C. (2019). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269-289. doi: 10.1175/BAMS-D-18-0023.1

Rauscher, S. A., & Ringler, T. D. (2014). Impact of Variable-Resolution Meshes on Midlatitude Baroclinic Eddies Using CAM-MPAS-A. *Monthly Weather Review*, *142*(11), 4256–4268. doi: 10.1175/MWR-D-13-00366.1

Rauscher, S. A., Ringler, T. D., Skamarock, W. C., & Mirin, A. A. (2013). Exploring a Global Multiresolution Modeling Approach Using Aquaplanet Simulations. *Journal of Climate*, *26*(8), 2432-2452. doi: 10.1175/JCLI-D-12-00154.1

Raymond, C., Horton, R. M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., ... others (2020). Understanding and managing connected extreme events. *Nature climate change*, *10*(7), 611–621. doi: https://doi.org/10.1038/s41558-020-0790-4

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. doi: 10.1175/2007JCLI1824.1

Rhoades, A. M., Huang, X., Ullrich, P. A., & Zarzycki, C. M. (2016). Characterizing Sierra Nevada Snowpack Using Variable-Resolution CESM. *Journal of Applied Meteorology and Climatology*, *55*(1), 173–196. doi: 10.1175/JAMC-D-15-0156.1

Rhoades, A. M., Jones, A. D., O'Brien, T. A., O'Brien, J. P., Ullrich, P. A., & Zarzycki, C. M. (2020). Influences of North Pacific Ocean Domain Extent on the Western U.S. Winter Hydroclimatology in Variable-Resolution CESM. *Journal of Geophysical Research: Atmospheres*, *125*(14), e2019JD031977. doi: 10.1029/2019JD031977

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., ... Zhou, Y. (2020). The Shifting Scales of Western U.S. Landfalling Atmospheric Rivers Under Climate Change. *Geophysical Research Letters*, *47*(17), e2020GL089096. doi: https://doi.org/10.1029/2020GL089096

Rhoades, A. M., Risser, M. D., Stone, D. A., Wehner, M. F., & Jones, A. D. (2021). Implications of warming on western United States landfalling atmospheric rivers and their flood damages. *Weather and Climate Extremes*, *32*, 100326. doi: https://doi.org/10.1016/j.wace.2021.100326

Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2018). Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM. *Climate Dynamics*, *50*(1), 261–288. doi: 10.1007/s00382-017-3606-0

Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A., Morrison, H., ... Collins, W. D. (2018). Sensitivity of Mountain Hydroclimate Simulations in Variable-Resolution CESM to Microphysics and Horizontal Resolution. *Journal of Advances in Modeling Earth Systems*, *10*(6), 1357–1380. doi: 10.1029/2018MS001326

Ricciotti, J. A., & Cordeira, J. M. (2022). Summarizing relationships among landfalling atmospheric rivers, integrated water vapor transport, and california watershed precipitation 1982–2019. *Journal of Hydrometeorology*, *23*(9), 1439 - 1454. doi: https://doi.org/10.1175/JHM-D-21-0119.1

Runge, J. (2023). Modern causal inference approaches to investigate biodiversity-ecosystem functioning relationships. *nature communications*, *14*(1), 1917. doi: https://doi.org/10.1038/s41467-023-37546-1

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., ... others (2019). Inferring causation from time series in Earth system sciences. *Nature communications*, *10*(1), 2553. doi: https://doi.org/10.1038/s41467-019-10105-3

Sakaguchi, K., Lu, J., Leung, L. R., Zhao, C., Li, Y., & Hagos, S. (2016). Sources and pathways of the upscale effects on the Southern Hemisphere jet in MPAS-CAM4 variable-resolution simulations. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1786-1805. doi: 10.1002/2016MS000743

Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic unit maps: US Geological Survey water supply paper 2294.* U.S. Geological Survey. doi: https://pubs.usgs.gov/wsp/wsp2294/

Shepherd, T. G. (2019). Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2225), 20190013. doi: 10.1098/rspa.2019.0013

Siirila-Woodburn, E., Rhoades, A., Hatchett, B., Huning, L., Szinai, J., Tague, C., ... Kaatz, L. (2021). Evidence of a low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth and Environment*, *2*, 800–819. doi: https://doi.org/10.1038/s43017-021-00219-y

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, *9*(1), 62–72. Retrieved from https://doi.org/10.1177/089443939100900106 doi: 10.1177/089443939100900106

Srivastava, A. K., Grotjahn, R., Ullrich, P. A., & Sadegh, M. (2021). Pooling data improves multimodel idf estimates over median-based idf estimates: Analysis over the susquehanna and florida. *Journal of Hydrometeorology*, *22*(4), 971 - 995. Retrieved from https://journals.ametsoc.org/view/journals/hydr/22/4/JHM-D-20-0180.1.xml doi: https://doi.org/10.1175/JHM-D-20-0180.1

Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., & Bair, E. H. (2023). Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets. *The Cryosphere*, *17*(2), 567–590. doi: 10.5194/tc-17-567-2023

Sugihara, G., May, R., Ye, H., hao Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting Causality in Complex Ecosystems. *Science*, *338*(6106), 496-500. doi: 10.1126/science.1227079

Tang, Q., Golaz, J.-C., Van Roekel, L. P., Taylor, M. A., Lin, W., Hillman, B. R., … Bader, D. C. (2022). The Fully Coupled Regionally Refined Model of E3SM Version 2: Overview of the Atmosphere, Land, and River. *Geoscientific Model Development Discussions*, *2022*, 1–64. Retrieved from https://gmd.copernicus.org/preprints/gmd-2022-262/ doi: 10.5194/gmd-2022-262

Tang, Q., Klein, S. A., Xie, S., Lin, W., Golaz, J.-C., Roesler, E. L., … Zheng, X. (2019). Regionally refined test bed in e3sm atmosphere model version 1 (eamv1) and applications for high-resolution modeling. *Geoscientific Model Development*, *12*(7), 2679–2706. Retrieved from https://gmd.copernicus.org/articles/12/2679/2019/ doi: 10.5194/gmd-12-2679-2019

Tarouilly, E., Li, D., & Lettenmaier, D. P. (2021). Western U.S. Superfloods in the Recent Instrumental Record. *Water Resources Research*, *57*(9), e2020WR029287. doi: https://doi.org/10.1029/2020WR029287

The NCAR Command Language (Version 6.6.2). (2022). *Boulder, Colorado: UCAR/NCAR/CISL/TDD.* (http://dx.doi.org/10.5065/D6WD3XH5)

Trenberth, E., Berry, C., & Buja, E. (1993). *Vertical interpolation and truncation of* *model-coordinate data.* doi: 10.5065/D6HX19NH

Ullrich, P. A. (2014). Understanding the treatment of waves in atmospheric models. Part 1: The shortest resolved waves of the 1D linearized shallow-water equations. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1426-1440. doi: https://doi.org/10.1002/qj.2226

Ullrich, P. A., Devendran, D., & Johansen, H. (2016). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part ii. *Monthly Weather Review*, *144*(4), 1529–1549.

Ullrich, P. A., & Taylor, M. A. (2015). Arbitrary-order conservative and consistent remapping and a theory of linear maps: Part i. *Monthly Weather Review*, *143*(6), 2419–2440.

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090. doi: 10.5194/gmd-10-1069-2017

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., . . . Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961-3012. doi: https://doi.org/10.1256/qj.04.176

Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., . . . Wood, E. F. (2022). High-Resolution Soil Moisture Data Reveal Complex Multi-Scale Spatial Variability Across the United States. *Geophysical Research Letters*, *49*(15), e2022GL098586. doi: https://doi.org/10.1029/2022GL098586

Wehner, M. F., Zarzycki, C. M., & Patricola, C. M. (2019). Estimating the human influence on tropical cyclone intensity as the climate changes. In J. Collins & K. J. Walsh (Eds.), *Hurricane risk* (pp. 235–260). Springer.

Whipple, A. A., & Viers, J. H. (2019). Coupling landscapes and river flows to restore highly modified rivers. *Water Resources Research*, *55*(6), 4512-4532. doi: https://doi.org/10.1029/2018WR022783

Whipple, A. A., Viers, J. H., & Dahlke, H. E. (2017). Flood regime typology for floodplain ecosystem management as applied to the unregulated Cosumnes River of California, United States. *Ecohydrology*, *10*(5), e1817. doi: https://doi.org/10.1002/eco.1817

Wing, O. E., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M.,

. . . Kousky, C. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, *12*(2), 156–162. doi: https://doi.org/10.1038/s41558-021-01265-6

Wu, C., Liu, X., Lin, Z., Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., . . . Rahimi-Esfarjani, S. R. (2017). Exploring a Variable-Resolution Approach for Simulating Regional Climate in the Rocky Mountain Region Using the VR-CESM. *Journal of Geophysical Research: Atmospheres*, *122*(20), 10,939–10,965. doi: 10.1002/2017JD027008

Xu, Z., Di Vittorio, A., Zhang, J., Rhoades, A., Xin, X., Xu, H., & Xiao, C. (2021). Evaluating Variable-Resolution CESM Over China and Western United States for Use in Water-Energy Nexus and Impacts Modeling. *Journal of Geophysical Research: Atmospheres*, *126*(15), e2020JD034361. doi: https://doi.org/10.1029/2020JD034361

Xu, Z., Rhoades, A. M., Johansen, H., Ullrich, P. A., & Collins, W. D. (2018). An Intercomparison of GCM and RCM Dynamical Downscaling for Characterizing the Hydroclimatology of California and Nevada. *Journal of Hydrometeorology*, *19*(9), 1485–1506. doi: 10.1175/JHM-D-17-0181.1

Yang, Q., Leung, L. R., Lu, J., Lin, Y.-L., Hagos, S., Sakaguchi, K., & Gao, Y. (2017). Exploring the effects of a nonhydrostatic dynamical core in high-resolution aquaplanet simulations. *Journal of Geophysical Research: Atmospheres*, *122*(6), 3245-3265. doi: https://doi.org/10.1002/2016JD025287

Yates, D., Szinai, J., & Jones, A. D. (2022). Modeling the Water Systems of the Western US to Support Climate-Resilient Electricity System Planning. *Authorea Preprints*. doi: 10.1002/essoar.10512623.1

Zarzycki, C. M., Jablonowski, C., & Taylor, M. A. (2014). Using Variable-Resolution Meshes to Model Tropical Cyclones in the Community Atmosphere Model. *Monthly Weather Review*, *142*(3), 1221–1239. doi: 10.1175/MWR-D-13-00179.1

Zarzycki, C. M., Jablonowski, C., Thatcher, D. R., & Taylor, M. A. (2015). Effects of localized grid refinement on the general circulation and climatology in the community atmosphere model. *Journal of Climate*, *28*(7), 2777 - 2803. doi: https://doi.org/10.1175/JCLI-D-14-00599.1

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, *44*(2), 1141-1149. doi: 10.1002/2016GL071606

Zhang, S., Zhang, K., Wan, H., & Sun, J. (2022). Further improvement and evaluation of nudging in the E3SM Atmosphere Model version 1 (EAMv1): simulations of the mean climate, weather events, and anthropogenic aerosol effects. *Geoscientific Model Development*, *15*(17), 6787–6816. doi: 10.5194/gmd-15-6787-2022

Zhou, Y., O'Brien, T. A., Ullrich, P. A., Collins, W. D., Patricola, C. M., & Rhoades, A. M. (2021). Uncertainties in Atmospheric River Lifecycles by Detection Algorithms: Climatology and Variability. *Journal of Geophysical Research: Atmospheres*, *126*(8), e2020JD033711. doi: https://doi.org/10.1029/2020JD033711

Supplemental Material

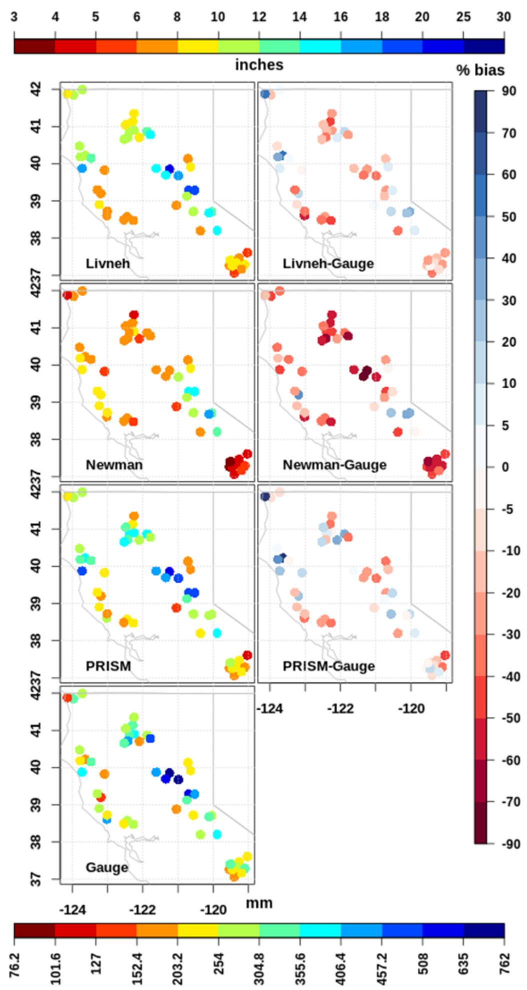

Figure S1. Percent bias in 4-day storm-total precipitation (31 December 1996 up to 4 January 1997), in three best-available reanalysis products compared against precipitation gauge stations. Leftmost sub-panel plots represent storm total precipitation and right sub-panel plots indicate percent biases. Metric (English) units are provided in the bottom (top) color bars.

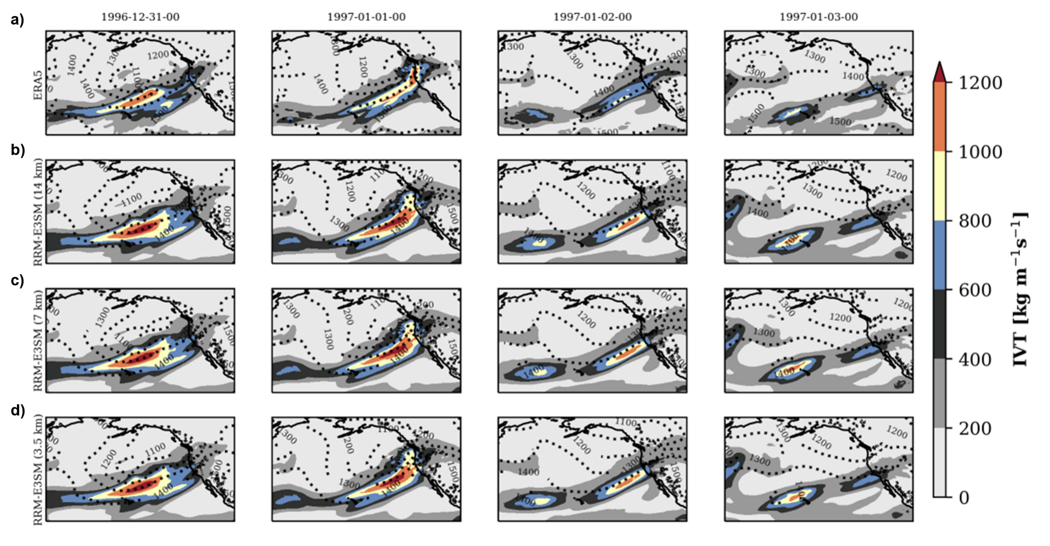

Figure S2. Same as Figure 3a, however, IVT is shown for each day of the 1997 flood.

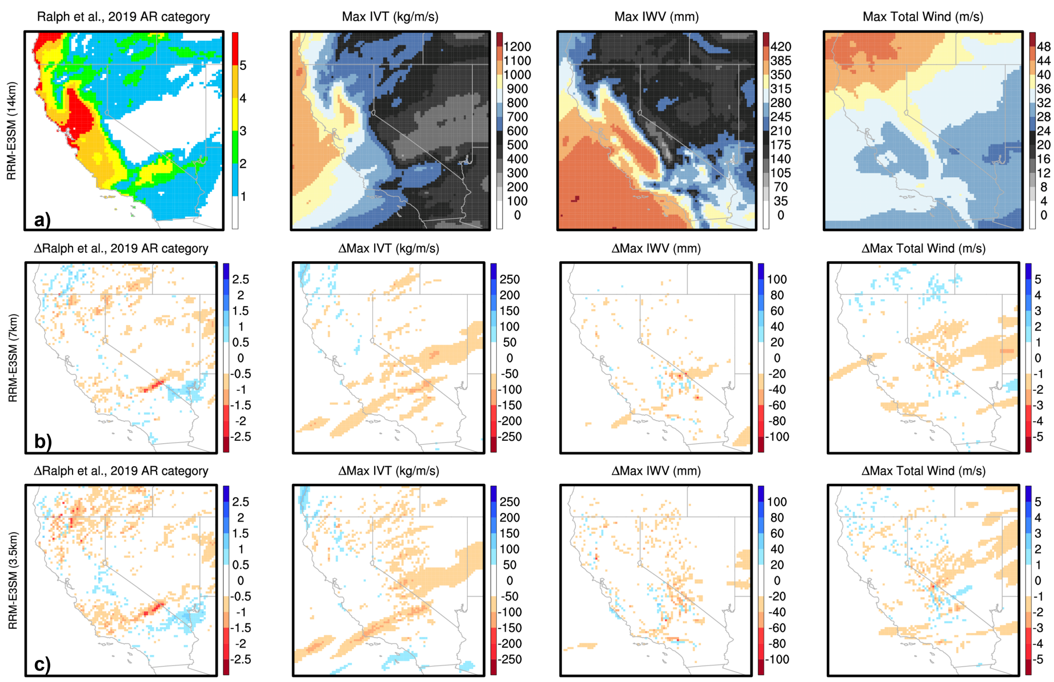

Figure S3. Same as Figure 4, however differences from RRM-E3SM (14km).

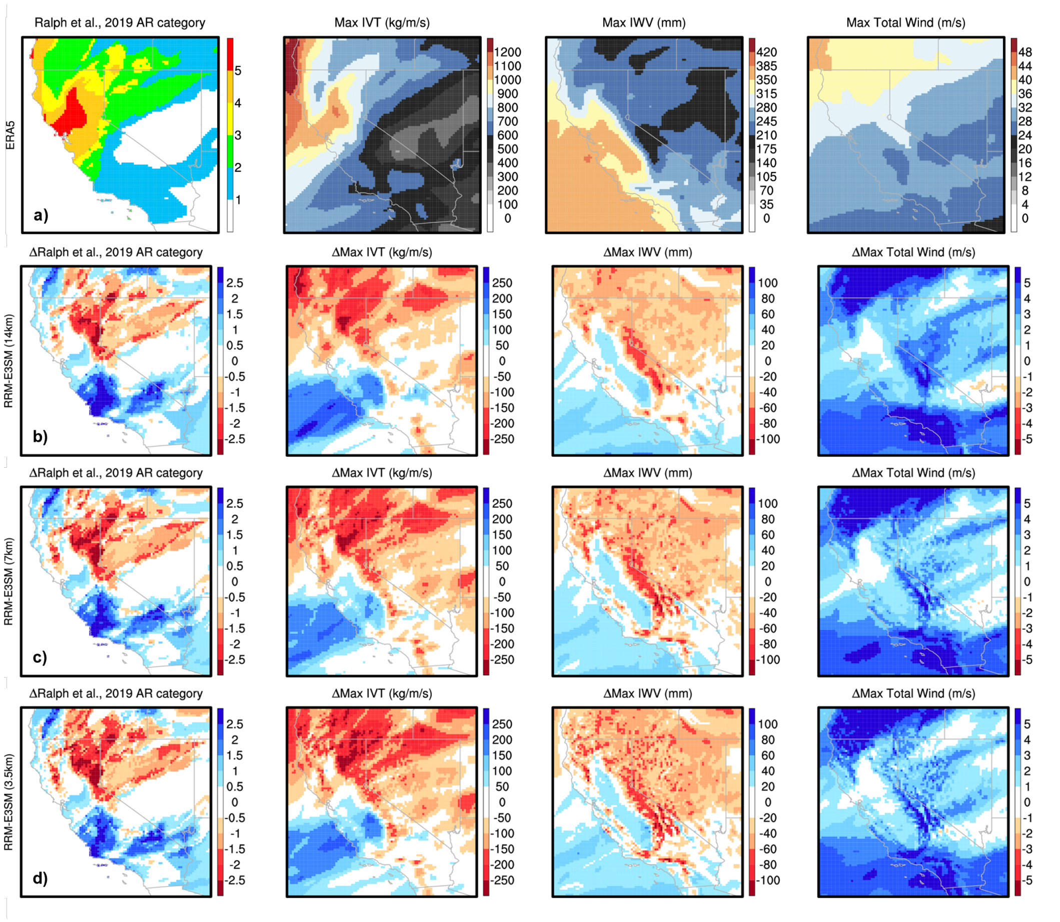

Figure S4. Same as Figure 4, however differences from ERA5.

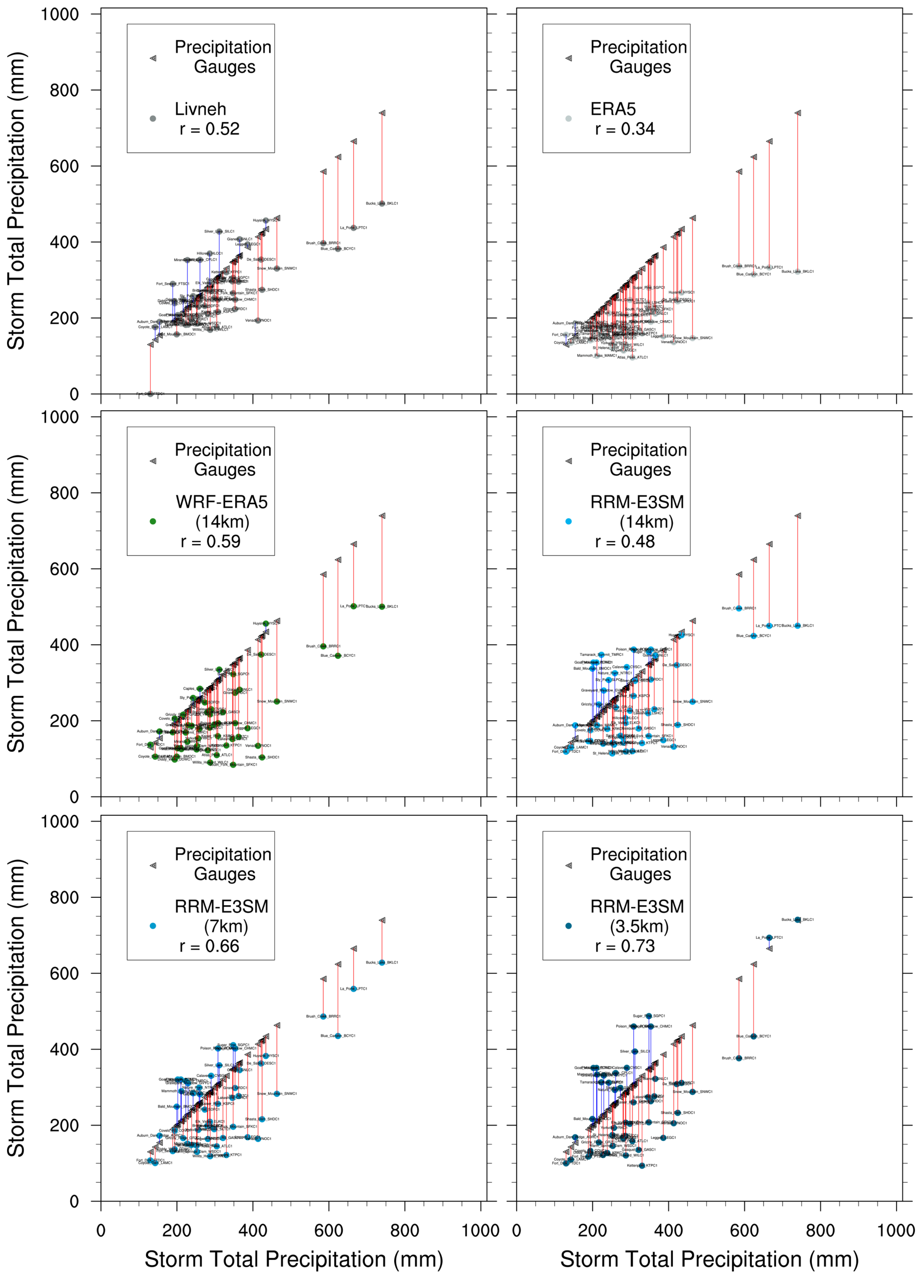

Figure S5. Scatter plot comparing storm total precipitation (31 December 1996 up to 4 January 1997) for the 52 NOAA precipitation gauges and the nearest grid cell within each of the reanalysis products and model simulations. Blue (Red) lines represent a grid cell that had a higher (lower) precipitation value than the nearest precipitation gauge. R values are provided in the legend. The name of the NOAA precipitation gauge is overlaid onto each of the reanalysis product and model simulation dots.

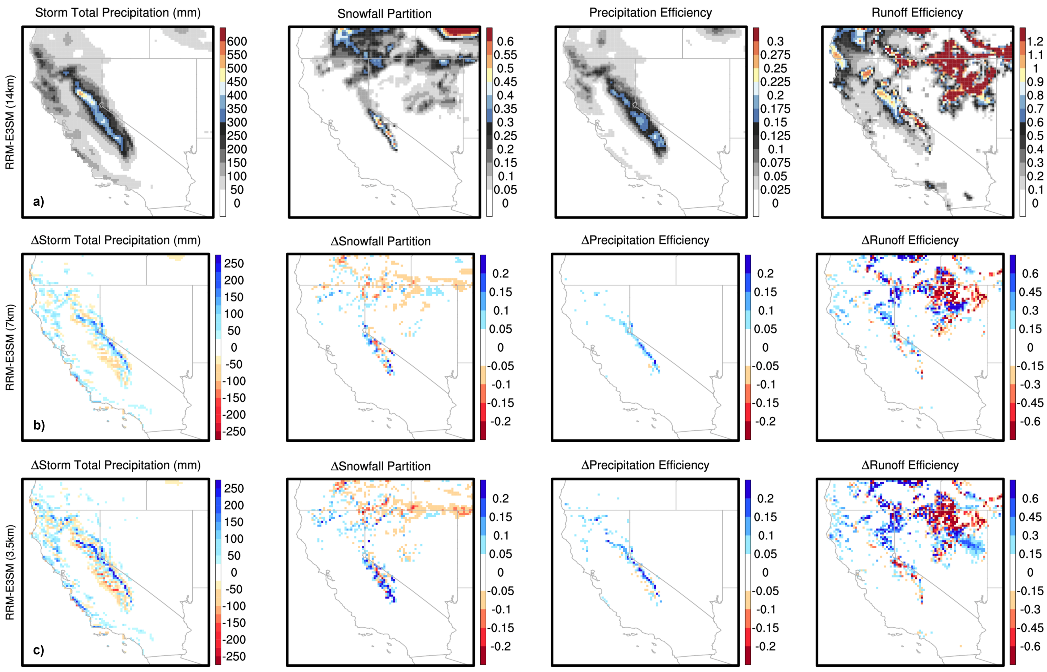

Figure S6. Same as Figure 10, however differences from RRM-E3SM (14km).

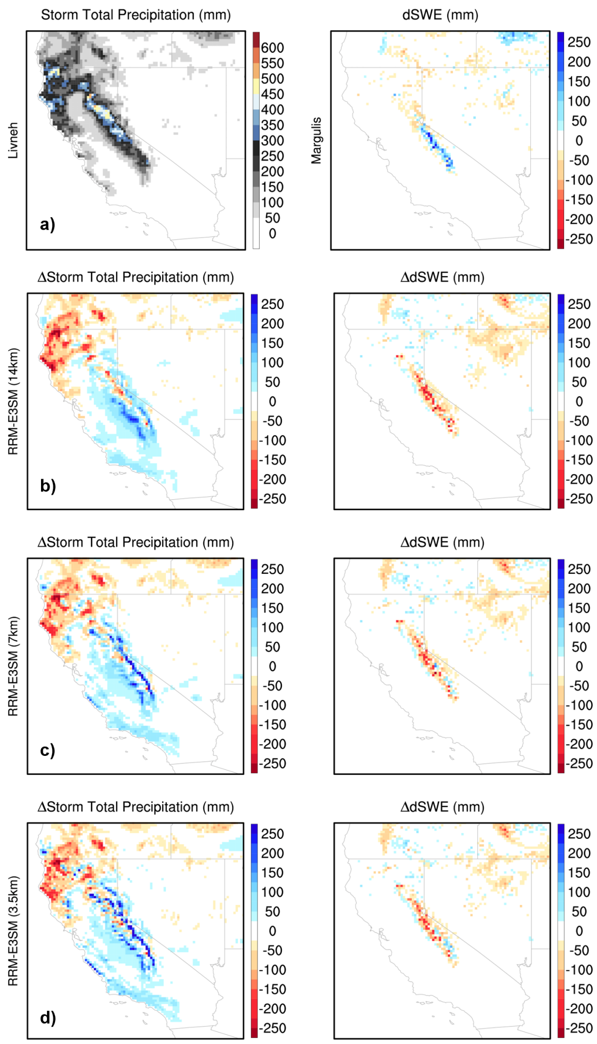

Figure S7. Same as Figure 5a and 9a, however differences from Livneh and Margulis.

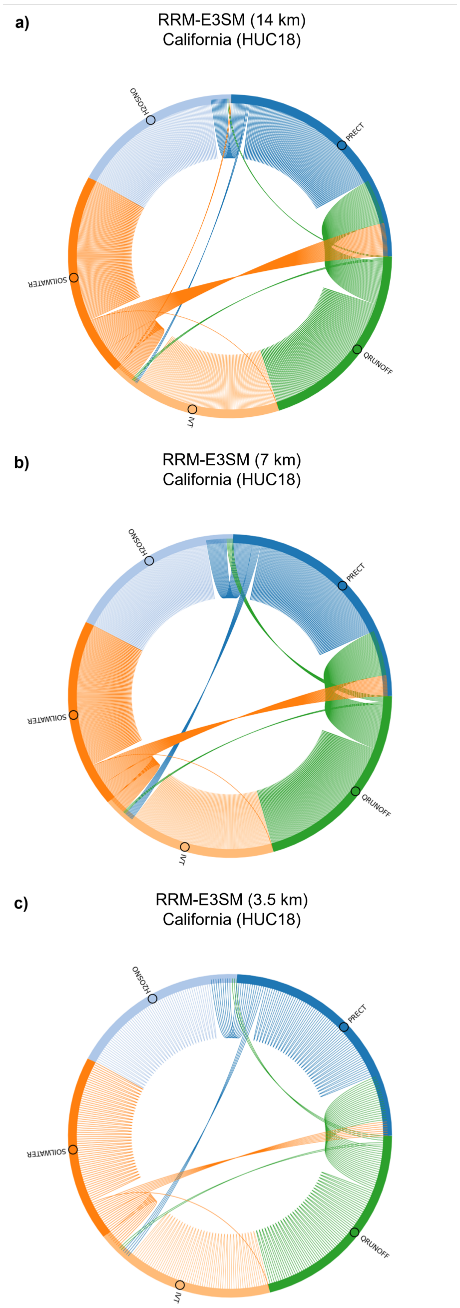

Figure S8. Chord diagrams depicting the causal inference estimates for the magnitude and direction of the impact of hydrometeorological variables on total runoff across California (Hydrologic Unit Code 18) for 31 December 1996 up to 4 January 1997. The four variables include integrated vapor transport (IVT), total precipitation (PRECT), snow water equivalent (SWE), and 10 cm soil moisture content (SOILWATER).

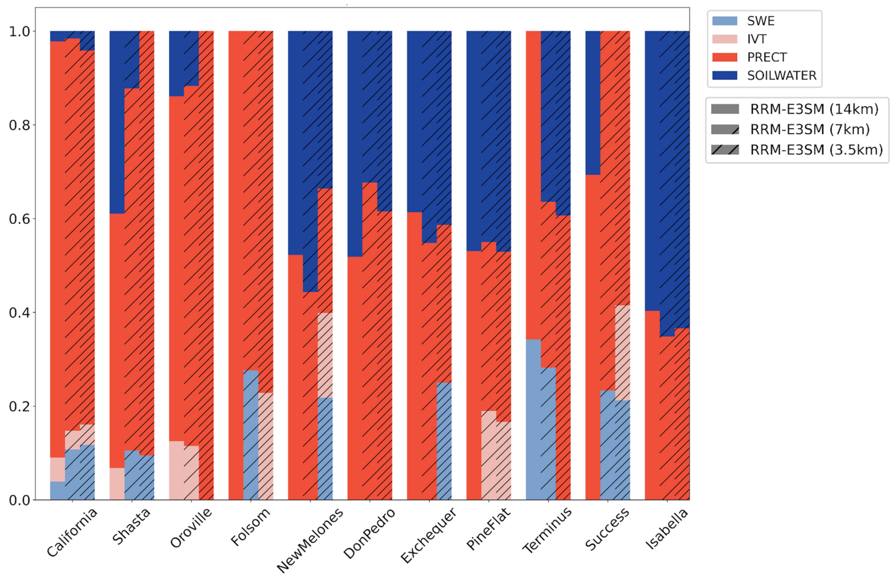

Figure S9. Same as Figure 12, however, stacked bar components are normalized by the total variance explained by the four hydrometeorological variables chosen.

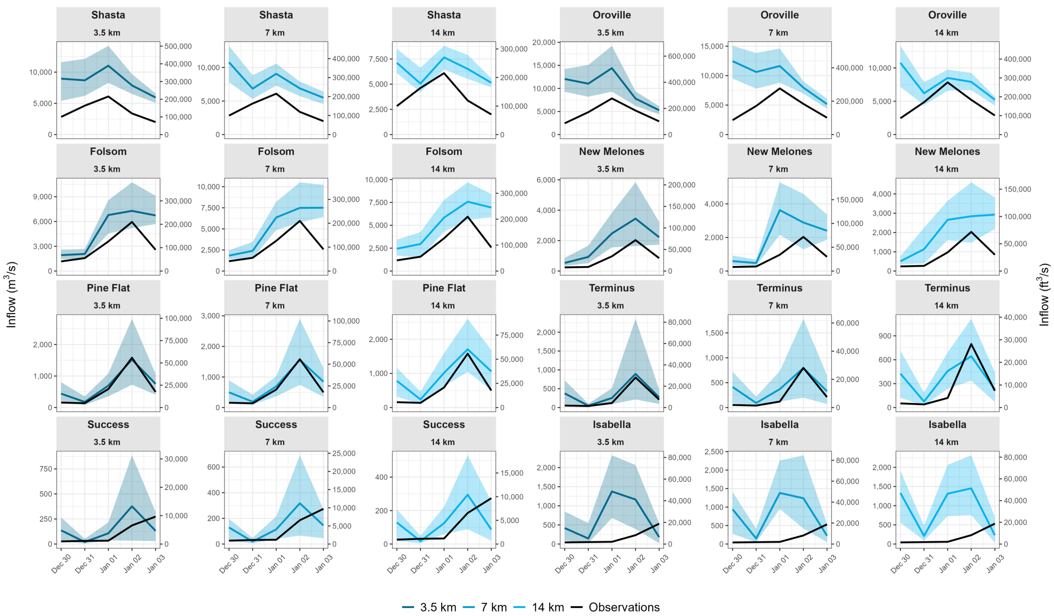

Figure S10. Same as Figure 11, however, the RRM-E3SM six-member forecast spread in reservoir inflows are shown (shaded region) around the ensemble mean (line).