Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts

PABLO A. MENDOZA

Department of Civil, Environmental, and Architectural Engineering, and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

MARTYN P. CLARK, NAOKI MIZUKAMI, ANDREW J. NEWMAN, MICHAEL BARLAGE, ETHAN D. GUTMANN, AND ROY M. RASMUSSEN

Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

BALAJI RAJAGOPALAN

Department of Civil, Environmental, and Architectural Engineering, and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

LEVI D. BREKKE

Bureau of Reclamation, Denver, Colorado

JEFFREY R. ARNOLD

U.S. Army Corps of Engineers, Seattle, Washington

(Manuscript received 3 June 2014, in final form 25 September 2014)

ABSTRACT

The assessment of climate change impacts on water resources involves several methodological decisions, including choices of global climate models (GCMs), emission scenarios, downscaling techniques, and hydrologic modeling approaches. Among these, hydrologic model structure selection and parameter calibration are particularly relevant and usually have a strong subjective component. The goal of this research is to improve understanding of the role of these decisions on the assessment of the effects of climate change on hydrologic processes. The study is conducted in three basins located in the Colorado headwaters region, using four different hydrologic model structures [PRMS, VIC, Noah LSM, and Noah LSM with multiparameterization options (Noah-MP)]. To better understand the role of parameter estimation, model performance and projected hydrologic changes (i.e., changes in the hydrology obtained from hydrologic models due to climate change) are compared before and after calibration with the University of Arizona shuffled complex evolution (SCE-UA) algorithm. Hydrologic changes are examined via a climate change scenario where the Community Climate System Model (CCSM) change signal is used to perturb the boundary conditions of the Weather Research and Forecasting (WRF) Model configured at 4-km resolution. Substantial intermodel differences (i.e., discrepancies between hydrologic models) in the portrayal of climate change impacts on water resources are demonstrated. Specifically, intermodel differences are larger than the mean signal from the CCSM–WRF climate scenario examined, even after the calibration process. Importantly, traditional single-objective calibration techniques aimed to reduce errors in runoff simulations do not necessarily improve intermodel agreement (i.e., same outputs from different hydrologic models) in projected changes of some hydrological processes such as evapotranspiration or snowpack.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Pablo A. Mendoza, Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, UCB 428, Boulder, CO 80309.
E-mail: pmendoza@colorado.edu

DOI: 10.1175/JHM-D-14-0104.1

1. Introduction

There is now general agreement in the scientific community that the rising levels of carbon dioxide in the atmosphere are modifying historical climate conditions (Stocker et al. 2014). One of the most relevant impacts of future climate change on society is changes in regional water availability for municipal, industrial, mining, irrigation, hydropower generation, and other activities (Xu 1999; Brekke et al. 2009; Wagener et al. 2010). This situation is particularly critical for the Colorado River basin (CRB) because of the susceptibility of runoff variations due to changes in precipitation and temperature, which stem from changes in evapotranspiration ET processes and snowpack accumulation–melt patterns (Christensen and Lettenmaier 2007). This vulnerability, together with the importance of the CRB for water resources supply for the growing regions of western and southwestern United States, has motivated many climate change studies in this area based on different modeling approaches and therefore resulting in a diverse set of conclusions (Milly et al. 2005; Christensen and Lettenmaier 2007; Hoerling and Eischeid 2007; Ray et al. 2008; Hoerling et al. 2009; Rasmussen et al. 2011, 2014; Miller et al. 2011, 2012; Vano et al. 2012, 2014).

The large uncertainty in estimates of hydrologic changes (i.e., changes in hydrologic variables obtained from hydrologic models) due to climate perturbation is not surprising for the hydrologic research community. In recent decades, many sources of uncertainty for quantifying climate change impacts on water resources have been identified (Chen et al. 2011), including: 1) selection of greenhouse gas emission scenarios, 2) choice of climate model(s), 3) specification of climate model initial conditions, 4) choice of meteorological forcing downscaling methods, 5) selection of hydrological model structures, and 6) choice of hydrological model parameter sets. Understanding risks associated with climate change requires estimating the uncertainty at each step of the modeling process (Xu 1999; Bergström et al. 2001; Wilby 2005; Wilby and Harris 2006; Graham et al. 2007; Chen et al. 2011; Vano et al. 2014). Among these elements, the choices of climate model (Murphy et al. 2004) and downscaling methods (Gutmann et al. 2012, 2014) have received significant attention, because recent studies have found that these are the main contributors to overall uncertainty (Wilby and Harris 2006; Chen et al. 2011).

Although a considerable number of past studies focused on the treatment of uncertainty in climate change projections, only a few have focused on hydrologic model structures and parameter uncertainty. For instance, Wilby (2005) explored parameter stability and identifiability using two hydrologic model structures, finding 1) that transferability of model parameters between wet and dry periods depends on the representativeness of the training period and 2) that model structure uncertainty on projected streamflow can be comparable to the uncertainty due to choice of emission scenario when the simplest model (low-flow period) is considered. Jones et al. (2006) applied three different models in 22 Australian catchments covering a wide range of climates and demonstrated that runoff variations due to changes in rainfall and evapotranspiration are clearly model dependent. Jiang et al. (2007) compared outputs from six hydrological models for mean annual and monthly changes in hydrologic variables due to perturbations of precipitation and temperature, finding 1) that differences across models depend on the climate scenario, the season, and the variable of interest and 2) that models without thresholds in soil moisture have larger differences in projected changes in soil storage. Poulin et al. (2011) used two different hydrological models to compare the effects of model structure against parameter equifinality on the uncertainty of hydrologic simulations, finding that model structure uncertainty dominates. More recently, Miller et al. (2012) found that hydrologic model choice has a large effect on the portrayal of climate change impact in the San Juan River basin. Vano et al. (2012) evaluated hydrologic changes due to perturbed climate scenarios using six hydrologic–land surface models in the CRB, demonstrating large intermodel differences in runoff changes due to shifts in precipitation and temperature. Surfleet et al. (2012) compared a large-scale approach, a basin-scale approach, and a site-specific approach in the Santiam River basin (United States), showing that differences in the portrayal of climate change impacts can be attributed to scale and the ability of the models to capture local hydrological processes.

Despite the increasing awareness of the implications of hydrologic model structures on the estimation of climate change impacts on hydrology, the effects of model representation of specific processes (e.g., evapotranspiration, snow accumulation and ablation, and percolation) on the overall hydrologic model response still remains unclear. In view of this, the main goal of this paper is to compare hydrologic changes obtained with different hydrologic model structures in terms of annual water balance, monthly simulated processes (e.g., ET, snowpack, and soil moisture), and signature measures of hydrologic behavior (e.g., runoff seasonality and long-term base flow) for uncalibrated and calibrated model simulations.

2. Study area

The headwaters of the CRB are snow dominated, with approximately 85% of the streamflow resulting from

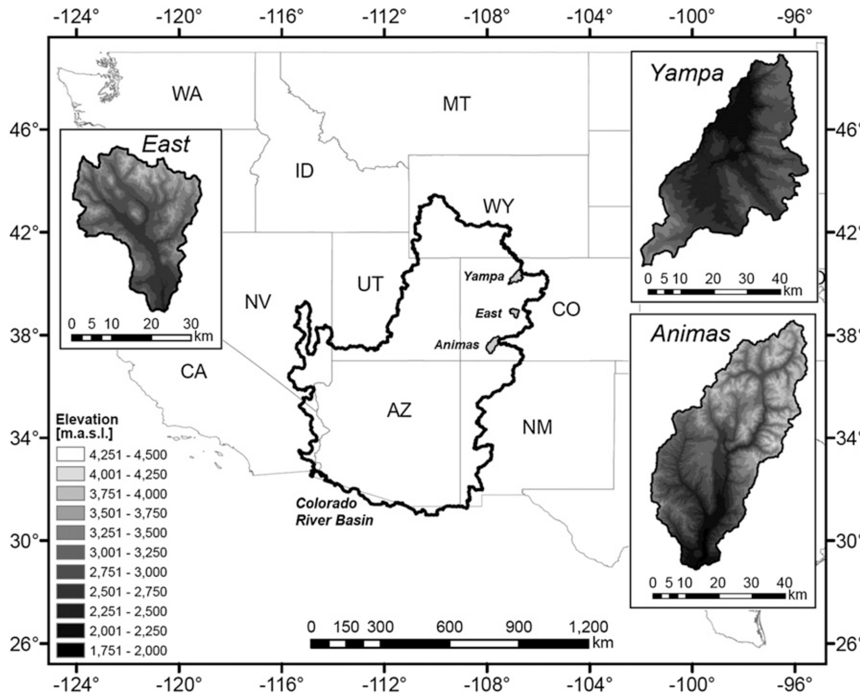

FIG. 1. Location of the basins of interest.

snowmelt. Changes in snowpack can therefore have a large impact on hydrologic processes within the Colorado headwaters (Miller and Piechota 2008). The water resources in the CRB are currently allocated to seven states and Mexico for consumption, irrigation, and hydropower, among other uses. The importance of the CRB for water management and decision making, together with strong evidence of a shift in the hydroclimatology over the past decades (e.g., Miller and Piechota 2008, 2011), has motivated several studies to generate streamflow projections under different future climate scenarios (e.g., Milly et al. 2005; Christensen and Lettenmaier 2007; Hoerling et al. 2009; Bureau of Reclamation 2012). We conduct this study over three basins in the Colorado headwaters region—Yampa River at Steamboat Springs, East River at Almont, and Animas River at Durango—whose location and elevation ranges are shown in Fig. 1. These basins are representative of the main hydroclimatic characteristics of other gauged, unregulated headwater basins throughout the upper Colorado River basin (not shown). Moreover, these catchments have been included in many past climate change studies (e.g., Wilby et al. 1999; Sankarasubramanian and Vogel 2002; Mastin et al. 2011; Milly and Dunne 2011) and, because of their relatively small size compared to the CRB, they offer a unique opportunity to perform extensive analysis involving thousands of model runs (e.g., sensitivity analysis and hydrologic model calibration), to evaluate different approaches in climate change impact assessment, and also to provide detailed understanding of physical processes in the headwaters of the CRB.

Table 1 summarizes the main hydroclimatic characteristics of the three basins for which historical data are available, over an 8-yr period (from October 2000 to September 2008). Mean basin precipitation ranges between 700 and 900 mm yr^{-1}, while mean basin elevation is above 2500 m MSL. Among these basins, the Yampa River at Steamboat Springs has the lowest runoff ratio (smallest runoff and largest precipitation amounts), and the East River at Almont has the highest runoff ratio. The land surface of the Yampa and Animas River basins is predominantly covered by deciduous forests (26% at Yampa and 23% at Animas) and evergreen forests (37% at Yampa and 39% at Animas), while the land surface of the East River basin is mainly covered by evergreen forests (29%) and grassland–herbaceous (26%).

3. Methods

a. Meteorological forcings

We use outputs from the regional Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008)

TABLE 1. Characteristics of the three study watersheds. Hydrologic variables correspond to the period from October 2000 to September 2008. Variables P, R, PE, RE, and DI denote basin-averaged mean annual values of precipitation, runoff, potential evapotranspiration, runoff efficiency, and dryness index, respectively. Values of PE are obtained from PRMS by using a Jensen–Haise formulation (Jensen et al. 1970).

Location	Area (km^2)	Mean basin elevation (m MSL)	Mean annual runoff ($mm\,yr^{-1}$)	Mean precipitation from WRF ($mm\,yr^{-1}$)	Mean annual PE ($mm\,yr^{-1}$)	Mean annual RE (R/P)	Mean annual DI (PE/P)
Yampa at Steamboat Springs	1468	2674	228	717	953	0.32	1.33
East at Almont	748	3127	327	782	757	0.42	0.97
Animas at Durango	1819	3098	365	883	885	0.41	1.00

to force our hydrological simulations. The datasets come from the WRF historical runs and pseudo global warming (PGW) simulations with horizontal grid spacing of 4 km described in Rasmussen et al. (2014). The model physics options used in that study included the Noah land surface model, version 3.2 (Noah LSM), with upgraded snow physics (Chen and Dudhia 2001; Barlage et al. 2010); the Thompson mixed-phase cloud microphysics scheme (Thompson et al. 2008); the Yonsei University planetary boundary layer (Hong et al. 2006); and the Community Atmosphere Model's (CAM) longwave and shortwave radiation schemes (Collins et al. 2006). In current climate, the WRF simulations have been validated against SNOTEL sites, and precipitation spatial variability, timing, and intensities are well represented by the model (Ikeda et al. 2010; Prein et al. 2013).

The PGW approach (Schär et al. 1996; Hara et al. 2008; Kawase et al. 2009) consists of adding a mean climate perturbation to the initial and 3-hourly boundary conditions, here taken from the North American Regional Reanalysis (NARR; Mesinger et al. 2006). The climate perturbation used was based on expected changes from the NCAR CCSM3 forced by the A1B scenario. This perturbation is generated by subtracting the current 10-yr (1995–2005) monthly climatology from a future 10-yr (2045–55) monthly climatology. A detailed description of this approach can be found in Rasmussen et al. (2011, 2014).

Meteorological data from WRF simulations are available at hourly time steps and a 4-km resolution for both historical and PGW conditions during the period from October 2000 to September 2008. The variables and temporal disaggregation used depend on specific hydrologic model requirements (Table 2). Figure 2 includes basin-averaged monthly precipitation and temperature from WRF for current and future climate scenarios over the period from October 2002 to September 2008. Note that PGW simulations reflect increases in precipitation during fall and winter and the beginning of spring and a decrease in precipitation during summer over all basins. On the other hand, the increase in temperature tends to be uniform throughout the year in all basins. These signals in precipitation and temperature changes are present at each individual water year (not shown), although monthly precipitation amounts can vary at the basins of interest from year to year.

The single choice of GCM, emission scenario, and the time period over which the climate perturbation was

TABLE 2. Summary of data sources and simulation setup used in this study. For the forcing variables, air temperature at 2 m and wind speed at 10 m are used for hydrologic simulations.

Model	Vegetation data	Soil data	Forcing variables	Spatial–temporal discretization
PRMS	USGS 1-km gridded vegetation type and density data (USDA 1992)	State soil geographic (STATSGO) 1-km gridded soils data (USDA 1994)	Daily precipitation; maximum and minimum daily temperature.	4 km and $\Delta t = 24$ h
VIC	University of Maryland 1-km Global Land Cover Classification (Hansen et al. 2000)	STATSGO 1-km gridded soils data (USDA 1994)	Precipitation, temperature, shortwave and longwave radiation, wind speed, relative humidity, and air pressure.	4 km and $\Delta t = 1$ h
Noah LSM and Noah-MP	National Land Cover Database, 2006 (Fry et al. 2011).	STATSGO 1-km gridded soils data (USDA 1994)	Precipitation, temperature, shortwave and longwave radiation, wind speed, relative humidity, and air pressure.	4 km and $\Delta t = 1$ h

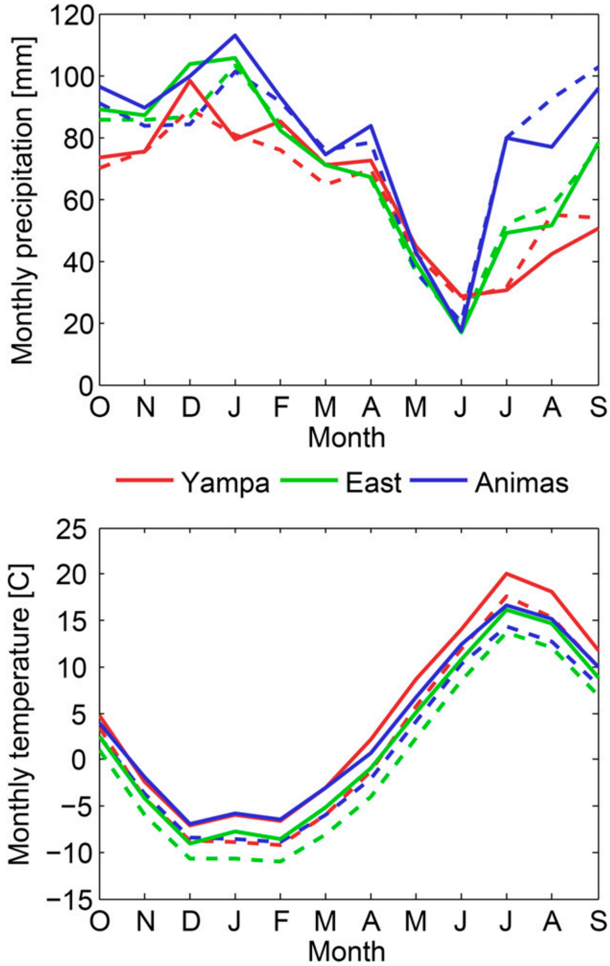

FIG. 2. Basin-averaged monthly (top) precipitation and (bottom) temperature values for CTRL (dashed lines) and PGW (solid lines) WRF outputs used in this study (period from October 2002 to September 2008).

obtained is certainly an important limitation for this study, since they affect the magnitude and direction of climatic shifts. Indeed, Vano et al. (2014) demonstrated the large effects of these decisions on long-term runoff projections over the upper CRB, including results from 19 GCMs and three emission scenarios (A2, A1B, and B1) obtained by Seager et al. (2007) and Christensen and Lettenmaier (2007). However, they also noted that higher future greenhouse gas emissions broadly translate to a warmer and, in most cases, drier climate, implying that a general decrease in runoff should be expected in this region. Although high-resolution climate models limit the number of scenarios that can be analyzed, they offer a more realistic representation of climate features that strongly depend on terrain complexity (Rasmussen et al. 2011, 2014), providing better meteorological fields for the assessment of climate change impacts on hydrology.

b. *Hydrologic–land surface models*

We choose four hydrologic–land surface models: the U.S. Geological Survey (USGS) Precipitation–Runoff Modeling System (PRMS; Leavesley et al. 1983; Leavesley and Stannard 1995), the Variable Infiltration Capacity model (VIC; Wood et al. 1992; Liang et al. 1994, 1996), the Noah land surface model (Noah LSM; Ek 2003; Mitchell et al. 2004), and the Noah LSM with multiparameterization options (Noah-MP; Niu et al. 2011; Yang et al. 2011). Our choice is based on the fact that the four models cover different degrees of complexity in terms of conceptualization of vegetation, soil, and seasonal snowpack (see Table 3 and Fig. 3 for further details) and also have different parameterizations for some hydrologic processes (different model equations for canopy storage, base flow, etc.). Additionally, these hydrologic model structures have been used in several research studies (e.g., Wilby et al. 1999; Haddeland et al. 2002; Hay et al. 2002; Hay and Clark 2003; Christensen and Lettenmaier 2007; Barlage et al. 2010; Yang et al. 2011; Cai et al. 2014). Our experimental design considers a hydrologic model spatial resolution (4 km) identical to that used in the WRF configuration of Rasmussen et al. (2014), though simulation time steps, forcing variables, and land cover data used for a priori parameter estimates vary depending on specific model requirements (see Table 2 for further details).

In this study, we use a single suite of physics options for Noah-MP, including a Ball–Berry-type model for canopy stomatal resistance, a CLM-type soil moisture factor for controlling stomatal resistance, the simple TOPMODEL-based runoff scheme (SIMTOP) for runoff and groundwater (Niu et al. 2005), a Monin–Obukhov similarity theory–based drag coefficient, supercooled liquid water and frozen soil permeability based on Niu and Yang (2006), a radiation transfer scheme equivalent to a ''mosaic'' model, a Canadian land surface scheme (CLASS) for snow surface albedo, a partitioning of precipitation into snowfall and rainfall based on Jordan (1991), a Noah-type lower boundary of soil temperature, and a semi-implicit snow–soil temperature time scheme. Readers are referred to Niu et al. (2011) for a full description of each model component.

c. *Experimental setup*

1) MODEL SIMULATIONS

All model simulations are carried out for the period from 1 October 2000 to 30 September 2008, using the first two years to initialize model states. As done for

TABLE 3. Overview of hydrologic model components used in this study.

Model	Snow accumulation and melt	Canopy storage	Moisture in the soil column/surface runoff	Base flow
PRMS	Two-layer energy–mass balance model. Snowpack energy balance is computed every 12 h.	The precipitation can be intercepted by and evaporated from the plant canopy. The precipitation that is not intercepted by the canopy layer (throughfall) is distributed to the watershed land surface. Interception of precipitation by the plant canopy is computed during a time step as a function of plant cover density and the storage available on the predominant plant cover type in each hydrologic response unit (HRU).	Surface runoff and infiltration are computed using a nonlinear variable-source-area method allowing for cascade flow.	The groundwater zone is conceptualized as a linear reservoir (i.e., base flow is computed as a linear function of groundwater storage).
VIC	Two-layer energy–mass balance model.	Water enters one-layer canopy reservoir and can leave as canopy evaporation, transpiration, or throughfall. Canopy throughfall occurs when additional precipitation exceeds the storage capacity of the canopy. Different vegetation classes are allowed within a unique grid cell via a mosaic approach, where energy and water balance terms are computed independently for each coverage class (vegetation and bare soil).	An infiltration capacity function is defined. Vertical movement of moisture through soil follows 1D Richards equation.	Defined as a function of the soil moisture in the third layer (Arno formulation). The function is linear below a soil moisture threshold and becomes nonlinear above that threshold.
Noah LSM	One-layer energy–mass balance model that simulates snow accumulation, sublimation, melting and heat exchange at the snow–atmosphere and snow–soil interfaces.	One canopy layer, simple canopy resistance. Simple Jarvis type of canopy resistance function, single linearized energy balance equation representing combined ground–vegetation surface, considering seasonal LAI and green vegetation fraction.	Surface runoff is computed as the difference between throughfall and a maximum infiltration rate. Vertical movement of moisture through soil layers follows 1D Richards equation.	Computed as the product of a scaling factor between 0 and 1 and the hydraulic conductivity of the bottom layer.
Noah-MP	Three-layer energy–mass balance model that represents percolation, retention, and refreezing of meltwater within the snowpack.	Snow interception includes loading–unloading, melt–refreeze capabilities, and sublimation of canopy-intercepted snow, along with detailed representation of transmission and attenuation of radiation through the canopy, within- and below-canopy turbulence, and different options to represent the biophysical controls on transpiration.	Surface runoff is an exponential function of depth to water table. Vertical movement of moisture through soil layers follows 1D Richards equation.	Base flow is parameterized as an exponential decaying function of the water-table level (SIMTOP).

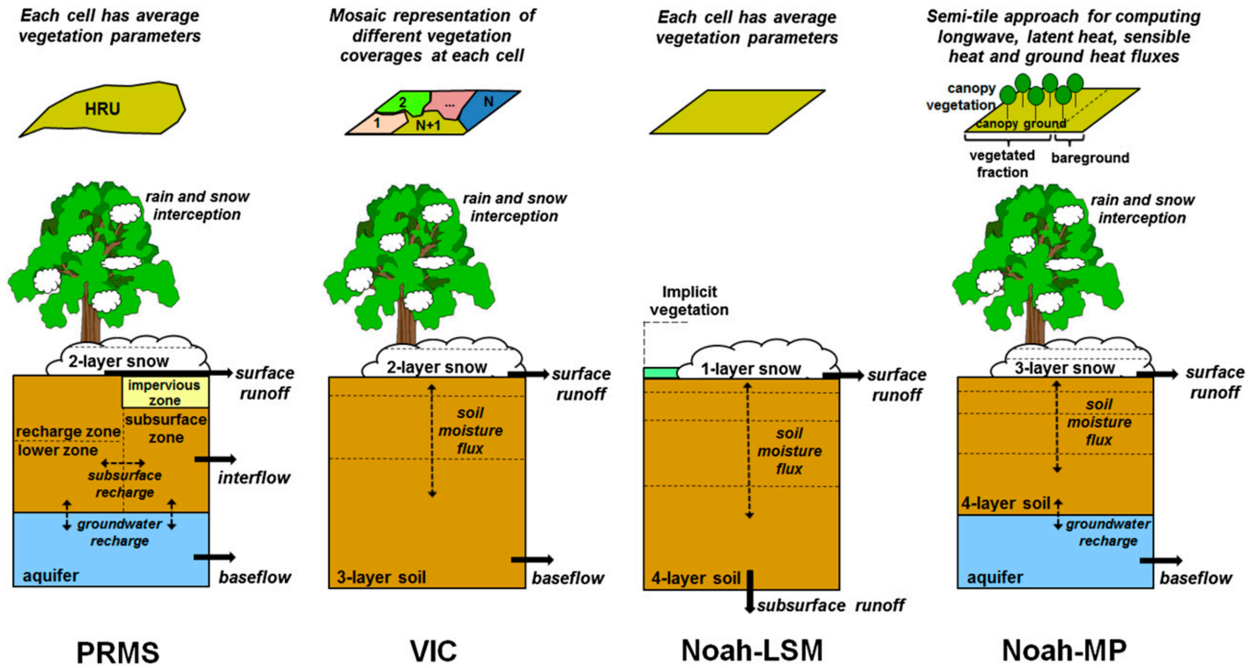

FIG. 3. Comparison of model architectures used in this study: PRMS, VIC, Noah LSM, and Noah-MP.

many past large-scale (i.e., from continental to global scale) hydrologic modeling experiments (e.g., Mitchell et al. 2004; Gerten et al. 2004; Xia et al. 2012), we first compute hydrologic changes using default parameter values obtained from the information sources described in Table 2. Therefore, a comparison of hydrologic change estimates obtained from uncalibrated (i.e., use of default parameters) and calibrated model simulations will provide a comprehensive assessment of the caveats behind traditional methodologies used for climate change impact evaluation.

We calibrate all the models for all basins with the University of Arizona shuffled complex evolution (SCE-UA) algorithm (Duan et al. 1992, 1993) by minimizing the root-mean-square error (RMSE) between observed and simulated daily streamflow for the period between 1 October 2002 and 30 September 2008. Given the short length of WRF reanalysis datasets and that our main priority is to analyze hydrologic change signals, we decided to perform calibration and compute hydrologic changes over the entire period from October 2002 to September 2008 instead of splitting it into calibration and validation datasets. In this study, runoff from hydrologic model simulations is obtained as the sum of surface runoff and base flow, including also interflow if the model is PRMS.

PRMS does not have an explicit river network routing scheme for streamflow; instead, it has a cascade module used to define connections for routing flow from upslope to downslope hydrologic response units and stream segments and among groundwater reservoirs (Markstrom et al. 2008). In VIC, Noah LSM, and Noah-MP, no horizontal routing of surface overland flow, subsurface flow, or channel flow is performed. Instead, basin-averaged runoff is taken as the average of the 1D (vertical) 4-km model grid cells' runoff. During the calibration process, we preserve the spatial variability of a priori model parameters (in case they are spatially distributed) through the adjustment of multiplier values that are applied for each parameter within the entire watershed. We adjust only those parameter multipliers identified as the most sensitive after performing a Distributed Evaluation of Local Sensitivity Analysis (DELSA; Rakovec et al. 2014). The reader is referred to the appendix for a list with the parameters included in the calibration of each model.

Once the calibration process is finished, hydrologic changes are computed for the period from October 2002 to September 2008 by forcing the models with the same meteorological datasets used for uncalibrated simulations.

2) EVALUATION METRICS

In this study we evaluate models using six signature measures (Stewart et al. 2005; Yilmaz et al. 2008) to present a comprehensive portrayal of model performance in terms of hydrologic functional behavior. First, we consider the runoff ratio RR as a measure of the overall water balance and therefore as a signature of the evapotranspiration model component:

$$\mathrm{RR} = R/P, \tag{1}$$

where R is the mean annual runoff and P is the mean annual precipitation.

The second metric selected is the centroid of the daily hydrograph for an average water year, or "center time" of runoff (CTR; Stewart et al. 2005), which is a measure of runoff seasonality:

$$\mathrm{CTR} = \frac{\sum_{i=1}^{N} t_i Q_i}{\sum_{i=1}^{N} Q_i}. \tag{2}$$

In the above equation, t_i is the number of days since 1 October, Q_i is the streamflow associated with t_i, and N is the total number of days in a water year.

Three signature measures are extracted from the flow duration curve (FDC). First, the FDC midsegment slope (FMS) represents the variability, or flashiness, of the flow magnitudes:

$$\mathrm{FMS} = \frac{\log(Q_{m_1}) - \log(Q_{m_2})}{m_1 - m_2}, \tag{3}$$

where $m_1 = 0.2$ and $m_2 = 0.7$, and thus, Q_{m_1} and Q_{m_2} are the flows with exceedance probabilities of 0.2 and 0.7, respectively. A steep slope of the FDC indicates flashiness of the streamflow response, whereas a flatter curve indicates a relatively damped response and a higher storage (Yadav et al. 2007; Casper et al. 2012). Second, the FDC high-segment volume (FHV) is a measure of the catchment response to high rainfall/snowmelt events:

$$\mathrm{FHV} = \sum_{h=1}^{H} Q_h, \tag{4}$$

where $h = 1, 2, \ldots, H$ are the flow indices into the array of flows with exceedance probabilities lower than 0.02. Third, the FDC low-segment volume (FLV) is the measure of the long-term base flow:

$$\mathrm{FLV} = \sum_{l=1}^{L} [\log(Q_l) - \log(Q_L)], \tag{5}$$

where $l = 1, 2, \ldots, L$ is the index into the array of flow values located within the low-flow segment (0.7–1.0 exceedance probabilities) of the FDC and L is the index for the minimum flow.

Finally, we choose the FDC median (FMM; Yilmaz et al. 2008) as a measure of midrange flows:

$$\mathrm{FMM} = \mathrm{median}[\log(\mathrm{FDC})], \tag{6}$$

where log(FDC) represents the array of sorted daily streamflow values in log space. The median is selected because it is less sensitive to a skewed distribution than the mean of the streamflow time series.

4. Results and discussion

a. Model performance

Figure 4 summarizes model performance for the period from October 2002 to September 2008 in terms of mean annual streamflow, monthly streamflow, and flow duration curves, for both uncalibrated and calibrated simulations. None of the hydrologic model structures considered in this study is able to reproduce seasonal runoff patterns or flow duration curves using default parameter values (Fig. 4a). Although this is not surprising and has been widely reported in the literature, many studies that seek to characterize the water balance at the continental scale make use of noncalibrated or semicalibrated land surface models (e.g., Mitchell et al. 2004; Xia et al. 2012). Importantly, the inclusion of a "classic" calibration process based on the minimization of the RMSE between simulated and observed total runoff still leaves inconsistencies across different model structures (Fig. 4b). Some models show large errors in mean annual runoff or seasonal runoff patterns even after calibration, and the FDC is not accurately represented by any model, particularly for low flows.

To assess how much model performance improves functional catchment behavior through a traditional single-objective calibration strategy, we analyze the differences between simulated and observed values of signature measures of hydrologic behavior for both uncalibrated and calibrated model simulations (Fig. 5). Parameter adjustment clearly improves the simulation of those signatures whose formulations are closer to the objective function used for calibration (in this case RMSE, which gives more relative importance to high flows). Consequently, calibration results in smaller intermodel differences in the runoff ratio, the response to large precipitation events (i.e., FHV), and midrange flows (i.e., FMM). On the other hand, intermodel differences in the runoff seasonality (i.e., CTR), the flashiness of runoff (i.e., FMS), and baseflow processes (i.e., FLV) are still pronounced after model calibration. Examples of this are Noah-MP and VIC at Yampa when looking at FMS, or Noah LSM at East and PRMS at Animas when evaluating baseflow processes (i.e., FLV),

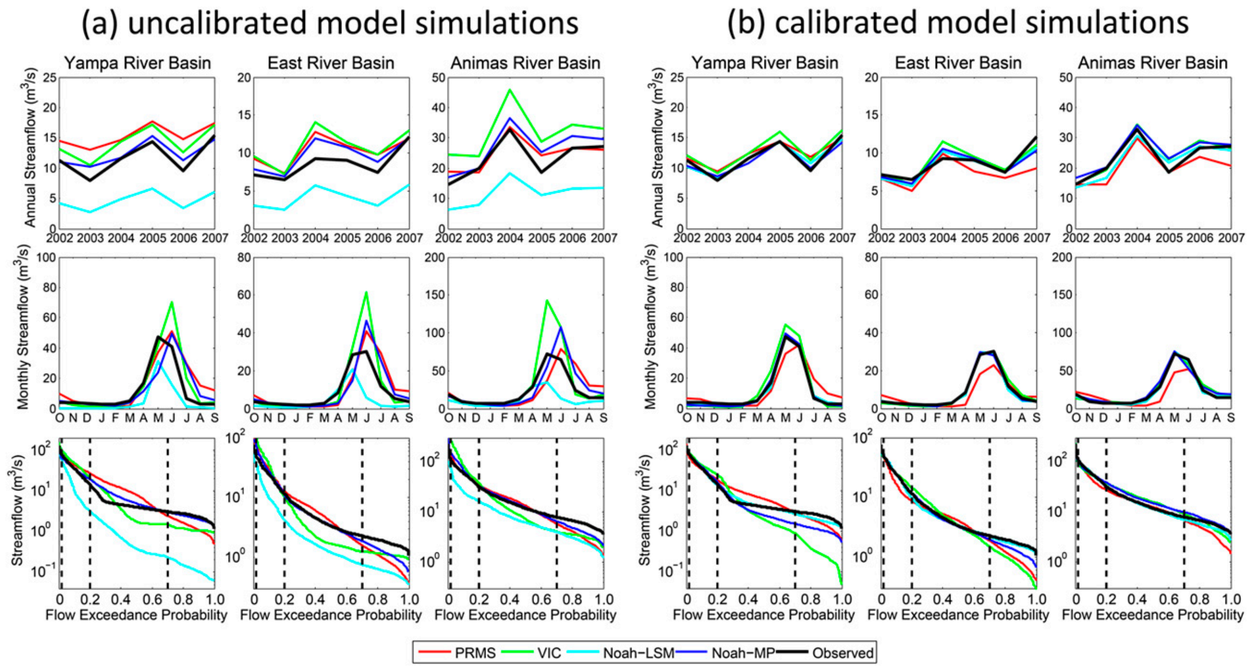

FIG. 4. Historical streamflow (a) uncalibrated and (b) calibrated simulation outputs for the period from October 2002 to September 2008 for all basins: mean annual streamflow for all water years (top), mean monthly flows (middle), and flow duration curves (bottom).

where calibration has actually degraded the signature measures. While a different objective function (e.g., based on the log of the flows) might improve other metrics (e.g., FLV), no single metric is likely to capture all catchment behaviors.

b. *Changes in annual water balance*

To what extent does parameter calibration decrease the uncertainty in projected changes in the overall water balance? To provide an initial answer to this question, we first analyze both uncalibrated and calibrated hydrologic model outputs in the runoff–ET space for a single climate scenario. In Fig. 6, the diagonal lines represent basin-averaged mean annual precipitation for current and future climate scenarios over a 6-yr average period (from October 2002 to September 2008). The intersection of these lines with the x axis indicates that all precipitation becomes runoff, while the intersection with the y axis indicates that the system converts all precipitation into ET. In the same figure, different symbols represent outputs coming from different hydrologic model structures for current climate (unfilled) and future climate (solid). A symbol located exactly on the diagonal lines represents a simulation with negligible changes in storage over the 6-yr simulation period, whereas symbols located below the 1:1 line denote increases in storage, and those above denote decreases in storage. Intermodel differences in precipitation partitioning are represented by the distance between different symbols (unfilled or solid), while the distance between a particular symbol (e.g., star for Noah-MP) for current (unfilled) and future (solid) climate scenarios represents the hydrologic change signal.

The results obtained from uncalibrated simulations (Fig. 6a) indicate that intermodel differences are much larger than the magnitude of hydrologic change signals. Furthermore, all the models have the same hydrologic change signal direction (increase in ET and decrease or negligible change in mean annual runoff) with the exception of Noah LSM, which projects increases in both runoff and ET (Fig. 7a). As expected, intermodel differences in runoff (Fig. 6b) decrease considerably (i.e., less variability along the x axis) and the direction of hydrologic change signal (Fig. 7b) is more consistent across models (i.e., less runoff and more ET for future climate scenario) after calibration, with the exceptions of VIC at Yampa and PRMS at East. Noah-MP stands out from the rest of the models because the direction and magnitude of the signal is not substantially altered after the calibration (cf. Figs. 7a,b). On the other hand, considerable shifts in projected runoff changes are obtained after calibrating PRMS at East (from -11 to $6\,\mathrm{mm\,yr^{-1}}$), VIC at Yampa (from -7 to $4\,\mathrm{mm\,yr^{-1}}$), and Noah LSM at all basins (from 8 to $-21\,\mathrm{mm\,yr^{-1}}$ at Yampa, from 6 to $-21\,\mathrm{mm\,yr^{-1}}$ at East, and from 12 to $-13\,\mathrm{mm\,yr^{-1}}$ at Animas). Moreover, an important

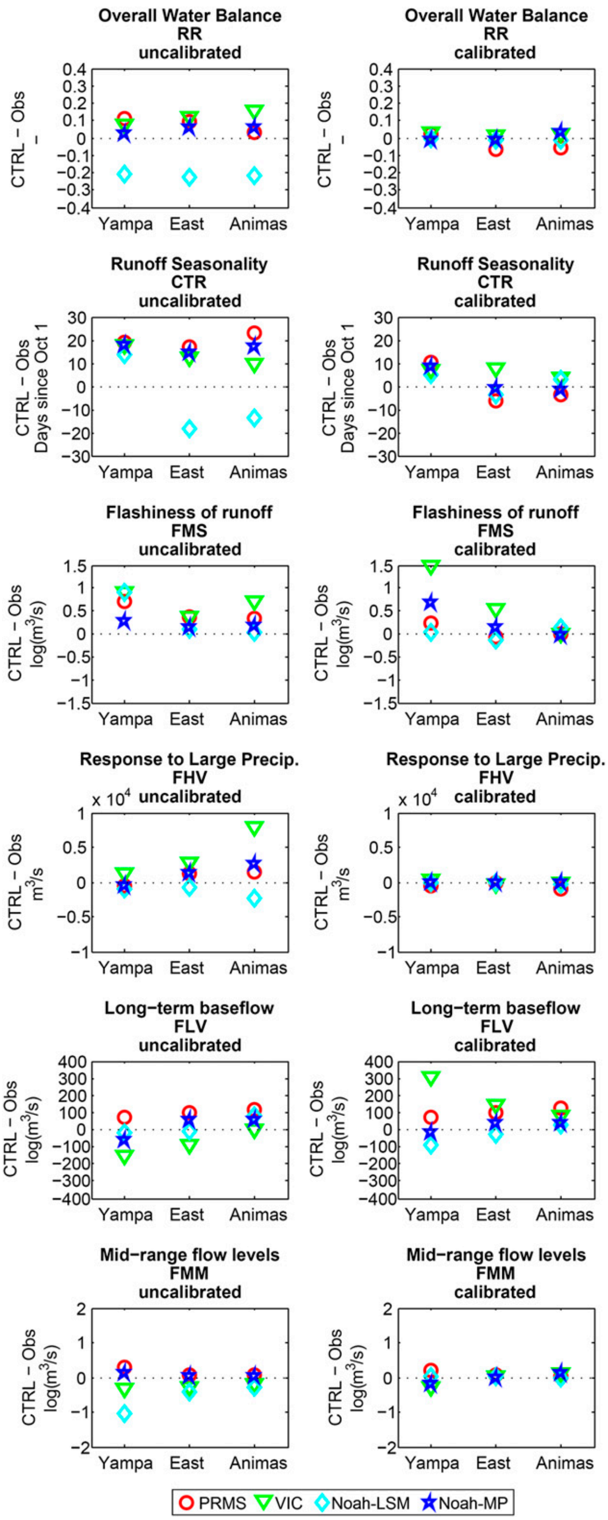

FIG. 5. Difference between simulated (CTRL) and observed (Obs) signature measures of hydrologic behavior (period from October 2002 to September 2008) obtained from (left) uncalibrated and (right) calibrated model runs.

result from Fig. 6b is that intermodel differences in precipitation partitioning into runoff and ET are still comparable or even larger than the magnitude of hydrologic change signal, even after model calibration.

Table 4 summarizes fractional hydrologic changes on an annual basis for both uncalibrated and calibrated model simulations over a 6-yr average period (from October 2002 to September 2008). A suite of different variables is included in order to illustrate how model structure selection and parameter calibration may affect the direction and magnitude of projected changes on hydrologic systems. For instance, in the East River basin, the magnitude of fractional changes in maximum snow water equivalent (SWE) increases with PRMS (from −0.10 to −0.19) and decreases with Noah-MP (from −0.09 to −0.04) after the calibration process. Another example is given by base flow at the Yampa River basin: fractional changes switch from positive to negative values after calibrating PRMS (from 0.03 to −0.04) and Noah LSM (from 0.08 to −0.08), but they shift from negative (−0.04) to positive (0.01) values if the model selected is VIC. Similarly, Table 4 illustrates the effects of calibration on fractional changes in total runoff (e.g., PRMS at East, VIC at Yampa, and Noah LSM at all basins), capturing (although in different units) the results from Fig. 7 that were previously discussed. The key result from Table 4 is that the intermodel differences in the hydrologic impacts of the CCSM–WRF climate scenario vary substantially across models (i.e., the differences in the columns of Table 4 for each basin), and the intermodel differences are larger than the mean multimodel change signal for most metrics.

c. *Monthly changes*

Figure 8 shows mean monthly runoff values obtained from all models for both uncalibrated and calibrated simulations over a 6-yr average period (from October 2002 to September 2008). As expected, the use of default parameters (Fig. 8a) translates into very different catchment responses under current and future climate scenarios, and these differences are also reflected in projected monthly changes [PGW minus control (CTRL)]. The largest and smallest changes in runoff are obtained from VIC and Noah LSM, respectively, and the seasonality of these shifts differs substantially across models. For instance, the Noah LSM simulates increases in runoff during February–April, extending to May for the Yampa River basin, and a decrease during May–June, while Noah-MP generates an increase in runoff during March–May and a decrease during June–September (Fig. 8a). Much more consistent results across models are obtained when parameter calibration is

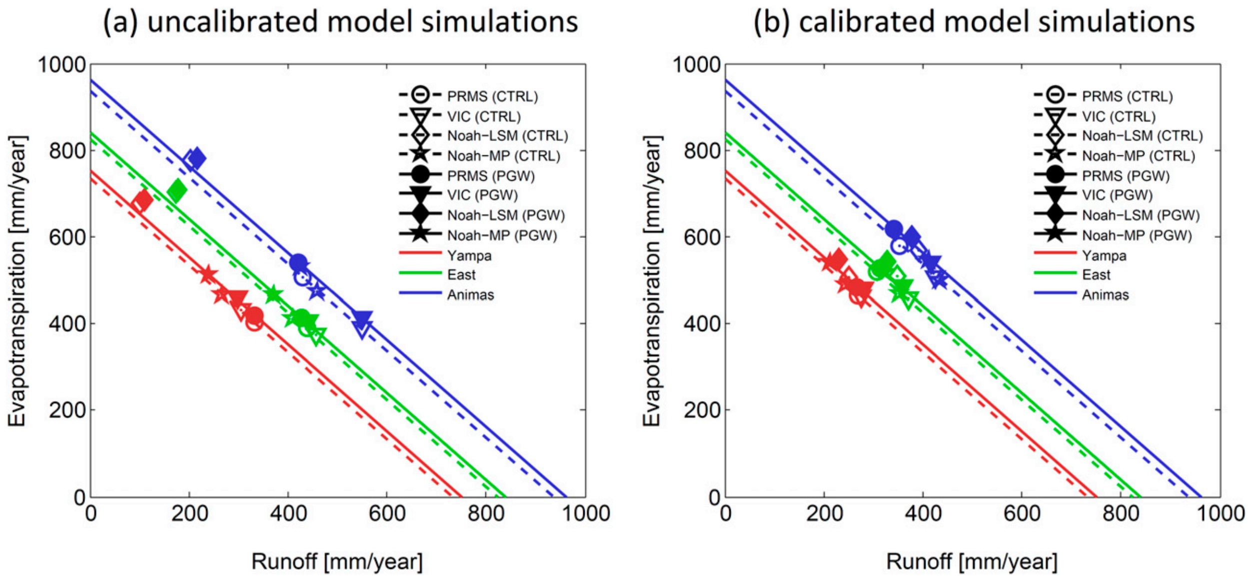

FIG. 6. Partitioning of current (CTRL) and future (PGW) basin-averaged mean annual precipitation (diagonal; mm yr^{-1}) into basin-averaged mean annual runoff (x axis; mm yr^{-1}) and evapotranspiration (y axis; mm yr^{-1}) across different model structures and basins for the period from October 2002 to September 2008. Results are displayed for (a) uncalibrated and (b) calibrated model simulations.

performed (Fig. 8b), and this is reflected in both the magnitude and seasonality of runoff variations. A key question that follows from here is whether intermodel similarities in runoff changes are due to intermodel agreement in changes of other water storages and fluxes.

With the aim to explore possible reasons for the (mis) match in projected runoff changes among different model structures, we analyze monthly changes in model states and fluxes obtained from both uncalibrated and calibrated runs (Fig. 9). The variables included in this analysis are ET, SWE, soil moisture, base flow, and surface runoff. To improve consistency in the comparison across models, we consider only the top two soil layers for the computation of soil moisture storage with VIC, Noah LSM, and Noah-MP, and the addition of interflow to surface runoff for PRMS. Figure 9a shows large differences in changes for ET, base flow, and surface runoff among models, while more consistent results in terms of seasonal cycles and amplitude are obtained for snowpack (except Noah LSM) and soil moisture. However, intermodel differences of soil moisture and surface runoff are preserved or emphasized after the calibration process (Fig. 9b). Furthermore, one can infer from the results displayed in Figs. 8b and 9b that the same runoff changes might be obtained using different hydrologic models due to very different mechanisms; that is, internal compensations of model structures and model parameter errors are adjusted through

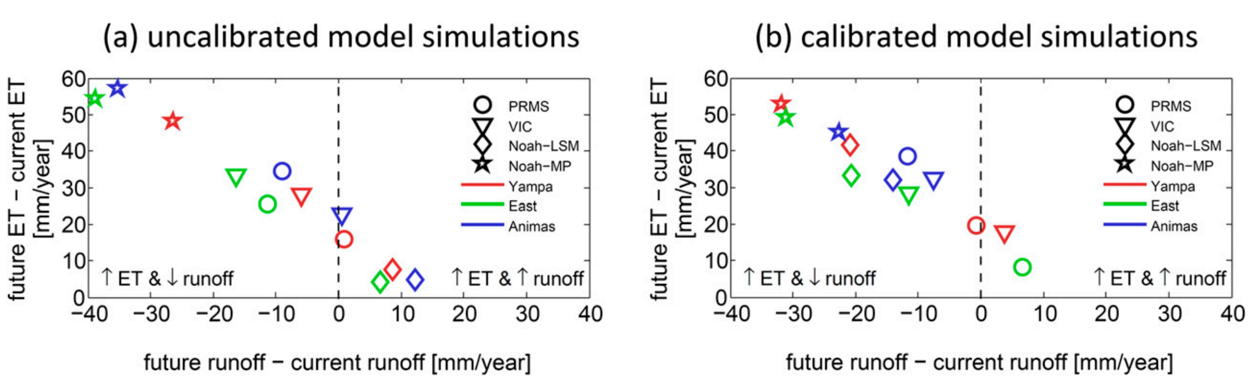

FIG. 7. Projected changes in basin-averaged mean annual runoff (x axis; mm yr^{-1}) and evapotranspiration (y axis; mm yr^{-1}) across different model structures and basins for the period from October 2002 to September 2008. Results are displayed for (a) uncalibrated and (b) calibrated model simulations.

TABLE 4. Values of fractional change [(PGW − current climate)/current climate] in basin-averaged total accumulated precipitation, peak SWE, accumulated ET, accumulated surface runoff, accumulated base flow, and accumulated total runoff (sum of surface runoff and base flow, including interflow if using PRMS) averaged for an average water year (from October 2002 to September 2008) obtained from both uncalibrated and calibrated model simulations. Also included are the changes in dates of maximum SWE for each basin/model, where the values represent CTRL minus PGW dates of maximum SWE. Mean values for each basin are given in boldface.

	Yampa					East					Animas				
Variable	PRMS	VIC	Noah LSM	Noah-MP	**Mean**	PRMS	VIC	Noah LSM	Noah-MP	**Mean**	PRMS	VIC	Noah LSM	Noah-MP	**Mean**
Total precipitation	0.02	0.02	0.02	0.02	—	0.02	0.02	0.02	0.02	—	0.03	0.03	0.03	0.03	—
Maximum SWE															
Uncalibrated	−0.12	−0.08	−0.12	−0.09	**−0.10**	−0.10	−0.10	−0.14	−0.09	**−0.11**	−0.12	−0.14	−0.14	−0.10	**−0.12**
Calibrated	−0.16	−0.10	−0.11	−0.12	**−0.12**	−0.19	−0.04	−0.09	−0.04	**−0.09**	−0.22	−0.06	−0.08	−0.06	**−0.11**
Date of maximum SWE															
Uncalibrated	25	32	7	13	**19.25**	18	13	3	12	**11.50**	12	31	46	31	**30.00**
Calibrated	25	2	7	0	**8.50**	25	4	4	6	**9.75**	12	1	5	0	**4.50**
Evapotranspiration															
Uncalibrated	0.04	0.07	0.01	0.10	**0.05**	0.07	0.09	0.01	0.13	**0.07**	0.07	0.06	0.01	0.12	**0.06**
Calibrated	0.04	0.04	0.08	0.11	**0.07**	0.02	0.06	0.07	0.10	**0.06**	0.07	0.06	0.06	0.09	**0.07**
Base flow															
Uncalibrated	0.03	−0.04	0.08	−0.08	**0.00**	−0.02	−0.05	0.05	−0.04	**−0.02**	−0.03	0.00	0.07	−0.04	**0.00**
Calibrated	−0.04	0.01	−0.08	−0.08	**−0.05**	0.01	−0.03	−0.04	−0.05	**−0.03**	−0.11	−0.02	−0.03	−0.04	**−0.05**
Surface runoff															
Uncalibrated	−0.01	0.01	0.14	−0.12	**0.01**	−0.03	0.00	−0.01	−0.15	**−0.05**	−0.01	0.01	−0.01	−0.12	**−0.03**
Calibrated	0.03	0.02	−0.08	−0.24	**−0.07**	0.03	−0.01	−0.14	−0.14	**−0.07**	0.01	0.00	−0.06	−0.09	**−0.03**
Total runoff															
Uncalibrated	0.00	−0.02	0.09	−0.10	**−0.01**	−0.03	−0.04	0.04	−0.09	**−0.03**	−0.02	0.00	0.06	−0.08	**−0.01**
Calibrated	0.00	0.01	−0.08	−0.13	**−0.05**	0.02	−0.03	−0.06	−0.09	**−0.04**	−0.03	−0.02	−0.04	−0.05	**−0.03**

calibration in a way that allows similar responses from different watershed models. The clearest example in this case study is observed in the East River basin, where monthly changes in runoff are very similar (Fig. 8b); nevertheless, VIC compensates very large variations in soil moisture with other variables such as ET and base flow, and PRMS does the same with large variations in ET, SWE, and surface runoff.

d. Projected changes in catchment behavior

Finally, we compare the effects of model choice and parameter adjustment on projected changes in

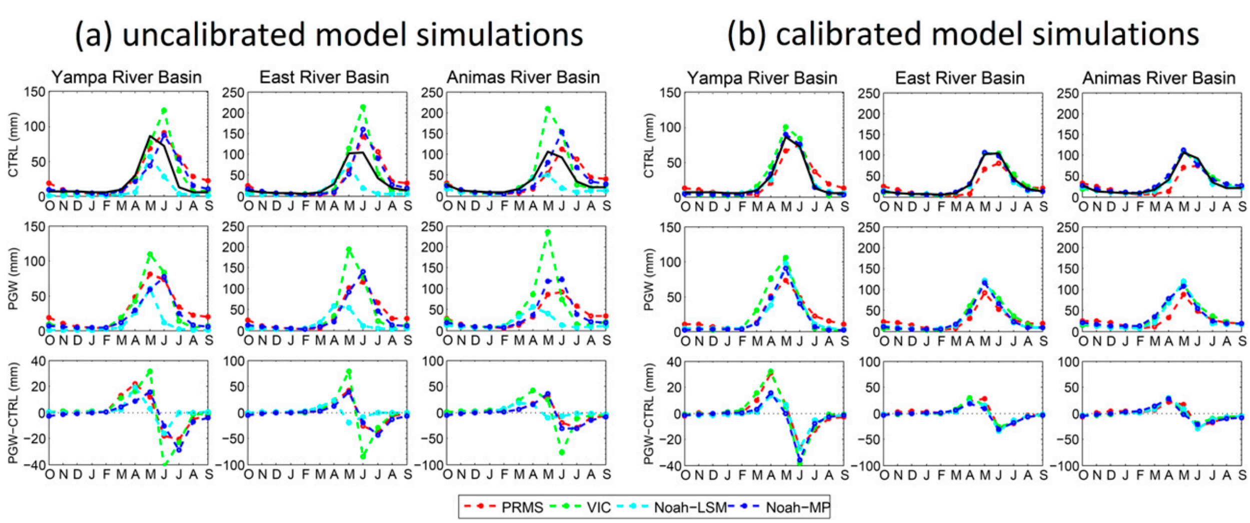

FIG. 8. Current (CTRL), future (PGW), and changes (PGW − CTRL) in basin-averaged monthly runoff for (a) uncalibrated and (b) calibrated model simulations over a 6-yr average (from October 2002 to September 2008). The black lines in CTRL represent historical observations.

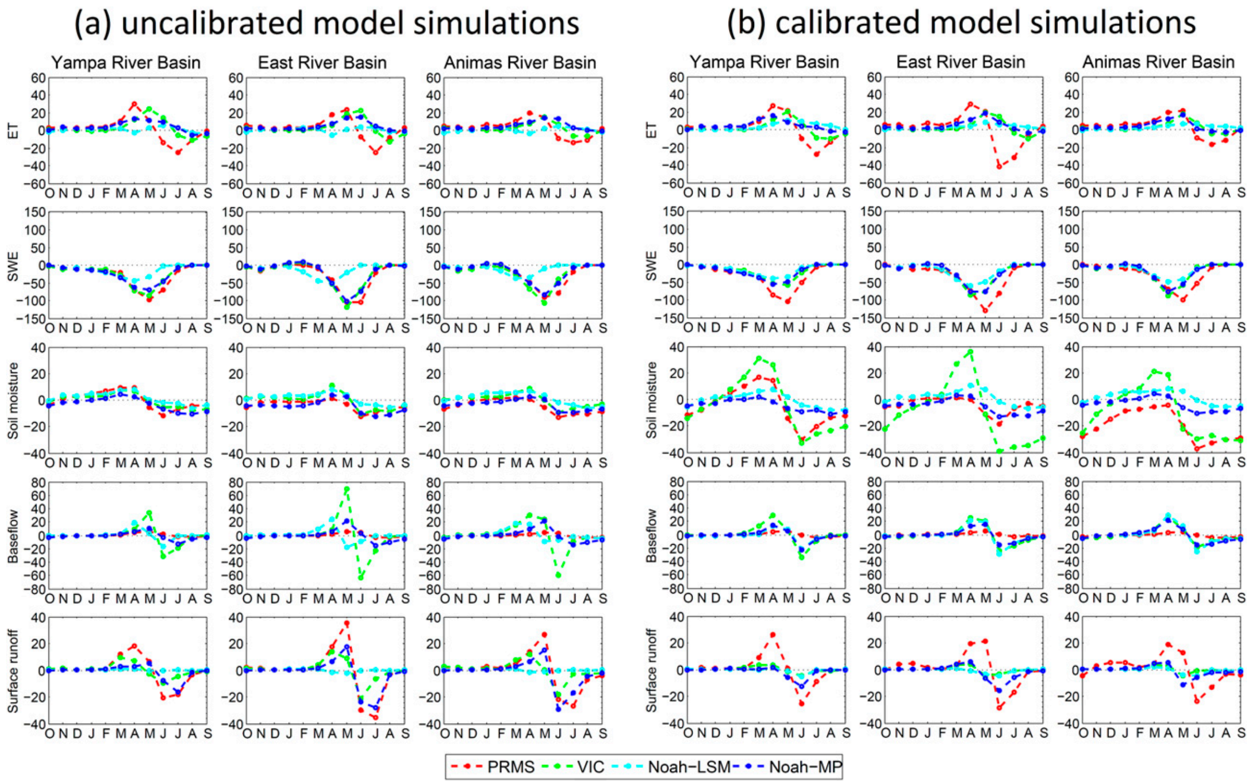

FIG. 9. Monthly changes (PGW − CTRL) in basin-averaged fluxes and states (mm) for (a) uncalibrated and (b) calibrated model simulations over a 6-yr average (from October 2002 to September 2008).

hydrologic signatures. Figure 10 illustrates differences between future (PGW) and current (CTRL) signature measures of hydrologic behavior for all models/basins, computed from both uncalibrated (Fig. 10, left) and calibrated (Fig. 10, right) model runs. The main result from Fig. 10 is that calibration helps to decrease the uncertainty associated with model choice in projected changes of those signatures closely related with the objective function selected. Clear examples of this are the response to large precipitation events (i.e., FHV) and midrange flow levels (i.e., FMM). However, the uncertainty due to model structure increases for some signatures and basins [e.g., runoff seasonality (i.e., CTR) at Yampa and East, flashiness of runoff (i.e., FMS) at Yampa, and baseflow processes (i.e., FLV) at Yampa and Animas]. Moreover, different hydrologic model structures can provide opposite changes (signal) of some signature metrics even after calibration (e.g., FLV and FMS).

It is interesting to see that for both uncalibrated and calibrated model outputs, the only consistent signal obtained with all models is a negative change in runoff seasonality (CTR), which is directly related with an expected decrease in snowpack under the PGW scenario (i.e., shorter accumulation season and earlier melt season). For the case of calibrated model simulations, a general reduction of high-flow volumes (FHV) occurs regardless of the model choice (except PRMS at Yampa). The results in Fig. 10 illustrate the strong interplay between model structure and model parameters and suggest the following hypothesis: different calibration approaches may lead to very different answers from those displayed in Fig. 10 (right) or, put differently, that subjective decisions on configuring and calibrating hydrologic models may have unexpected and underappreciated impacts on the portrayal of climate change impacts. Current work is focused on this problem in order to get a better comprehension of uncertainties introduced by model structure selection and different parameter estimation strategies.

5. Conclusions

This study aims to improve our understanding of the effects of hydrologic model choice on the portrayal of climate change impacts. Specifically, we assess the effects of model structure selection on: 1) historical performance in terms of hydrologic signature measures and 2) hydrologic changes due to a climate

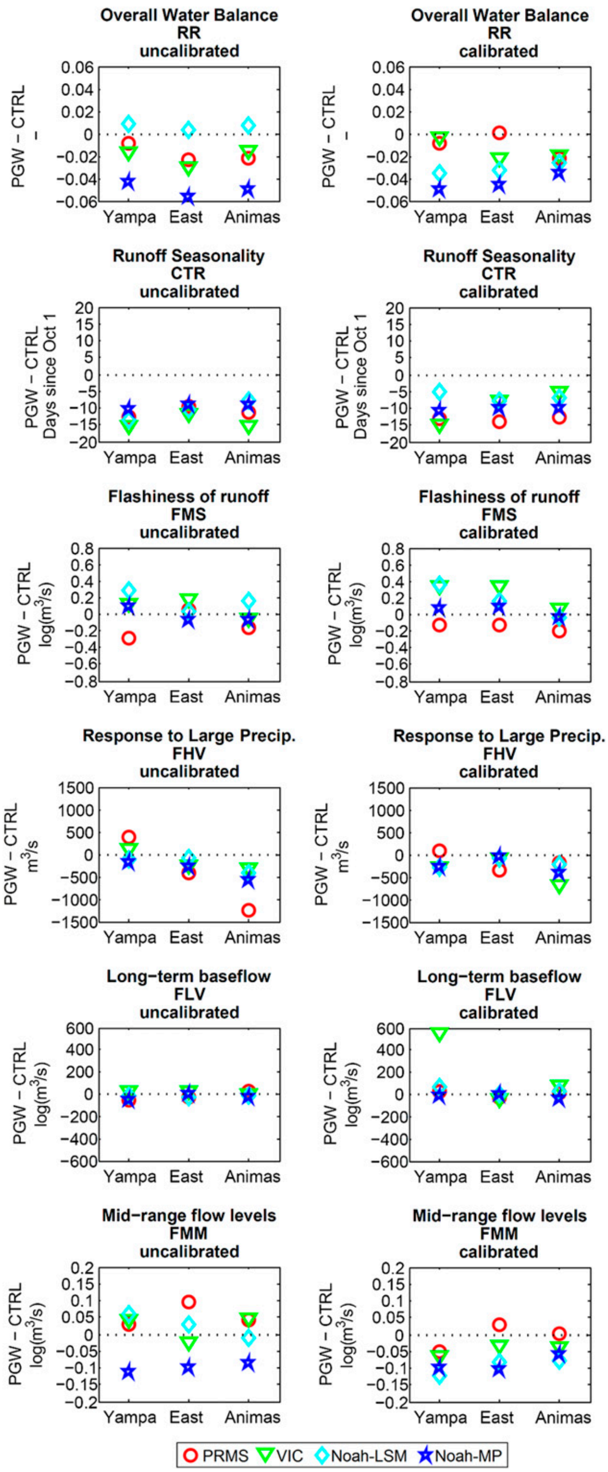

FIG. 10. Impact of climate change on signature measures of hydrologic behavior for both (left) uncalibrated and (right) calibrated model simulations over a 6-yr average (from October 2002 to September 2008).

perturbation, with focus on the overall water balance and catchment processes. Because several efforts aimed to characterize future changes on the hydrology at the continental or global scales have made use of hydrologic–land surface models with little or no calibration, we include in our analysis a comparison between uncalibrated and calibrated model outputs. Our main findings are as follows:

- Intermodel differences in portrayal of climate change impacts are substantial, even after calibration. These differences reflect on projected changes in overall water balance, monthly changes in individual simulated processes, and signature measures of hydrologic behavior.
- In this paper, better values for specific process evaluation metrics (i.e., signature measures) were obtained over the historical period from October 2002 to September 2008 only if their mathematical formulation was close to the RMSE between simulated and observed runoff (i.e., the calibration objective function).
- Consequently, single-objective calibration procedures constrain intermodel differences in climate change impacts for hydrologic metrics that are closely related to the objective function. In this study, calibration improved intermodel agreement on future projected changes of the response to large precipitation events, and midrange flow levels. However, intermodel agreement decreased when evaluating the change of other metrics related with flashiness of runoff and baseflow processes.
- Although traditional calibration methods certainly improve intermodel agreement in projected changes of the overall water balance (i.e., partitioning of precipitation into ET and runoff), intermodel differences in the runoff–ET space are comparable and even larger than the hydrologic change signal for the scenario examined here.
- Single-objective calibration approaches aimed to reduce errors in runoff simulations do not necessarily enhance intermodel agreement in projected changes of some hydrological processes such as ET or snowpack. Moreover, identical changes in runoff might be obtained with different hydrologic model structures for very different reasons, indicating that the calibration process is compensating structural and parameter errors to give us ''good'' runoff simulations, but not to correctly reproduce catchment processes.

The main conclusion from this study is that subjective decisions in the selection of hydrologic model structures and parameters have large effects on the portrayal of climate change impacts. Moreover, these effects may directly impact adaptation strategies. For instance, 1) the diversity of projected changes in runoff amounts and timing affects reservoir operations such as release schedules and magnitudes (Miller et al. 2012); 2) uncertainty in responses to large precipitation events propagates to flood frequency estimates, which are required for design and safety assessment of infrastructure (Raff et al. 2009); 3) uncertainties in ET projections relate with irrigation demands and should therefore be considered in agricultural adaptation plans; and 4) the diverse responses obtained in terms of long-term base flow may impact future drought risk evaluation (Wilby and Harris 2006) and policies related with minimum instream flow requirements (Vano et al. 2014).

The implication of our findings is that previous studies evaluating the impacts of climate change on water resources may be overconfident. Moving forward, it is necessary to have a much more comprehensive assessment of the myriad of uncertainties in climate risk assessments; in particular, to improve characterization of uncertainties in hydrologic modeling applications.

Acknowledgments. The authors thank Kyoko Ikeda for her help with WRF reanalysis datasets. We are grateful to Marketa Elsner, Roland Viger, Adam Carheden, and Tor Mohling for their advice and assistance for setting up the models. We also thank Olda Rakovec and Mary Hill for their help with the implementation of DELSA. This study was supported through a cooperative agreement with the Bureau of Reclamation (USBR), through a contract from the U.S. Army Corps of Engineers (USACE), and through a graduate fellowship from the Cooperative Institute for Research and Environmental Sciences (CIRES).

APPENDIX

Parameters Included in the Calibration Process

The model parameters included in the calibration process are selected based on background sensitivity analysis performed for each hydrologic–land surface model. In this study, we use the DELSA (Rakovec et al. 2014) method to quantify parameter sensitivity, using the RMSE between observed and simulated streamflow as objective function. In DELSA, the assessment of parameter sensitivity is based on local gradients of the model performance index with respect to model parameters at multiple points throughout the parameter space. A number of soil, vegetation, runoff, and snow parameters were considered in DELSA for each model: 17 for PRMS, 34 for VIC, 17 for Noah LSM, and 30 for Noah-MP.

Based on the sensitivity analysis results, the numbers of parameters calibrated are 8 for PRMS, 9 for VIC, 11 for Noah LSM, and 14 for Noah-MP. These parameters are listed in Tables A1–A4.

TABLE A1. Summary of PRMS parameters considered for calibration. If the parameter is distributed, calibration is performed on the basis of multipliers. Although description and units refer to actual parameters, the values in boldface represent the multiplier values (instead of actual parameter values). For parameter jh_coef, the range is provided for a multiplier applied to each monthly value.

Parameter	Description	Units	Distributed	Calibration range	
				Min	Max
jh_coef	Monthly Jensen–Haise air temperature coefficient	F	No	0.36	2.86
fastcoef_lin	Linear flow routing coefficient for fast interflow	day^{-1}	No	0	10
fastcoef_sq	Nonlinear flow routing coefficient for fast interflow	$\mathrm{in.}^{-1}\,\mathrm{day}^{-1}$	No	0	1.25
pref_flow_den	Decimal fraction of the soil zone available for preferential flow	—	No	0	5
soil_moist_max	Maximum volume of water per unit area in the capillary reservoir	in.	Yes	**0**	**2.87**
snarea_curve	Snow area depletion curve values	—	Yes	**0**	**1**
tmax_allsnow	Monthly maximum air temperature at which precipitation is all snow for the HRU	F	No	−10	40
tmax_allrain	Monthly minimum air temperature at an HRU that results in all precipitation during a day being rain	F	No	0	90

TABLE A2. Summary of VIC parameters considered for calibration. If the parameter is distributed, its calibration is performed on the basis of multipliers. Although description and units refer to actual parameters, the values in boldface represent the multiplier values (instead of actual parameter values).

				Calibration range	
Parameter	Description	Units	Distributed	Min	Max
binfilt	Variable infiltration curve parameter	—	No	0.001	0.4
Ds	Fraction of Dsmax where nonlinear base flow begins	—	No	10^{-5}	1
Dsmax	Maximum velocity of base flow	mm day^{-1}	Yes	**0.01**	**2**
Ws	Fraction of maximum soil moisture where nonlinear base flow occurs	—	No	9×10^{-4}	1
depth2	Thickness of soil layer 2	m	Yes	**0.5**	**6**
depth3	Thickness of soil layer 3	m	Yes	**0.5**	**6**
newalb	New snow albedo	—	No	0.7	0.99
albaa	Base in snow albedo function (accumulation)	—	No	0.88	0.99
albtha	Base in snow albedo function (melt)	—	No	0.66	0.98

TABLE A3. Summary of Noah LSM parameters considered for calibration. If the parameter is distributed, its calibration is performed on the basis of multipliers. Although description and units refer to actual parameters, the values in boldface represent the multiplier values (instead of actual parameter values).

				Calibration range	
Parameter	Description	Units	Distributed	Min	Max
maxsmc	Soil porosity	$m^3 m^{-3}$	Yes	**0.88**	**1.18**
satdk	Saturated soil hydraulic conductivity	$m s^{-1}$	Yes	**0.41**	**1.39**
quartz	Soil quartz content	—	Yes	**0.29**	**1.37**
refdk	Used with refkdt to compute runoff parameter kdt	—	No	2×10^{-8}	2×10^{-4}
fxexp	Bare soil evaporation exponent	—	No	0.2	4
refkdt	Surface runoff parameter	—	No	0.1	10
czil	Zilitinkevich parameter	—	No	0.05	8
cmcmax	Maximum canopy water capacity used in canopy evaporation	m	No	5×10^{-5}	2
rsmax	Maximum stomatal resistance	$s m^{-1}$	No	2	10
lvcoef	Livneh coefficient for adjusting snow albedo	—	No	0	1
slope	Linear coefficient used to compute subsurface runoff	—	No	0.2	1

TABLE A4. Summary of Noah-MP parameters considered for calibration. If the parameter is distributed, its calibration is performed on the basis of multipliers. Although description and units refer to actual parameters, the values in boldface represent the multiplier values (instead of actual parameter values).

				Calibration range	
Parameter	Description	Units	Distributed	Min	Max
maxsmc	Soil porosity	$m^3 m^{-3}$	Yes	**0.88**	**1.18**
wind_rp	Empirical canopy wind parameter	m^{-1}	Yes	**0.7**	**1.3**
slope_ps	Slope of conductance-to-photosynthesis relationship	—	Yes	**0.7**	**1.3**
laimss	Monthly LAI, one sided (spring–summer)	—	Yes	**0.7**	**1.3**
fff	Runoff decay factor	m^{-1}	No	1	8
rsbmx	Baseflow coefficient	mm s^{-1}	No	0.5	8
timean	Grid cell mean topographic index	—	No	7.35	13.65
mexp	Exponent used in the curves for the melting season	—	No	0.5	3
z0sno	Snow surface roughness length	m	No	0.0002	0.02
snow_iwc	Liquid water holding capacity for snowpack	$m^3 m^{-3}$	No	0.02	0.06
swemx	New snow mass to fully cover old snow	mm	No	0.1	20
albmin	Minimum snow albedo	—	No	0.44	0.66
albmax	Maximum snow albedo	—	No	0.68	1
albdecay	Exponent in snow decay albedo relationship	h^{-1}	No	0.001	0.1

REFERENCES

Barlage, M., and Coauthors, 2010: Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *J. Geophys. Res.,* **115,** D22101, doi:10.1029/2009JD013470.

Bergström, S., B. Carlsson, M. Gardelin, G. Lindstrom, A. Petterson, and M. Rummukainen, 2001: Climate change impacts on runoff in Sweden—assessments by global climate models, dynamical downscaling and hydrological modelling. *Climate Res.,* **16,** 101–112, doi:10.3354/cr016101.

Brekke, L. D., E. P. Maurer, J. D. Anderson, M. D. Dettinger, E. S. Townsley, A. Harrison, and T. Pruitt, 2009: Assessing reservoir operations risk under climate change. *Water Resour. Res.,* **45,** W04411, doi:10.1029/2008WR006941.

Bureau of Reclamation, 2012: Colorado River basin water supply and demand study: Technical report B—Water supply assessment. Tech. Rep., 103 pp.

Cai, X., Z.-L. Yang, C. H. David, G.-Y. Niu, and M. Rodell, 2014: Hydrological evaluation of the Noah-MP land surface model for the Mississippi River basin. *J. Geophys. Res. Atmos.,* **119,** 23–38, doi:10.1002/2013JD020792.

Casper, M. C., G. Grigoryan, O. Gronz, O. Gutjahr, G. Heinemann, R. Ley, and A. Rock, 2012: Analysis of projected hydrological behavior of catchments based on signature indices. *Hydrol. Earth Syst. Sci.,* **16,** 409–421, doi:10.5194/hess-16-409-2012.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129,** 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, J., F. P. Brissette, A. Poulin, and R. Leconte, 2011: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resour. Res.,* **47,** W12509, doi:10.1029/2011WR010602.

Christensen, N. S., and D. P. Lettenmaier, 2007: A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River basin. *Hydrol. Earth Syst. Sci.,* **11,** 1417–1434, doi:10.5194/hess-11-1417-2007.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate,* **19,** 2122–2143, doi:10.1175/JCLI3761.1.

Duan, Q. Y., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.,* **28,** 1015–1031, doi:10.1029/91WR02985.

——, V. K. Gupta, and S. Sorooshian, 1993: Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.,* **76,** 501–521, doi:10.1007/BF00939380.

Ek, M. B., 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Fry, J., and Coauthors, 2011: Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.,* **77,** 858–864.

Gerten, D., S. Schaphoff, U. Haberlandt, W. Lucht, and S. Sitch, 2004: Terrestrial vegetation and water balance—Hydrological evaluation of a dynamic global vegetation model. *J. Hydrol.,* **286,** 249–270, doi:10.1016/j.jhydrol.2003.09.029.

Graham, L. P., J. Andréasson, and B. Carlsson, 2007: Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—A case study on the Lule River basin. *Climate Change,* **81,** 293–307, doi:10.1007/s10584-006-9215-2.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson, 2012: A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J. Climate,* **25,** 262–281, doi:10.1175/2011JCLI4109.1.

——, T. Pruitt, M. P. Clark, L. Brekke, J. R. Arnold, D. A. Raff, and R. M. Rasmussen, 2014: An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resour. Res.,* **50,** 7167–7186, doi:10.1002/2014WR015559.

Haddeland, I., B. V. Matheussen, and D. P. Lettenmaier, 2002: Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model. *Water Resour. Res.,* **38,** 1124, doi:10.1029/2001WR000854.

Hansen, M. C., R. S. DeFries, J. R. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.,* **21,** 1331–1364, doi:10.1080/014311600210209.

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.

Hay, L. E., and M. P. Clark, 2003: Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *J. Hydrol.,* **282,** 56–75, doi:10.1016/S0022-1694(03)00252-X.

——, ——, R. Wilby, W. J. Gutowski Jr., G. H. Leavesley, Z. Pan, R. W. Arritt, and E. S. Takle, 2002: Use of regional climate model output for hydrologic simulations. *J. Hydrometeor.,* **3,** 571–590, doi:10.1175/1525-7541(2002)003<0571:UORCMO>2.0.CO;2.

Hoerling, M., and J. Eischeid, 2007: Past peak water in the southwest. *Southwest Hydrol.,* **6,** 18–19. [Available online at www.swhydro.arizona.edu/archive/V6_N1/feature2.pdf.]

——, D. Lettenmaier, D. Cayan, and B. Udall, 2009: Reconciling projections of Colorado River streamflow. *Southwest Hydrol.,* **8,** 20–21. [Available online at www.swhydro.arizona.edu/archive/V8_N3/feature2.pdf.]

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **134,** 2318–2341, doi:10.1175/MWR3199.1.

Ikeda, K., and Coauthors, 2010: Simulation of seasonal snowfall over Colorado. *Atmos. Res.,* **97,** 462–477, doi:10.1016/j.atmosres.2010.04.010.

Jensen, M., D. Rob, and C. Franzoy, 1970: Scheduling irrigations using climate–crop–soil data. *J. Irrig. Drain. Div.,* **96,** 25–38.

Jiang, T., Y. D. Chen, C. Xu, X. Chen, X. Chen, and V. P. Singh, 2007: Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang basin, south China. *J. Hydrol.,* **336,** 316–333, doi:10.1016/j.jhydrol.2007.01.010.

Jones, R. N., F. H. S. Chiew, W. C. Boughton, and L. Zhang, 2006: Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models. *Adv. Water Resour.,* **29,** 1419–1429, doi:10.1016/j.advwatres.2005.11.001.

Jordan, R., 1991: A one-dimensional temperature model for a snow cover: Technical documentation for SNTERERM.89. Special

Rep. 91-16, Cold Region Research and Engineers Laboratory, U.S. Army Corps of Engineers, Hanover, NH, 61 pp.

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.

Leavesley, G. H., and L. G. Stannard, 1995: The precipitation–runoff modeling system-PRMS. *Computer Models of Watershed Hydrology,* V. P. Singh, Ed., Water Resources Publications, 281–310.

——, R. W. Lichty, B. M. Troutman, and L. G. Saindon, 1983: Precipitation–runoff modeling system: User's manual. Water-Resources Investigations Rep. 83-4238, 206 pp. [Available online at http://pubs.usgs.gov/wri/1983/4238/report.pdf.]

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.,* **99,** 14 415–14 428, doi:10.1029/94JD00483.

——, E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global Planet. Change,* **13,** 195–206, doi:10.1016/0921-8181(95)00046-1.

Markstrom, S. L., R. G. Niswonger, R. S. Regan, D. E. Prudic, and P. M. Barlow, 2008: GSFLOW—Coupled Ground-Water and Surface-Water Flow Model Based on the Integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005). USGS Techniques and Methods 6-01, 240 pp.

Mastin, M. C., K. J. Chase, and R. W. Dudley, 2011: Changes in spring snowpack for selected basins in the United States for different climate-change scenarios. *Earth Interact.,* **15,** doi:10.1175/2010EI368.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360, doi:10.1175/BAMS-87-3-343.

Miller, W. P., and T. C. Piechota, 2008: Regional analysis of trend and step changes observed in hydroclimatic variables around the Colorado River basin. *J. Hydrometeor.,* **9,** 1020–1034, doi:10.1175/2008JHM988.1.

——, and ——, 2011: Trends in western U.S. snowpack and related upper Colorado River basin streamflow. *J. Amer. Water Resour. Assoc.,* **47,** 1197–1210, doi:10.1111/j.1752-1688.2011.00565.x.

——, ——, S. Gangopadhyay, and T. Pruitt, 2011: Development of streamflow projections under changing climate conditions over Colorado River basin headwaters. *Hydrol. Earth Syst. Sci.,* **15,** 2145–2164, doi:10.5194/hess-15-2145-2011.

——, R. A. Butler, T. Piechota, J. Prairie, K. Grantz, and G. DeRosa, 2012: Water management decisions using multiple hydrologic models within the San Juan River basin under changing climate conditions. *J. Water Resour. Plann. Manag.,* **138,** 412–420, doi:10.1061/(ASCE)WR.1943-5452.0000237.

Milly, P. C. D., and K. A. Dunne, 2011: On the hydrologic adjustment of climate-model projections: The potential pitfall of potential evapotranspiration. *Earth Interact.,* **15,** doi:10.1175/2010EI363.1.

——, ——, and A. V. Vecchia, 2005: Global pattern of trends in streamflow and water availability in a changing climate. *Nature,* **438,** 347–350, doi:10.1038/nature04312.

Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.,* **109,** D07S90, doi:10.1029/2003JD003823.

Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature,* **430,** 768–772, doi:10.1038/nature02771.

Niu, G.-Y., and Z. Yang, 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *J. Hydrometeor.,* 937–953, doi:10.1175/JHM538.1.

——, ——, R. E. Dickinson, and L. E. Gulden, 2005: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *J. Geophys. Res.,* **110,** D21106, doi:10.1029/2005JD006111.

——, 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.

Poulin, A., F. Brissette, R. Leconte, R. Arsenault, and J.-S. Malo, 2011: Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *J. Hydrol.,* **409,** 626–636, doi:10.1016/j.jhydrol.2011.08.057.

Prein, A. F., G. J. Holland, R. M. Rasmussen, J. Done, K. Ikeda, M. P. Clark, and C. H. Liu, 2013: Importance of regional climate model grid spacing for the simulation of heavy precipitation in the Colorado headwaters. *J. Climate,* **26,** 4848–4857, doi:10.1175/JCLI-D-12-00727.1.

Raff, D., T. Pruitt, and L. Brekke, 2009: A framework for assessing flood frequency based on climate projection information. *Hydrol. Earth Syst. Sci.,* **13,** 2119–2136, doi:10.5194/hess-13-2119-2009.

Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet, 2014: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models. *Water Resour. Res.,* **50,** 409–426, doi:10.1002/2013WR014063.

Rasmussen, R., and Coauthors, 2011: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *J. Climate,* **24,** 3015–3048, doi:10.1175/2010JCLI3985.1.

——, and Coauthors, 2014: Climate change impacts on the water balance of the Colorado headwaters: High-resolution regional climate model simulations. *J. Hydrometeor.,* **15,** 1091–1116, doi:10.1175/JHM-D-13-0118.1.

Ray, A., J. Barsugli, K. Averyt, K. Wolter, M. Hoerling, N. Doesken, B. Udall, and R. Webb, 2008: Climate change in Colorado: A synthesis to support water resources management and adaptation. *Intermountain West Climate Summary,* Vol. 4, No. 8, Western Water Assessment, CIRES, Boulder, CO, 2–4. [Available online at http://wwa.colorado.edu/climate/iwcs/archive/IWCS_2008_Nov.pdf.]

Sankarasubramanian, A., and R. M. Vogel, 2002: Annual hydroclimatology of the United States. *Water Resour. Res.,* **38,** 1083, doi:10.1029/2001WR000619.

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

Seager, R., and Coauthors, 2007: Model projections of an imminent transition to a more arid climate in southwestern North America. *Science,* **316,** 1181–1184, doi:10.1126/science.1139601.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2005: Changes toward earlier streamflow timing across western North America. *J. Climate,* **18,** 1136–1155, doi:10.1175/JCLI3321.1.

Stocker, T. F., and Coauthors, Eds., 2014: *Climate Change 2013: The Physical Science Basis.* Cambridge University Press, 1535 pp.

Surfleet, C. G., D. Tullos, H. Chang, and I.-W. Jung, 2012: Selection of hydrologic modeling approaches for climate change assessment: A comparison of model scale and structures. *J. Hydrol.,* **464–465,** 233–248, doi:10.1016/j.jhydrol.2012.07.012.

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.

USDA, 1992: Forest land distribution data for the United States. Global Change Master Directory, NASA, digital media. [Available online at http://gcmd.nasa.gov/records/GCMD_USDA_FSSFES_FOREST.html.]

——, 1994: State Soil Geographic (STATSGO) data base: Data use information. USDA Misc. Publ. 1492, 113 pp.

Vano, J. A., T. Das, and D. P. Lettenmaier, 2012: Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *J. Hydrometeor.,* **13,** 932–949, doi:10.1175/JHM-D-11-069.1.

——, and Coauthors, 2014: Understanding uncertainties in future Colorado River streamflow. *Bull. Amer. Meteor. Soc.,* **95,** 59–78, doi:10.1175/BAMS-D-12-00228.1.

Wagener, T., and Coauthors, 2010: The future of hydrology: An evolving science for a changing world. *Water Resour. Res.,* **46,** W05301, doi:10.1029/2009WR008906.

Wilby, R. L., 2005: Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrol. Processes,* **19,** 3201–3219, doi:10.1002/hyp.5819.

——, and I. Harris, 2006: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resour. Res.,* **42,** W02419, doi:10.1029/2005WR004065.

——, L. E. Hay, and G. H. Leavesley, 1999: A comparison of downscaled and raw GCM output: Implications for climate change scenarios in the San Juan River basin, Colorado. *J. Hydrol.,* **225,** 67–91, doi:10.1016/S0022-1694(99)00136-5.

Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, 1992: A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res.,* **97,** 2717–2728, doi:10.1029/91JD01786.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.,* **117,** D03110, doi:10.1029/2011JD016051.

Xu, C., 1999: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches. *Prog. Phys. Geogr.,* **23,** 229–249, doi:10.1177/030913339902300204.

Yadav, M., T. Wagener, and H. Gupta, 2007: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Adv. Water Resour.,* **30,** 1756–1774, doi:10.1016/j.advwatres.2007.01.005.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.,* **116,** D12110, doi:10.1029/2010JD015140.

Yilmaz, K. K., H. V. Gupta, and T. Wagener, 2008: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resour. Res.,* **44,** W09417, doi:10.1029/2007WR006716.