Significant and inevitable end-of-21st-century advances in surface runoff timing in California's Sierra Nevada

Marla Schwartz[1], Alex Hall, Fengpeng Sun, Daniel Walton, Neil Berg

University of California, Los Angeles

Department of Atmospheric and Oceanic Sciences

Los Angeles, California

Key Points

1. End-of-21st-century near-surface warming leads to statistically significant advances in surface runoff timing in the Sierra Nevada Mountains for all plausible forcing scenarios and all GCMs. Thus a detectable change in runoff timing is inevitable.
2. The 2000–2750m elevation band is associated with the greatest runoff timing advances, due in large part to snow albedo feedback.
3. Even when greenhouse gas emissions are curtailed, the runoff change is still climatically significant when compared to natural variability. If greenhouse gas emissions continue unabated, a truly dramatic change in surface hydrology is anticipated.

[1]*Corresponding author address: Marla Schwartz, Math Science Bldg Rm 7229, UCLA, Department of Atmospheric and Oceanic Sciences, Los Angeles CA 90095.*
Email: marla@atmos.ucla.edu

Abstract

Using hybrid dynamical/statistical downscaling, we project 3-km resolution end-of-21st-century runoff timing changes over California's Sierra Nevada Mountains for all available global climate models (GCMs) from phase 5 of the Coupled Model Intercomparison Project (CMIP5). All four Representative Concentration Pathways (RCPs) adopted by the Intergovernmental Panel on Climate Change's Fifth Assessment Report are examined. These multi-model, multi-scenario projections allow for quantification of ensemble-mean runoff timing changes and associated range of possible outcomes due to both intermodel variability and choice of forcing scenario. Under a "business-as-usual" forcing scenario (RCP8.5), warming leads to a shift toward much earlier snowmelt-driven surface runoff in 2091–2100 compared to 1991–2000, with advances of as much as 80 days projected in the 35-model ensemble-mean. For a realistic "mitigation" scenario (RCP4.5), the ensemble-mean change is smaller but still large (up to 30 days). For all plausible forcing scenarios and all GCMs, the simulated changes are statistically significant, so that a detectable change in runoff timing is inevitable. Even for the mitigation scenario, the ensemble-mean change is approximately equivalent to one standard deviation of the natural variability at most elevations. Thus even when greenhouse gas emissions are curtailed, the runoff change is climatically significant. For the business-as-usual scenario, the ensemble-mean change is approximately two standard deviations of the natural variability at most elevations, portending a truly dramatic change in surface hydrology by the century's end if greenhouse gas emissions continue unabated.

1. Introduction

Over half of California's developed water comes from small streams in the ecologically-sensitive Sierra Nevada (SN; USDA Forest Service 2009). Understanding future streamflow changes in this region is therefore critical to ensuring enough freshwater resources for humans and ecosystems in the coming decades. Recent warming has already produced detectable changes in the timing, magnitude, and variability of SN streamflow (Aguado et al. 1992, Dettinger and Cayan 1995, Cayan et al. 2001, Regonda et al. 2005, Stewart et al. 2005, McCabe and Clark 2005, Maurer et al. 2007, Hidalgo et al. 2009, Kim and Jain 2011). Stewart et al. (2005) found that from 1948 to 2000, a majority of SN rivers exhibited earlier timing of roughly 10–30 days during the snowmelt season. McCabe and Clark (2005) found a similar result for 84 streamflow gauges in the Western U.S., with increased April–July temperatures largely accounting for the advancement of runoff timing at most sites. Finally, Cayan et al. (2001) found that the first major pulse of snowmelt at high-elevation stream gauges in the Western U.S. advanced by about 10 days between 1948 and 1995.

While observed shifts in SN runoff timing have been well documented, few studies have produced quantitative estimates of its future changes and associated uncertainty. One reason for this is that runoff timing in this region is influenced by a complex interplay of climatic and geographic factors that are poorly resolved in coarse-resolution (~100-km) global climate models (GCMs). GCMs lack important spatial structure in local climatic factors that are dominant controls on runoff timing and its spatial distribution, such as temperature (T) and snowpack. Additionally, GCM resolution is too low to adequately represent physical watershed characteristics (e.g. elevation, slope, and vegetation type and coverage) that can also profoundly influence runoff timing and its spatial distribution.

These limitations have motivated efforts to regionalize GCM climate change signals through a variety of downscaling methods (Giorgi et al. 1994, Snyder et al. 2002, Timbal et al. 2003, Hayhoe et al. 2004, Leung et al. 2004, Tebaldi et al. 2005, Duffy et al. 2006, Cabré et al. 2010, Salathé et al. 2010, Pierce et al. 2013a). In this study, we rely on dynamical downscaling to simulate SN hydroclimate. We use a high-resolution regional climate model (RCM) to explicitly simulate complex fine-scale physical processes (Caldwell et al. 2009, Salathé et al. 2008, Salathé et al. 2010, Arritt and Rummukainen 2011, Pierce et al. 2013a). Our RCM framework resolves much of SN's fine-scale topography, the associated orographic precipitation (*P*), and demarcations between solid and liquid forms of *P*. These processes are crucial for accurate representations of accumulated wintertime snowpack and spring/summertime runoff. Moreover, the RCM more credibly simulates the strength of the snow albedo feedback (SAF) over high elevations, which has an intricate spatial structure and is also a critical influence on local warming and runoff timing.

Previous studies have used RCMs to project future runoff timing changes in the SN. Rauscher et al. (2008) used the ICTP Regional Climate Model RegCM3 (Pal et al. 2007) to investigate future changes in snowmelt-driven runoff over the Western U.S. under the A2 emissions scenario (as described in the Special Report on Emissions Scenarios; Nakicenovic et al. 2000). They found that increases in January–March *T* of approximately 3–5° C could cause runoff to occur as much as two months earlier in the late 21st-century compared to a baseline period (1961–1989). Future runoff timing projections in Rauscher et al. (2008) are only for a small number of GCMs, yielding limited information about most-likely outcomes and the associated model spread. This study also relied on a single forcing scenario, making it impossible to evaluate the consequences of societal choices regarding future greenhouse gas emissions.

Regionalizing a large number of GCM simulations is necessary to quantify ensemble-mean and uncertainty statistics for a single forcing scenario, let alone multiple forcing scenarios. However, this is impractical due to the high computational cost of RCMs. This shortcoming of RCMs highlights the need to develop a technique to project high-resolution future runoff timing in a way that fully samples the GCMs and forcing scenarios without a heavy computational burden. Stewart et al. (2004) provide an example of a more computationally feasible method using a statistically-based technique, i.e. relying on regression equations describing relationships among historical *P, T,* and runoff timing to project future runoff timing. However, they present results for only one climate model under one forcing scenario. Moreover, as with nearly all statistical techniques, their reliance on relationships derived from historical variability involves the so-called "stationarity" assumption, which may not be valid: It is possible those relationships may not hold in the future, especially for sustained changes in *T* that far exceed those observed during the historical period.

The lack of a high-resolution multi-model, multi-scenario analysis of end-of-21st-century runoff timing changes over the SN serves as the primary motivation for this study. Here is a brief overview of our technique. First, we produce a historical or "baseline" simulation for the region by dynamically downscaling reanalysis data covering the final decade of the 20th century. Next, five GCMs from phase 5 of the Coupled Model Intercomparison Project (CMIP5, Taylor et al. 2012) are dynamically downscaled under the Representative Concentration Pathway 8.5 (RCP8.5) forcing scenario (van Vuuren et al. 2011). Then, output from the dynamical simulations is used to build a simple statistical model of runoff timing that emulates the dynamical model behavior. This model takes advantage of dynamical downscaling's physical credibility but is computationally efficient, allowing us to produce a large ensemble of runoff timing projections. Using the statistical model, we project runoff timing changes for all available CMIP5 models and forcing scenarios associated with the IPCC Fifth Assessment Report (Van Vuuren et al. 2011). This allows for quantification of ensemble-mean future runoff timing changes in the SN and its associated uncertainty due to intermodel GCM spread, as well as the consequences associated with choice of forcing scenarios. Thus we can assess the degree to which runoff timing changes occur no matter which model or forcing scenario is chosen, and are therefore inevitable. Through comparison of the climate change signals with natural variability in the baseline simulation, we can also assess the statistical and climatic significance of the change signals. Because our technique involves both dynamical and statistical downscaling, we call it hybrid dynamical-statistical downscaling, or simply hybrid downscaling.

This paper is organized as follows: Section 2 describes the dynamical downscaling model configuration, and provides an observational evaluation of its performance. Section 2 also presents dynamically-downscaled end-of-21st-century changes to runoff timing. Section 3 describes the statistical runoff timing model and its evaluation. Section 4 describes statistically-based runoff timing projections for the full CMIP5 GCM ensemble under for all forcing scenarios. This section quantifies ensemble-mean runoff timing changes, ranges due to intermodel variability, and consequences stemming from choice of forcing scenario. Section 5 contains a discussion of the importance of SAF to the results, and compares the runoff timing changes projected in this study to those associated with other downscaled data products that do not include SAF. Finally, section 6 summarizes the major findings of this study and their implications.

2. Dynamical model set-up, evaluation and results

2a. Dynamical model set-up

Dynamical downscaling is performed using the Weather Research and Forecasting (WRF) model version 3.5 (Skamarock et al. 2008). WRF is coupled to the community Noah land surface model with multi-parameterization options (Noah-MP, Niu et al. 2011). Three one-way nested domains are used to represent the complex topography of California and the SN as accurately as possible (Fig. 1a). The outermost domain spans the entire U.S. West Coast and adjacent Pacific Ocean at 27-km horizontal resolution. The middle domain, at 9-km resolution, covers all of California. The innermost domain, at 3-km resolution, spans the eastern edge of the Central Valley to the leeside of the California SN (Fig. 1b); this domain is the focus of this study.

In each domain, all variables within five grid cells from the horizontal lateral boundary are relaxed toward the corresponding values at the boundaries. To provide a better representation of surface and boundary layer processes, the model's vertical resolution is enhanced near the surface, with 30 out of 43 total sigma-levels below 3-km. WRF parameterization testing was done to optimize the model's performance in hydroclimate simulations, with the aim of improving the realism of simulated SN snowpack and streamflow processes. The package of physical parameterizations consists of the New Thompson microphysics scheme (Thompson et al. 2008), Dudhia shortwave radiation scheme (Dudhia 1989), Rapid Radiative Transfer Model longwave (RRTM) longwave radiation scheme (Mlawer et al. 1997), MYNN Level 2.5 surface/boundary layer scheme (Nakanishi and Niino 2006), and Old Kain-Fritsch cumulus convection scheme (Kain and Fritsch 1990). Spectral nudging of temperature, zonal and meridional winds, and geopotential height is employed above the boundary layer (roughly 850 hPa) over the outermost 27-km resolution domain.

Climate changes signals are produced from a single baseline simulation and five future simulations. The baseline simulation spans October 1991 to September 2001 (water years 1992–

2001; hereinafter "$WY_{1992–2001}$") and is a dynamical downscaling of the National Centers for Environmental Prediction's 6-hourly North America Regional Reanalysis (NARR; Mesinger et al. 2006). NARR is a relatively coarse-resolution (32-km) reanalysis dataset that provides lateral boundary forcings and initial conditions for the outermost WRF domain in Fig. 1a. The baseline simulation allows us to evaluate the model's ability to simulate regional runoff timing through a comparison to observational data (section 2b) and serves as a climate state against which we can compare future climate simulations to measure change.

Using the same model configuration as the baseline, we perform a five-member ensemble of dynamical downscaling experiments to simulate a future end-of-21st-century climate. The simulations go from October 2091 to September 2101 (water years 2092–2101, hereinafter "$WY_{2092–2101}$"). We dynamically downscale GCM experiments forced by RCP8.5. Out of all available CMIP5 GCMs forced by RCP8.5, we select five (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR). These GCMs approximately sample the range of end-of-21st-century near-surface T and P changes over California (see Walton et al. 2016, Fig. 2).

To produce boundary conditions for the future WRF simulations, we add a perturbation reflecting the mean change in GCM climatology to NARR data for $WY_{1992–2001}$, following Schar et al. (1996), Hara et al. (2008), Kawase et al. (2009) and Rasmussen et al. (2011). To calculate these GCM climate changes, we first quantify the differences in GCM monthly climatology between the historical and RCP8.5 experiments (2081–2100 average minus 1981–2000 average). Differences are calculated for temperature, humidity, zonal and meridional winds, and geopotential height. Then, for each of the five dynamically-downscaled GCMs, we perturb the baseline 6-hourly NARR reanalysis data for each month by the corresponding monthly mean climatological change. The perturbed NARR fields then serve as WRF boundary conditions for five future climate simulations. This method allows us to assess how $WY_{1992–2001}$ would transpire if the mean climate were altered to reflect the climate changes projected by each of the five GCMs. It allows us to quantify how the climate change signals simulated in the GCMs are expressed at the regional scale, without the future simulations being subject to significant biases in mean state often found in GCMs. For additional information on model setup, parameterizations and design of future simulations, the reader is referred to Walton et al. (2016).

2b. Baseline runoff timing climatology and model evaluation

We first evaluate WRF's ability to simulate surface runoff timing during the baseline period. As a measure of runoff timing, we consider the date in the water year (October 1 – September 30; hereinafter WY) by which 50% of the cumulative WY surface runoff has occurred (R_{50}). R_{50} is widely used as a metric of snowmelt timing (Regonda et al. 2005, Moore et al. 2007, Rauscher et al. 2008, Hidalgo et al. 2009, Wenger et al. 2010, Ashfaq et al. 2013). R_{50} is similar to the center timing of streamflow used in Stewart et al. (2004), Stewart et al. (2005) and McCabe and Clark (2005), but is found to be less sensitive to outliers in streamflow (Moore et al. 2007). Moreover, Regonda et al. (2005) suggest that R_{50} is a more reliable indicator of snowmelt timing (in its relation to climatic variability and change) than the day of peak flow. In this paper, we use R_{50} both for model evaluation and as a metric to diagnose future changes to runoff timing.

Fig. 2 presents the baseline ($WY_{1992–2001}$) climatological date of R_{50} in the 3-km domain (seen in Fig. 1b). Climatological R_{50} generally occurs after March 1 throughout the SN and shifts to even later in the WY as both elevation and the fraction of precipitation falling as snow (*S/P*) increase. At lower elevations in the Northern SN where the annual *S/P* (not shown) ranges from 0.6 to 0.8, climatological R_{50} generally occurs before the start of summer. However, mid to high elevations over the Central and Southern SN have a higher *S/P* ratio (0.8 to 0.95), which leads to snowmelt-driven surface runoff throughout the summer months, pushing back climatological R_{50}. For example, R_{50} in the mountains just southwest of Mono Lake typically occurs as late as the beginning of July. Throughout the Central Valley, Owens Valley and western Great Basin Desert along the California-Nevada border, annual *P* is low, and any *P* typically falls as rain (*S/P* < 0.2). So surface runoff timing matches *P* timing.

For this study, we consider surface runoff timing changes at locations where surface runoff is mostly generated by snowmelt. The March 1 R_{50} cutoff date segregates snowfall-dominated grid points from rain-dominated regions or locations with minute climatological *P*. The black contour in Fig. 2 denotes locations with climatological baseline R_{50} occurring on or after March 1, indicating snowmelt-dominated runoff. The average baseline climatological *S/P* within the contoured region is 0.86, also indicative of a snowfall-dominated regime. Within the contoured region in Fig. 2, the median and mean climatological percentages of total water-year runoff occurring from April–July are 78% and 69%, respectively. This is consistent with other snowmelt-dominated watersheds in western North America examined by Stewart et al. (2005). We consider only grid points with climatological baseline R_{50} on or after March 1 for the rest of the study. We also exclude inland water locations in our analyses. From here forward, the term "domain-average" shall refer to an average over this restricted zone.

The dynamical model's ability to reproduce runoff timing variations during the baseline period can be assessed by comparing simulated R_{50} to observations obtained from the United States Geological Survey Hydro-Climatic Data Network-2009 (USGS HCDN-2009, http://waterdata/usgs.gov/nwis/). The USGS HCDN-2009 is a network of streamflow gauges having the following characteristics: (1) natural streamflow least affected by direct human activities, (2) accurate measurement records, and (3) at least 20 years of complete and continuous records through WY 2009 (Slack et al. 1993; Lins 2012). We obtained daily, quality-controlled streamflow data from 11 stations for which data was available within our study domain for the baseline period. The station locations are indicated in Fig. 1b with blue circles, and information associated with each station is summarized in Table 1. The 11 stations represent a variety of elevations, drainage areas and USGS eight-digit Hydrologic Unit Codes across SN creeks and rivers.

The scatter plot in Fig. 3 presents observed versus simulated climatological R_{50} for each of the 11 stations. Simulated climatological R_{50} is taken to be the average R_{50} of the grid points upstream of a gauge within that gauge's USGS Hydrologic Unit. This is equivalent to assuming instantaneous transport of water from the grid cell to the stream gauge location. A portion of the biases in this evaluation is likely due to this admittedly primitive river routing scheme. Each gauge's data point is also colored by the corresponding interannual correlation coefficient. For each gauge, simulated R_{50} is very well correlated with the observed R_{50}, with temporal r ranging from 0.75 to 0.96. (The gauge-average is 0.87.) Fig. 3 also demonstrates that observed and simulated R_{50} dates are well-correlated spatially (r = 0.62) across all gauges. The root-mean-square error between observed and simulated climatological R_{50} is 12.2 days. Overall, the degree of agreement between simulated and observed R_{50} dates indicates that the dynamical model is able to capture the main features of spatial and temporal R_{50} variability across the SN. In section 3, we also evaluate the realism of the dynamical model's sensitivity of R_{50} to spring temperatures, a key parameter of the statistical model we develop to project future R_{50}.

2c. Dynamically-downscaled changes in runoff timing

Fig. 4 (row 1) presents the dynamically-downscaled WRF end-of-21st-century change ($WY_{2092–2101}$ minus $WY_{1992–2001}$) in R_{50} (ΔR_{50}) under the RCP8.5 forcing scenario for the five GCMs. For all simulations, advances in R_{50} are projected at all locations with substantial climatological baseline snowmelt-driven surface runoff. GFDL-CM3 (Fig. 4b) and IPSL-CM5A-LR (Fig. 4d) project the largest advances, with domain-average advances greater than 60 days. Advances in domain-average mean R_{50} for CNRM-CM5 (Fig. 4a) and INM-CM4 (Fig. 4c) are smaller, but are still nearly 6 weeks earlier.

For all dynamically-downscaled GCMs, advances in mean R_{50} tend to be greater on western-facing mountain slopes. This spatial pattern can be explained by mean near-surface (2-meter) springtime (March-May) warming projections (ΔT_{MAM}). Fig. 5 (row 1) presents WRF dynamically-downscaled end-of-21st-century ΔT_{MAM} under RCP8.5. For each of the five GCMs, somewhat stronger warming is projected on the western-facing mountain slopes near the springtime freezing line. These regions have the strongest SAF (Walton et al. 2016) and greatest April 1st snow water equivalent (SWE) loss (Sun et al. 2016). This warming leads to decreases in annual-mean *S/P* and earlier snowmelt, which together result in large advances in mean R_{50} in those areas. Another feature of the spatial patterns of R_{50} advances is relatively small changes at the highest elevations in the Southern SN. Despite significant future warming (Walton et al. 2016), these areas remain well above the freezing line during the accumulation season. As a result, changes to *S/P* and snow accumulation are small at the highest elevations, and the weak advances in R_{50} (10-20 days) at those locations are primarily due to earlier snowmelt.

Intermodel differences in ΔR_{50} can largely be explained by differences in ΔT_{MAM}, as ΔR_{50} appears to be strongly negatively related to ΔT_{MAM}. GFDL-CM3 and IPSL-CM5A-LR project large ΔT_{MAM}. Domain-average spring warmings are 6.0 °C and 6.9 °C, respectively, and some locations warm more than 7 °C. This strong warming explains the sizable advances in mean R_{50} for GFDL-CM3 and IPSL-CM5A-LR. Weaker ΔT_{MAM} in INM-CM4 and CNRM-CM5 (domain-average 3.6 °C and 3.7 °C, respectively) corresponds to smaller mean R_{50} advances. (MPI-ESM-LR is moderate in both ΔT_{MAM} and ΔR_{50}.) This link suggests ΔT_{MAM} might be a reasonable predictor for ΔR_{50}, a hypothesis that will be explored in the description of the statistical ΔR_{50} model in section 3. In section 3, we also consider mean P changes as a predictor for ΔR_{50}. However, P timing hardly changes in the downscaled WRF simulations, so intermodel differences in R_{50} advances are likely not attributable to P changes, as we will show.

3. ΔR_{50} Statistical model description and evaluation

The previous section focused on projections of changes to mean R_{50} for only a single time slice, a single forcing scenario, and for only five GCMs. This information is insufficient to fully quantify the range of possible outcomes due to intermodel spread and choice of forcing scenario. To project ΔR_{50} for all available CMIP5 GCMs and all forcing scenarios, we adopt a hybrid downscaling approach, developing a computationally efficient statistical ΔR_{50} model that is designed to emulate the dynamical model. In this section, we describe and evaluate this statistical model.

As noted above, there is a negative relationship between WRF ΔR_{50} (Fig. 4, row 1) and ΔT_{MAM} (Fig. 5, row 1). To quantify this relationship, Fig. 6a shows the correlation coefficient for each grid point between dynamical ΔR_{50} and ΔT_{MAM}. The correlation values reflect a blend of intermodel and interannual variability, as they are calculated from annually-averaged ΔR_{50} and ΔT_{MAM} values from all five models. This produces a sample size of 50 for each grid point (5 models × 10 water years). There is a very strong anti-correlation between ΔR_{50} and ΔT_{MAM}, with a spatially-averaged value of r = -0.82. That ΔT_{MAM} would be a predictor for ΔR_{50} is physically sensible, as climatological baseline R_{50} for many mountainous locations falls in March–May (Fig. 2), and March–May runoff accounts for a significant portion of annual runoff throughout much of the SN. Thus we aim to build a statistical modeling framework that projects ΔR_{50} given ΔT_{MAM}. Below we discuss our choice of ΔT_{MAM} as a predictor further, and other predictors we considered.

The first step is to linearly regress dynamically-downscaled ΔR_{50} onto dynamically-downscaled ΔT_{MAM} for each pair of coordinates (i,j) in the 3-km resolution domain with climatological baseline R_{50} on or after March 1. As with the corresponding correlation coefficient shown in Fig. 6a, the slope (α) of this linear regression is determined by intermodel and interannual variability, i.e. 50 data points (10 water years × 5 models) for each (i,j) pair. Fig. 6b presents the spatial pattern of α, the average expected advance in mean R_{50} timing per degree March–May near-surface warming. In calculating α, we force the linear relationship to go through (0,0), i.e. it has no intercept. This is an expression of the physical constraint that one would not expect a change in R_{50} timing without a change in T_{MAM}. The domain-average α is -10.2 days/°C, but Northern SN and mid-elevation western slopes are much more sensitive, with projected R_{50} changes of more than -19 days/°C. The strong sensitivity at these mid-elevation locations is due to both warming-driven S/P decreases and earlier snowmelt, which conspire to advance R_{50}. The sensitivity at higher elevations is lower because the T_{MAM} increases lead mostly just to earlier snowmelt. Moreover, these more sensitive regions correspond well to regions of greatest projected April 1st SWE decreases (Sun et al. 2016) and greatest SAF-enhanced warming and snow cover loss (Walton et al. 2016). After determining α, we then predict ΔR_{50} with following equation:

$$\Delta R_{50,GCM,i,j} \cong \alpha_{i,j} * \Delta T_{MAM,GCM,i,j} \quad (1)$$

It is possible to evaluate the realism of α as simulated by the dynamical model. The relationship between purely interannual R_{50} and T_{MAM} anomalies is linear to a very good approximation in both observations and the WRF simulation. Fig. 7 presents a scatter plot of observed annual T_{MAM} anomalies vs. R_{50} anomalies over $WY_{1916–2014}$. Observed interannual R_{50} and T_{MAM} variations in California are very anti-correlated (r = -0.67). The observed linear sensitivity of $WY_{1916–2014}$ R_{50} to T_{MAM} is -9.46 days/°C. In WRF, the domain-average slope of the linear regression of $WY_{1992–2001}$ R_{50} onto T_{MAM} is -11.4 days/°C (close to the domain-average α of -10.2 days/°C). The approximate agreement between observations and the WRF simulation provides crucial support for the realism of the WRF simulation of streamflow timing and a statistical model based on the linear relationships between ΔR_{50} and ΔT_{MAM}. This form of model evaluation, focusing on sensitivity parameters key to climate change response, is likely more relevant than the general model evaluation of temporal and spatial variability in streamflow presented in section 2b.

One source of error in the statistical ΔR_{50} model (Eq. 1) arises from approximating ΔR_{50} as linear function of ΔT_{MAM}. Though this error source must be small because the linear correlation coefficients between the two variables are very high (Fig. 6a), we can evaluate it by statistically projecting ΔR_{50} with the dynamically-downscaled ΔT_{MAM} under RCP8.5 (Fig. 5, row 1) as input. Row 2 of Fig. 4 presents this statistical ΔR_{50} projection, which can be compared to dynamically-downscaled ΔR_{50} (Fig. 4, row 1). Overall, the approximated values of ΔR_{50} (Fig. 4, row 2) almost perfectly mirror the dynamically-downscaled values (Fig. 4, row 1). The approximated spatial patterns are highly correlated with their dynamical counterparts (r > 0.84 for all GCMs). The mean absolute errors (MAE, calculated by averaging the absolute value of the errors over the region of interest) are less than 11 days for all models, small compared to domain-average advances in R_{50}

that range between 39 and 66 days. This comparison lends further credibility to the choice to model ΔR_{50} as a linear function of ΔT_{MAM}.

To apply the statistical ΔR_{50} model to all GCMs and forcing scenarios, we rely on projections of ΔT_{MAM} from Walton et al. (2016, hereinafter "W2016"). W2016 produced 3-km horizontal resolution monthly near-surface warming projections for our study domain for all available CMIP5 GCMs under forcing scenarios RCP8.5, 6.0, 4.5 and 2.6. W2016 also used a hybrid dynamical-statistical technique to downscaling warming that relies on two large-scale GCM predictors (regional-mean warming and east-west warming contrast) and a representation of SAF's significant contribution to elevational variations in warming. Fig. 5 (row 2) presents end-of-21st-century hybrid downscaled ΔT_{MAM} under RCP8.5 from W2016. As discussed in detail in W2016, this method captures the spatial pattern and approximate magnitude of ΔT_{MAM} for each of the 5 dynamically-downscaled GCMs (Fig. 5, row 1), including the warming enhancement due to SAF at mid-elevations and in the Northern SN.

To assess the error associated with the use of W2016's hybrid downscaled ΔT_{MAM} as input to our statistical ΔR_{50} model, we compare the dynamically-downscaled ΔR_{50} projections under RCP8.5 (Fig. 4, row 1) to those calculated by the statistical ΔR_{50} model (Eq. 1), now with the hybrid downscaled ΔT_{MAM} projections of W2016 as input (Fig. 4, row 3). Overall, the spatial correlations between these ΔR_{50} patterns and WRF's dynamically downscaled patterns are very high ($r > 0.83$) and the MAE values are low compared to the magnitude of ΔR_{50}, indicating that the use of hybrid downscaled ΔT_{MAM} input reproduces dynamically-downscaled ΔR_{50} projections reasonably well. Still, we note some minor discrepancies. For GFDL-CM3, INM-CM4 and IPSL-CM5A-LR, hybrid projections of ΔT_{MAM} by W2016 (Fig. 5, row 2) underestimate the dynamically-downscaled ΔT_{MAM} somewhat (Fig. 5, row 1). As a result, using hybrid downscaled ΔT_{MAM} leads to an underestimate of the magnitude of the dynamically-downscaled ΔR_{50}. Similarly, W2016 slightly overestimates ΔT_{MAM} for CNRM-CM5 and MPI-ESM-LR, which results in a small overestimation of mean R_{50} advances for those GCMs.

Precipitation (P) changes (especially its seasonality) may also affect future runoff timing, suggesting it ought to be included in our statistical model as a co-predictor. Previous studies have found only modest projected changes in mean P, which are also small compared to natural variability (Pierce et al. 2013b, Cayan et al. 2008, Duffy et al. 2006). Nevertheless, testing was done to include mean wet-season P (December-March) changes as a co-predictor along with ΔT_{MAM} in the statistical ΔR_{50} model. Less than 4% improvement was seen in the model-average MAE, compared to using ΔT_{MAM} alone. Additional testing was done to determine if ΔT or ΔP averaged over other months produced a more skillful model than one that relies on only ΔT_{MAM}, but again, no value was gained. Including changes to April 1st SWE as a co-predictor also added no value. Overall, this indicates that advances in ΔR_{50} are nearly entirely driven by spring warming, consistent with previous studies of observed and projected runoff timing changes over the SN and Western U.S. (e.g. Stewart et al. 2004).

We also note that the dynamical downscaling framework imposes identical interannual variability levels between the baseline and future time slices. Possible changes to interannual variability modes in the 21st-century, for example the El Niño-Southern Oscillation phenomenon (Cai et al. 2014), could impact overall P levels and timing through atmospheric teleconnections, a factor not fully accounted for in GCMs or in this study. However, given the very large magnitude of changes in mean runoff timing driven by warming alone, it is difficult to see how our main conclusions would be significantly different if El Niño-driven changes in P do occur.

4. Results for the full GCM ensemble and all forcing scenarios

Using the statistical ΔR_{50} model (Eq. 1) with the W2016 hybrid downscaled ΔT_{MAM} as input, we now generate projections of mean changes in end-of-century R_{50} for all available CMIP5 GCMs under four forcing scenarios: RCPs 2.6, 4.5, 6.0 and 8.5. Fig. 8 (row 1) presents ensemble-mean changes in R_{50} for RCPs 2.6, 4.5, 6.0, and 8.5. The spatial patterns of ΔR_{50} are qualitatively similar for each forcing scenario, with the magnitudes increasing with forcing scenario strength. While all locations show some advance, the largest are found at elevations between 2000–2750m and are generally on the western slope of the SN. In some locations, ensemble-mean R_{50} is projected to advance by more than 80 days under RCP8.5. For RCP8.5, the ensemble-mean domain-average ΔR_{50} is -49.7 days (Fig. 8d), which is very close to that of the five-model dynamically downscaled ensemble (-51.7 days). This supports the idea that the five GCMs we select for dynamical downscaling approximately represent the GCM ensemble.

Ensemble-mean R_{50} changes are substantial when compared with the interannual variability of the baseline period. To provide a more statistically stable estimate of baseline interannual variability, we extend the baseline simulation to span $WY_{1982–2001}$. This 20-year simulation uses the same modeling framework described in section 2a. Fig. 8 (row 2) presents z-scores associated with the ensemble-mean changes in R_{50} in Fig. 8 (row 1). The z-score is calculated by dividing the mean R_{50} change by the standard deviation of R_{50} for the extended baseline period ($WY_{1982–2001}$), and therefore represents how far outside the baseline $WY_{1992–2001}$ R_{50} distribution an average future R_{50} is. For all scenarios, the z-score indicates a significant shift. Under RCP2.6 and 4.5 (Fig. 8e–f), for example, the domain-average ensemble-mean z-scores are -0.60 and -0.93, respectively. Under RCP6.0 (Fig. 8g), the domain-average ensemble-mean z-score (-1.18) translates to a future mean R_{50} equivalent to the 12th percentile of baseline R_{50} distribution.

Ensemble-mean R_{50} changes compared to the baseline's interannual variability are dramatic for RCP8.5 (Fig. 8h), as the domain-average z-score is -1.84, approximately the 3rd percentile of the baseline R_{50} distribution. In fact, under RCP8.5, the ensemble-mean domain-average R_{50} is projected to be earlier than that of any baseline year of the extended baseline simulation. For RCP4.5, RCP6.0, and RCP8.5 especially, the ensemble-mean R_{50} changes correspond to a substantial change in the runoff climatology.

Figure 9 shows the elevational profile of ΔR_{50} for the ensemble-mean (thick solid line) under the four RCPs. Elevations are binned every 100m, and ΔR_{50} for a given elevation bin is the spatial average across grid cells within the bin. Light gray shading represents the standard deviation of R_{50} over $WY_{1982–2001}$ at each elevation, a measure of interannual variability. Under RCP8.5 (Fig. 9d), ensemble-mean ΔR_{50} has a greater than one standard deviation advance for all elevations above 1500m. Ensemble-mean ΔR_{50} is outside of one standard deviation in the 2000–3100m elevation band under RCP6.0 (Fig. 9c), and is near or less than one standard deviation for RCP4.5 and RCP2.6 (Fig. 9a-b).

Ensemble projections also allow for the quantification of uncertainty in R_{50} projections due to GCM spread. Thick dashed lines in Fig. 9 represent the 10th and 90th percentiles of the GCM spread in ΔR_{50} when calculated with hybrid downscaled ΔT_{MAM}. For all forcing scenarios, GCM spread is greatest in the 2000–3000m elevation band, which reflects the larger spread in ΔT_{MAM} projections at those elevations (Walton et al. 2016). Under RCP8.5 however, despite an intermodel ΔR_{50} range of more than 30 days at some elevations, the advance in R_{50} is well outside one standard deviation of interannual variability for all models in the 2000–3200m elevation band.

To shed light on the statistical significance of these changes, we perform a one-tailed t-test that assesses the likelihood that a 10-year sample with a given mean shift in R_{50} could be drawn from the same population as the baseline $WY_{1982–2001}$ R_{50} distribution. To do this, we assume the future period is a 20-year sample. The sample size is $n = 20$, so nineteen degrees of freedom are used. The region outside of the dark gray shading in Fig. 9 represents changes in mean R_{50} timing that are significant at the 5% level for each elevation. RCP2.6 ensemble-mean changes are significant at the 5% level in the 1800–3300m elevation bin, but the changes are not significant for all GCMs. As we discuss in Section 6, RCP2.6 is not likely to be a plausible forcing scenario. Under RCP4.5, 6.0 and 8.5 (Fig. 9b-d), ΔR_{50} is significantly different at the 5% level from the baseline mean for all elevations and all GCMs. (A minor exception can be found under RCP4.5 at the lowest elevations for the GCMs giving the least warming.) Under RCP8.5, ΔR_{50} is significantly different at the 1% level from the baseline mean for all elevations and all GCMs (not shown). Estimates of recent global greenhouse gas emissions indicate they are closely approaching and possibly exceeding the RCP8.5 pathway (Le Quéré et al. 2015). Should emissions continue to follow RCP8.5, it is therefore very likely that future advances in runoff timing will be dramatically different from internal climate variability at all elevations.

The spatial structure of runoff timing advances also leads to an increase in the spatial homogeneity of R_{50} across the SN. Fig. 10 presents the distribution of R_{50} dates for (a) the end-of-20th-century and (b) end-of-21st-century under RCP8.5, binned by 5-day intervals. A striking change in the shape of the distribution is seen. In the baseline (Fig. 10a), R_{50} dates are fairly evenly distributed between the months of March to June, with roughly one quarter of R_{50} dates occurring in each of those four months (25.9% in March, 21.0% in April, 26.5% in May and 25.9% in June). However, there is significantly less diversity in projected end-of-21st-century R_{50} spatial patterns, with over half the gridpoints having projected R_{50} dates in February (Fig. 10b). Clearly, warming produces a strong tendency for the center of runoff timing to more closely match the center of precipitation timing. We are not aware of any assessment of increasing homogeneity of runoff timing across the SN or other regions with snowmelt-dominated runoff. This important consequence of warming must be considered in water resources planning and flood protection.

5. Importance of snow albedo feedback to ΔR_{50} projections

Both WRF dynamically-downscaled and hybrid-downscaled ΔT_{MAM} projections explicitly include warming enhancement due to SAF and its intricate spatial structure (Fig. 5). This mechanism has the largest effect at mid-elevations, which is likely also linked to larger runoff timing changes there. Here we quantify the importance of using warming patterns that include SAF to ΔR_{50} outcomes. For this exercise, we consider three methods of projecting ΔT_{MAM} that do not consider SAF effects in the SN, at least not explicitly: linear interpolation of GCM output, Bias Correction and Constructed Analogs (BCCA; Hidalgo et al. 2008; Maurer and Hidalgo, 2008) and Bias Correction with Spatial Disaggregation (BCSD; Wood et al. 2002; Wood et al. 2004; Maurer, 2007). BCCA and BCSD are two commonly used statistical downscaling techniques. Linear interpolation is a simple and naïve method of downscaling GCM output that represents a baseline measure of downscaling skill against which the other methods can be compared. BCCA and BCSD T projections were obtained online from the archive of Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections [Reclamation, 2013]. BCCA T projections are available as daily maximum and minimum T at 1/8 degree resolution; we average these together to produce monthly average T. Similar processing was applied to BCSD maximum and minimum T, which are available as monthly averages.

Fig. 11 presents the end-of-21st-century ΔT_{MAM} under RCP8.5 averaged over five GCMs (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR) downscaled using

5 methods: (a) WRF dynamical downscaling, (b) hybrid downscaling, (c) linear interpolation, (d) BCCA and (e) BCSD. Both dynamical downscaling (Fig. 11a and Fig 5a-e) and hybrid downscaling (Fig. 11b and Fig. 5f-j) reveal warming amplification due to snow cover loss and SAF at mid-elevations and in the Southern SN. However, warming patterns produced through linear interpolation (Fig. 11c), BCCA (Fig. 11d) and BCSD (Fig. 11e) do not feature such a warming enhancement. We note that warming signals produced through BCSD downscaling are nearly identical to those produced using linear interpolation. This similarity arises because BCSD applies the same bias correction to both the baseline and future time periods. For a comprehensive analysis of the difference in warming patterns that arise through these downscaling methods, the reader is referred to W2016.

Next, we analyze patterns of runoff timing that arise from the ΔR_{50} statistical model (Eq. 1) calculated with the five methods of downscaled ΔT_{MAM} in Fig. 11 as input. Fig. 12a presents ΔR_{50} estimated based on WRF dynamically-downscaled ΔT_{MAM} averaged over the five GCMs, while Fig. 12b-e show the differences between outcomes in Fig. 12a and those produced with ΔT_{MAM} from the other four downscaling methods. Using W2016's hybrid downscaled ΔT_{MAM} model as input to the ΔR_{50} statistical model produces outcomes (Fig. 12b) very similar to those produced with WRF dynamically downscaled ΔT_{MAM} as input (domain-average MAE is only 3.03 days). However, ΔR_{50} outcomes produced using linearly interpolated, BCCA and BCSD (Fig. 12c-e) ΔT_{MAM} systematically underestimate the magnitude of ΔR_{50} in WRF (Fig. 12a), with domain-average differences of 7.67, 13.97 and 8.41 days, respectively. The differences are greatest in the Northern SN and at mid-elevations on the western slopes, where linear interpolation, BCCA and BCSD systematically underestimate warming because they do not include warming amplification due to SAF. For example, ΔR_{50} outcomes produced using BCCA ΔT_{MAM} are 20–30 days less than those produced using WRF's ΔT_{MAM} at these locations. At the highest elevations (>3000 m), WRF's ΔT_{MAM} (Fig. 11a) roughly agrees with that of linear interpolation (Fig. 11c) and BCSD (Fig. 11e). This approximate agreement in ΔT_{MAM}, together with a weaker linear sensitivity of ΔR_{50} to ΔT_{MAM} at the highest elevations (Fig. 6b), are the main reasons ΔR_{50} calculations based on the various data sets of ΔT_{MAM} are within 10 days of one another at the highest elevations.

The impact of downscaling technique is also seen in Fig. 9, where thin black (green) lines show the elevational profile of ensemble-mean ΔR_{50} calculated with BCSD-downscaled (BCCA-downscaled) spring warming as input. As mentioned before, the elevational profile of ΔR_{50} calculated with BCSD-downscaled spring warming is nearly identical to that produced using linearly interpolated GCM spring warming. For each RCP, using BCSD or BCCA downscaled ΔT_{MAM} as input significantly underestimates the magnitude of the R_{50} advance at elevations below 2700m compared to that calculated using W2016's hybrid downscaled ΔT_{MAM} (solid colored lines). This is partly due to an underestimation of mid-elevation (2000–2700m) warming that stems from the inability of BCCA and BCSD to incorporate SAF effects.

Several regional climate adaptation planning agencies and tools rely on BCCA or BCSD downscaled projections. For example, Cal-Adapt (http://cal-adapt.org/), which was developed based on the 2009 California Climate Adaptation Strategy and provides access to climate data produced by California's scientific community, employs BCSD to downscale T, P and SWE to 1/8 degree spatial resolution. Though BCCA and BCSD do not directly simulate runoff timing, their T, P and SWE projections can serve as input to a hydrologic or land surface model (such as Noah-MP) to simulate runoff and estimate the sensitivity of runoff timing to spring temperature. Because SAF is very likely a key feature of future warming in the SN, hydroclimate projections based on BCSD/BCCA are associated with an underestimation of future runoff timing advances, especially at mid-elevations.

6. Summary and implications

We develop a statistical model for the date in the water year by which 50% of the cumulative surface runoff has occurred (R_{50}), and create multi-model, multi-scenario projections of high-resolution changes to Sierra Nevada runoff timing for the end-of-the-21st-century. Projections are based on linear relationships between end-of-21st-century springtime warming and runoff timing changes according to five dynamically downscaled GCMs. These linear relationships are very similar to those found in observations. Hybrid downscaled T that explicitly accounts for SAF (Walton et al. 2016) is then used to project runoff timing changes for all GCMs under forcing scenarios RCP2.6, 4.5, 6.0, and 8.5. Evaluation of the statistical model for runoff timing projections shows that it is able to successfully reproduce dynamical solutions and can credibly produce outcomes for any GCM given only its regionalized spring warming.

Projections reveal that future warming in the Sierra Nevada leads to strong shifts toward more liquid precipitation and earlier snowmelt. Together, these hydroclimatic changes significantly advance surface runoff, particularly at mid-elevations (2000–2750m). R_{50} advances of over 80 days are projected at some mid-elevation locations in the 35-model ensemble-mean under RCP8.5. Strong R_{50} advances are projected at mid-elevations even for a forcing scenario associated with curtailed greenhouse gas emissions (RCP4.5), where ensemble-mean R_{50} advances in the 2000–2750m elevation band are nearly 40 days. The larger changes at mid-elevations are driven in part by SAF. The absence of this mechanism in other downscaled data products implies a significant underestimate of runoff timing changes at these elevations.

Given estimates of recent global greenhouse gas emissions (Le Quéré et al. 2015), RCP2.6 involves greenhouse gas reductions that have not occurred since the RCP forcing scenarios were created in 2005. The reductions associated with RCP2.6 in the coming decades are likewise unlikely to occur. Thus we only consider RCP4.5, RCP6.0, and RCP8.5 to be the plausible forcing scenarios. With the minor exception of a few GCMs at elevations below 1800m under RCP4.5, R_{50} advances are significant at the 5% level for all elevations for all GCMs and all three of these forcing scenarios (Fig. 9). Therefore, detectable changes in runoff timing in the SN are inevitable.

In addition to testing the statistical significance of R_{50} advances, we compare the magnitude of R_{50} changes to the standard deviation of interannual variations in R_{50} in the baseline period to assess their climatic significance. Under RCP4.5, ensemble-mean R_{50} advances in the 2000–2750m elevation band are greater than one standard deviation of baseline interannual variability at mid-elevations, and are nearly one standard deviation elsewhere (Fig. 9b). Thus even when greenhouse gas emissions are curtailed, the runoff change is climatically significant. For RCP8.5, ensemble-mean R_{50} advances are roughly two standard deviations of baseline interannual variability at mid-elevations and greater than one standard deviation elsewhere (Fig 9d). Thus if greenhouse gas emissions continue unabated, a truly dramatic change in surface hydrology is anticipated by century's end. It is important to keep in mind that the dramatic advances in R_{50} timing examined here are at the level of individual grid points in the regional model, and that information about R_{50} changes at streamflow gauges or the watershed-level is beyond the scope of this study.

Another important finding of this study is that our projected R_{50} advances are much larger (especially at elevations below 2700m) than those implied by other commonly-used downscaled data warming products (e.g. BCCA and BCSD), because these other downscaling methods miss crucial warming amplification due to SAF. Additionally, we find one new consequence of warming-driven advances in runoff timing—an increase in the homogeneity of runoff timing dates across the SN.

Significant and inevitable runoff timing advances have major implications for California's water resource infrastructure. The current infrastructure assumes SN snowpack melts gradually throughout the dry season, and it is unclear whether it can accommodate such drastic changes to runoff timing. Reservoir operational rule curves specify the monthly target water level for each reservoir and are crucial for both flood control/protection and storage. The rule curves were developed in the mid-1900s when most of California's dams were built, and the historical data used to inform them generally reflects the hydroclimate of the first half of the 20^{th}-century (Willis et al. 2011). Given significant changes to snowmelt runoff timing found in this study, at a minimum it will be necessary to revise rule curves to avoid detrimental and wasteful water releases. It may also be necessary to find alternative storage, such as in groundwater reservoirs. Changes to runoff timing will also have important consequences for water rights tied to specific seasons or months. Lastly, shifts in runoff timing have implications beyond California's water resources, including for aquatic ecosystem vitality, soil moisture change in riparian areas and recreational activities throughout the SN. Long-term climate and streamflow observations throughout the Sierra Nevada will continue to be crucial for detection and attribution of anthropogenic runoff timing changes.

Acknowledgments

Funding for this work was provided by the Metabolic Studio in partnership with the Annenberg Foundation (Grant #12-469: Climate Change Projections in the Sierra Nevada) and the US

Department of Energy (Grant #DE-SC0014061: Developing Metrics to Evaluate the Skill and Credibility of Downscaling).

6 References

Aguado, E., D. Cayan, L. Riddle, and M. Roos, 1992: Climatic fluctuations and the timing of West Coast streamflow, J. Clim., 5, 1468–1483.

Arritt, R. W., and M. Rummukainen, 2011: Challenges in regional-scale climate modeling, Bull. Amer. Meteor. Soc., 92, 365–368, doi:10.1175/2010BAMS2971.1.

Ashfaq, M., S. Ghosh, S.C. Kao, L. C. Bowling, P. Mote,D. Touma, S. A. Rauscher, and N. S. Diffenbaugh, 2013: Near-term acceleration of hydroclimatic change in the western U.S., J. Geophys. Res. Atmos., 118,676–10,693, doi:10.1002/jgrd.50816.

Cabré, M., S. A. Solman, and M. N. Nuñez, 2010: Creating regional climate change scenarios over southern South America for the 2020's and 2050's using the pattern scaling technique: validity and limitations, Clim. Change, 98, 449–469. doi:10.1007/s10584-009-9737-5.

Cai, W., and Coauthors, 2014: Increasing frequency of extreme El Niño events due to greenhouse warming. Nature Clim. Change, 4, 111–116, doi:10.1038/nclimate2100.

Caldwell, P. M., H.-N. S. Chin, D. C. Bader, and G. Bala, 2009: Evaluation of a WRF based dynamical downscaling simulation over California. Clim. Change, 95, 499–521, doi:10.1007/s10584-009-9583-5.

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson, 2001: Changes in the onset of spring in the western United States, Bull. Am. Meteorol. Soc., 82, 399–416.

Cayan, D. R., E. P. Maurer, M. D. Dettinger, M. Tyree, and K. Hayhoe, 2008: Climate change scenarios for the California region. Climatic Change, 87, 21–42, doi:10.1007/s10584-007-9377-6.

Dettinger, M. D., and D. R. Cayan, 1995: Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California, J. Clim., 8, 606–623.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, J. Atmos. Sci., 46, 3077–3107.

Duffy, P. B., and Coauthors, 2006: Simulations of present and future climates in the western United States with four nested regional climate models, J. Clim., 19, 873–895.

Giorgi, F., C. S. Brodeur, and G. T. Bates, 1994: Regional climate change scenarios over the United States produced with a nested regional climate model, J. Clim., 7, 375–399, doi:10.1175/1520-0442(1994)007<0375:RCCSOT>2.0.CO;2

Hara, M., T. Yoshikane, H. Kawase, and F. Kimura, 2008: Estimation of the Impact of Global Warming on Snow Depth in Japan by the Pseudo-Global-Warming Method. Hydrol. Res. Lett., 2, 61–64, doi:10.3178/hrl.2.61.

Hayhoe, K. and Coauthors, 2004: Emissions pathways, climate change, and impacts on California. Proc. Natl. Acad. Sci. U. S. A., 101, 12422–12427, doi:10.1073/pnas.0404500101.

Hidalgo H. G., M. D. Dettinger, D. R. Cayan, 2008: Downscaling with constructed analogues: daily precipitation and temperature fields over the Unites States, California Energy Commission technical report CEC-500-2007-123, 48 pp.

Hidalgo HG, Das T, Dettinger MD, Cayan DR, Pierce DW, Barnett TP, Bala G, Mirin A, Wood AW, Bonfils C, Santer BD, Nozawa T, 2009: Detection and attribution of streamflow timing changes to climate change in the Western United States. Journal of Climate 22:3838–3855.

IPCC [Intergovernmental Panel on Climate Change]. 2014. Climate Change 2014: Impacts, adaptation and vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernment Panel on Climate Change. Field C. B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization, J. Atm. Sci, 47(23), 2784–2802, doi:10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.

Kawase, H., T. Yoshikane, M. Hara, F. Kimura, T. Yasunari, B. Ailikun, H. Ueda, and T. Inoue, 2009: Intermodel variability of future changes in the Baiu rainband estimated by the pseudo global warming downscaling method. J. Geophys. Res., 114, D24110, doi:10.1029/2009JD011803.

Kim, J., and S. Jain, 2010: High-resolution streamflow trend analysis applicable to annual decision calendars: A western United States case study, Clim. Change, 102, 699–707, doi:10.1007/s10584-010-9933-3.

Le Quéré, C., R. Moriarty, R. M. Andrew and others, 2015: Global Carbon Budget 2014. Earth System Science Data, 7: 47–85. doi:10.5194/essd-7-47-2015.

Leung, L. R., Y. Qian, X. Bian, W. M., Washington, J. Han, and J. O. Roads, 2004: Mid-century ensemble regional climate change scenarios for the western United States, Clim. Change, 62(1-3), 75–113, doi:10.1023/B:CLIM.0000013692.50640.55.

Lins, H.F, 2012: USGS Hydro-Climatic Data Network 2009 (HCDN-2009). U.S. Geological Survey Fact Sheet. 2012-3047. http://pubs.usgs.gov/fs/2012/3047.

Maurer, E. P. (2007), Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios, Clim. Change, 82, 309–325, doi:10.1007/s10584-006-9180-9.

Maurer, E. P., and H. G. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods, Hydrol. Earth Syst. Sci., 12, 551–563, doi:10.5194/hess-12-551-2008.

McCabe, G. and M. Clark, 2005: Trends and variability in snowmelt runoff in the western United States, J. Hydrometeorol., 6, 476–482.

Mesinger, F., and Coauthors, 2006: North American regional reanalysis, Bull. Amer. Meteor. Soc., 87(3), 343–360, doi:10.1175/BAMS-87-3-343.

Mlawer, E. J., S. J., Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated ‐ k model for the longwave, J. Geophys. Res., 102(D14), 16663–16682, doi:10.1029/97JD00237.

Moore, J. N., J. T. Harper, and M. C. Greenwood, 2007:Significance of trends toward earlier snowmelt turnoff, Columbia and Missouri Basin headwaters, western United States. Geophys. Res. Lett.,34, L16402. doi:10.1029/2007GL031022.

Nakanishi, M., and H. Niino (2006), An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog, Boundary-Layer Meteor., 119, 397–407, doi:10.1007/s10546-005-9030-8.

Niu, G.-Y., and Coauthors (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116(D12109), doi:10.1029/2010JD015139.

Pal J.S., Giorgi F, Bi X.Q., Elguindi N and coauthors, 2007: Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. Bull. Amer. Meteor. Soc., 88, 1395–1409, doi:10.1175/BAMS-88-9-1395.

Pierce, D. W., T. Das, D. R. Cayan, E. P. Maurer, N. Miller, Y. Bao, and M. Tyree, 2013a: Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling, Clim. Dyn., 40(3-4), 839–856, doi:10.1007/s00382-012-1337-9.

Pierce, D. W., and Coauthors, 2013b: Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. Climate Dyn., 40, 839– 856, doi:10.1007/s00382-012-1337-9.

Qu, X. and A. Hall, 2014: On the persistent spread of snow-albedo feedback, Clim. Dyn., 42(1–2), 69–81, doi:10.1007/s00382-013-1945-z.

Rasmussen, R., C. Liu, K. Ikeda, D. Gochis, D. Yates, F. Chen, and E. Gutmann, 2011: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate, J. Clim., 24(12), 3015–3048, doi:10.1175/2010JCLI3985.1.

Rauscher, S. A., J. S. Pal, N. S. Diffenbaugh, and M. M. Benedetti, 2008: Future changes in snowmelt-driven runoff timing over the western US, Geophys. Res. Lett., 35, L16703, doi:10.1029/2008GL034424.

Reclamation, 2013: Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections: Release of Downscaled CMIP5 Climate Projections, Comparison with preceding Information, and Summary of User Needs. Prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 47 pp.

Regonda, S. K., B. Rajagopalan, M. Clark, and J. Pitlick, 2005: Seasonal cycle shifts in hydroclimatology over the western United States, J. Clim., 18, 372–384.

Salathé Jr, E. P., L. R. Leung, Y. Qian, and Y. Zhang, 2010: Regional climate model projections for the State of Washington, Clim. Change, 102(1-2), 51–75, doi:10.1007/s10584-010-9849-y.

Salathé Jr., E. P., R. Steed, C. F. Mass, P. H. Zahn, 2008: A High-Resolution Climate Model for the U.S. Pacific Northwest: Mesoscale Feedbacks and Local Responses to Climate Change, J. Clim., 21, 5708–5726, doi:10.1175/2008JCLI2090.1.

Schär, C., C. Frei, D. Lüthi, and H. C. Davies, 1996: Surrogate climate-change scenarios for regional climate models. Geophys. Res. Lett., 23(6), 669–672. doi:10.1029/96GL00265.

Skamarock, W.C., Klemp J.B., Dudhia J., Gill D.O., Barker D.M., Duda M.G., Huang X-Y, Wang W., Powers J.G., 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note, NCAR/TN-475+STR.

Slack, J. R., A. M. Lumb, and J. M. Landwehr, 1993: Hydroclimatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variation, 1874–1988, U.S. Geol. Surv. Water Resour. Invest. Rep. [CD-ROM], 93-4076.

Snyder, M. A., J. L. Bell, L. C. Sloan, P. B. Duffy, and B. Govindasamy, 2002: Climate responses to a doubling of atmospheric carbon dioxide for a climatically vulnerable region, Geophys. Res. Lett., 29(11), doi:10.1029/2001GL014431.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2004: Changes in snowmelt runoff timing in western North America under a 'Business as Usual' climate change scenario, Clim. Change, 63, 217–332.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2005: Changes toward earlier streamflow timing across western North America, J. Clim., 18, 1136–1155.

Sun, F., A. Hall, M. Schwartz, N. Berg and D. Walton, 2016: Inevitable end-of-century loss of spring snowpack over California's Sierra Nevada. In preparation.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design, Bull. Amer. Meteor. Soc., 93(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, J. Clim., 18(10), 1524–1540, doi:10.1175/JCLI3363.1.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, Mon. Wea. Rev., 136, 5095–5115, doi:10.1175/2008MWR2387.1.

Timbal, B., A. Dufour, and B. McAvaney, 2003: An estimate of future climate change for western France using a statistical downscaling technique, Clim. Dyn., 20(7-8), 807–823, doi:10.1007/s00382-002-0298-9.

USDA Forest Service, 2009: The importance of streamflow in California's southern Sierra Nevada mountains: Kings River Experimental Watersheds. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 2 p http://www.fs.fed.us/psw/topics/water/kingsriver/documents/brochures_handouts/Stream_Discharge_PUBLIC.pdf

van Vuuren, D.P., J. Edmonds, and Coauthors, 2011: The representative concentration pathways: an overview. Climatic Change, 109(1–2): 5–31. doi: 10.1007/s10584-011-0148-z.

Walton, D.W., A. Hall, N. Berg, M. Schwartz and F. Sun, 2016: Incorporating snow albedo feedback into downscaled temperature and snow cover projections for California's Sierra Nevada. Journal of Climate, in press. http://research.atmos.ucla.edu/csrl/publications/Hall/Walton_JCLI-D-16-0168_R2.pdf

Washington, W. M., J.W. Weatherly, G.A. Meehl, A.J. Semtner, T.W. Bettge, A.P. Craig, W.G. Strand, J.M. Arblaster, Wayland, V. B., James, R., and Zhang, Y., 2000: 'Parallel Climate Model (PCM) Control and Transient Simulations, Clim. Dyn. 16, 755–774.

Wenger, S. J., C. H. Luce, A. F. Hamlet, D. J. Isaak, andH. M. Neville, 2010: Macroscale hydrologic modeling of ecologically relevant flow metrics,Water Resour.Res., 46,W09513, doi:10.1029/2009WR008839.

Willis, A.D., J. R. Lund, E.S. Townsley, and B. A. Faber, 2011: Climate change and flood operations in he Sacramento Basin, California. San Francisco Estuary Watershed Science, 9 (2), 1-17.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys. Res., 107(D20), 4429, doi:10.1029/2001JD000659.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Clim. Change 62,189–216, doi:10.1023/B:CLIM.0000013685.99609.9e.

Table 1: Summary of information associated with observational streamflow gauges from the United States Geological Survey Hydro-Climatic Data Network-2009 used to evaluate the baseline simulation.

USGS HCDN-2009 ID	Station Name	Hydrologic Unit Code	Latitude	Longitude	Drainage area (sq. km)
10308200	East Fork Carson River below Markleeville Creek	16050201	38.714	-119.764	716.4
10336645	General Creek near Meeks Bay, CA	16050101	39.051	-120.118	19.6
10336660	Blackwood Creek near Tahoe City, CA	16050101	39.107	-120.162	29.8
10336676	Ward Creek at State Highway 89, near Tahoe Pines, CA	16050101	39.132	-120.157	24.7
10336740	Logan House Creek near Glenbrook, NV	16050101	39.066	-119.935	5.5
11230500	Bear Creek near Lake Thomas A. Edison, CA	18040006	37.339	-118.973	135.5
11237500	Pitman Creek below Tamarack Creek, CA	18040006	37.198	-119.213	59.8
11264500	Merced River at Happy Isles Bridge, near Yosemite, CA	18040008	37.731	-119.558	468.0
11266500	Merced River at Pohono Bridge, near Yosemite, CA	18040008	37.716	-119.666	833.1
11315000	Cole Creek near Salt Springs Dam, CA	18040012	38.519	-120.212	54.0
11427700	Duncan Canyon Creek near French Meadows, CA	18020128	39.135	-120.478	25.5

Figure Captions

Fig. 1: a) Model setup, with three nested WRF domains at resolutions of 27, 9, and 3 km (from the outermost to innermost domain). Topography (m) is shown at the resolution of the 27km domain in color and black lines show boundaries for US states. (b) Topography (m) of the innermost domain (3-km resolution) of the regional simulation, with the state borders of California and Nevada in black. Blue circles show the locations of 11 USGS-HCDN 2009 streamflow gauges used for model evaluation.

Fig. 2: Baseline (October 1991–September 2001) climatological date of R_{50}, which represents the date in the water year (October 1–September 30) by which 50% of the cumulative surface runoff has occurred. The black contour outlines grid points with climatological R_{50} occurring on or after March 1st.

Fig. 3: Observed versus WRF-simulated climatological R_{50} at 11 USGS streamflow gauges (water years 1992–2001). Simulated R_{50} is estimated as the average R_{50} of grid points upstream of a gauge within its watershed. Colors indicate the correlation coefficient between the time series of WRF-simulated and observed values of R_{50}. The line y = x is shown in black.

Fig. 4: End-of-21st-century change ($WY_{2092–2101}$ average minus $WY_{1992–2001}$ average) in R_{50} (unit: days) under the RCP8.5 emissions scenario for CNRM-CM5, GFDL-CM3, inmcm4, IPSL-CM5A-LR, and MPI-ESM-LR produced from three methods. Row 1: Dynamically-downscaled WRF output. Row 2: Statistical projection using dynamically-downscaled WRF spring near-surface warming (ΔT_{MAM}) as input. Row 3: Statistical projection using Walton et al. (2016)'s hybrid dynamical-statistical downscaled ΔT_{MAM} as input. Results are shown for locations with climatological baseline R_{50} on or after March 1st, and green through blue shades represent advances in R_{50}. Black text shows domain-average in R_{50}. Blue text in rows 2–3

denotes the mean absolute error compared to row 1. Green text in rows 2–3 denotes the spatial correlation with row 1 for each GCM.

Fig. 5: End-of-21st-century change in near-surface temperature averaged over March–May (ΔT_{MAM}, unit: °C) under the RCP8.5 forcing scenario for CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR. Row 1: WRF dynamically-downscaled output. Row 2: Hybrid dynamical-statistical downscaled output from Walton et al. (2016). Black text shows domain-average ΔT_{MAM}.

Fig. 6: (a) Correlation coefficient between the 5-model dynamically-downscaled end-of-21st-century change in R_{50} timing (ΔR_{50}) and near-surface March–May warming (ΔT_{MAM}). (b) Slope of the linear regression of the 5-model dynamically-downscaled ΔR_{50} onto the 5-model dynamically-downscaled ΔT_{MAM}. Unit: days/°C. Black text denotes the domain average value.

Fig. 7: Scatter plot of observed near-surface temperature anomalies (unit: °C) averaged over March–May (T_{MAM}) and observed R_{50} anomalies (unit: days) over water years 1916–2014. The blue line is the linear regression of $\mathrm{WY}_{1916\text{–}2014}$ R_{50} onto T_{MAM}. Blue text denotes the slope of this linear regression as well as the correlation coefficient. MAM 2-m temperature anomalies are calculated from the National Oceanic and Atmospheric Administration's National Climatic Data Center’s nClimDiv statewide temperature database (ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/state-readme.txt), which includes monthly-mean maximum and minimum temperature aggregated at statewide levels for the United States for January 1895 to the present. Monthly maximum and minimum temperatures are averaged together to calculate monthly mean temperature. MAM temperature anomalies presented here are calculated from the detrended MAM temperature time series for California. R_{50} anomalies

are calculated from the detrended gauge-averaged R_{50} time series from available observations at the 11 USGS-HCDN streamflow gauges in Table 1 (described in section 2b).

Fig. 8: Row 1: Ensemble-mean statistical projections of end-of-21st-century change in R_{50} (unit: days) under emissions scenarios (a) RCP2.6, (b) RCP4.5, (c) RCP6.0 and (d) RCP8.5. Row 2 (e-h): The associated z-score for the ensemble-mean change in R_{50}, which is calculated by dividing the mean R_{50} change by the standard deviation of R_{50} of a 20-year baseline (water years 1982–2001). Black text denotes the domain average value. The number of GCMs included in the ensemble-mean is denoted in the title.

Fig. 9: Statistical projections of end-of-21st-century change in R_{50} as a function of elevation (binned every 100m) under emissions scenarios RCP2.6, 4.5, RCP6.0 and RCP8.5. Solid colored lines represent the ensemble-mean R_{50} change calculated with hybrid dynamical-statistical spring warming as input, while dashed colored lines represent the 10th and 90th percentiles of this GCM distribution. Light gray shading denotes the standard deviation of R_{50} for the extended baseline period (water years 1982–2001). The region outside of the dark gray shading denotes mean changes in R_{50} that are significant at the 5% level according to a one-tailed t-test. Thin black (green) lines represent the ensemble-mean R_{50} change calculated with BCSD-downscaled (BCCA-downscaled) spring warming as input. Results are shown for locations with climatological baseline R_{50} on or after March 1st. The number of GCMs included in the hybrid-downscaled GCM ensemble is denoted in the title.

Fig. 10: Distribution of R_{50} dates for (a) the end-of-20th-century baseline (water years 1992–2001) and (b) end-of-21st-century (water years 2092–2101) under RCP8.5. R_{50} dates are binned in 5-day intervals. We consider gridpoints with snowmelt-dominated runoff in the baseline, so the distribution is calculated based on gridpoints within the contoured region in Fig. 2. The

distribution of R_{50} dates in (a) is based on the WRF dynamical downscaling simulation; the distribution of R_{50} dates in (b) is based on the 35-model ensemble-mean statistical R_{50} projection. Text within each subplot denotes the percent of gridpoints with R_{50} dates falling in each of the months of February through June.

Fig. 11: End-of-21st-century change ($WY_{2092\text{-}2101}$ average minus $WY_{1992\text{-}2001}$ average) in near-surface March–May temperature (ΔT_{MAM}, unit: °C) under the RCP8.5 forcing scenario averaged over five GCMs (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR) downscaled using 5 methods: (a) WRF dynamical downscaling, (b) Walton et al. (2016)'s statistical downscaling, (c) linear interpolation, (d) BCCA, and (e) BCSD. Black text denotes domain-average warming within black contoured region. Red text in b-e denotes the spatial correlation with (a) within the black contoured region.

Fig. 12: Statistical projection of end-of-21st-century change in R_{50} (unit: days) under the RCP8.5 forcing scenario calculated with spring warming from **(a)** WRF dynamical downscaling, **(b)** hybrid dynamical-statistical downscaling, **(c)** linear interpolation of GCM output, **(d)** BCCA statistical downscaling, and **(e)** BCSD statistical downscaling. Results are averaged over five GCMs (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR). In **b-e,** **a** is subtracted to highlight differences. Results are shown for locations with climatological baseline R_{50} on or after March 1st, and black text denotes domain-average value.

Significant and inevitable end-of-21st-century advances in surface runoff timing in California's Sierra Nevada

Marla Schwartz, Alex Hall, Fengpeng Sun, Daniel Walton, Neil Berg

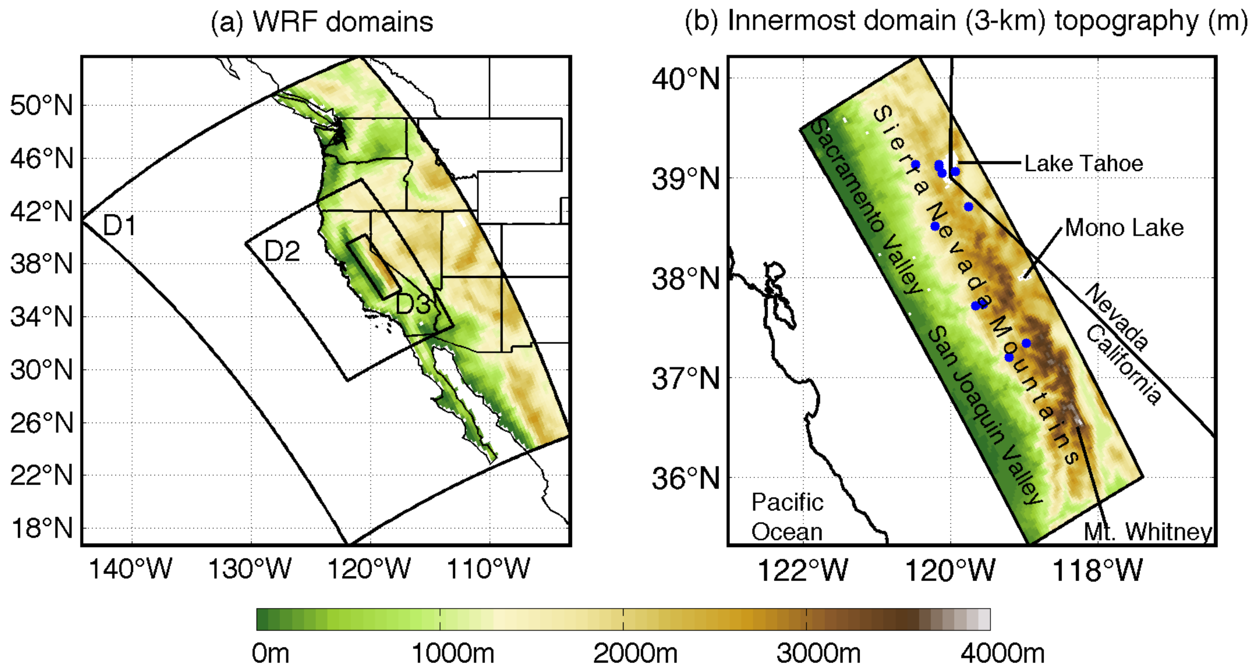

Fig. 1: a) Model setup, with three nested WRF domains at resolutions of 27, 9, and 3 km (from the outermost to innermost domain). Topography (m) is shown at the resolution of the 27km domain in color and black lines show boundaries for US states. (b) Topography (m) of the innermost domain (3-km resolution) of the regional simulation, with the state borders of California and Nevada in black. Blue circles show the locations of 11 USGS-HCDN 2009 streamflow gauges used for model evaluation.

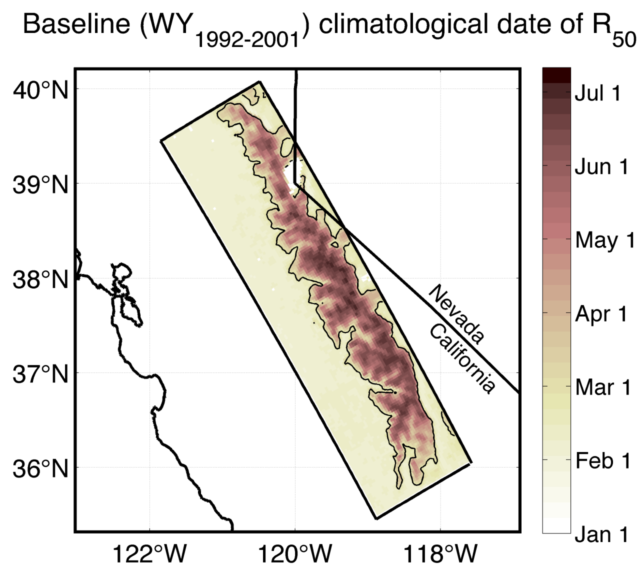

Fig. 2: Baseline (October 1991–September 2001) climatological date of R_{50}, which represents the date in the water year (October 1–September 30) by which 50% of the cumulative surface runoff has occurred. The black contour outlines grid points with climatological R_{50} occurring on or after March 1st.

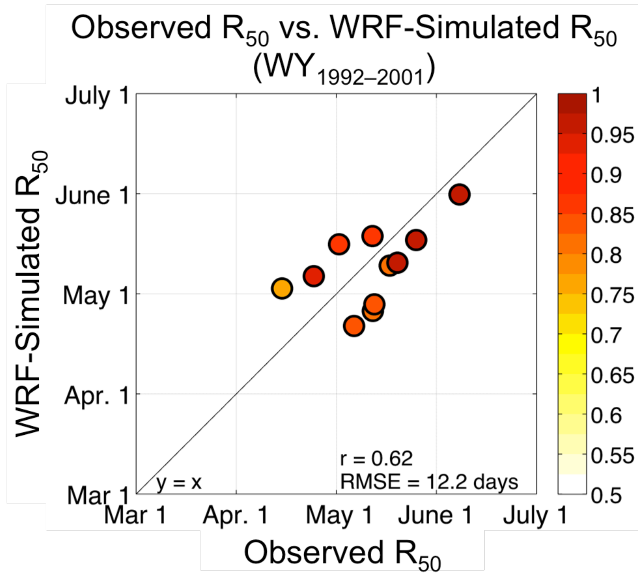

Fig. 3: Observed versus WRF-simulated climatological R_{50} at 11 USGS streamflow gauges (water years 1992–2001). Simulated R_{50} is estimated as the average R_{50} of grid points upstream of a gauge within its watershed. Colors indicate the correlation coefficient between the time series of WRF-simulated and observed values of R_{50}. The line y = x is shown in black.

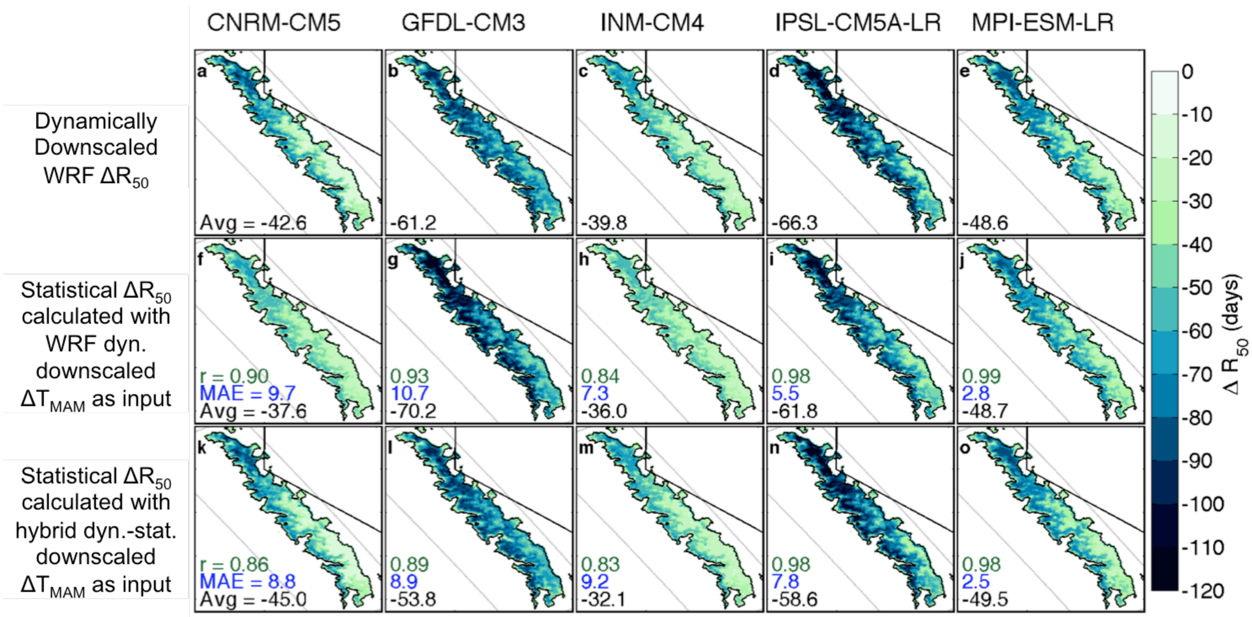

Fig. 4: End-of-21st-century change ($WY_{2092–2101}$ average minus $WY_{1992–2001}$ average) in R_{50} (unit: days) under the RCP8.5 emissions scenario for CNRM-CM5, GFDL-CM3, inmcm4, IPSL-CM5A-LR, and MPI-ESM-LR produced from three methods. Row 1: Dynamically-downscaled WRF output. Row 2: Statistical projection using dynamically-downscaled WRF spring near-surface warming (ΔT_{MAM}) as input. Row 3: Statistical projection using Walton et al. (2016)'s hybrid dynamical-statistical downscaled ΔT_{MAM} as input. Results are shown for locations with climatological baseline R_{50} on or after March 1st, and green through blue shades represent advances in R_{50}. Black text shows domain-average in R_{50}. Blue text in rows 2–3 denotes the mean absolute error compared to row 1. Green text in rows 2–3 denotes the spatial correlation with row 1 for each GCM.

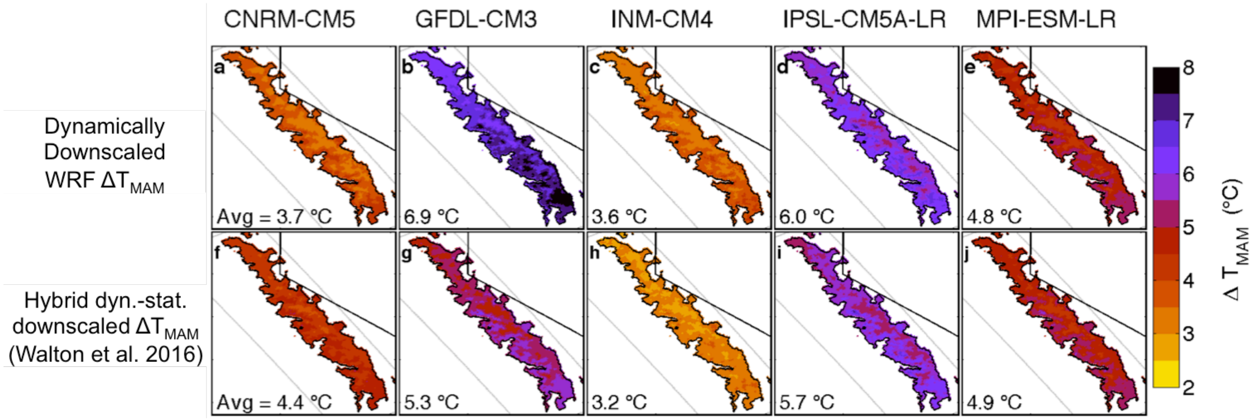

Fig. 5: End-of-21st-century change in near-surface temperature averaged over March–May (ΔT_{MAM}, unit: °C) under the RCP8.5 forcing scenario for CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR. Row 1: WRF dynamically-downscaled output. Row 2: Hybrid dynamical-statistical downscaled output from Walton et al. (2016). Black text shows domain-average ΔT_{MAM}.

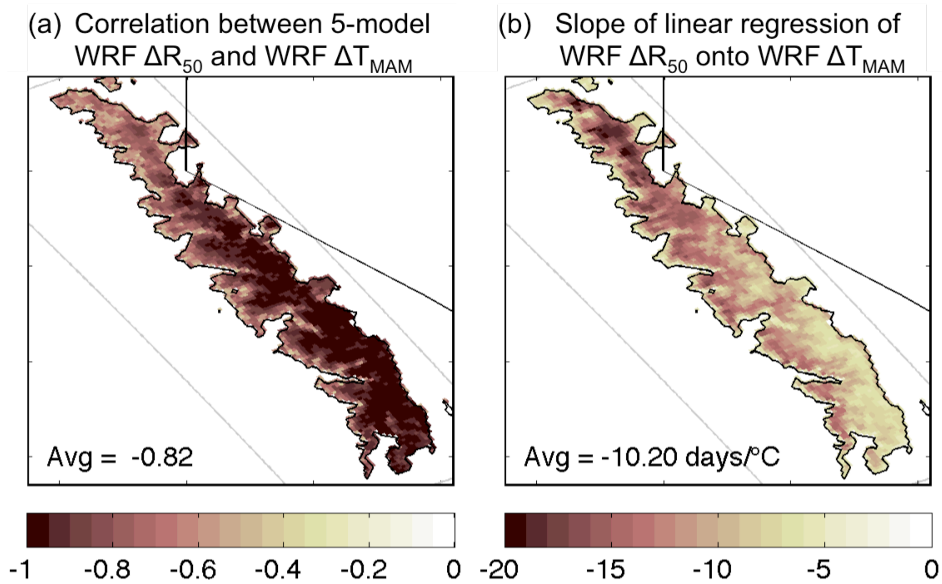

Fig. 6: (a) Correlation coefficient between the 5-model dynamically-downscaled end-of-21st-century change in R_{50} timing (ΔR_{50}) and near-surface March–May warming (ΔT_{MAM}). (b) Slope of the linear regression of the 5-model dynamically-downscaled ΔR_{50} onto the 5-model dynamically-downscaled ΔT_{MAM}. Unit: days/°C. Black text denotes the domain average value.

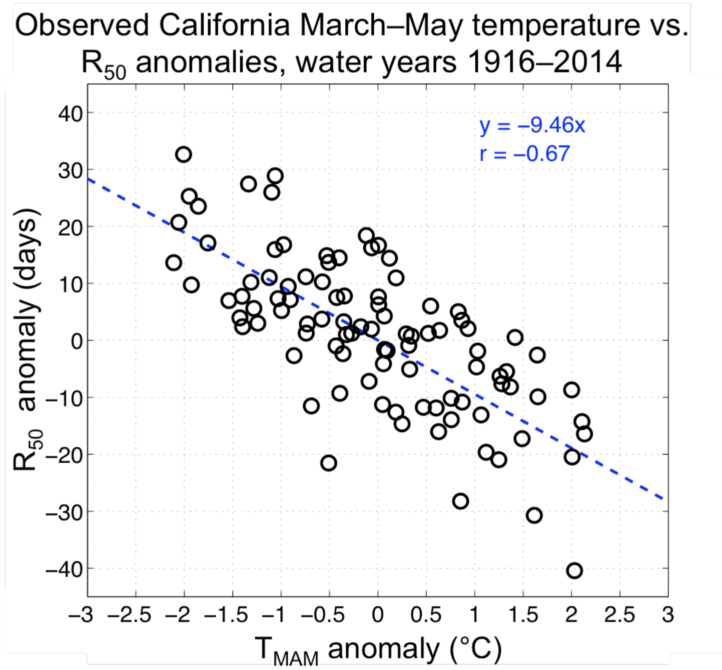

Fig. 7: Scatter plot of observed near-surface temperature anomalies (unit: °C) averaged over March–May (T_{MAM}) and observed R_{50} anomalies (unit: days) over water years 1916–2014. The blue line is the linear regression of $\mathrm{WY}_{1916\text{–}2014}$ R_{50} onto T_{MAM}. Blue text denotes the slope of this linear regression as well as the correlation coefficient. MAM 2-m temperature anomalies are calculated from the National Oceanic and Atmospheric Administration's National Climatic Data Center's nClimDiv statewide temperature database (ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/state-readme.txt), which includes monthly-mean maximum and minimum temperature aggregated at statewide levels for the United States for January 1895 to the present. Monthly maximum and minimum temperatures are averaged together to calculate monthly mean temperature. MAM temperature anomalies presented here are calculated from the detrended MAM temperature time series for California. R_{50} anomalies are calculated from the detrended gauge-averaged R_{50} time series from available observations at the 11 USGS-HCDN streamflow gauges in Table 1 (described in section 2b).

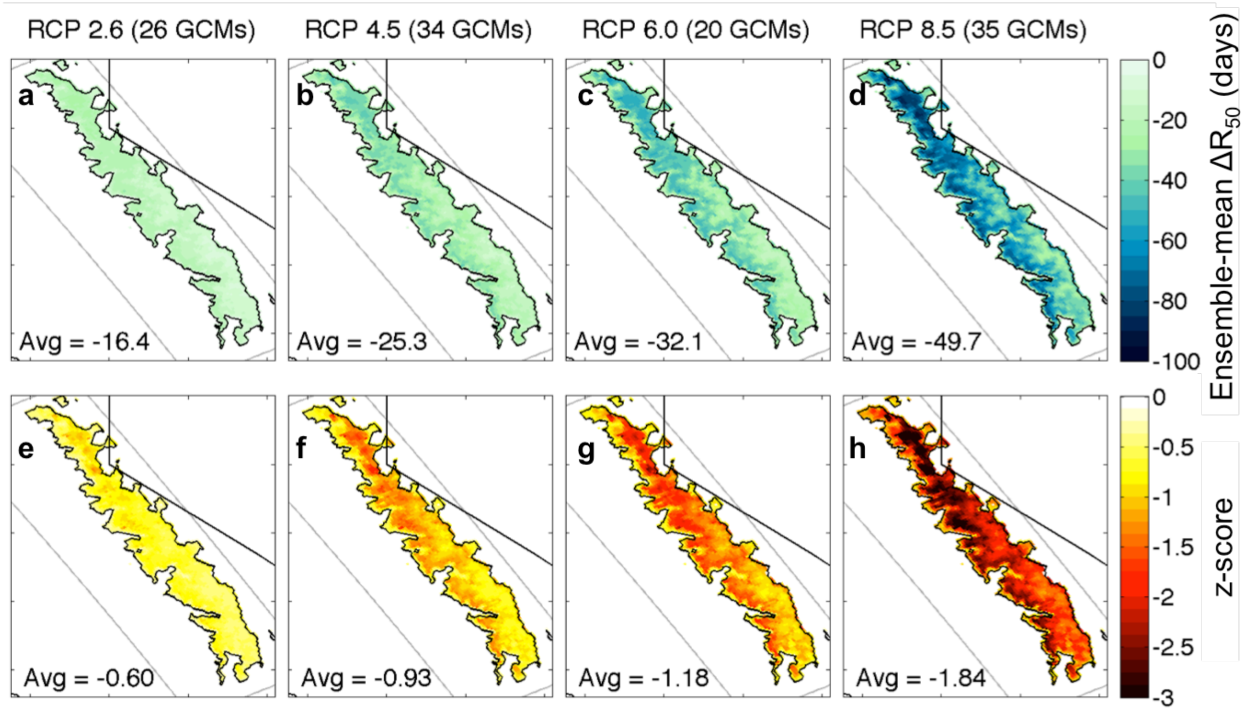

Fig. 8: Row 1: Ensemble-mean statistical projections of end-of-21st-century change in R_{50} (unit: days) under emissions scenarios (a) RCP2.6, (b) RCP4.5, (c) RCP6.0 and (d) RCP8.5. Row 2 (e-h): The associated z-score for the ensemble-mean change in R_{50}, which is calculated by dividing the mean R_{50} change by the standard deviation of R_{50} of a 20-year baseline (water years 1982–2001). Black text denotes the domain average value. The number of GCMs included in the ensemble-mean is denoted in the title.

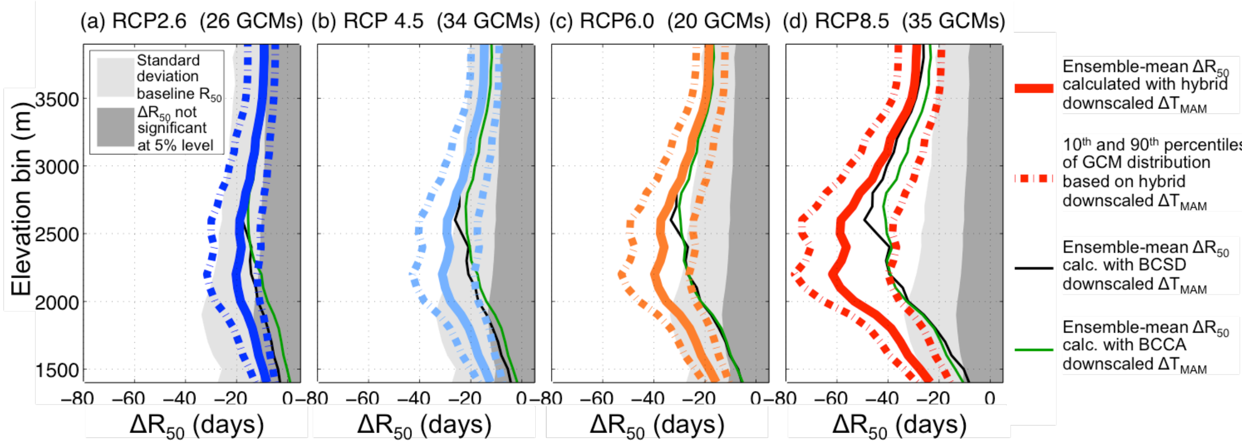

Fig. 9: Statistical projections of end-of-21st-century change in R_{50} as a function of elevation (binned every 100m) under emissions scenarios RCP2.6, 4.5, RCP6.0 and RCP8.5. Solid colored lines represent the ensemble-mean R_{50} change calculated with hybrid dynamical-statistical spring warming as input, while dashed colored lines represent the 10th and 90th percentiles of this GCM distribution. Light gray shading denotes the standard deviation of R_{50} for the extended baseline period (water years 1982–2001). The region outside of the dark gray shading denotes mean changes in R_{50} that are significant at the 5% level according to a one-tailed t-test. Thin black (green) lines represent the ensemble-mean R_{50} change calculated with BCSD-downscaled (BCCA-downscaled) spring warming as input. Results are shown for locations with climatological baseline R_{50} on or after March 1st. The number of GCMs included in the hybrid-downscaled GCM ensemble is denoted in the title.

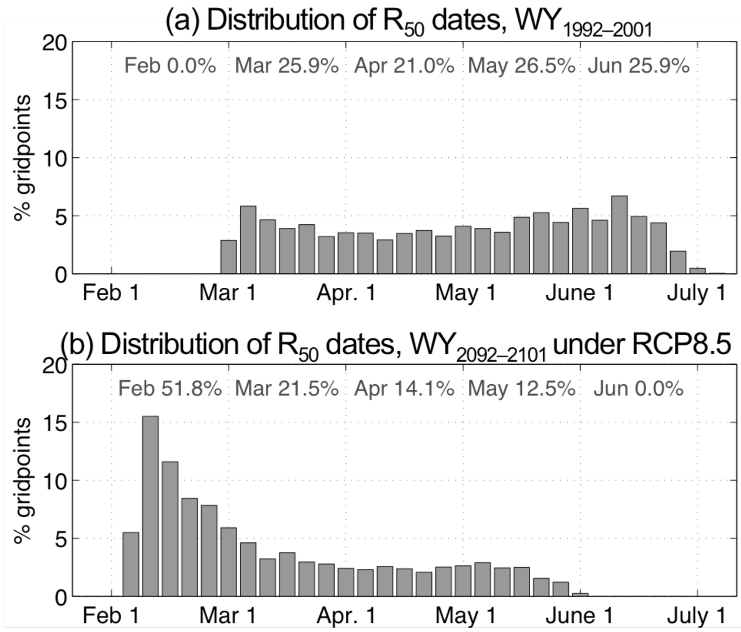

Fig. 10: Distribution of R_{50} dates for (a) the end-of-20th-century baseline (water years 1992–2001) and (b) end-of-21st-century (water years 2092–2101) under RCP8.5. R_{50} dates are binned in 5-day intervals. We consider gridpoints with snowmelt-dominated runoff in the baseline, so the distribution is calculated based on gridpoints within the contoured region in Fig. 2. The distribution of R_{50} dates in (a) is based on the WRF dynamical downscaling simulation; the distribution of R_{50} dates in (b) is based on the 35-model ensemble-mean statistical R_{50} projection. Text within each subplot denotes the percent of gridpoints with R_{50} dates falling in each of the months of February through June.

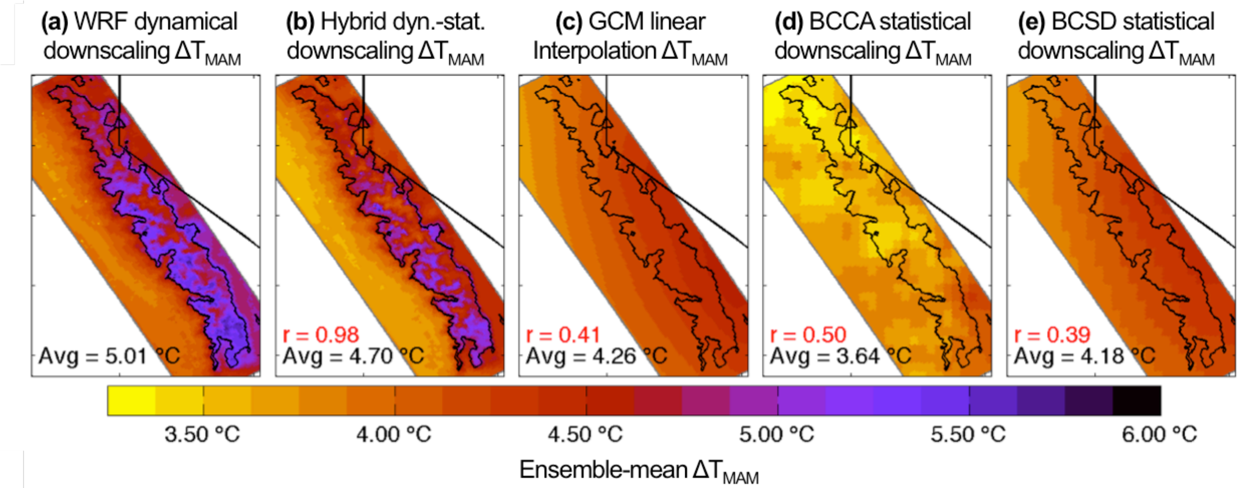

Fig. 11: End-of-21st-century change ($WY_{2092-2101}$ average minus $WY_{1992-2001}$ average) in near-surface March–May temperature (ΔT_{MAM}, unit: °C) under the RCP8.5 forcing scenario averaged over five GCMs (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR) downscaled using 5 methods: (a) WRF dynamical downscaling, (b) Walton et al. (2016)'s statistical downscaling, (c) linear interpolation, (d) BCCA, and (e) BCSD. Black text denotes domain-average warming within black contoured region. Red text in b-e denotes the spatial correlation with (a) within the black contoured region.

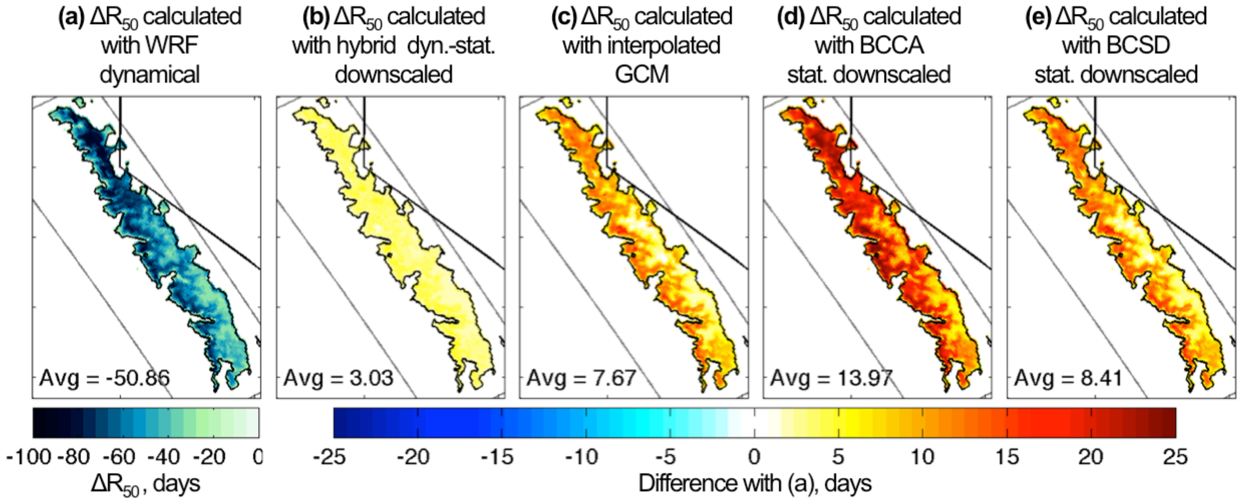

Fig. 12: Statistical projection of end-of-21st-century change in R_{50} (unit: days) under the RCP8.5 forcing scenario calculated with spring warming from **(a)** WRF dynamical downscaling, **(b)** hybrid dynamical-statistical downscaling, **(c)** linear interpolation of GCM output, **(d)** BCCA statistical downscaling, and **(e)** BCSD statistical downscaling. Results are averaged over five GCMs (CNRM-CM5, GFDL-CM3, INM-CM4, IPSL-CM5A-LR, and MPI-ESM-LR). In **b-e, a** is subtracted to highlight differences. Results are shown for locations with climatological baseline R_{50} on or after March 1st, and black text denotes domain-average value.