RESEARCH ARTICLE

Key Points:
- When suitably parameterized, the WRF regional climate model skillfully simulates mountain runoff
- Inclusion of a simple aquifer correction improves accuracy of runoff into 10 key California reservoirs
- WRF runoff provides skill when driving an operational reservoir model, though errors remain

Supporting Information:
- Supporting Information S1

Correspondence to:
T. M. Pavelsky,pavelsky@unc.edu

Tailoring WRF and Noah-MP to Improve Process Representation of Sierra Nevada Runoff: Diagnostic Evaluation and Applications

Nataniel M. Holtzman[1,2] (iD), Tamlin M. Pavelsky[1,3] (iD), Jonathan S. Cohen[4] (iD), Melissa L. Wrzesien[3] (iD), and Jonathan D. Herman[4] (iD)

[1]Institute for the Environment, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [2]Now at Department of Earth System Science, Stanford University, Stanford, CA, USA, [3]Department of Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [4]Department of Civil and Environmental Engineering, University of California, Davis, CA, USA

Abstract Watersheds at the western margin of the Sierra Nevada mountains in California are regulated by large dams providing crucial water supply, flood control, and electricity generation. Runoff in these basins is snowmelt dominated and therefore vulnerable to alteration due to climate change. Regional climate models coupled to land surface models can be used to study the hydrologic impacts of climate change, but there is little evidence that they accurately simulate watershed-scale runoff in complex terrain. This study evaluates capabilities of the Weather Research and Forecasting (WRF) regional climate model, coupled to the Noah-multiparameterization (MP) land surface model, to simulate runoff into nine Sierra Nevada reservoirs over the period 2007–2017. Default parameterizations lead to substantial inaccuracy in results, including median bias of 61%. Errors can be traced to process representations; specifically, we modify the representation of snowflake formation in the Thompson microphysics scheme and subsurface runoff generation in the Noah-MP land surface model, including a correction representing effects of groundwater storage. The resulting parameterization improves Nash-Sutcliffe efficiency to above 0.7 across all basins and reduces median bias to 21%. To assess capabilities of the modified WRF/Noah-MP system in supporting analysis of human-altered hydrology, we use its streamflow projections to force a reservoir operations model, results of which maintain high accuracy in predicting reservoir storage and releases (mean Nash-Sutcliffe efficiency > 0.41). This diagnostic analysis indicates that coupled climate and land surface models can be used to study climate change effects on reservoir systems in mountain regions via dynamical downscaling, when adequate physical parameterizations are used.

Plain Language Summary In California, reservoirs are a critical component of hydroelectric power generation and flood control, in addition to storing water used by farms and cities. Much of the water that flows into the reservoirs comes as snowmelt from the Sierra Nevada mountains each spring. Wintertime snow accumulation and springtime snowmelt are likely to be impacted by rising temperatures and shifting precipitation patterns. However, little research has addressed whether regional climate models (RCMs) can accurately simulate runoff for California's reservoirs. Here we make multiple changes to a commonly used RCM, including improvements to how water is processed in the soil and how groundwater is represented. These changes improve the modeled runoff estimates when compared to values based on observations. We use 11 years of modeled runoff to drive a reservoir operations model. Though the modeled water system behavior using RCM runoff is not as realistic as a version of the model using observed runoff, the results predict many of the same variations in reservoir storage. The combined RCM and reservoir operations model can be used to project how reservoir inflows and storage may be impacted by future climate change.

1. Introduction

In many parts of the world, mountain regions function as natural reservoirs, storing water as snow during winter and releasing it in the warm season (Viviroli et al., 2007). This function is especially important in areas such as California, where little precipitation falls in the summer, and mountain snowmelt becomes the primary source of summer streamflow. Many of the rivers draining the west side of the Sierra Nevada

mountains in California flow into reservoirs used for generating hydroelectric power and providing a consistent water supply for agriculture and urban areas (Waliser et al., 2011). Since these basins are snowmelt dominated, they are vulnerable to climate change altering their flow timing, and such changes have been described extensively in the recent scientific literature. Over the past few decades, the fraction of precipitation in western North America that falls as snow has been decreasing (Knowles et al., 2006; Pederson et al., 2011). Many rivers, including some in the Sierra Nevada, are transitioning from snowmelt dominated to rain dominated (Fritze et al., 2011; Kampf & Lefsky, 2016). Barnett et al. (2008) attributed most of the decrease in peak snow water equivalent (SWE) and movement toward earlier river flow centroid time across the western United States to anthropogenic climate change.

Trends toward increased temperature and decreased snowfall are expected to continue throughout the 21st century, particularly in mountain regions, due to snow albedo and other feedbacks (Pepin et al., 2015; Sun et al., 2018; Walton et al., 2017). Regardless of changes in total precipitation amount, in the future, snow will likely melt earlier in the Sierra Nevada. When snow melts earlier, it also tends to melt more slowly (Musselman et al., 2017), leading to lower flows overall as there is less water partitioned to subsurface lateral flow and more water accessible to evapotranspiration (Barnhart et al., 2016). Decreased snow fraction of precipitation is also associated with decreased streamflow (Berghuijs et al., 2014). Thus, it is important to understand how hydroclimatic changes in California may impact reservoir inflow.

Since global climate models (GCMs) are generally too coarse to accurately represent regions with complex topography (Schwartz et al., 2017), numerous studies have dynamically downscaled GCM estimates of runoff for western North America, where many watersheds have headwaters in the mountains. However, previous work estimates runoff on a monthly, or even annual, time step (Christensen & Lettenmaier, 2007; Harding et al., 2012; Pagán et al., 2016; Rasmussen et al., 2014), which does not capture extreme flow events that impact human systems. It is therefore insufficient for modeling reservoir operations (Naz et al., 2016). Other studies focus on runoff timing rather than runoff amount (e.g., Schwartz et al., 2017).

Here we present what is, to our knowledge, the first attempt to evaluate daily runoff in the mountains of the western United States from a coupled land surface model (LSM) and regional climate model (RCM), which can be used to drive operational reservoir models. We assess the capability of the Noah-multiparameterization (MP) LSM coupled with the Weather Research and Forecasting (WRF) RCM to simulate daily basin-scale runoff in California. We compare simulated runoff to full natural flow (FNF, or what the river discharge would be absent upstream human modifications) estimates based on historical observations, assessing model accuracy under default parameter settings over two 1-year periods. Based on this diagnostic analysis, we incorporate several process-based modifications to precipitation schemes within WRF and to runoff generation schemes in Noah-MP, which we test using an 11-year simulation experiment. Finally, we compare the performance of the resulting simulations and FNF as inputs to the Operation of Reservoirs in California (ORCA) reservoir operations model.

Our goal is to modify the WRF/Noah-MP system so that, when properly parameterized, it can be used for dynamical downscaling of GCMs to characterize climate change impacts on runoff, reservoir operations, and water resources management in California. Unlike what is often done with hydrologic models (Refsgaard, 1997), we do not attempt to perform a rigorous calibration of any parameters. Such an undertaking for a coupled RCM/LSM would require much more computing power than we had access to. Instead, our philosophy was to perturb a few parameters in a small number of ways, within realistic ranges and based on knowledge of the Sierra Nevada area. The final parameterized version of WRF/Noah-MP that developed represents the aggregate of those perturbations that produced the most accurate outputs in terms of runoff and snow. Our research does not seek to completely optimize WRF for the Sierra Nevada; in terms of calibration it is more qualitative than quantitative. Rather, we demonstrate that the model's default parameter values produce inaccurate results and that those results can be markedly improved by nudging the parameters in the right direction, within the bounds of physically realistic parameter values.

2. Methods

2.1. Experimental Design

Our geographic setup for WRF v3.8.1 is the same used by Wrzesien et al. (2015): a 9-km resolution inner nest covering California within a 27-km resolution outer domain (Figure 1). Initial and 3-hourly boundary
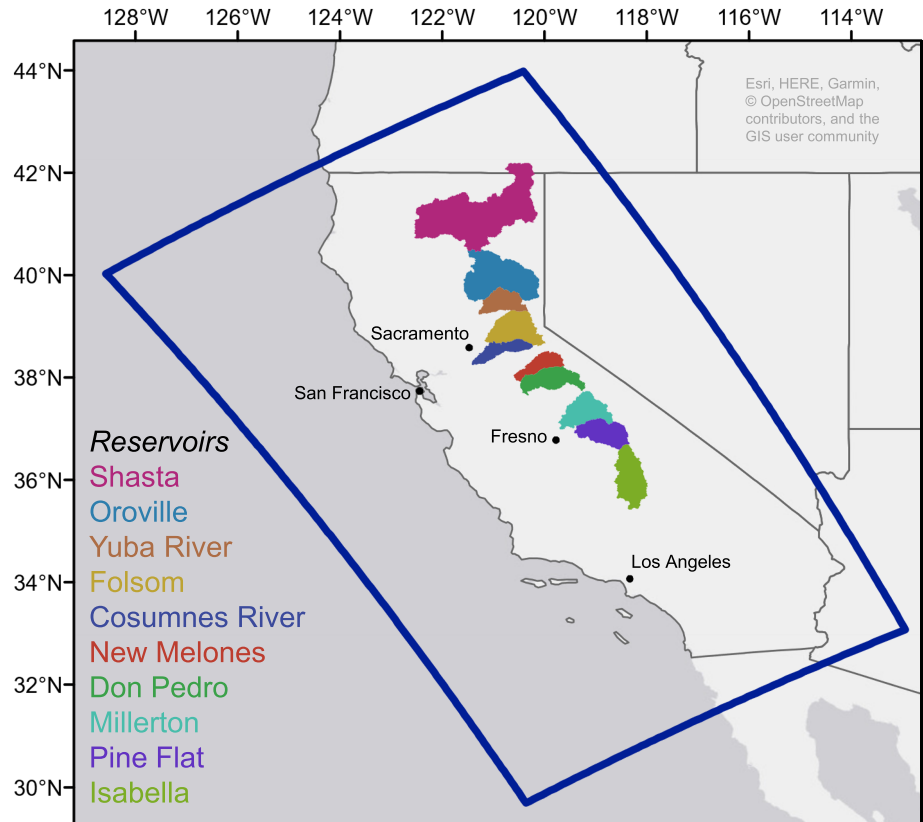
Figure 1. Watersheds of eight reservoirs and two stream gauges in California. Boundary of 9-km inner WRF domain shown as blue box. Major cities shown for reference.

meteorological conditions were taken from the North American Regional Reanalysis (NARR) at 32-km resolution (Mesinger et al., 2006). NARR was used as the boundary condition because of its high resolution compared to other available reanalyses. Daily sea surface temperature within the model domain was taken from the National Centers for Environmental Prediction Real Time Global analysis at 0.5° resolution (Thiébaux et al., 2003). The physics parametrizations chosen included the Kain-Fritch convection scheme, Thompson microphysics scheme and Noah-MP LSM, based on their demonstrated accuracy at simulating snow accumulation and ablation in the Sierra Nevada (Hughes et al., 2017; Wrzesien et al., 2015). We do not expect the choice of convective scheme to have a strong impact on snowfall accumulation in the Sierra Nevada, because most precipitation there occurs in the wet season of winter and early spring and is not convectively driven. Within Noah-MP, we used the original Noah runoff scheme with free drainage, which is the default scheme beginning in WRF v3.9.

In our setup, WRF outputs surface and subsurface runoff at 3-hr time steps for each grid cell. To evaluate the simulation results against observational data, we identified 10 locations at which suitable flow data are available, spanning the length of the Sierra Nevada. One of the locations is a U.S. Geological Survey gauge station on the Cosumnes River, which has no large dams. The other nine locations are served by California Data Exchange Center (CDEC) stations, for which daily FNF was available. FNF represents the discharge for a location without human modifications to the river system. It is derived from actual gauged discharge by appropriately adding or subtracting upstream diversions and storage changes in upstream reservoirs (Huang, 2016). Since there are numerous reservoirs upstream of the stations we chose, and WRF simulations do not include human modifications such as reservoir storage, we must compare to FNF and not to unadjusted gauged discharge. Eight of the CDEC stations we used are located at dams and directly measure the FNF upstream of reservoirs. We also used one CDEC station, Yuba River at Smartville, that is less than 1 km downstream of Englebright Lake. Table 1 provides more information on the 10 stations and their watersheds.

Table 1
Characteristics of the 10 Basins in the Study, Listed in Order From North to South, Along With the Effect of WRF/Noah-MP Modifications on Runoff Bias Over WY 2007

Station	Basin area (km^2)	Mean FNF (m^3/s)	Annual specific runoff (mm/yr)	Median elevation in WRF (m)	Default WRF runoff % bias	Modified WRF runoff % bias
Shasta	17,848	6,819	502	1,455	−5	−25
Oroville	9,043	5,192	754	1,609	28	−7
Yuba	3,397	2,686	1,039	1,317	59	30
Folsom	4,671	3,332	937	1,412	46	13
Cosumnes	1,369	430	413	857	53	30
New Melones	2,302	1,366	780	1,875	63	−2
Don Pedro	3,898	2,443	823	1,869	72	2
Millerton	4,121	2,121	676	2,268	110	3
Pine Flat	3,902	2,036	686	2,532	106	−6
Isabella	5,175	830	211	2,117	128	−47

We used the HydroSHEDS flow direction raster at 3-arc sec ((1/1,200)°) resolution (Lehner et al., 2008) to delineate each station's watershed (Figure 1), which we then transferred to the WRF inner domain grid. By summing the WRF subsurface and surface runoff output over each watershed and averaging, we estimate runoff on a daily time scale. We first ran WRF for one water year (WY), 1 October 2006 to 30 September 2007
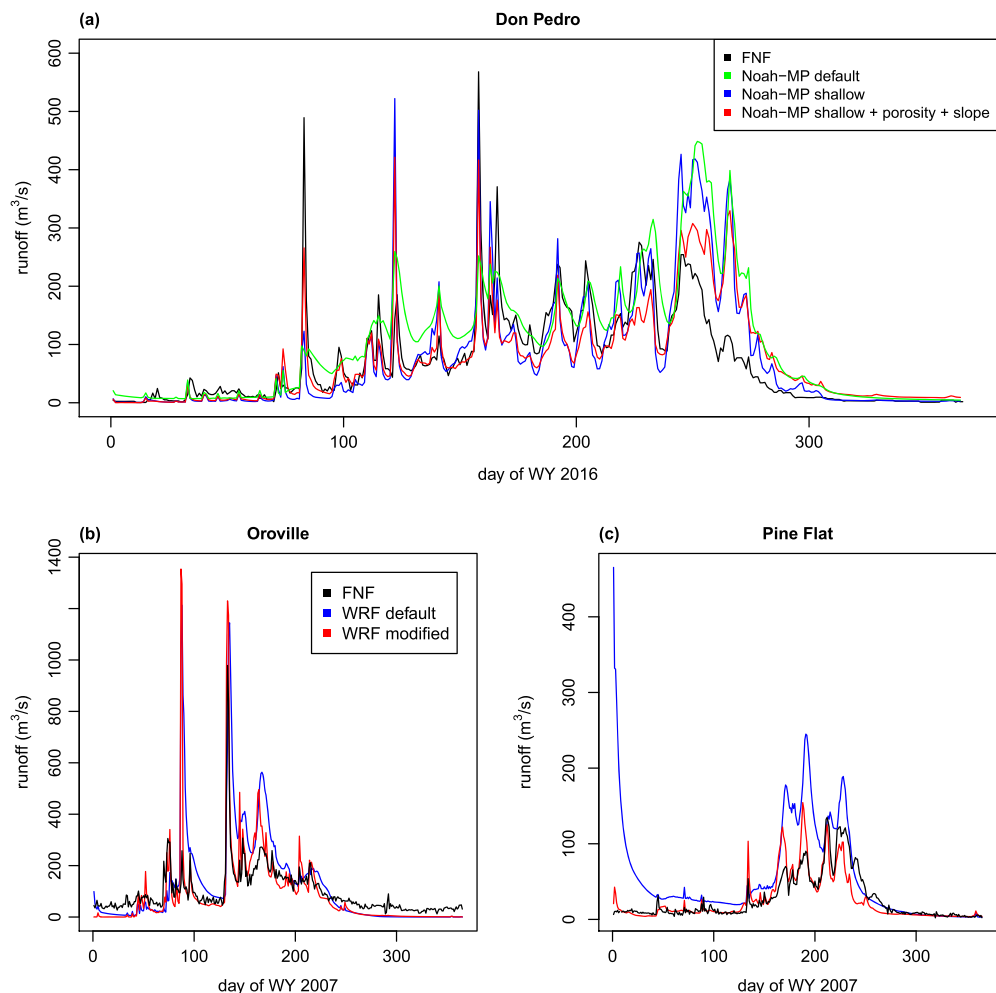
Figure 2. Modifications to Noah-MP improve WRF runoff accuracy. (a) Off-line Noah-MP-only simulation forced with WRF outputs. (b and c) Fully coupled WRF/Noah-MP.

Table 2
Summary of Modifications Made to WRF/Noah-MP System

Category	Default assumption	Modification
Snow capacitance in the Thompson microphysics scheme	Varied linearly from 0.15 at −1.5 °C to 0.5 at −30 °C	Constant value of 0.2
Soil type	Default soil texture map with large initial soil moisture values	Sand and ice grid cells changed to sandy loam
Soil porosity for sandy loam	0.434	0.52
Slope parameter to calculate subsurface flow	0.1	0.5
Depth of subsurface runoff generation	Deepest soil layer, from 1 to 2 m below the surface	Second-from-top soil layer, from 10 to 30 cm below the surface
Rain-snow partitioning	A function of 2-m air temperature (opt_snf = 1)	Allow the microphysics scheme to partition the precipitation (opt_snf = 4)

with the default values for all physics parameters. We compared the resulting model output to FNF values to assess accuracy (Table 1 and Figure 2). To improve snowfall and runoff accuracy, we made changes to the microphysics scheme and the runoff scheme, which are described in section 3.1. With the resulting modified setup, we ran WRF for WY 2007–2017. In lieu of a routing scheme, we imposed a 6-hr delay on runoff for each station; for all 10 stations in two WYs tested, 6 hr produced a Nash-Sutcliffe efficiency (NSE) within 0.05 of the maximum value tested (Nash & Sutcliffe, 1970), while longer or shorter delay times produced lower NSE values. NSE is an accuracy metric with values ranging from negative infinity for an inaccurate model to 1 for a perfect model. An NSE value greater than 0 indicates a model that is more accurate than simply outputting a constant value equal to the observed mean at all times. The WRF-Hydro modeling system could alternatively be used to implement more physically based flow routing, but we thought that its added complexity was not necessary given the size of the basins we used (all less than 20,000 km^2). We used several metrics (percent bias, correlation, and NSE) and visualizations of both runoff amount and runoff timing to assess simulated runoff accuracy relative to FNF (Table 3 and Figures 2–4). We tested whether simulated runoff accuracy in certain basins can be further improved by using a statistical correction that approximates the effects of groundwater storage. Finally, we studied the accuracy of using our WRF outputs as inputs to models of reservoir management, described in section 2.2. A flowchart summarizing our methods is shown in Figure S1 in the supporting information.

2.2. Water Resources Model

The newly developed ORCA model simulates operations of Shasta, Oroville, and Folsom reservoirs in the foothills of the Sierra Nevada at the edge of the Sacramento Valley, as well as management of water supply infrastructure in the Sacramento-San Joaquin Delta. Written in Python, the model is open source and can be found at the GitHub (https://github.com/jscohen4/orca). It uses daily hydrologic input data that can be obtained from the CDEC: reservoir inflows and FNF, basin-averaged SWE, temperature, and basin-averaged precipitation. The model uses monthly SWE values to forecast future reservoir inflows in the dry season, mimicking the seasonal streamflow forecasting strategy used in the California water system. Model results are generated at a daily time step. The primary results of interest for this study are reservoir releases and water supply pumping through the Sacramento-San Joaquin Delta ("Delta pumping").

ORCA simulates how major reservoirs in the Central Valley Project and State Water Project operate in conjunction with water supply exports south of the Delta, primarily for agricultural water supply. These exports are strongly influenced by available storage in Shasta, Oroville, and Folsom reservoirs, the forecasted seasonal inflow for each reservoir, and the ability to meet environmental requirements in the Delta. Reservoir releases are determined based on minimum environmental flow requirements, flood control, and water supply demand. ORCA can be used to explore flood control, water supply, and water system planning under many hydrologic scenarios.

In this study we evaluated ORCA simulations for WYs 2007–2018, comparing hydrologic inputs from WRF and CDEC (historical), including temperature, precipitation, and SWE. Because ORCA uses SWE data from CDEC-reported snow pillows or snow courses, which are at a much smaller spatial scale than the 9-km WRF
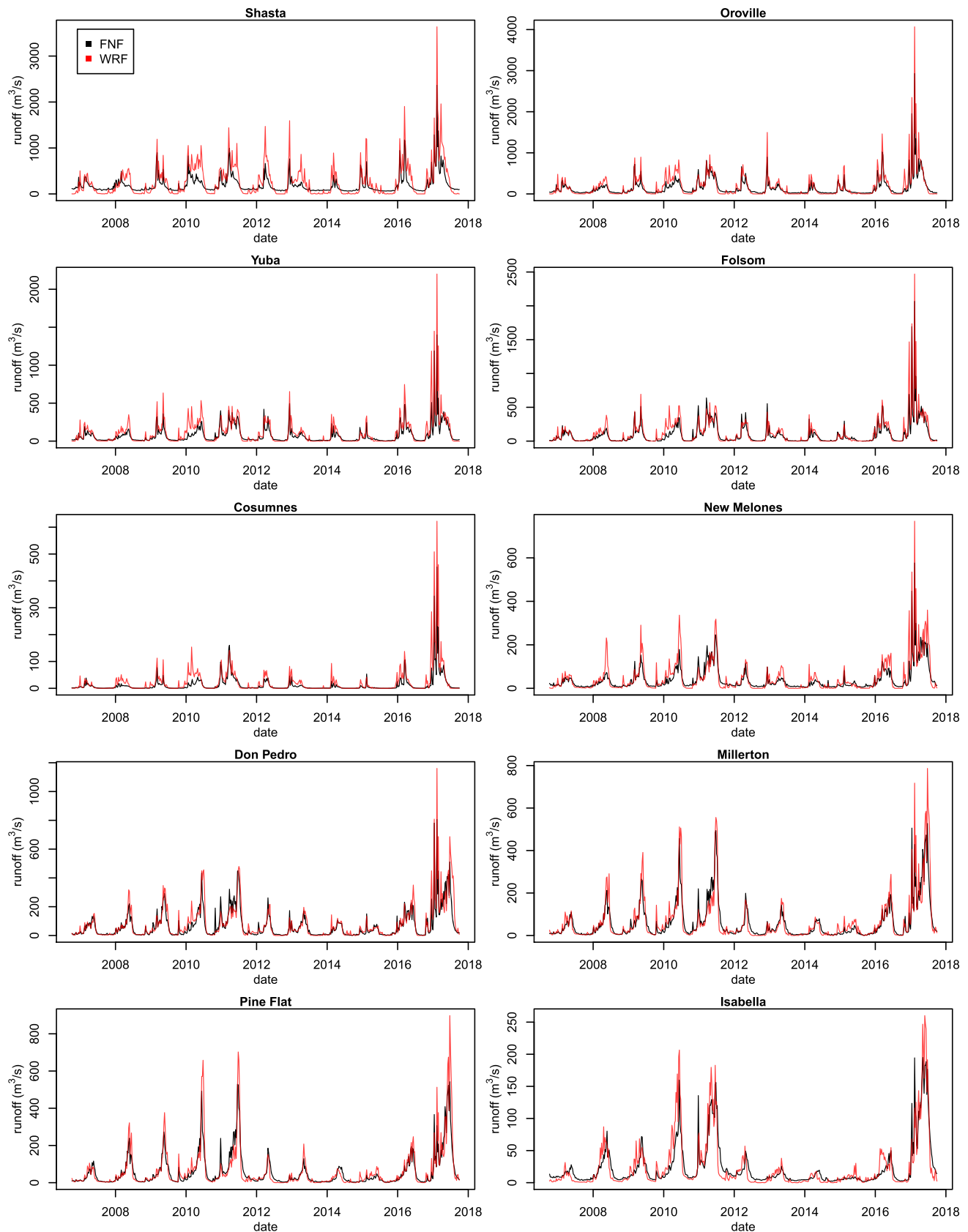
Figure 3. Weekly runoff from modified WRF compared to FNF.

Table 3
Metrics for Accuracy of WRF/Noah-MP Modeled Runoff, With and Without a Linear Correction That Imitates an Aquifer Model as Discussed in Section 3.2

	Without aquifer correction					With linear aquifer correction				
	Daily			Weekly			Daily			Weekly	
Station	% Bias	Correlation	NSE	Correlation	NSE	% Bias	Correlation	NSE	Correlation	NSE
Shasta	34	0.85	−0.37	0.91	−0.04	33	0.85	0.58	0.90	0.69
Oroville	22	0.89	0.34	0.96	0.68	21	0.89	0.67	0.96	0.87
Yuba	42	0.87	0.17	0.94	0.45	41	0.87	0.45	0.94	0.66
Folsom	25	0.87	0.57	0.95	0.79	24	0.87	0.62	0.95	0.82
Cosumnes	73	0.79	−0.12	0.94	0.42	73	0.80	−0.04	0.94	0.48
New Melones	22	0.86	0.39	0.92	0.57	20	0.87	0.70	0.92	0.81
Don Pedro	17	0.85	0.57	0.91	0.70	16	0.85	0.65	0.91	0.77
Millerton	10	0.90	0.67	0.92	0.73	8	0.90	0.76	0.92	0.81
Pine Flat	9	0.89	0.63	0.91	0.68	7	0.90	0.74	0.91	0.79
Isabella	2	0.89	0.70	0.92	0.77	−1	0.90	0.81	0.93	0.86

*Note.* Stations listed from north to south.

grid, the WRF SWE outputs were downscaled using the following procedure: We first identified the WRF grid cell that contains a snow station used by ORCA. Next, by linearly regressing maximum annual observed SWE and modeled SWE, we calculated an empirical factor that allows us to transform WRF grid cell SWE to a time series similar in magnitude to the station observation. By applying the empirical factor to the modeled daily SWE time series, we downscale it to the station level.

2.3. Evaluation Data Sets

We use two gridded data sets to evaluate WRF estimates of SWE and precipitation. For SWE, we compare against the Sierra Nevada Snow Reanalysis (SNSR) (Margulis et al., 2016). By constraining modeled snow depth and snow density by Landsat observations of fractional snow-covered area, the SNSR produces estimates of SWE at a spatial resolution of 90 m. As of the time of this study, SWE estimates of the Sierra Nevada mountain range are available from the SNSR from 1985 through 2016. To evaluate precipitation, we use the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al., 1994; PRISM Climate Group, 2018), which is derived by interpolating a combination of gauges and satellite precipitation products. Daily precipitation estimates from PRISM are available from 1981 to 2018; here we use the 4-km PRISM data to evaluate WRF estimates of precipitation.

3. Diagnostic Evaluation of WRF/Noah-MP

Using the default WRF/Noah-MP parameterizations, we observed large positive biases in runoff for nine of the 10 stations (Table 1). To assess whether snow accumulation errors contributed to this bias, we compared WRF snow on 1 April 2016 with the SNSR (a relatively average snow accumulation year where 1 April SWE is within <10% of peak SWE), finding a 20% bias in total grid-cell-wise peak SWE over our basins of interest. This bias agrees with the findings of Wrzesien et al. (2017), who also noted that WRF overestimates SWE in the Sierra Nevada compared to the best available data sets. The SWE bias was especially large in high elevation grid cells and those with large values of SWE, which are important for generating late-season runoff. Although WRF and Noah-MP represent many processes using physically based equations, parameterization based on empirical data is still common, such as for defining soil porosity for each soil class (Clapp & Hornberger, 1978). Past studies of snowpack and runoff using WRF and Noah-MP have maintained default parameter values (e.g., Rasmussen et al., 2014; Schwartz et al., 2017) when running the model to assess its accuracy. However, many of these parameters should vary in space and have significant spread in acceptable values. In order to address the bias observed in snowpack and runoff, we identified parameters in WRF and Noah-MP that controlled snowfall and runoff processes: snow capacitance in the Thompson microphysics scheme, the matchup between initial soil moisture values from NARR and soil porosity, the soil porosity, the slope value used to calculate subsurface flow, and the depth at which runoff is generated in the subsurface. We also consider how rain-snow partitioning impacts snowfall and snow accumulation. All changes to WRF/Noah-MP are summarized in Table 2.
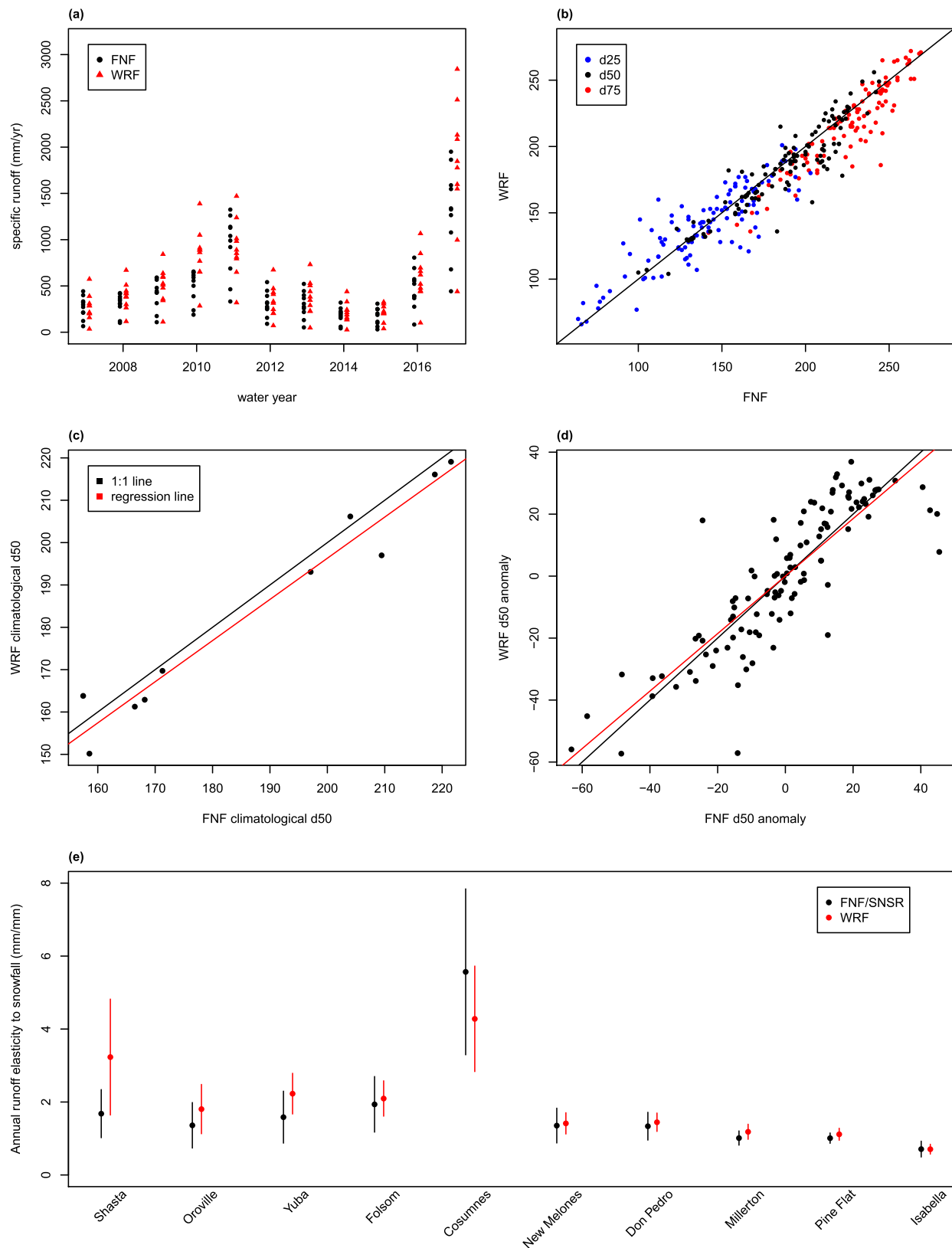
Figure 4. (a) annual total specific runoff in WRF compared to FNF for 10 stations, (b) WRF runoff timing compared to FNF for each lake and WY, (c) WRF runoff timing compared to FNF averaged over 11 WY for each lake, (d) anomaly of (b) minus (c), and (e) WRF runoff elasticity to snowfall compared to FNF and SNSR with 2-standard-error intervals from regression

According to the default option in Noah-MP, precipitation is partitioned between rain and snow through a function of 2-m air temperature that does not vary spatially. However, the optimal temperature threshold to perform this partitioning has been found to vary by region (Jennings et al., 2018). Thus, we tested allowing the microphysics scheme from WRF to partition the precipitation, which is referred to as "opt_snf = 4" in the WRF namelist. This change reduced the peak SWE bias in 2016 compared to SNSR from 20% to 16%. To further reduce the snow accumulation bias, we altered the value of snow capacitance within the Thompson microphysics scheme (Thompson et al., 2008). Snow capacitance is a dimensionless value associated with a snowflake's shape that describes how efficiently the snowflake can attract water vapor. In the original parameterization, the snow capacitance varied linearly with temperature from 0.15 at −1.5 °C to 0.5, the value for a sphere, at −30 °C. Based on the assumption that heavily rimed snow or graupel crystals are not perfect spheres, we instead used a constant capacitance of 0.2, which is a typical value for aggregates of snow crystals (Westbrook et al., 2008). Changing the snow capacitance reduced 2016 SWE bias over our basins of interest to 11%.

Even with these changes to snowpack, positive runoff biases remained, as shown in Table 1 and Figure 2. Some of this bias occurs at the beginning of the WY in our evaluation runs (e.g., Figure 2c). A small number of grid cells produced extremely high runoff at the beginning of the simulation, prior to any precipitation. These anomalous cells had a soil type of either sand or ice. In these cases, the initial soil moisture from NARR was too large for the assigned soil type to hold. To remedy this issue, we changed the soil type of all cells that were sand, loamy sand, or ice to sandy loam, as this was the most common soil type (74% of grid cells in the basins of interest were sandy loam or loam). To further increase the runoff accuracy, we made changes to the subsurface runoff scheme. We did not change the surface (infiltration excess) runoff scheme because in our area of interest it generates relatively little runoff compared to the subsurface. With the free drainage option (opt_run = 3), Noah-MP calculates subsurface runoff as a product of soil hydraulic conductivity and a slope parameter (Skamarock et al., 2008). However, when Noah-MP is coupled with WRF v3.8.1, a single value of the slope parameter is used for the entire domain. We chose 0.5 as the slope parameter, instead of the default value of 0.1 (which is simply the first value in the list of possible values provided with Noah-MP), as 0.5 gave better results in an off-line version of Noah-MP (e.g., NSE of 0.59 instead of 0.49 for WY 2016 in Don Pedro Lake, comparing versions with modified runoff depth and porosity as discussed below, only differing in the slope parameter).

By default, subsurface runoff in Noah-MP is generated at the lowest soil layer, which extends from 1 to 2 m deep. At higher elevations in the Sierra Nevada, the depth to weathered bedrock is often less than 2 m, so modeled runoff should come from closer to the surface (Dahlgren et al., 1997; Meyer et al., 2007). A recent high-resolution study of a catchment in the Alps found that simulated snowmelt runoff accuracy was maximized when the Alpine3D model generated subsurface runoff at a soil depth of 30 cm (Wever et al., 2017). We decided to test implementing a similar scheme in Noah-MP; although soils in the Sierra Nevada are certainly not the same as those in the Alps, they are likely similarly shallow in many places. We modified the Noah-MP runoff scheme to produce subsurface runoff from the second-from-top soil layer, which extends from 10 to 30 cm. This change lowered melt-season runoff and produced hydrographs with both peak runoff values and recessional limbs closer to the FNF (Figure 2).

In Noah-MP, soil permeability is proportional to porosity and to a power function of soil moisture content. The porosity of each soil type in Noah-MP is assigned a single value, which is determined by the mean soil type value from the range identified by Clapp and Hornberger (1978). However, real soils exhibit a wide range of porosities (and other hydraulic characteristics) within each soil type, as indicated by the standard deviations that accompany the mean values. To reduce runoff, we increased the porosity of the predominant soil type, sandy loam, by 1 standard deviation from 0.434 to 0.52. Figure 2a shows the effects on runoff into Don Pedro Lake of the major changes we made to Noah-MP. In order to allow efficient testing of different parameterizations, we used an off-line version of Noah-MP forced with WRF atmospheric outputs for WY 2016, an average WY. The NSEs relative to FNF are 0.19, 0.37, and 0.59 for Noah-MP with default settings, Noah-MP with shallow runoff generation, and the fully modified Noah-MP (shallow runoff generation, higher slope parameter, and higher porosity), respectively. The NSE for the same station and year in the fully modified version of WRF coupled with Noah-MP is 0.66, which demonstrates the advantage of coupling the atmospheric and LSMs. For all stations except Shasta, bias decreased by between 21 and more than 100 percentage points from the default to the fully modified version of Noah-MP within WRF (Table 1) for WY 2007,

the first year of our planned multiyear run. Thus, the changes we implemented make WRF/Noah-MP runoff much more accurate in the Sierra Nevada region.

4. Results

4.1. Modified WRF/Noah-MP Model Accuracy

Comparison of simulated runoff time series from our modified WRF/Noah-MP against FNF for 10 stations over WY 2007–2017, shown in Figure 3, suggests that the runoff simulations have considerable skill. An interactive visualization of the same data is available at the website (https://nholtzman.shinyapps.io/pre-sapp_v2). Model accuracy compared to FNF is summarized in the left half of Table 3. The runoff bias generally decreases from north to south, which may be related to the increasing proportion of precipitation that falls as snow in the higher-elevation southern regions of the Sierra Nevada. Cosumnes, the station where our model is by far the least accurate, has both the smallest basin area and the lowest basin elevation (Table 1). To further improve performance in northern and lower-elevation parts of the Sierra Nevada, it may be necessary to make changes to WRF affecting rainfall, not just snowfall as we did in this study. Our revised Noah-MP subsurface runoff model does not model groundwater, a weakness that impairs the modeled runoff accuracy, particularly in the northern three reservoirs where FNF has higher baseflow in the dry season than WRF. Furthermore, in the southern Sierra Nevada, WRF runoff recedes faster than FNF; both the absence of groundwater and the possibility of modeled snow melting too quickly may contribute to this difference.

To assess the accuracy of modeled runoff timing without regard to absolute runoff amount, we calculated the days at each station when 25%, 50%, and 75% of each WY's runoff had occurred (Figures 4b–4d), which we refer to as d25, d50, and d75, respectively. The climatological 11-year averages of d50 are similar, with a root-mean-square error of 6.0 days, indicating that WRF realistically captures differences between watersheds in overall runoff timing. Subtracting the climatological d50 value from the lake-year data gives the d50 anomaly for each lake year; for the d50 anomaly values, there is a correlation between WRF and FNF with a slope that is not significantly different from 1 (0.93 with a standard error of 0.05), suggesting that WRF models interannual variability in runoff timing well. We can repeat the analysis for runoff centroid date, or the day of the WY with the center of mass of streamflow (Stewart et al., 2004). WRF produces earlier centroids than FNF by 5.7 days on average, while WRF d50 dates are 3.3 days earlier that the corresponding days obtained from FNF. Analysis of d25 shows similar patterns to d50. However, d75 shows more bias, with WRF dates 8.5 days earlier than FNF. This bias is related to WRF flow levels receding too quickly, which may be related to the absence of groundwater storage in the model or to modeled snow melting too quickly. The number of days it takes for SWE aggregated over all 10 basins to melt to half of its maximum value in each year is on average 8.2 days earlier in WRF than in SNSR, with a standard deviation of 22 days (averaged over WYs 2007 to 2016). Thus, inaccurate snowmelt rate might play a role in runoff timing bias for particular years, but there is no clear overall snowmelt rate bias.

In addition to runoff timing, we also compare annual runoff amounts between WRF and FNF for all stations (Figure 4a), with varying results. In some WYs (2010 and 2017), WRF has a clear positive bias in all reservoirs, while in other years WRF is much more accurate. With only eleven years of simulations, it is difficult to discern a pattern in these annual differences. Interestingly, WYs 2010 and 2011 have nearly the same WRF runoff amounts, and the same WRF rain and snow, but very different FNF runoff amounts. Compared to observation-based data, WRF produces too much precipitation in WY 2010. WRF snow accumulation biases relative to SNSR are 19% and −5% in WY 2010 and 2011, respectively, while WRF total precipitation biases relative to the PRISM data set are 33% and −4%. WY 2010 was an especially dry one in California, and it is possible that the NARR forcing did not capture this dry extreme well. We also compare WRF outputs to a combination of FNF and SNSR to assess runoff elasticity to snowfall (Figure 4e). For each basin, we fit a linear relationship between annual runoff and annual SWE accumulation, using the 10 WYs, 2007–2016. WY 2017 was not used because it is not included in SNSR. We define the elasticity as the slope of this relationship. There is good agreement between the elasticities found in WRF and those derived from FNF and SNSR.

The Airborne Snow Observatory (ASO) is another source of remotely sensed snow data over part of the Sierra (Painter et al., 2016). Most of the currently available ASO data are from the Tuolumne basin, over which our WRF output shows a positive bias on the order of 50% compared to ASO SWE (Figure S2).

However, we believe that this basin is not representative of WRF performance over the whole Sierra; the SWE bias is negative in the majority of other years for the other ASO basins. SNSR SWE agrees well with ASO in the year 2016, showing a correlation of 0.99 and a −5.6% bias. So we have high confidence in using SNSR to evaluate WRF over basins and years where ASO data is not available. Figure S3 shows that the bias of WRF SWE relative to SNSR is much closer to 0 when aggregated over all 10 basins we studied, compared to when it is calculated over just the drainage basin of Lake Don Pedro (which contains the portion of the Tuolumne River Basin observed using ASO).

4.2. Adding a Simple Groundwater Model

One possible reason for the error characteristics described in section 3.1 is that WRF/Noah-MP does not adequately model groundwater storage and release. Including such a capability could increase the runoff accuracy of our modeling system for all parts of the Sierra Nevada. Although the high-mountain areas that experience the most snowfall may have little potential for groundwater storage, the stations studied here are all below 400 m in elevation, so it is reasonable to expect significant groundwater storage upstream of the stations. Though Noah-MP includes an optional water table parametrization, this scheme may take many model years to spin up (Niu et al., 2007). To avoid this issue, we implemented a more conceptually simple groundwater model. Based on schemes found in the Regional Hydro-Ecologic Simulation System and other modeling systems (Tague & Band, 2004), our parametrization models aquifer storage volume but does not explicitly model a water table.

In each time step of our scheme, a fraction b of infiltrating water bypasses the soil column and enters the aquifer; the aquifer then discharges a small fraction k of its volume. Details of the implementation are described in supporting information section S1. For WY 2007 there were four stations (Oroville, Yuba, New Melones, and Isabella) where the fully coupled WRF/Noah-MP model with aquifer parametrization improved on the soil-only configuration by simulating Fall 2007 FNF baseflow with less than 25% bias, while at these stations the soil-only version simulated very low baseflow with a negative bias of over 62% (Figure 5). In one station (Shasta), the aquifer configuration still did not produce enough baseflow; b and k may need to be larger to accurately represent the aquifers in this basin. For the remaining five stations, baseflow from the soil-only WRF simulation was similar to FNF, and adding the aquifer produced too much baseflow. Decreasing b and k in these basins, representing less influence of from groundwater, might improve accuracy. Adding the aquifer model did not have much effect on the incorrectly rapid recession events in the southern Sierra Nevada, which may stem from errors with snow melt rate. Thus, a simple groundwater model shows promise for simulating persistent baseflow, but its parameters need to be set on a regional basis, perhaps based on bedrock types. The northern Sierra Nevada is underlain by volcanic rocks that are more permeable than the granite in the southern portion of the range, so it would make sense to assign the northern Sierra Nevada a larger b value. However, we leave the detailed parameterization of such a model to future efforts.

For each station, the modeled runoff with the aquifer model was approximately a linear function of the modeled runoff without the aquifer model. The regression slopes were the same (0.83) for all 10 stations, while the y intercepts were different, presumably due to the different initial aquifer value in each grid cell. After accounting for the differing intercepts, the overall correlation of daily modeled station runoff with and without the aquifer was 0.9987. This strong linear relationship suggests that the addition of the aquifer model can be approximated statistically without rerunning WRF. The advantage of this approach is that the parameters of the relationship for each station can be easily estimated by comparing to the station FNF. Estimating the parameters of the aquifer model in Noah-MP would be more difficult: While the spatial variation might generally correspond with rock types as discussed above, the parameters are still conceptual. They do not clearly correspond to physical quantities measurable in nature, and actual values would have to be empirically determined by comparing to FNF. Such a calibration for an off-line Noah-MP implementation is much more computationally expensive than calibrating the parameters of a linear relationship as a postprocessing step. Thus, we chose to approximate the effect of the aquifer model for the full 11-year run by applying a simple linear relationship.

To account for long-term memory in the aquifer, we let the intercept of this relationship vary by introducing a term proportional to the previous year's mean runoff. For the first year of the simulation, the 11-year runoff
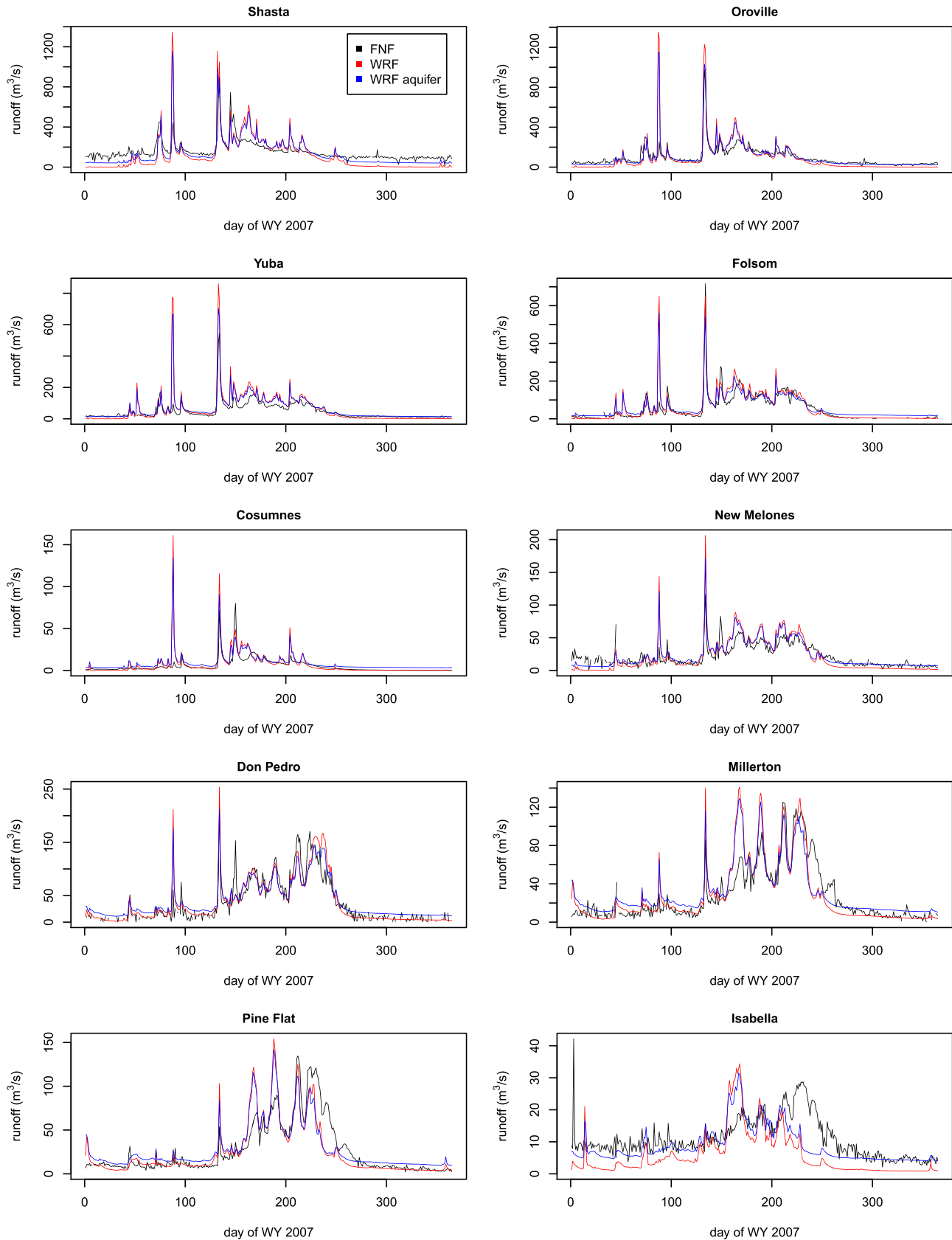
Figure 5. Results of fully coupled WRF/Noah-MP for WY 2007 with aquifer model.

Table 4
Parameters of Aquifer Correction, Derived From Linear Regression

Station	c	d	1-c-d
Shasta	0.59	0.09	0.32
Oroville	0.79	0.05	0.15
Yuba	0.86	0.08	0.07
Folsom	0.95	0.05	0.00
Cosumnes	0.97	0.03	0.00
New Melones	0.73	0.14	0.13
Don Pedro	0.90	0.00	0.10
Millerton	0.87	0.11	0.01
Pine Flat	0.86	0.12	0.03
Isabella	0.73	0.24	0.04

climatology from the WRF simulation is used as the previous year. This term conceptually represents the elasticity of annual minimum flow to annual mean flow. The full linear relationship is of the form

$$q_{aq} = cq + d\overline{q_{365}} + (1-c-d)\overline{q} \qquad (1)$$

In equation (1), q_{aq} is daily runoff after the linear transformation has been applied, q is the daily modeled runoff without the aquifer model, $\overline{q_{365}}$ is the mean q over the previous 365 days, $\overline{q}$ is the climatological mean modeled runoff, and c and d are parameters to be estimated for each lake. The right-hand side of equation (1) is constructed so that the overall mean of the modeled runoff is unaffected by the transformation. We estimated the c and d parameters through linear regression by comparing with daily FNF for each station. Details about the regression are described in supporting information section S2. The resulting parameters are shown in Table 4.

For all 10 stations, the aquifer correction does not dramatically change the correlation between modeled runoff and FNF. For the stations with c parameters nearly equal to 1, the correction has only a small effect on modeled flow and accuracy. For stations with lower c parameters, the correction uniformly improves the NSE (Table 3), though the original correlations were already high (≥ 0.9). Figures 6a and 6b show the effect of the aquifer correction on modeled runoff into Lakes Shasta and Isabella, the most northern and southern lakes in our study: The low flows increase, and the high flows decrease. For most years in Lake Shasta and some years in Lake Isabella, the postcorrection low flows are higher than the FNF low flows because overall WRF runoff for those years has a high bias already. We can understand the elasticity of low flow to total flow by studying the annual minimum runoff (approximated as the minimum weekly mean runoff in the second half of the WY) as a function of the annual mean runoff, normalized by the climatological annual mean runoff. From this perspective, the aquifer-corrected WRF output agrees with FNF, while the uncorrected WRF output is too low for the lakes where the correction is important (Figures 6c and 6d).

4.3. A Water Resources Model Forced With WRF Outputs

When outputs from WRF, including the linear aquifer correction, are used as inputs for the ORCA water resources model, the results are mixed. Figure 7 compares model output from simulations forced with WRF (blue), CDEC FNF (yellow), and observed (black). Discrepancies between the CDEC simulations and observed data can be attributed to errors in the water resources model, while discrepancies between the CDEC and WRF simulations can be attributed to errors in WRF. The primary errors of interest are in simulated streamflow, but also in SWE, which is used in the ORCA model as an input to seasonal forecasts for reservoir operations. In general, NSEs are lower for all ORCA simulations when WRF is the input (Figure 7). This comparison enables us to use ORCA as a diagnostic tool to pinpoint how relatively small errors in WRF might propagate to produce large errors in reservoir management, with implications for water supply, flood control, and environmental benefits.

We first focus on results for Shasta Reservoir, where the greatest model differences occur in simulation of operations during the 2012–2016 drought (Figure 7). This discrepancy can be explained by two factors in the WRF output. First, WRF overestimates streamflow into Lake Shasta during the winter season (Figure 3). However, SWE values during this period are relatively accurate compared to SNSR (bias = 12.0 mm, RMSE = 29.1 mm), especially for WYs 2014 and 2015. Low snowpack leads to a dry seasonal forecast in the ORCA model, so the reservoir is operated to maintain high storage levels by curtailing water releases despite the biased inflow. Thus, excess inflows cause a high bias in Lake Shasta storage during this time period. Oroville Reservoir shows similar issues for 2008–2011 (Figure 7). For WY 2011, WRF snowpack is less than SNSR (bias = −31.8 mm, RMSE = 56.0 mm), causing ORCA to misclassify the WY as dry. This leads to unnecessary curtailments in the model followed by unexpected higher inflows.

In ORCA simulations driven by WRF/Noah-MP, water supply pumping through the Sacramento-San Joaquin Delta is biased low during the wetter period of 2009–2012 (Figure 7). Over this period, WRF-estimated SWE in the Oroville basin is low compared to SNSR, causing unnecessary curtailments of Oroville releases. Agricultural water supply would be the first objective to be lost because of this
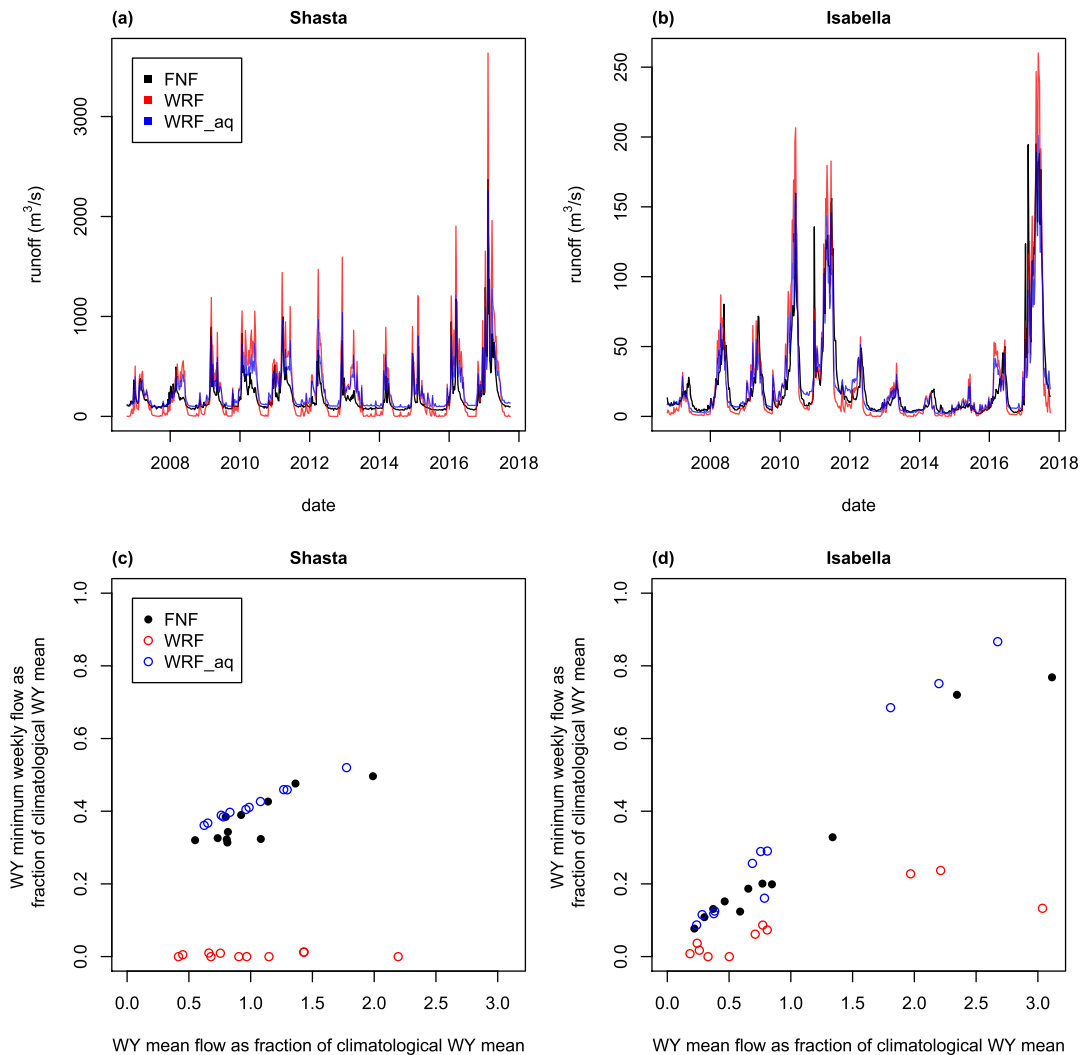
Figure 6. Effect of linear correction imitating aquifer (labeled as WRF_aq) on weekly hydrograph (a, b) and on relationship between annual mean and annual minimum flows (c, d).

curtailment, thus creating more restrictions on pumping. The biased inflows, along with misrepresentation of pumping requirements based on WY type and reservoir inflow forecasts, lead to reduced Delta pumping during this period.

Though there are differences between the simulations forced by WRF and CDEC, the WRF results are promising. Despite differences in scale of many orders of magnitude (9-km WRF grid cells versus point CDEC observations), the WRF/ORCA simulation captures seasonality for both Shasta and Oroville storage and for Delta pumping. WRF/ORCA does overestimate storage compared to historical operations for some time periods, but for other points during the study period, such as 2010 through 2013 for Shasta and 2013 and 2016 for Oroville, the WRF/ORCA simulation demonstrates skill in reproducing historical observations, indicated by NSE values well above 0. This suggests that WRF estimates can then be used to project how future climate conditions may impact water management for the region.

In addition to predicting reservoir inflows and Delta pumping, combining WRF and ORCA allows us to study how errors propagate between the two models. Errors in the WRF-forced ORCA model serve as an opportunity to understand how small errors in hydrologic and atmospheric model output may unexpectedly lead to greater errors in water supply planning. The techniques in this study can help future work formally categorize, analyze, and understand error propagation between hydrologic and planning models.
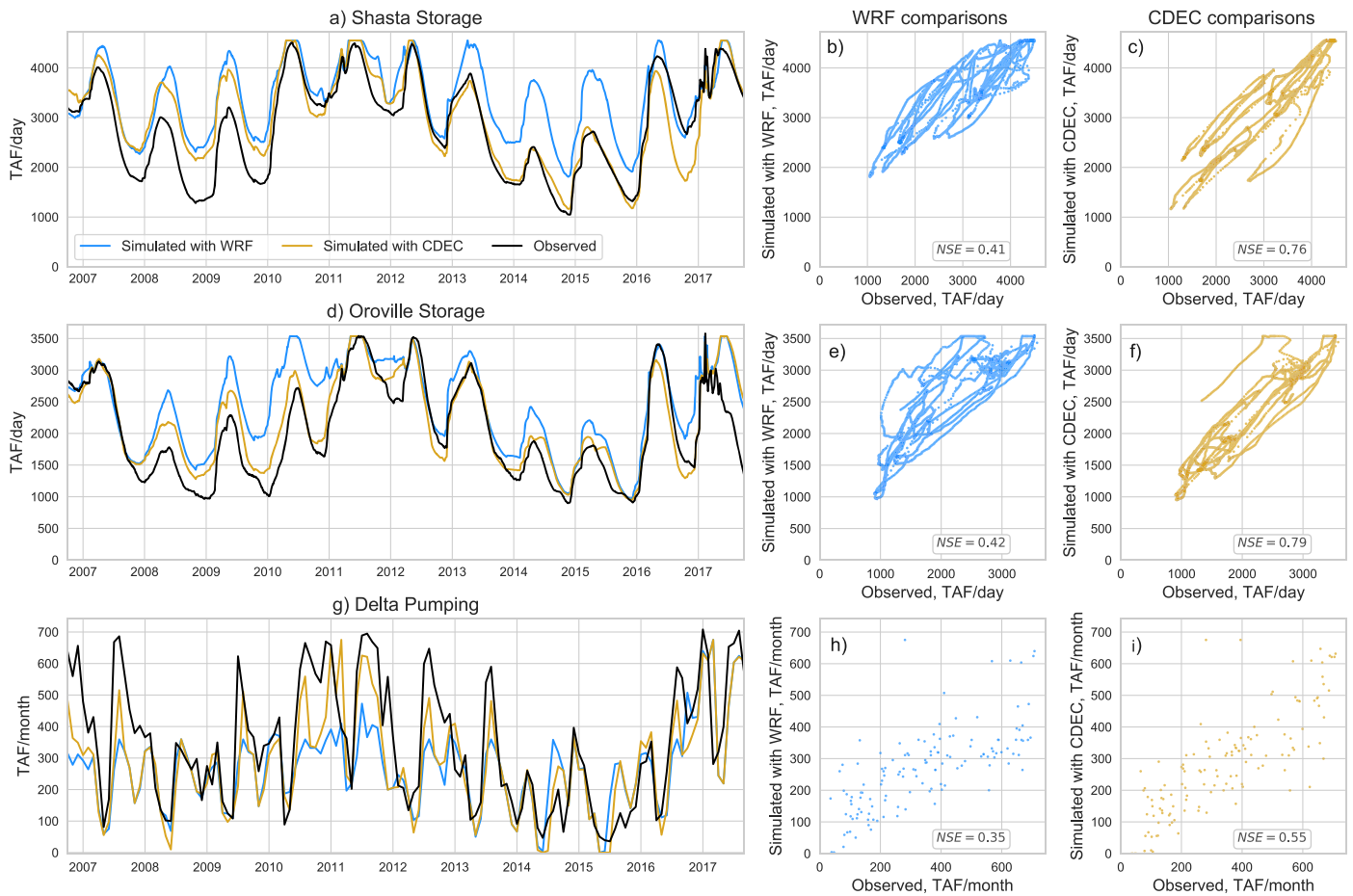
Figure 7. Outputs of WRF-forced and CDEC-forced ORCA simulations compared to historical operations for Shasta storage (a-c), Oroville storage (d-f), and Delta pumping exports (g-i). Subplots show both time series representations of outputs and scatterplots of each simulation versus historical time step value plotted against each other to visualize model output accuracies. Note that the storage outputs are shown on a daily time step, while the pumping outputs are averaged to a monthly time step.

5. Discussion and Conclusion

The primary outcome of this work is that the WRF/Noah-MP modeling system can accurately simulate runoff into Sierra Nevada reservoirs when it is suitably parameterized; we also introduce a simple groundwater model to improve baseflow simulation. We have shown that a spatial resolution of 9 km is adequate in the Sierra Nevada for modeling daily runoff into individual reservoirs. There has been little previous work using RCMs to simulate runoff at fine temporal and spatial scales. Past work has often relied on bias correction of model outputs, which can distort the modeled effects of climate change (Liersch et al., 2018). Instead, we sought to alter process representations within the model itself, which may make the model more suitable for simulating future changes in runoff. The runoff parametrization changes we made (related to soil depth and porosity) do not necessarily apply directly to other mountain regions, as runoff processes may be influenced by spatially variable soil and geologic properties. It is possible that, in the future, suitable parameterizations could be developed based on high-resolution soil and bedrock geology maps. The optimal precipitation parametrization may also vary by region for complex reasons related to local atmospheric dynamics. However, our findings suggest that, in mountainous regions, default parameterizations may require modification in order to suitably simulate runoff generation processes.

The default parametrizations within a LSM are not necessarily the most appropriate for simulating runoff in mountain areas. Using the default settings of Noah-MP, runoff in the Sierra Nevada was biased high, and variability was underestimated, so we modified parametrizations that describe soil hydraulic properties

and effective soil depth. The parameterizations of snow formation and accumulation on the ground also produced high runoff bias in the default configuration. Our changes made model runoff more accurate, but further work is necessary to determine whether these modifications fully match the intermediate physical processes they represent. For example, though it is unlikely that runoff generation in the Sierra Nevada occurs at 1- to 2-m depth, as in the default Noah-MP parameterization, it is unclear if approximately 30-cm depth in the soil column is the most physically accurate parameterization. Furthermore, more detailed experiments are needed to disentangle errors in modeled precipitation from errors in modeled runoff generation. Due to limited computational resources, we did not systematically test each change we made to WRF/Noah-MP in isolation and in each possible combination.

For rivers with significant baseflow, modeling runoff generation in the soil alone is not sufficiently accurate; some representation of an aquifer is needed. We have implemented a simple aquifer model into Noah-MP, and we present a simple statistical correction that can be applied as a postprocessing step to runoff generated from a model without an aquifer. However, the simple conceptual model may not accurately replicate the long-term multiyear dynamics of low flows. More research is needed on developing aquifer models that can represent these dynamics accurately, can be calibrated easily, and can be integrated in a climate model without requiring long model spin up. Schlemmer et al. (2018) have presented such a model, although its runoff output has not been extensively validated. Their model considers subgrid orography, an important influence on subsurface runoff that we neglected here. The WRF-Hydro model can also be used to incorporate groundwater, overland flow, and streamflow routing into WRF, with promising results in mountain regions (Arnault et al., 2019; Rummler et al., 2019). The Terrestrial Systems Modeling Platform is another example of integrating groundwater into an RCM/LSM (Kollet et al., 2018; Shrestha et al., 2018). One more approach to improving hydrologic predictions from an RCM is to take a distributed hydrologic model that has already been calibrated for the region of interest and couple that to an RCM (Larsen, Christensen, et al., 2016; Larsen, Refsgaard, et al., 2016).

When simulated runoff from our modified WRF/Noah-MP system is used as an input to the ORCA water resources model, simulated planning objectives including reservoir storage and pumping in the Sacramento/San Joaquin Delta are reasonably accurate for most periods. Seasonality is generally simulated correctly, and NSE values are well above 0, indicating considerable skill. Periods where WRF-forced ORCA output varies significantly from observation-forced output likely result from snowpack and runoff bias in WRF compared to observations. Further refinement of LSMs is needed to produce runoff outputs that are fully compatible with water resource models. However, the research we described here is a first step in that direction. To advance on previous efforts to dynamically downscale runoff in the Sierra Nevada (Pagán et al., 2016; Schwartz et al., 2017), our work explicitly validates daily runoff magnitude, and thus is better suited to interfacing with water resources models. As such, we have presented a more extensive validation of runoff from a RCM, which in future work could be used to downscale GCM projections with more confidence.

References

Arnault, J., Wei, J., Rummler, T., Fersch, B., Zhang, Z., Jung, G., et al. (2019). A joint soil-vegetation-atmospheric water tagging procedure with WRF-Hydro: Implementation and application to the case of precipitation partitioning in the upper Danube River basin. *Water Resources Research*, *55*, 6217–6243. https://doi.org/10.1029/2019WR024780

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., et al. (2008). Human-induced changes in the hydrology of the western United States. *Science*, *319*(5866), 1080–1083. https://doi.org/10.1126/science.1152538

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., & Schneider, D. (2016). Snowmelt rate dictates streamflow. *Geophysical Research Letters*, *43*, 8006–8016. https://doi.org/10.1002/2016GL069690

Berghuijs, W. R., Woods, R. A., & Hrachowitz, M. (2014). A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, *4*(7), 583–586. https://doi.org/10.1038/nclimate2246

Christensen, N. S., & Lettenmaier, D. P. (2007). A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin. *Hydrology and Earth System Sciences*, *11*(4), 1417–1434. https://doi.org/10.5194/hess-11-1417-2007

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water Resources Research*, *14*(4), 601–604. https://doi.org/10.1029/WR014i004p00601

Dahlgren, R. A., Boettinger, J. L., Huntington, G. L., & Amundson, R. G. (1997). Soil development along an elevational transect in the western Sierra Nevada, California. *Geoderma*, *78*(3–4), 207–236. https://doi.org/10.1016/S0016-7061(97)00034-7

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Fritze, H., Stewart, I. T., & Pebesma, E. (2011). Shifts in Western North American snowmelt runoff regimes for the recent warm decades. *Journal of Hydrometeorology*, *12*(5), 989–1006. https://doi.org/10.1175/2011JHM1360.1

Harding, B. L., Wood, A. W., & Prairie, J. R. (2012). The implications of climate change scenario selection for future streamflow projection in the Upper Colorado River Basin. *Hydrology and Earth System Sciences*, *16*(11), 3989–4007. https://doi.org/10.5194/hess-16-3989-2012

Huang, G. (2016, March). Estimates of natural and unimpaired flows for the Central Valley of California: Water years 1922–2014. California Department of Water Resources.

Hughes, M., Lundquist, J. D., & Henn, B. (2017). Dynamical downscaling improves upon gridded precipitation products in the Sierra Nevada, California, Climate Dynamics. https://doi.org/10.1007/s00382-017-3631-z

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1148. https://doi.org/10.1038/s41467-018-03629-7

Kampf, S. K., & Lefsky, M. A. (2016). Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. *Water Resources Research*, *52*, 407–422. https://doi.org/10.1002/2015WR017784

Knowles, N., Dettinger, M. D., & Cayan, D. R. (2006). Trends in snowfall versus rainfall in the western United States. *Journal of Climate*, *19*(18), 4545–4559. https://doi.org/10.1175/JCLI3850.1

Kollet, S., Gasper, F., Brdar, S., Goergen, K., Hendricks-Franssen, H. J., Keune, J., et al. (2018). Introduction of an experimental terrestrial forecasting/monitoring system at regional to continental scales based on the Terrestrial Systems Modeling Platform (v1.1.0). *Water*, *10*(11), 1697. https://doi.org/10.3390/w10111697

Larsen, M. A. D., Christensen, J. H., Drews, M., Butts, M. B., & Refsgaard, J. C. (2016). Local control on precipitation in a fully coupled climate-hydrology model. *Scientific Reports*, *6*(1), 1–9. https://doi.org/10.1038/srep22927

Larsen, M. A. D., Refsgaard, J. C., Jensen, K. H., Butts, M. B., Stisen, S., & Mollerup, M. (2016). Calibration of a distributed hydrology and land surface model using energy flux measurements. *Agricultural and Forest Meteorology*, *217*, 74–88. https://doi.org/10.1016/j.agrformet.2015.11.012

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, *89*(10), 93. https://doi.org/10.1029/2008EO100001

Liersch, S., Tecklenburg, J., Rust, H., Dobler, A., Fischer, M., Kruschke, T., et al. (2018). Are we using the right fuel to drive hydrological models? A climate impact study in the Upper Blue Nile. *Hydrology and Earth System Sciences*, *22*(4), 2163–2185. https://doi.org/10.5194/hess-22-2163-2018

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. https://doi.org/10.1175/JHM-D-15-0177.1

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. https://doi.org/10.1175/BAMS-87-3-343

Meyer, M. D., North, M. P., Gray, A. N., & Zald, H. S. J. (2007). Influence of soil thickness on stand characteristics in a Sierra Nevada mixed-conifer forest. *Plant and Soil*, *294*(1–2), 113–123. https://doi.org/10.1007/s11104-007-9235-3

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219. https://doi.org/10.1038/nclimate3225

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models Part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Naz, B. S., Kao, S.-C., Ashfaq, M., Rastogi, D., Mei, R., & Bowling, L. C. (2016). Regional hydrologic response to climate change in the conterminous United States using high-resolution hydroclimate simulations. *Global and Planetary Change*, *143*, 100–117. https://doi.org/10.1016/j.gloplacha.2016.06.003

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research*, *112*, D07103. https://doi.org/10.1029/2006JD007522

Pagán, B. R., Ashfaq, M., Rastogi, D., Kendall, D. R., Kao, S. C., Naz, B. S., et al. (2016). Extreme hydrological changes in the southwestern US drive reductions in water supply to Southern California by mid century. *Environmental Research Letters*, *11*(9), 94026. https://doi.org/10.1088/1748-9326/11/9/094026

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (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. https://doi.org/10.1016/j.rse.2016.06.018

Pederson, G. T., Gray, S. T., Woodhouse, C. A., Betancourt, J. L., Fagre, D. B., Littell, J. S., et al. (2011). The unusual nature of recent snowpack declines in the North American Cordillera. *Science*, *333*(6040), 332–335. https://doi.org/10.1126/science.1201570

Pepin, N., Bradley, R. S., Diaz, H. F., Baraër, M., Caceres, E. B., Forsythe, N., et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, *5*(5), 424–430. https://doi.org/10.1038/nclimate2563

PRISM Climate Group (2018). Oregon State University. http://prism.oregonstate.edu, created 10 Aug 2018.

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., et al. (2014). Climate change impacts on the water balance of the Colorado headwaters: High-resolution regional climate model simulations. *Journal of Hydrometeorology*, *15*(3), 1091–1116. https://doi.org/10.1175/JHM-D-13-0118.1

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, *198*(1–4), 69–97. https://doi.org/10.1016/S0022-1694(96)03329-X

Rummler, T., Arnault, J., Gochis, D., & Kunstmann, H. (2019). Role of lateral terrestrial water flow on the regional water cycle in a complex terrain region: Investigation with a fully coupled model system. *Journal of Geophysical Research: Atmospheres*, *124*, 507–529. https://doi.org/10.1029/2018JD029004

Schlemmer, L., Schär, C., Lüthi, D., & Strebel, L. (2018). A groundwater and runoff formulation for weather and climate models. *Journal of Advances in Modeling Earth Systems*, *10*, 1809–1832. https://doi.org/10.1029/2017MS001260

Schwartz, M., Hall, A., Sun, F., Walton, D., & Berg, N. (2017). Significant and inevitable end-of-21st-century advances in surface runoff timing in California's Sierra Nevada. *Journal of Hydrometeorology*, *18*(12), 3181–3197. https://doi.org/10.1175/JHM-D-16-0257.1

Shrestha, P., Sulis, M., Simmer, C., & Kollet, S. (2018). Effects of horizontal grid resolution on evapotranspiration partitioning using TerrSysMP. *Journal of Hydrology*, *557*, 910–915. https://doi.org/10.1016/j.jhydrol.2018.01.024

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF Version 3. (NCAR Tech. Note NCAR/TN-475+STR, 113 pp.). https://doi.org/10.5065/D68S4MVH

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2004). Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. *Climatic Change*, *62*(1-3), 217–232.

Sun, F., Berg, N., Hall, A., Schwartz, M., & Walton, D. (2018). Understanding end-of-century snowpack changes over California's Sierra Nevada. *Geophysical Research Letters*, *46*, 933–943. https://doi.org/10.1029/2018GL080362

Tague, C. L., & Band, L. E. (2004). RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling. *Earth Interactions*, *8*(19), 1–42. https://doi.org/10.1175/1087-3562(2004)82.0.co;2

Thiébaux, J., Rogers, E., Wang, W., & Katz, B. (2003). A new high-resolution blended real-time global sea surface temperature analysis. *Bulletin of the American Meteorological Society*, *84*(5), 645–656. https://doi.org/10.1175/BAMS-84-5-645

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, *43*, W07447. https://doi.org/10.1029/2006WR005653

Waliser, D., Kim, J., Xue, Y., Chao, Y., Eldering, A., Fovell, R., et al. (2011). Simulating cold season snowpack: Impacts of snow albedo and multi-layer snow physics. *Climatic Change*, *109*(1), 95–117. https://doi.org/10.1007/s10584-011-0312-5

Walton, D. B., Hall, A., Berg, N., Schwartz, M., & Sun, F. (2017). Incorporating snow albedo feedback into downscaled temperature and snow cover projections for California's Sierra Nevada. *Journal of Climate*, *30*(4), 1417–1438. https://doi.org/10.1175/JCLI-D-16-0168.1

Westbrook, C. D., Hogan, R. J., & Illingworth, A. J. (2008). The capacitance of pristine ice crystals and aggregate snowflakes. *Journal of the Atmospheric Sciences*, *65*(1), 206–219. https://doi.org/10.1175/2007JAS2315.1

Wever, N., Comola, F., Bavay, M., & Lehning, M. (2017). Simulating the influence of snow surface processes on soil moisture dynamics and streamflow generation in an alpine catchment. *Hydrology and Earth System Sciences*, *21*(8), 4053–4071. https://doi.org/10.5194/hess-21-4053-2017

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Howat, I. M., Margulis, S. A., & Huning, L. S. (2017). Comparison of methods to estimate snow water equivalent at the mountain range scale: A case study of the California Sierra Nevada. *Journal of Hydrometeorology*, *18*(4), 1101–1119. https://doi.org/10.1175/JHM-D-16-0246.1

Wrzesien, M. L., Pavelsky, T. M., Kapnick, S. B., Durand, M. T., & Painter, T. H. (2015). Evaluation of snow cover fraction for regional climate simulations in the Sierra Nevada. *International Journal of Climatology*, *35*(9), 2472–2484. https://doi.org/10.1002/joc.4136