The DOI for this manuscript is doi: 10.1175/JHM-D-13-083.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Elsner, M., S. Gangopadhyay, T. Pruitt, L. Brekke, N. Mizukami, and M. Clark, 2014: How does the Choice of Distributed Meteorological Data Affect Hydrologic Model Calibration and Streamflow Simulations? J. Hydrometeor. doi:10.1175/JHM-D-13-083.1, in press.

How does the Choice of Distributed Meteorological Data Affect Hydrologic Model Calibration and Streamflow Simulations?

Marketa M. Elsner[1]
Technical Service Center, US Bureau of Reclamation, Denver, CO

Subhrendu Gangopadhyay
Technical Service Center, US Bureau of Reclamation, Denver, CO

Tom Pruitt
Technical Service Center, US Bureau of Reclamation, Denver, CO

Levi Brekke
Technical Service Center, US Bureau of Reclamation, Denver, CO

Naoki Mizukami
NCAR Research Applications Laboratory, Boulder, CO

Martyn Clark
NCAR Research Applications Laboratory, Boulder, CO

[1] *Corresponding author address*: Marketa M. Elsner, Technical Service Center, U.S. Bureau of Reclamation, Bldg 67 5th Floor West, PO Box 25007 (86-68210),
Denver, CO 80225-0007.
E-mail: melsner@usbr.gov

Abstract

Spatially distributed historical meteorological forcings (temperature and precipitation) are commonly incorporated into modeling efforts for long-term natural resources planning. For water management decisions, it is critical to understand the uncertainty associated with the different choices made in hydrologic impact assessments (e.g., choice of hydrologic model, choice of forcing dataset, calibration strategy, etc.). This paper evaluates differences among four commonly used historical meteorological datasets and their impacts on streamflow simulations produced using the Variable Infiltration Capacity (VIC) model. The four meteorological datasets examined here have substantial differences, particularly in minimum and maximum temperatures in high elevation regions such as the Rocky Mountains. The temperature differences among meteorological forcing datasets are generally larger than the differences between calibration and validation periods. Of the four meteorological forcing datasets considered, there are substantial differences in calibrated model parameters and simulations of the water balance. However, no single dataset is superior to the others with respect to VIC simulations of streamflow. Also, optimal calibration parameter values vary across case study watersheds and select meteorological datasets, suggesting that there is enough flexibility in the calibration parameters to compensate for the effects of using select meteorological datasets. Evaluation of runoff sensitivity to changes in climate indicates that the choice of meteorological dataset may be as important in characterizing changes in runoff as climate change, supporting consideration of multiple sources of uncertainty in long-term planning studies.

1. Introduction

Use of sophisticated physical process models informed by statistically or dynamically downscaled climate change scenarios is increasingly becoming an integral part of long term natural resources planning. For example, the proposed listing of the North American Wolverine in 2013 as threatened under the Endangered Species Act (Federal Register, Vol. 78, No. 23) relied, in part, on work done by McKelvey et al. (2011) to evaluate the impacts of climate change on this distinct population, which depends heavily on contiguous snowpack. In addition, Wenger et al. (2011) identified opportunities for mitigation efforts to revive populations of trout species in the interior western United States based on an analysis of future climate change impacts. Finally, Bentz et al. (2010) utilized population models driven by projected climate scenarios to identify regions in North America with a high potential for bark beetle outbreak. For environmental management decisions highlighted by these studies, as well as water management decisions, understanding the uncertainty associated with various underlying modeling application choices is critical.

In an assessment of climate change impacts on water resources, modeling application choices may include historical and projected future climate datasets, model structure, and model calibration metrics, objective function, and calibration scheme. With respect to choice of historical meteorological forcings, studies have shown that the dataset choice may cause as much sensitivity in the resulting water balance as the choice of land surface model (Guo et al. 2006), if not more (Mo et al. 2012). Hossain and Anagnostou (2005) and Maggioni et al (2012) investigated the relative impact of model and rainfall forcing errors in hydrologic simulations by land surface models and found that both together contribute a large amount of the uncertainty in soil moisture estimates. Precipitation appears to cause the greatest sensitivity in runoff (Materia et al. 2009; Nasonova et al. 2011) and that sensitivity is not consistent across watersheds (Xue et al. 1991). Precipitation estimates are strongly dependent on the method used to interpolate the data, particularly in regions in the western United States where climate stations, upon which the datasets are based, are sparse (Mo et al. 2012). Mizukami et al. (2013) compared model simulations forced by two meteorological datasets (developed using different methodologies) and found that differences in shortwave radiation estimates have a large impact on hydrologic states and fluxes, particularly at higher elevation, influencing snow melt and runoff timing as well as evapotranspiration.

Other studies indicate that model structure may influence hydrologic model simulations. For example, Bohn et al (2013) found that the Thorton and Running (1999) approach for deriving meteorological forcings based on precipitation and temperature have inconsistent biases across large spatial domains. Clark et al. (2008) found that model structure is just as important as the choice of model parameters. Finally, Vano et al. (2012) found that hydrologic model structure significantly influences runoff sensitivities to changes in precipitation and temperature (i.e. imposed changes in climate).

Further, other studies suggest that calibration method may also affect hydrologic modeling results. Streamflow simulations may not be sensitive to calibration approach; however intermediate states such as potential evapotranspiration may differ substantially (Hay et al. 2000). Also, calibration parameters may not be stationary in time and simulation errors may increase with the time lag between calibration and simulation periods, as found by Merz et al. (2011) in their analysis of 273 catchments in Austria. With respect to climate change studies, Wilby (2005) found that the uncertainty in changes in projected future streamflow due to the choice of calibration period is similar to the uncertainty due to future greenhouse gas emissions scenarios. Also, Vaze et al. (2010) found that results from a hydrologic model calibrated over an average or wet climatic period are suitable for climate change impact studies where the difference between historical and predicted future rainfall is within about 15%.

Results from the previously mentioned studies suggest that hydrologic model calibration may be significantly impacted by choice of meteorological forcing dataset. Numerous meteorological forcing datasets have been developed over parts of the United States and they commonly consist of daily precipitation, temperature (minimum and maximum), and wind speed, at a minimum. Historical datasets are often developed based on interpolated data from National Weather Service daily cooperative observer (Co-op) stations (corrected for elevation) with specific needs in mind. For example, historical datasets developed by Maurer et al. (2002) and Livneh et al. (2013) (spanning 1915-2000 and 1915-2011, respectively) encompass the continental United States (CONUS) and their methodology focuses on the accuracy of spatial patterns and variability. The dataset developed by Wood and Lettenmaier (2006) (spanning 1915-2005 over the CONUS) was used as the basis of a west-wide seasonal hydrologic forecast system, which relied on a stations with real-time observations. Datasets by Hamlet and Lettenmaier (2005), Elsner et al. (2010), and Littell et al. (2011) (all spanning 1915-2006 and covering parts of the western United States) were developed with the objective of evaluating long-term climate trends and evaluating implications of climate change. For hydrologic model applications such as the Variable Infiltration Capacity (VIC) Model (Liang et al. 1994; Liang et al. 1996), additional meteorological forcings (i.e. humidity and radiative fluxes) need to be estimated from the diurnal temperature range and precipitation (e.g. using the approach of Thornton and Running 1999) or taken from other sources such as reanalysis products.

There is an increasing number of historical datasets based on reanalysis products such as the National Centers for Environmental Prediction, North American Regional Reanalysis (NARR; Mesinger et al. 2006). For example, the North American Land Data Assimilation System (NLDAS) Phase 2 (Xia et al. 2012) dataset is primarily derived from NARR data and this dataset is used by Mizukami et al. (2013) in their analysis of model sensitivities to meteorological forcings in mountainous terrain. Abatzoglou (2011) developed a 4-km gridded historical climate dataset based on the NLDAS Phase 2 dataset and the monthly 800 meter PRISM product (Daly et al. 2008).

Although there have been an increasing number of scientific studies exploring uncertainties associated with hydrologic model application choices, these uncertainties are still not well understood. Further, natural resource managers are increasingly using datasets and modeling tools, like those previously described, in long-term planning. Federal natural resource management and conservation agencies, including among others the Bureau of Reclamation (Reclamation), U.S. Geological Survey, U.S. Fish and Wildlife Service, U.S. Forest Service, National Oceanic and Atmospheric Association (NOAA), and Pacific Northwest National Laboratory, all have mandates for incorporating climate change into long-term planning. Climate projections originate from GCMs at coarse scale in space and time and are typically downscaled, either statistically or dynamically using a regional climate model, so that they may be useful for planning studies (e.g. Wood et al. 2004, Salathe et al. 2007, Christensen and Lettenmaier 2007, Maurer et al. 2007, among others). Statistically downscaled climate projections, arguably the type of projections most commonly used in long term planning studies, rely on historical meteorological datasets as the basis for downscaling. Numerous archives of statistically downscaled climate projections available for various domains within the western

United States utilize different historical datasets. For example, archives of hydro-climate scenarios developed for the Pacific Northwest (Hamlet et al. 2013), as well as major western United States river basins (Littell et al. 2011) at 1/16th degree spatial resolution, rely on historical datasets developed by Elsner et al. (2010) and Littell et al. (2011) as the basis for downscaling. In another example, Maurer et al (2007) developed an archive of statistically downscaled hydro-climate scenarios covering the CONUS plus contributing areas of Canada, which have served as a consistent dataset used by Reclamation in numerous basin studies pursuant to the SECURE Water Act of 2009 (Public Law 111-11), and rely on the historical dataset developed by Maurer et al. (2002) at 1/8th degree spatial resolution as its basis.

Greater understanding of the implications associated with using a particular historical dataset is important not only for historical hydrologic studies, but also for characterizing the uncertainty associated with projected future hydrologic conditions. In summary, this paper seeks to answer two questions:

(1) Is there an optimal distributed meteorological forcing dataset to be used in simulating streamflow through the VIC hydrological model?

(2) How does the choice of distributed meteorological data affect hydrologic model calibration and sensitivity analysis, particularly with respect to changes in climate?

In the following section, we describe the study approach. The study analysis is organized in two sections. First, we compare four meteorological forcing datasets commonly used in natural resource studies. Second, we discuss hydrologic model calibrations, using each of the four compared datasets, and resulting simulations. We conclude with a discussion of key findings in the context of various uncertainties in long-term natural resources planning studies.

2. Approach

2.1 Historical Meteorological Forcing Datasets

We compile and compare four spatially distributed meteorological datasets that differ in their use of station observations, handling of temporal inhomogeneities, spatial extent, spatial resolution, and temporal coverage. The four historical gridded meteorological datasets were developed by: 1) Maurer et al. (2002) – hereafter called the Maurer dataset; 2) Wood and Lettenmaier (2006) – hereafter called the Wood-Lettenmaier dataset; 3) Abatzoglou (2011) – hereafter called the Abatzoglou dataset; and 4) Elsner et al. (2010), expanded by Littell et al. (2011) – hereafter called the Elsner-Littell dataset (datasets are summarized in Table 1). We compare precipitation and temperature (maximum, minimum, and diurnal range) from these datasets over a common time period (water years 1980-1999), spatial resolution (1/8 degree), and domain, generally the United States portions of four major western hydrologic regions, including the Pacific Northwest (Columbia River Basin plus coastal drainages in Oregon and Washington); California; the Great Basin; the Colorado River Basin; and, the Missouri River basin west of 93 degrees west longitude (Fig. 1). The Maurer and Wood-Lettenmaier datasets have a native spatial resolution of 1/8 degree and use a common grid, consistent with the North American Land Data Assimilation System (NLDAS, Mitchell et al. 2004). The Abatzoglou and Elsner-Littell datasets were aggregated from their native resolution (4-km and 1/16th degree, respectively) to the same common 1/8 degree grid, using a local area averaging approach. Consequences of aggregating precipitation and temperature from these datasets are not explored in this study. However, we may speculate reduced error in precipitation and temperature aggregated from finer scale to 1/8 degree due to the fact that coarse station observations are the basis for development of both Abatzoglou and Elsner-Littell datasets. In addition, Gangopadhyay et al. (2004) evaluated the impacts of spatial aggregation on precipitation forecast skill in the context of statistically downscaled precipitation estimates. They found that spatial averaging either had little effect or increased the skill of downscaled precipitation estimates. Additional studies may be needed to evaluate the issue of scale of meteorological data for watersheds smaller than those considered in this study (the smallest of which is 1,792 square kilometers). Distinguishing characteristics of the four datasets are summarized in Table 1. We refer to their associated publications for details regarding the purpose and applications of each dataset, and the approaches taken in developing them.

The Maurer, Wood-Lettenmaier, and Elsner-Littell gridded precipitation fields are primarily based on the Co-op Station Network (along with similar networks in Canada and Mexico), interpolated to a grid using the SYMAP algorithm (Shepard 1984). The Maurer dataset only includes stations with more than 20 years of data from 1949-2000. The Wood-Lettenmaier dataset only includes stations that have both long term records and report in real time (through 2005). These stations have at least 45 years of record and at least 80% coverage of the period between 1915 and 2005 (Wood 2008). The Elsner-Littell dataset follows the approach of Hamlet and Lettenmaier (2005) and only includes stations with at least 5 years of data and at least one continuous year from 1915-2006. The dataset is then corrected for temporal inhomogeneities by use of monthly Historical Climatology Network (HCN) data (and Canadian equivalent). Precipitation fields from all three of the above mentioned datasets incorporate a correction to monthly climatologies from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al. 2008) albeit for slightly different time periods (1961-1990 for Maurer and Wood-Lettenmaier datasets and 1971-2000 for Elsner-Littell dataset). The Abatzoglou precipitation fields are derived from NLDAS Phase 2 data (Xia et al. 2012), comprised of gage data (Co-op stations included), radar, and reanalysis data (at 32-km spatial resolution). The Abatzoglou dataset applies a secondary correction to the monthly 800 meter PRISM timeseries.

Temperature (minimum and maximum) fields in the Maurer, Wood-Lettenmaier, and Elsner-Littell datasets are also obtained from Co-op stations (station mix as described for precipitation) and are lapsed (at -6.5 degrees Celsius [C] per km) to the mean grid cell elevation. The Elsner-Littell dataset, however, applies a secondary correction of average temperature to the PRISM climatologies (preserving the range between minimum and maximum temperature in Co-op station data). The Abatzoglou temperature fields are based on NLDAS Phase 2 and a secondary correction to monthly 800 meter PRISM timeseries.

The Maurer, Wood-Lettenmaier, and Elsner-Littell datasets rely on wind speeds from NLDAS Phase 1, which are downscaled wind fields from the National Centers for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR) reanalysis products (Kalanay et al. 1996). The wind speeds in the Abatzoglou dataset are taken from the NLDAS Phase 2, which is based on the NCEP North American Regional Reanalysis (NARR). Barsugli et al. (2012) found that, in Colorado, NARR windspeeds are substantially greater than NCEP-NCAR windspeeds at higher elevations and that NARR windspeeds more closely compare with available observations. They also demonstrate that choice of windspeed data may impact resulting streamflow simulations. However, we choose not to compare differences in wind speed in this study, in part because there is less confidence overall in gridded windspeed data, and the use of the Abatzoglou dataset with NARR windspeeds helps to demonstrate the sensitivity of simulated streamflow to changes in meteorological forcings.

2.2 Case Study Watersheds

We investigate the implications of model calibration using each of these datasets on seven case study watersheds across the domain, namely: 1) Animas River at Durango, CO (USGS ID 09361500, hereafter called ANIMS); 2) Dolores River near Cisco, UT (USGS ID 09180000, hereafter called DOLOR); 3) Green River at Green River, UT (USGS ID 09315000, hereafter called GREEN); 4) Missouri River at Toston, MT (USGS ID 06054500, hereafter called MISSO); 5) Sacramento River at Bend Bridge near Red Bluff, CA (USGS ID 11377200, hereafter called SACRB); 6) Salt River near Chrysotile, AZ (USGS ID 09497500, hereafter called SALTC); and, 7) Snake River near Heise, ID (USGS ID 13037500, hereafter called SNAKE).

Specifically, we explore whether calibration of a hydrologic model using one meteorological dataset yields significantly different calibration parameters than a model calibrated using a different meteorological dataset. Further, we explore whether a hydrologic model calibrated to one meteorological dataset yields significantly different results when forced with a different meteorological dataset. Lastly, we explore the sensitivity of runoff to changes in climate (as represented by differences between calibration and validation periods) using the four calibrated models. Direct comparisons of the distributed meteorological datasets and evaluation of hydrologic model simulations over the case study watersheds allows us to better understand the implications of these datasets with respect to long-term planning studies.

2.3 Modeling Scheme

To represent physical hydrologic processes in the seven case study watersheds, we apply the VIC hydrologic model. The VIC model has been widely used in large scale hydrologic studies across the globe and to explore the implications of climate change on water and other resources including forests, agriculture, fish and wildlife (e.g. Christensen and Lettenmaier 2007, Elsner et al. 2010, Wenger et al. 2011). It was employed in the same studies for which three of the four comparison datasets were developed, with the exception of the Abatzoglou dataset. The VIC model was also used to validate the datasets developed as part of the NLDAS project (Mitchell et al. 2004, Xia et al. 2012). The model configuration used here is consistent with that used in the Reclamation's West-wide Climate Risk Assessment (Reclamation 2011). Namely, we apply VIC model version 4.0.7 (also used by Elsner et al. 2010 and Hamlet et al. 2013) to simulate surface runoff and baseflow per model grid cell. We then apply the Lohmann et al. (1998) model to route surface runoff and baseflow to select locations, producing simulated natural streamflow. Natural flows are defined as streamflow that would exist in the absence of diversions and return flows resulting from human activities. Hydrologic model simulations are performed in water balance mode using a daily time step water balance and 1-hour time step internal snow model.

VIC model calibration is conducted using the multi-objective complex evolution approach developed by Yapo et al. (1998). The user may define the calibration parameters, and the objectives (calibration metrics) on which to base the objective function. Pareto sets are theoretically equal in terms of their objective functions. As such, one set of parameters was generally chosen manually from the Pareto optimal set. Bennett et al. (2012) showed that the choice of model parameter set within the Pareto optimal set had minimal impact on resulting hydrologic simulations in analyzed watersheds of British Columbia. Calibrations are repeated up to seven times to ensure parameters were globally optimal and to account for lack of convergence in some calibrations. Calibration metrics include three error statistics computed between simulated and reconstructed natural streamflow, which is considered the best estimate of observed natural conditions. The objective function for calibration is computed based on three metrics: the Nash-Sutcliffe Efficiency computed using monthly flows (NSE_{mon}), the root mean squared error of monthly flows divided by the observed mean monthly flow ($RMSE_{mon}$), and the normalized error in mean monthly flow volume ($VolErr_{mon}$). These metrics were chosen to reduce errors in seasonal timing and magnitude of flow (NSE_{mon} and $RMSE_{mon}$) as well as reduce error in annual flow volume ($VolErr_{mon}$). All three metrics generally have values between 0 and 1; however, $VolErr_{mon}$ is generally quite low, effectively giving the NSE_{mon} and $RMSE_{mon}$ metrics relatively greater weight. The NSE_{mon} function emphasizes the high–peak flow periods and therefore produces parameters that optimize hydrograph performance during the seasonal peak (Bennett et al. 2012, Clark et al. 2008). The $VolErr_{mon}$ strictly emphasizes volume conservation over the calibration period and is not responsive to errors in streamflow timing or seasonality (Bennett et al. 2012).

We evaluate the sensitivity of streamflow to variations in common VIC model calibration parameters over the seven case study watersheds in order to determine the most appropriate calibration parameter set. Model parameters considered for calibration are summarized in Table 2. Sensitivity is evaluated based on perturbation experiments spanning the accepted range of each parameter. The three calibration metrics described above are computed for each perturbation experiment and metrics are compared across case study watersheds. Parameter sensitivity may be dependent on watershed, making it difficult to apply a stringent threshold for each calibration watershed. Therefore, for a single parameter, if the majority (i.e. more than half) of the metrics for all calibration watersheds varies by less than 10 percent, that parameter is considered insensitive. Based on this sensitivity analysis, the following parameters were chosen: Ds, Ws, Dsmax, D2, and D3. Ds, Ws, and Dsmax are parameters that define the shape of the baseflow curve (Liang et al. 1994). D2 and D3 consist of the depth of the middle and deepest of three model soil layers. Other parameters, including the parameter defining the shape of the variable infiltration capacity curve (bi), wind speed attenuation through the canopy, snow roughness, radiation attenuation in the canopy, and routing flow velocity, were found to minimally contribute to VIC model sensitivity and were not modified during calibration (Table 2). Choosing appropriate calibration parameters, while limiting the number, allows for successful and more computationally efficient model calibrations (Kampf and Burges 2007).

2.4 Evaluation Methods

Model simulations are performed over seven case study watersheds to evaluate the implications of using different meteorological datasets on simulated streamflow. Case study watersheds represent each of the major western United States watersheds under Reclamation's purview and vary in size, elevation, aspect, and climatic conditions. The time period of model calibration and validation is dictated by the length of record of available observed reconstructed natural streamflow and meteorological data, but is also chosen to include a range of hydrologic conditions. Table 3 summarizes the characteristics of each case study watershed and identifies their model calibration/validation periods.

To evaluate the implications of VIC model calibration on simulated streamflow, we employ a procedure where the VIC model is calibrated for each of the case study watersheds and using one of the four select meteorological datasets. Each calibrated model is then forced with the remaining three meteorological datasets. Resulting simulated mean monthly hydrographs for each watershed are compared with reconstructed natural streamflow.

The sensitivity of runoff to changes in climate is also explored using the calibrated simulations by partitioning the validation period for each case study watershed (generally a 10 year period, but 5 years for MISSO and 7 years for ANIMS; see Table 3) into cool-wet and warm-dry water years. Cool-wet and warm-dry validation years were selected based on their computed difference (in percent and degrees C, respectively) from the median of annual precipitation and temperature over the simulation period, 1980-1999 water years. Since the change in climate between calibration and validation periods for most case study watersheds (except MISSO) is as great as the change in climate between meteorological datasets, we choose these two converse year types to help demonstrate the greatest potential change in runoff sensitivity due to dataset choice and provide context for potential implications. Unique groups of years were selected and averaged to generate mean annual precipitation, temperature, and runoff for each case study watershed and meteorological forcing dataset. However, some years were commonly classified as cool-wet and warm-dry for most watersheds and meteorological forcing datasets (e.g. water year 1982 was a common cool-wet year, while 1981 was a common warm-dry year). For each calibrated model, change in mean annual runoff between calibration period and each of the two validation year types (as a function of change in climate - mean annual precipitation and temperature) is computed to determine whether runoff sensitivity changes with change in climate or meteorological forcing dataset.

3. Comparison of Spatially Distributed Meteorological Data

Four meteorological forcing datasets (Maurer, Wood – Lettenmaier, Abatzoglou, and Elsner – Littell) are compared across a common study domain (see purple dashed line in Fig. 1) and time period (1980-1999 water years). The datasets are compared with respect to precipitation (Prcp) and temperature (minimum [Tmin], maximum [Tmax], and diurnal range [Tran]). Across the common domain, datasets are compared based on their means, standard deviation, and correlations. Similar analyses are performed over a longer period (1950-1999 water years), with the exception of the Abatzoglou dataset (which begins in 1979), and comparable results are found and, therefore, not presented. In addition to a comparison across the study domain, the datasets are compared over seven case study watersheds based on monthly values over calibration, validation, and simulation periods. For both sets of comparisons, statistics are computed using monthly and annual totals for precipitation and daily averages over the month or year for temperature.

3.1 Differences in Meteorological Forcings across Study Domain

Figures 2 through 7 illustrate monthly and annual statistics for all four variables. Values are presented as comparisons of the Abatzoglou (A), Elsner-Littell (EL), and Wood-Lettenmaier (WL) datasets to the Maurer (M) dataset. The Maurer dataset is commonly used in statistical downscaling efforts and is the baseline historical dataset used in Reclamation's West Wide Climate Risk Assessment (Reclamation 2011). It is therefore used as the basis for comparison of the remaining three datasets. Figures 2 and 3 show percent differences in precipitation statistics between datasets (computed over 1980-1999 water years), while Figures 4 through 7 show absolute differences in temperature statistics in degrees C. Boxplots in Figures 2 and 4 through 6 compare annual values across VIC grid cells, where the boxes represent the 25th, 50th, and 75th percentile values, while the whiskers represent the 5th and 95th percentiles. Monthly statistics were similarly analyzed, but the results are not presented here, as they are consistent with annual statistics overall. However, notable differences between monthly and annual statistics are discussed. Figures 3 and 7 illustrate how precipitation and temperature (Tmax, Tmin, and Tran) vary spatially in winter and summer, represented by January and July, respectively.

Results show considerable differences in precipitation among datasets, both in terms of distribution of statistics (Fig 2) and spatial differences (Fig. 3). In particular, note that over 50% of grid cells in Fig. 2 have differences in precipitation greater than 10%, as can be seen by the difference between 25^{th} and 75^{th} percentile values. Although there are considerable differences in some parts of the domain (Fig. 3), the medians of precipitation difference are close to zero (5 percent or less). Monthly analysis shows greater a distribution of differences in July than other months, likely corresponding with a smaller magnitude of precipitation occurring in much of the western United States in summer. In January, the Maurer dataset generally has more precipitation (median negative difference on the order of 5-10 percent) in the northern portion of the domain (defined as north of the California-Oregon border at 42 degrees N latitude) and less precipitation (median positive difference on the order of 0-5 percent) in the southern portion of the domain, compared with the alternate datasets (Fig. 3). In July, the Maurer dataset generally has less precipitation than the compared datasets in all regions. The exceptions include a median negative difference in California of about 38 percent comparing it with the Wood-Lettenmaier dataset, and of about 4 percent comparing it with the Elsner-Littell dataset. To put these results in context, consider that many future climate projections suggest changes in precipitation within +/- 10% by the 2050s (Reclamation 2011). The differences are notable, despite the expectation that the Wood-Lettenmaier dataset is more similar to the Maurer dataset with respect to precipitation, than either the Elsner-Littell or Abatzoglou datset, due to the use of the same PRISM dataset for secondary corrections, namely the 1961-1990 climatology. PRISM climatologies cannot be directly compared and, by extension, cannot be attributed as the sole source of differences between datasets because their products incorporate data improvements and station networks and underlying data are not consistent between products.

There are also considerable differences in temperature among datasets (Fig. 4 showing meand annual maximum temperature, Fig. 5 showing mean annual minimum temperature, Fig. 6 showing mean annual diurnal temperature range, and Fig. 7 showing spatial differences for

January and July). Specifically, the Elsner-Littell dataset shows differences in mean annual maximum temperature greater than 1 degree C for approximately 25% of grid cells (Fig.), while Elsner-Littell and Abatzoglou datasets show differences in mean annual minimum temperature greater than 1 degree C for approximately 25% of grid cells (Fig. 4), with the Abatzoglou dataset showing minimum temperature differences in the daily mean greater than 2 degrees C for approximately 25% of grid cells. Monthly analysis shows the greatest distribution of differences occurs in the cool season (approximately September to March). Temperature differences are most pronounced in high elevation areas, especially throughout the Rocky Mountains (Fig. 7). Additionally, the Abatzoglou dataset has a generally lower diurnal temperature range than the Maurer dataset, particularly during July. As described in section 2.1, the datasets differ in their corrections of temperature by elevation. Maurer and Wood-Lettenmaier datasets impose a constant lapse rate (-6.5 degrees C per km) in the gridding of temperature from station observations, while the Abatzoglou and Elsner-Littell datasets incorporate corrections to finer scale PRISM temperature climatologies (described in section 2.1), causing substantial differences in daily mean minimum and maximum temperatures, particularly at higher elevations. A lapse rate of -6.5 degrees C per km appears to be too high for temperature based on recommended lapse rates in complex terrain (e.g. Blandford et al. 2008; Minder et al. 2010). Blandford et al. (2008) found that this lapse rate may be applicable to maximum temperature, but grossly overestimates actual lapse rates for daily minimum and mean temperature. Mizukami et al. (2013) further discuss the significant implications of the use of a constant lapse rate on the diurnal temperature range and empirical estimates of shortwave radiation.

In comparison of standard deviation between three datasets (Abatzoglou, Elsner-Littell, Wood-Lettenmaier) to the Maurer dataset, it is evident that the Wood-Lettenmaier dataset has more similar variability than the other datasets for Prcp (Fig. 2). However for temperature, (Tmin, Tmax, and Tran) the variability is generally comparable (see Figs. 4 through 6). Correlation between datasets across the entire study domain is highest between Abatzoglou and Maurer datasets for precipitation and temperature (Tmin, Tmax, and Tran) and generally lowest between Elsner-Littell and Maurer datasets, which is interesting provided Abatzoglou and Elsner-Littell datasets both apply temperature corrections based on PRISM climatologies. It may be speculated that for the Elsner-Littell dataset, the use of monthly HCN (and Canadian equivalent) station data to correct for temporal inhomogeneities in precipitation and temperature, due to the use of relatively short station records (minimum of 5 years, with one year of continuous data), may alter daily precipitation values enough to cause the lower correlations between the Elsner-Littell and Maurer datasets for precipitation and temperature (and generally lower correlations between Elsner-Littell and other datasets as well, although results are not shown). The Abatzoglou dataset, which is based on a combination of CPC daily gage data and National Weather Service Stage II radar, does not incorporate a similar monthly correction factor using HCN station data.

3.2 Differences in Meteorological Forcings across Basins

Figure 8 summarizes differences in mean annual precipitation and temperature (average [Tavg], Tmax, and Tmin) between Abatzoglou, Elsner-Littell, Wood-Lettenmaier and the reference Maurer dataset. These differences are shown for calibration, validation, and overall simulation periods, and over the seven case study watersheds, which span a range of geographic regions and elevations. The figure informs analysis of hydrologic model calibration and simulations (section 4). Case study watersheds are presented in order of mean watershed elevation; the watershed with the lowest mean elevation (SACRB) is on the far left of each figure panel, while the watershed with the highest mean elevation (ANIMS) is on the far right.

Calibration and validation periods (as well as overall simulation period which includes both) for each case study watershed are generally similar in climate, with precipitation differences generally less than 10% and temperature differences less than 0.5 degrees C. The MISSO watershed is the exception, where mean annual precipitation over the calibration and validation periods differ by 18-20%.

Interestingly, substantial differences are evident between alternate meteorological forcing datasets and the Maurer reference dataset. Figure 8 shows that for temperature, the differences among datasets are larger than the differences between calibration and validation periods, with differences up to 3 degrees C. For precipitation, the differences among meteorological forcing datasets are comparable with differences between calibration and validation periods, with differences generally less than 10% with the exception of the MISSO basin (as previously described).

Specifically, Abatzoglou and Elsner-Littell datasets have higher daily average temperature than Maurer and Wood-Lettenmaier datasets for all case study watersheds, with the differences in daily average temperature are primarily driven by differences in the daily minimum for the Abatzoglou dataset and daily maximum for the Elsner-Littell dataset. Mean annual precipitation between the four meteorological forcing datasets is within +/- 10% in each of the case study watersheds. Higher elevation watersheds (SNAKE and ANIMS watersheds) exhibit the greatest difference in temperature between these datasets for reasons described in section 3.

4. Hydrologic Model Simulations for Case Study Watersheds

Hydrologic model calibrations and simulations for seven case study watersheds are evaluated to improve our understanding of potential impacts of meteorological forcings on model calibration parameters.

4.1 Differences in Calibrated Parameters and Model Performance

Each of the seven case study watersheds is calibrated through implementation of an automated multiple objective approach using the VIC hydrologic model. Table 4 summarizes the resulting optimal parameter values. In general, there does not appear to be a relationship between optimal parameters and either watershed or meteorological dataset. This suggests that different parameter combinations may result in similar objective function values for a given watershed and meteorological forcing dataset. Alternatively, it may suggest that optimal parameter combinations may not coincide with the best representations of model physics, but instead are compensating for biases in forcing data and weaknesses in model structure.

Model performance during calibration and validation periods does not depend on the choice of meteorological dataset (Table 5). The NSE_{mon}, which is used as a hydrologic metric to evaluate model simulations of seasonal flow volume and timing and the characteristic shape of the hydrograph, is above 0.70 for all but one model calibration (MISSO watershed calibrated using the Elsner-Littell dataset), indicating a good fit between simulated and reconstructed natural streamflow (NSE_{mon} may vary between $-\infty$ and 1, with 1 being perfect). Calibration of SNAKE and SACRB watersheds result in the highest NSE_{mon} (between 0.93 and 0.98 for SNAKE and between 0.92 and 0.95 for SACRB), consistently across models calibrated with each meteorological dataset. Calibration of DOLOR and MISSO result in the lowest NSE_{mon}

values, but still close to or above 0.70 (between 0.76 and 0.78 for DOLOR and between 0.69 and 0.80 for MISSO). Similar results are evident for $RMSE_{mon}$. There is not one meteorological dataset that results in model calibrations with more optimal (higher) NSE_{mon} values, indicating that the quality of the datasets are comparable or there is enough flexibility in the model parameters to compensate for differences among forcing datasets.

4.2 Assessment of Compensatory Errors

We evaluate the forcing of calibrated models for the case study watersheds (to each of the four meteorological forcing datasets) with alternate forcing datasets (Fig. 9) to understand the influence of meteorological datasets on streamflow, as well as of the sensitivity of model simulations to calibration. In Fig. 9, the meteorological dataset listed in the legend title for each panel is the “base” meteorological dataset used for model calibration. The red solid line in each panel illustrates the resulting mean monthly hydrograph from “base” calibrated simulations, having corresponding dataset and calibration parameters. The colored dashed lines illustrate mean monthly hydrographs from simulations using the calibrated parameters from the base simulation along with alternate meteorological datasets. The solid black line in each panel illustrates the mean monthly reconstructed natural streamflow hydrograph.

For the ANIMS watershed, simulated flow resulting from models calibrated with Abatzoglou and Wood-Lettenmaier datasets (second and fourth panels from left) are closer to reconstructed natural streamflow than flow resulting from models calibrated with the other datasets (see calibration statistics in Table 5). Also, models calibrated with Elsner-Littell and Maurer datasets (first and third panels from left), when forced with the Abatzoglou dataset, perform better than the calibrated models themselves (e.g. NSE_{mon} improved from 0.70 to 0.81 in the Elsner-Littell calibrated model and from 0.84 to 0.87 in the Maurer calibrated model).

However, in each of the simulations, model calibration and meteorological dataset combination do little to change the magnitude of flows during the low flow period (autumn and winter).

For the DOLOR watershed, forcing calibrated models using alternate meteorological datasets does not improve existing errors in the calibrated models in flow magnitude during autumn and winter months. The model calibrated using the Wood-Lettenmaier meteorological dataset (fourth panel from left) more closely captures the mean reconstructed natural streamflow seasonal peak magnitude and has the best calibration error statistics of the four calibrated DOLOR models.

For the GREEN watershed, each of the calibrated models results in mean monthly hydrographs that closely correspond with reconstructed natural streamflow and the forcing of these models with alternate datasets does not significantly change the results. It may be speculated that the relative insensitivity of simulated streamflow to forcing dataset or calibration parameters in the GREEN watershed is likely due the relatively large size of the GREEN watershed compared with other case study watersheds as well as its hydrologic characteristics. The GREEN watershed (approximately 116,000 square kilometers) is approximately three times larger than the next largest case study watershed, MISSO (approximately 40,000 square kilometers). Compensatory errors have a greater tendency to negate each other in a larger watershed, resulting in simulations that closely correspond with reconstructed natural flow. For example, errors in interpolated meteorological station data are more likely to impact a small watershed that may have few or no stations within it. Also, GREEN is a snowmelt dominant watershed, which reduces the relative effects of other processes on the water balance (such as effects of subsurface flow).

For the MISSO watershed, each of the calibrated models results in mean monthly hydrographs that do not correspond well with reconstructed natural streamflow with respect to the seasonal peak. It appears that over this watershed the Elsner-Littell and Abatzoglou datasets yield similar flows because, in the left most panel (model calibrated with Elsner-Littell dataset), the simulated flows from the Abatzoglou-forced model closely correspond with the Elsner-Littell optimal calibrated flows (red line). Using an analogous comparison, it appears that the Maurer and Wood-Lettenmaier datasets yield similar flows, as seen in the panel third from left, where the flows resulting from the Abatzoglou-forced model closely correspond with the Maurer optimal calibrated flows.

For the SACRB and SALTC watersheds, it appears that simulated flows using a model forced by the Elsner-Littell dataset differs noticeably from others. In the top left panel, simulated flows using the Elsner-Littell calibrated model and forced with alternate datasets all show significantly lower mean seasonal peaks. Similarly, results from each of the other calibrated models show the Elsner-Littell forced flows have significantly higher seasonal peaks.

For the SNAKE watershed, it appears that simulated flows using a model forced by the Maurer dataset differs noticeably from others, similarly to the comparison described above for SACRB and SALTC. Unique differences in mean monthly hydrographs for each basin suggest that there may be compounding effects of forcing dataset, model calibration, and physical representation of important watershed processes.

4.3 Sensitivity of the Portrayal of Climate Impacts to Calibrated Parameters

In a final analysis, we evaluate the sensitivity of runoff change to observed historical changes in precipitation and temperature (combined) using calibrated models forced with the four meteorological datasets in attempt to differentiate changes in sensitivity due to changes in climate and to choice of dataset. Figure 10 summarizes the results for each case study watershed, with panels ordered column wise by lowest mean elevation (SACRB) to highest mean elevation (ANIMS). Each panel shows change in mean annual water year precipitation (percent) versus change in mean annual runoff (percent) between the calibration period and select years in the validation period. The size of each plotted symbol represents the corresponding magnitude (absolute value) of change in annual temperature (degrees C). The diamonds in each figure panel correspond with cool-wet validation years, while circles correspond with warm-dry validation years. Individual points represent results for one calibrated model simulation corresponding with the forcing dataset used for calibration. For all basins but MISSO and ANIMS, the computed change in precipitation between calibration years and cool-wet validation years is generally positive, while the change between calibration and warm-wet validation years is generally negative.

Figure 10 illustrates that precipitation is the primary driver of runoff change, which is consistent with conclusions of Materia et al. (2009), Nasonova et al. (2011), and Xue et al. (1991). Generally, increases in precipitation correspond with greater increases in runoff, similar to findings by Elsner et al. (2010) and Vano et al. (2012) which indicate about a 12-20% increase and a 20-30% increase in runoff for a 10% increase in precipitation for watersheds in Washington and the Colorado River basin, respectively.

The figure also shows that precipitation change and corresponding changes in runoff can be substantially different between datasets, on the order of, or greater than, projected changes in precipitation by the 2050s. The expectation would be that changes in precipitation and runoff from different calibrated models (and correspondingly different meteorological forcings) would cluster in two distinct groups corresponding to warm-dry and cool-wet regimes. Such clustering is evident for the SACRB, for example. However, some watersheds have substantial differences (SALTC, for example), indicating that the choice of meteorological dataset may be as important in characterizing changes in runoff as is climate change.

Anomalies to the above generalizations regarding changes between calibration years and select validation years exist for the MISSO and ANIMS watersheds. In the MISSO watershed, the computed change in precipitation is positive between calibration years and both sets of validation years. For this watershed, as noted previously in the comparison of forcing datasets, all validation years were wetter than the calibration years, hence showing positive change precipitation, even in so-called warm-dry years (see also Fig. 8). For the ANIMS watershed, no validation years were classified as cool-wet for the Abatzoglou or Maurer datasets, so changes could not be computed. Plotted changes in precipitation and temperature for cool-wet validation years for the Elsner-Littell and Wood-Lettemnaier datasets show slightly less precipitation (by approximately 3 percent), despite the cool-wet classification, along with negative and positive changes in runoff (respectively). We speculate that the increased runoff with reduced precipitation, computed for the simulations using the Wood-Lettenmaier dataset, is an anomalous result of averaging mean annual values across select validation years.

5. Discussion

By comparing four spatially distributed meteorological forcing datasets and conducting experiments based on combinations of forcings and calibrated VIC hydrologic models, we seek to determine whether there is an optimal forcing dataset to be used by hydrologic models to simulate streamflow, and whether the choice of dataset affects VIC model calibration and portrayal of climate sensitivity.

The meteorological datasets considered (Abatzoglou, Elsner-Littell, Maurer, Wood-Lettenmaier) have substantial differences, particularly in minimum and maximum temperatures in higher elevation regions, which are primarily attributed to the approach taken to adjust temperature by elevation when interpolating station data to a grid. Temperature influences derived forcings within the VIC hydrologic model, such as radiation, and, consequently, the accumulation and ablation of the mountain snowpack. Therefore differences in minimum and maximum temperature may significantly affect the simulated water balance.

The temperature differences among meteorological forcing datasets are generally larger than the differences between calibration and validation periods. For precipitation, the differences among datasets are comparable with differences between calibration and validation periods, with the exception of the MISSO basin where the calibration and validation periods differ by 18-20%.

Although there are substantial differences among these datasets, no single dataset is superior to the others with respect to VIC simulations of streamflow. Also, there is no apparent relationship between optimal calibration parameter values and meteorological dataset or watershed, suggesting that the quality of the datasets is comparable or there is enough flexibility in the model parameters to compensate for differences among forcing datasets and potential biases in process representation.

The model calibration analysis shows that choice of forcing dataset influences VIC model calibration with respect to calibration parameters and resulting streamflow, in particular seasonal streamflow peaks. For example, in the ANIMS watershed, the Abatzoglou dataset results in better model performance according to the chosen calibration metrics, even when the model was calibrated to another dataset. In the SACRB watershed, the Elsner-Littell dataset results in significantly different mean monthly hydrographs than models using other datasets.

Finally, regarding exploration of runoff sensitivity to portrayal of climate impacts, we find that precipitation change and corresponding changes in runoff can be substantially different between datasets, on the order of, or greater than, projected climate change by the 2050s. This indicates that the choice of meteorological dataset may be as important in characterizing changes in runoff as climate change. Further, choice of meteorological forcing dataset will influence statistical downscaling of projected climate scenarios from coarser scale (in space and time) GCMs, thereby influencing the uncertainty associated with downscaled climate projections.

This work supports previous findings, suggesting that there are significant differences in meteorological forcing datasets, downscaling of global climate projections, hydrologic model constructs, and model calibration schemes, all of which may impact the portrayal of climate change impacts in long term natural resources planning studies. This work, along with other mentioned studies, supports the argument that consideration of uncertainties in modeling frameworks is as important as consideration of an ensemble of future climate projections in long-term planning studies. Further studies exploring the sensitivity of other hydrologic variables beyond streamflow (i.e. snowpack, evapotranspiration, etc.) to choice of meteorological forcing dataset, changes in runoff sensitivity due to hydrologic model calibration, as well as studies using ensembles of approaches and techniques (including additional hydrologic models), will enhance understanding of uncertainties and are critical for identifying best practices for applications.

Acknowledgments

The authors would like to thank Dr. Alan Hamlet (Assistant Professor at University of Notre Dame) and Dr. Andrew Wood (Research Scientist, NCAR Research Applications Laboratory) for their insights into the development of the meteorological forcing datasets, as well as collaborators at the National Center for Atmospheric Research (Ethan Gutmann and Pablo Mendoza) for their constructive comments in the later stages of this study. The authors would also like to thank two anonymous reviewers for their valuable comments. Finally, the authors would like to acknowledge the Bureau of Reclamation Research and Development Office for financially supporting this study.

References

Abatzoglou, J. T. 2011: Development of gridded surface meteorological data for ecological applications and modeling. *Int. J. Climatol.*, **33**, 121–131.

Barsugli, J.J., M.M. Elsner, and A.F. Hamlet, 2012. *Building a Stronger and More Extensive Hydrologic Foundation for Environmental Flow and Climate Change Research Across the Colorado River Basin.* Final project report, prepared for The Nature Conservancy.

Bennett, K.E., A.T. Werner, M. Schnorbus. 2012: Uncertainties in Hydrologic and Climate Change Impact Analyses in Headwater Basins of British Columbia. *J. Climate*, **25**, 5711–5730.

Bentz, B.J., J. Régnière, C.J. Fettig, E.M. Hansen, J.L. Hayes, J.A. Hicke, R.G. Kelsey, J.F. Negrón, and S.J. Seybold. 2010. Climate change and bark beetles of the western United States and Canada: direct and indirect effects. *BioScience*, 60 (8), 602-613..

Blandford, T. R., K. S. Humes, B. J. Harshburger, B. C. Moore, V. P. Walden, and H. Ye, 2008: Seasonal and Synoptic Variations in Near-Surface Air Temperature Lapse Rates in a Mountainous Basin, Journal of Applied Meteorology and Climatology, 47, 249-261.

Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models, *Agric. For. Meteorol.*, **176**, 38-49.

Christensen N. 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 System Sci.*, **11**, 1417-1434

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay, 2008: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resour. Res.*, **44**, W00B02.

Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, G.H. Taylor, J. Curtis, and P.A. Pasteris, 2008: Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. Int. J. Climatol., 28, 2031-2064.

Elsner, M.M., L. Cuo, N. Voisin, J.S. Deems, A.F. Hamlet, J.A. Vano, K.E.B. Mickelson, S.Y. Lee, and D.P. Lettenmaier, 2010: Implications of 21st Century climate change for the hydrology of Washington State. *Climatic Change*, **102 (1-2)**, 225-260.

Gangopadhyay, S., M. Clark, K. Werner, D. Brandon, and B. Rajagopalan, 2004: Effects of Spatial and Temporal Aggregation on the Accuracy of Statistically Downscaled Precipitation Estimates in the Upper Colorado River Basin. *J. Hydrometeor*, **5**, 1192–1206.

Guo, Z., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao, 2006: Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 2. Sensitivity to external meteorological forcing. *J. Geophys. Res.: Atmos.*, **111**, n/a-n/a.

Hamlet, A. F. and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental U.S. *J. Hydrometeorol.*, **6 (3)**, 330-336.

Hamlet, A. F., M.M. Elsner, G. Mauger, S. Y. Lee, and I. M. Tohver, 2013: An Overview of the Columbia Basin Climate Change Scenarios Project: Approach, Methods, and Summary of Key Results. *Atmos.-Ocean*, **51(4)**, 392-415.

Hay, L. E., R. L. Wilby, and G. H. Leavesley, 2000: A COMPARISON OF DELTA CHANGE AND DOWNSCALED GCM SCENARIOS FOR THREE MOUNFAINOUS BASINS IN THE UNITED STATES. *J. Amer. Water Resour. Assoc.*, **36**, 387–397.

Hay, L.E., G. H. Leavesley, M. P. Clark, S. L. Markstrom, R. J. Viger, and M. Umemoto, 2006: Step Wise, Multiple Objective Calibration of a Hydrologic Model for a Snowmelt Dominated Basin. *J Amer. Water Res. Assoc.*, **42**, 877–890.

Hossain, F., and E.N. Anagnostou. (2005). Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture. *Advances in Water Res.*, **28 (12)**, 1336-1350.

Kalaney, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, and R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, and J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.* **77**, 437–471.

Kampf, S. K., and S. J. Burges 2007: A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resour. Res.*, **43**, W05423.

Liang X, E. F. Wood, D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: evaluation and modifications. *Glob. Planet Change*, **13**, 195–206.

Littell, J. S., M. M. Elsner, G. S. Mauger, E. R. Lutz, A. F. Hamlet, and E. P. Salathé. 2011: *Regional Climate and Hydrologic Change in the Northern U.S. Rockies and Pacific Northwest: Internally Consistent Projections of Future Climate for Resource Management.* Preliminary project report, USFS JVA 09-JV-11015600-039. Prepared by the Climate Impacts Group, University of Washington, Seattle.

Livneh B, E.A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. Andreadis, E.P. Maurer, and D.P. Lettenmaier, 2013: A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States: Updates and extensions. *J. Climate*, (in review).

Lohmann, D., E. Raschke, B. Nijssen, and D.P. Lettenmaier, 1998: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol. Sci. J.*, **43**, 131-141.

Maggioni, V., E. N. Anagnostou, and R. H. Reichle.(2012) The impact of model and rainfall forcing errors on characterizing soil moisture uncertainty in land surface modeling. *Hydrol. Earth Sys. Sci.* **16 (10)**, 3499-3515.

Materia, S., P. A. Dirmeyer, Z. Guo, A. Alessandri, A. Navarra, 2010: The Sensitivity of Simulated River Discharge to Land Surface Representation and Meteorological Forcings. *J. Hydrometeor.*, **11**, 334–351.

Maurer, E.P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237-3251.

Maurer, E. P., L. Brekke, T. Pruitt, and P. B. Duffy, 2007: Fine-resolution climate projections enhance regional climate change impact studies. *Eos Trans. AGU*, **88(47)**, 504-504.

McKelvey, K.S., J.P. Copeland, M.K. Schwartz, J.S. Littell, K.B. Aubry, J.R. Squires, S.A. Parks, M.M. Elsner, and G.S. Mauger, 2011: Climate change predicted to shift wolverine distributions, connectivity, and dispersal corridors. *Ecol. Appl.*, **21(8)**, 2882-2897.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383-1394.

Merz, R., J. Parajka, and G. Blo¨schl, 2011: Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resour. Res.*, **47**, W02531.

Mesinger, F, G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360.

Mitchell, K.E., D. Lohmann, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, B.A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R.W. Higgins, R.T. Pinker, J.D. Tarpley, D.P. Lettenmaier, C.H. Marshall, J.K. Entin, M. Pan, W. Shi, V. Koren, J. Meng, B. H. Ramsay, and A.A. Bailey, 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.

Mizukami, N., M. P., Clark, A. Slater, L. Brekke, M. M. Elsner, J. Arnold, S. Gangopadhyay, 2013: Evaluation of different methods to construct large-scale hydrologic model forcing data in mountainous regions. *J. Hydrometeorol.*, **15,** 474–488.

Mo, K. C., L.-C. Chen, S. Shukla, T. J. Bohn, and D. P. Lettenmaier, 2012: Uncertainties in North American Land Data Assimilation Systems over the Contiguous United States. *J. Hydrometeorol.*, **13**, 996-1009.

Nasonova, O. N., Y. M. Gusev, and Y. E. Kovalev, 2011: Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrol. Proc.*, **25**, 1074-1090.

Reclamation, 2011: *West-Wide Climate Risk Assessments: Bias-Corrected and Spatially Downscaled Surface Water Projections, Technical Memorandum No. 86-68210-2011-01*, prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 138pp.

Salathé, E.P., P.W. Mote, and M.W. Wiley, 2007: Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States Pacific Northwest. *Int. J. of Climatol.*, **27(12)**, 1611-1621.

Shepard, D.S.,1984: Computer mapping: the SYMAP interpolation algorithm. In: Willmott GL, Reidel CJ (eds) Spatial statistics and Models Gaille, pp 133–145.

Thornton, P. E., and S. W. Running, 1999: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation, *Agric. For. Meteorol.*, **93**, 211 –228.

Vano, J. A., T. Das, and D. P. Lettenmaier, 2012: Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature, *J. Hydrometeorol.*, **13(3)**, 932-949.

Vaze, J., D. A. Post, F. H. S. Chiew, J.-M. Perraud, N. R. Viney, and J. Teng, 2010: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, *J. Hydrol.*, **394(3–4)**, 447-457.

Wenger, S. J., D. J. Isaak, C. H. Luce, H. M. Neville, K. D. Fausch, J. B. Dunham, D.C. Dauwalter, M. K. Young, M. M. Elsner, B. E. Rieman, A. F. Hamlet, and J. E. Williams, 2011: Flow regime, temperature and biotic interactions drive differential declines of trout species under climate change. *Proc. Natl. Acad. Sci. U. S. A.*, **108(34)**, 14175-14180.

Wilby, R. L. 2005: Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrol. Process.*, **19**, 3201–3219.

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, *Climatic Change*, **62(1-3)**, 189-216.

Wood, A. W. and D.P. Lettenmaier, 2006: A testbed for new seasonal hydrologic forecasting approaches in the western U.S. *Bull. Amer. Meteor. Soc.*, **87(12)**, 1699-1712.

Wood, A. W. 2008: The University of Washington Surface Water Monitor: An experimental platform for national hydrologic prediction, *Proc. Amer. Meteor. Soc. Annual Meeting, New Orleans*, 13 p.

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, D. Mocko, 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.

Xue, Y., P. J. Sellers, J. L. Kinter, and J. Shukla, 1991: A simplified biospheremodel for global climate studies. *J. Climate*, **4**, 345–364.

Yapo, P.O., Hoshin Vijai Gupta, Soroosh Sorooshian (1998). Multi-objective global optimization for hydrologic models. *J. Hydrol.*, **204(1–4)**, 83–97.

Tables

TABLE 1. Summary of differences in development of spatially distributed meteorological datasets. Notes: HCN is Historical Climatology Network; AHCCD is Adjusted Historical Canadian Climate Database; Prcp is precipitation; Tmax is maximum temperature; Tmin is minimum temperature; CONUS is continental United States; PRISM is Parameter-elevation Regressions on Independent Slopes Model.

Name	References	Spatial Extent	Native Spatial Resolution	Temporal Coverage	Distinguishing Characteristics
Maurer (M)	Maurer et al. 2002	CONUS plus Canadian portions of Columbia and Missouri basins	1/8 degree	1949-2000	Gridded Co-op station data (w/ more than 20 years data); Prcp scaled to PRISM climatology (1961-1990); Temp lapsed to grid cell elevation (-6.5degrees C per km);
Wood-Lettenmaier (WL)	Wood and Lettenmaier 2006; Wood (2008)	Major Western US watersheds, including Canadian portions	1/8 degree	1915-2005	Gridded Co-op station data (w/ more than 45 years data and 80% coverage); Index Station Method applied to data post 2004; Prcp scaled to PRISM climatology (1961-1990); Temp lapsed to grid cell elevation (-6.5degrees C per km);
Abatzolou (A)	Abatzoglou 2011	CONUS	4-km	1979-2010	NLDAS Phase 2 – Prcp, Tmin, Tmax interpolated & scaled to PRISM monthly timeseries
Elsner-Littell (EL)	Elsner et al. 2010; Littell et al. 2011	Major Western US watersheds, including Canadian portions	1/16 degree	1915-2006	Gridded Co-op station data (w/ more than 5 years data); HCN and AHCCD station data used to correct temporal inhomogeneities; Temp lapsed to grid cell elevation (-6.5degrees C per km); Prcp & Tavg scaled to PRISM climatology (1971-2000).

TABLE 2. Summary of VIC model parameters considered for calibration. Parameters were evaluated using perturbation experiments and those chosen for calibration are noted by “X”.

Considered Model Calibration Parameters	Parameter Units	Description	Parameter Range	Sensitive
bi	NA	Variable infiltration curve parameter	0 - 0.4	
Ds	fraction	Fraction of Dsmax where nonlinear baseflow occurs	0.00001 - 1	X
Dsmax	mm/day	Maximum velocity of baseflow	0.1 - 30	X
Ws	fraction	Fraction of max. soil moisture were nonlinear baseflow occurs	0.05 - 1	X
D2	mm	Middle soil depth	0.1 - 1.0	X
D3	mm	Lowest soil depth	0.5 - 2.5	X
wind_atten	fraction	Defines windspeed profile through canopy	0 - 1	
snow_rough	m	Surface roughness of snowpack	0 - 1	
rad_atten	fraction	Defines shortwave radiation through canopy	0.1 - 0.6	
Velocity	m/s	streamflow routing velocity	0.5 - 2.5	

TABLE 3. Summary of case study watersheds.

Name (ID)	Description	Size, sqkm (No. VIC cells)	Calibration Period (water years)	Validation Period (water years)	Mean Annual P (mm)	Mean Annual T (deg C)	Mean Annual Flow (cms)
ANIMS (1)	Animas River at Durango, CO (USGS ID 09361500)	1792 (21)	1993-1999	1986-1992	900 - 978	0.7 - 2.3	24
DOLOR (2)	Dolores River near Cisco, UT (USGS ID 09180000)	11,862 (103)	1990-1999	1980-1989	552 - 591	6.0 - 6.9	38
GREEN (3)	Green River at Green River, UT (USGS ID 09315000)	116,162 (816)	1990-1999	1980-1989	423 - 450	4.0 - 5.1	226
MISSO (4)	Missouri River at Toston, MT (USGS ID 06054500)	39,993 (346)	1985-1989	1980-1984	589 - 644	2.5 - 3.4	189
SACRB (5)	Sacramento River at Bend Bridge near Red Bluff, CA (USGS ID 11377200)	23,051 (230)	1990-1999	1980-1989	888 - 958	8.6 - 9.8	351
SALTC (6)	Salt River near Chrysotile, AZ (USGS ID 09497500)	7,379 (72)	1990-1999	1980-1989	603 - 643	9.7 - 9.8	23
SNAKE (7)	Snake River near Heise, ID (USGS ID 13037500)	14,898 (144)	1990-1999	1980-1989	825 - 897	0.4 - 2.1	208

TABLE 4. Summary of optimal VIC model calibration parameters according to meteorological dataset.

	Name	Abatzoglou	Elsner-Littell	Maurer	Wood-Lettenmaier
Ds (fraction)	ANIMS	0.00378	0.08373	0.99968	0.04988
	DOLOR	0.00283	0.01581	0.15324	0.00072
	GREEN	0.00961	0.01700	0.04588	0.02679
	MISSO	0.00922	0.01157	0.04193	0.02575
	SACRB	0.36768	0.32754	0.35505	0.39765
	SALTC	0.00295	0.00010	0.05496	0.07888
	SNAKE	0.02216	0.04806	0.50982	0.05335
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
Ws (fraction)	ANIMS	0.16129	0.36299	0.71606	0.12935
	DOLOR	0.53094	0.43526	0.51216	0.41465
	GREEN	0.51271	0.63571	0.78918	0.64319
	MISSO	0.15003	0.21649	0.38934	0.36493
	SACRB	0.93204	0.99933	0.99326	0.14904
	SALTC	0.60128	0.44716	0.50946	0.71447
	SNAKE	0.14693	0.25227	0.79960	0.48482
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
Dsmax (mm/d)	ANIMS	4.896	7.014	6.061	29.738
	DOLOR	14.710	5.539	3.511	27.741
	GREEN	4.367	4.624	1.658	2.264
	MISSO	25.537	24.344	6.713	7.378
	SACRB	3.073	3.244	2.630	0.603
	SALTC	17.930	0.878	29.179	1.240
	SNAKE	29.982	29.831	2.940	29.546
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
D2 (mm)	ANIMS	0.9955	0.9944	0.8166	0.9727
	DOLOR	0.9899	0.9998	0.9165	0.4979
	GREEN	0.9744	0.7994	0.8084	0.9682
	MISSO	0.9835	0.9988	0.9965	0.9963
	SACRB	0.3642	0.9633	0.2295	0.6505
	SALTC	0.5753	0.1748	0.9997	0.2929
	SNAKE	0.3403	0.4892	0.1002	0.2622
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
D3 (mm)	ANIMS	0.7107	1.1179	0.6296	1.4930
	DOLOR	1.3021	0.9797	0.5711	2.4071
	GREEN	1.1892	1.7170	2.1575	1.1475
	MISSO	2.4738	2.4913	1.7045	2.0047
	SACRB	1.5519	2.0709	1.7005	0.9533
	SALTC	0.7808	0.9315	0.5011	0.6348
	SNAKE	1.5616	1.2289	1.3075	1.1267

TABLE 5. Summary of VIC model calibration (validation) statistics according to calibration parameter, watershed, and meteorological dataset.

	Name	Abatzoglou	Elsner-Littell	Maurer	Wood-Lettenmaier
NSE_{mon}	ANIMS	0.87 (0.78)	0.70 (0.75)	0.82 (0.73)	0.87 (0.80)
	DOLOR	0.78 (0.74)	0.76 (0.70)	0.76 (0.75)	0.78 (0.79)
	GREEN	0.95 (0.93)	0.89 (0.88)	0.94 (0.92)	0.93 (0.91)
	MISSO	0.74 (0.87)	0.69 (0.84)	0.80 (0.91)	0.80 (0.91)
	SACRB	0.95 (0.94)	0.92 (0.86)	0.92 (0.91)	0.94 (0.93)
	SALTC	0.85 (0.56)	0.84 (0.71)	0.77 (0.65)	0.83 (0.65)
	SNAKE	0.98 (0.91)	0.93 (0.86)	0.93 (0.87)	0.96 (0.95)
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
$RMSE_{mon}$	ANIMS	0.41 (0.48)	0.61 (0.51)	0.47 (0.56)	0.40 (0.46)
	DOLOR	0.62 (0.68)	0.65 (0.72)	0.65 (0.66)	0.63 (0.60)
	GREEN	0.25 (0.29)	0.35 (0.37)	0.27 (0.30)	0.28 (0.31)
	MISSO	0.33 (0.27)	0.36 (0.30)	0.29 (0.23)	0.29 (0.22)
	SACRB	0.24 (0.24)	0.29 (0.37)	0.29 (0.29)	0.25 (0.26)
	SALTC	0.58 (0.77)	0.60 (0.63)	0.72 (0.69)	0.62 (0.69)
	SNAKE	0.16 (0.28)	0.27 (0.34)	0.28 (0.33)	0.20 (0.20)
	Name	**Abatzoglou**	**Elsner-Littell**	**Maurer**	**Wood-Lettenmaier**
$VolErr_{mon}$	ANIMS	0.00 (0.01)	0.00 (0.02)	0.00 (0.26)	0.00 (0.01)
	DOLOR	0.21 (0.22)	0.00 (0.11)	0.21 (0.35)	0.03 (0.18)
	GREEN	0.00 (0.01)	0.01 (0.07)	0.00 (0.00)	0.00 (0.09)
	MISSO	0.00 (0.05)	0.00 (0.12)	0.00 (0.05)	0.00 (0.00)
	SACRB	0.00 (0.00)	0.00 (0.10)	0.00 (0.10)	0.05 (0.07)
	SALTC	0.01 (0.06)	0.00 (0.20)	0.00 (0.17)	0.01 (0.12)
	SNAKE	0.00 (0.02)	0.00 (0.03)	0.00 (0.13)	0.01 (0.04)

List of Figures

FIG 1. Overview map of study domain (2-digit HUC scale) and case study watersheds. Case study watersheds include: 1) Animas River at Durango, CO (USGS ID 09361500); 2) Dolores River near Cisco, UT (USGS ID 09180000); 3) Green River at Green River, UT (USGS ID 09315000); 4) Missouri River at Toston, MT (USGS ID 06054500); 5) Sacramento River at Bend Bridge near Red Bluff, CA (USGS ID 11377200); 6) Salt River near Chrysotile, AZ (USGS ID 09497500); and, 7) Snake River near Heise, ID (USGS ID 13037500). The purple dashed line indicates the common domain used for meteorological dataset comparison.

FIG 2. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three precipitation (Prcp) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles. Light dashed lines represent change of +/-10 percent.

FIG 3a-b. Spatial comparison of percent difference in monthly mean precipitation (Prcp) - January, top [A]; July, bottom [B]- comparing Wood-Lettenmaier, Elsner-Littell, and Abatzoglou datasets with respect to the Maurer dataset. Positive difference indicates higher monthly precipitation, while negative median difference indicates lower monthly precipitation.

FIG 4. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three maximum temperature (Tmax) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al.

(2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

FIG 5. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three minimum temperature (Tmin) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

FIG 6. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three diurnal temperature range (Tran) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

FIG 7a-b. Spatial comparison of difference (in degrees C) in monthly mean temperature (maximum [Tmax], minimum [Tmin], and diurnal range [Tran]) – January, top [A]; July, bottom [B] – comparing Wood-Lettenmaier, Elsner-Littell, and Abatzoglou datasets with respect to the Maurer dataset. Positive difference indicates high monthly temperature, while negative difference indicates lower monthly temperature.

FIG 8. Summary of differences in mean annual precipitation and temperature (Tavg, Tmax, and Tmin) between Abatzoglou, Elsner-Littell, Wood-Lettenmaier and the reference Maurer dataset.

Differences are shown over the seven case study watersheds and over 3 simulation periods: full simulation – 1980-1999 water years, calibration period, and validation period.

FIG 9. Summary of simulated flows based on calibrated models for seven case study watersheds (to each of the four meteorological forcing datasets) forced with alternate forcing datasets. In each figure panel, EL, A, M, and WL in the legend title (i.e. top row of legend above the line) indicate the base meteorological dataset used for model calibration. The black line represents mean monthly reconstructed natural streamflow at the watershed outlet. The red line represents resulting mean monthly streamflow from "base" calibrated simulations, having corresponding dataset and calibration parameters. The colored dashed lines represent mean monthly streamflow from simulations using calibrated parameters from the base simulation along with alternate meteorological datasets.

FIG 10. Change in mean annual precipitation (Prcp) vs. change in mean annual runoff (RO), computed between the calibration period and selected warm-dry years (circles) and cool-wet years (diamonds) in the validation period. Size of shapes represents the relative magnitude (absolute value) of corresponding change in mean annual temperature.

Figures

FIG 1. Overview map of study domain (2-digit HUC scale) and case study watersheds. Case study watersheds include: 1) Animas River at Durango, CO (USGS ID 09361500); 2) Dolores River near Cisco, UT (USGS ID 09180000); 3) Green River at Green River, UT (USGS ID 09315000); 4) Missouri River at Toston, MT (USGS ID 06054500); 5) Sacramento River at Bend Bridge near Red Bluff, CA (USGS ID 11377200); 6) Salt River near Chrysotile, AZ (USGS ID 09497500); and, 7) Snake River near Heise, ID (USGS ID 13037500). The purple dashed line indicates the common domain used for meteorological dataset comparison.

FIG 2. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three precipitation (Prcp) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles. Light dashed lines represent change of +/-10 percent.

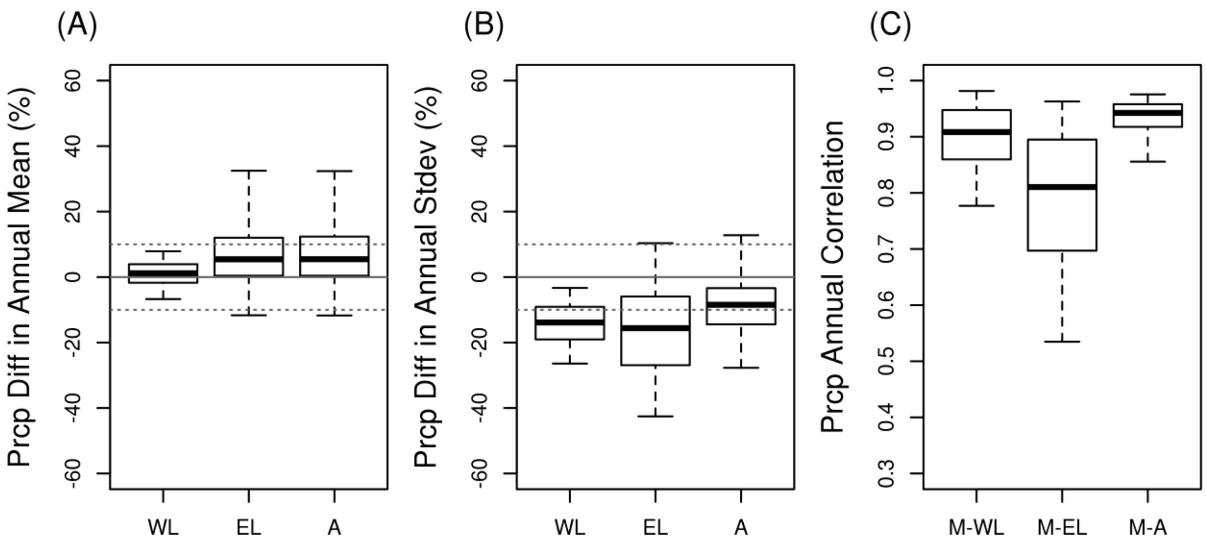


FIG 3a-b. Spatial comparison of percent difference in monthly mean precipitation (Prcp) - January, top [A]; July, bottom [B]- comparing Wood-Lettenmaier, Elsner-Littell, and Abatzoglou datasets with respect to the Maurer dataset. Positive difference indicates higher monthly precipitation, while negative median difference indicates lower monthly precipitation.

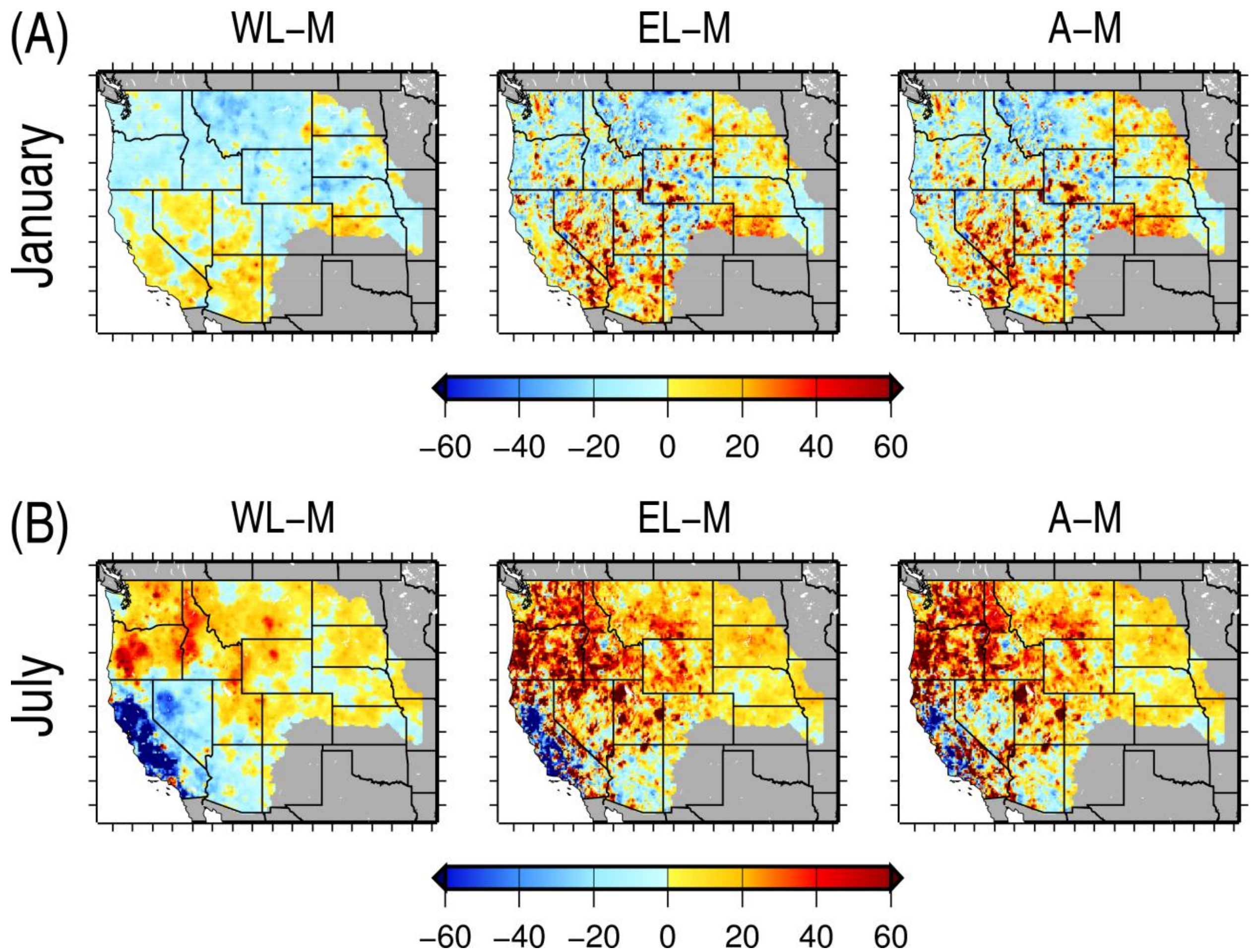


FIG 4. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three maximum temperature (Tmax) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

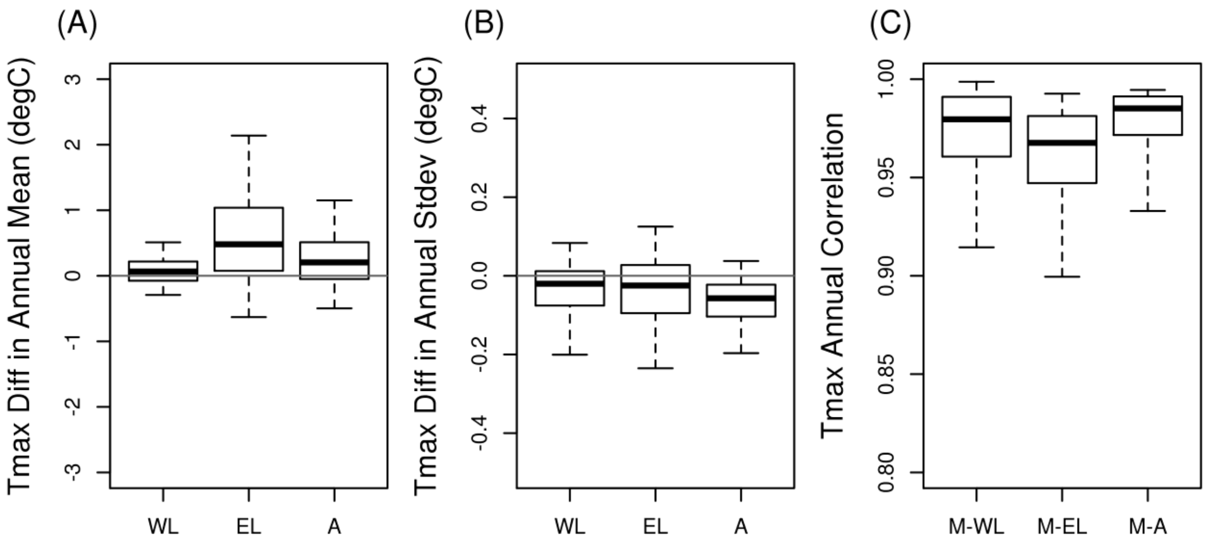


FIG 5. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three minimum temperature (Tmin) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

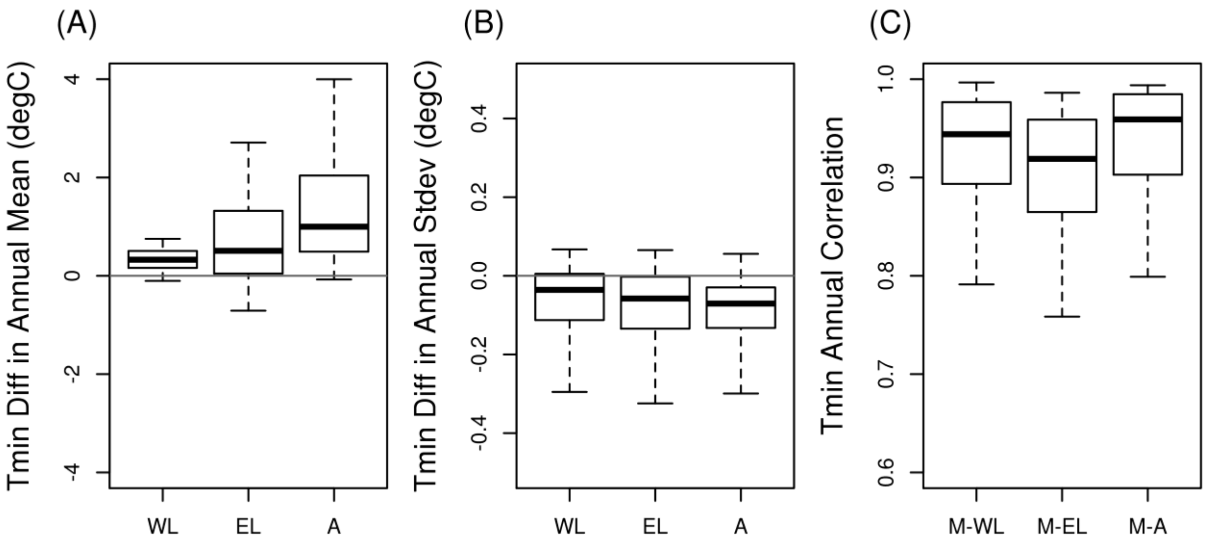


FIG 6. Percent differences of (A) annual means, (B) standard deviations, and (C) correlation coefficients between each of the three diurnal temperature range (Tran) datasets (A = Abatzoglou; EL=Elsner-¬-Littell; WL=Wood-¬-Lettenmaier) and the reference dataset, i.e., Maurer et al. (2002). The boxes represent the 25th, 50th, 75th percentiles, while the whiskers represent the 5th and 95th percentiles.

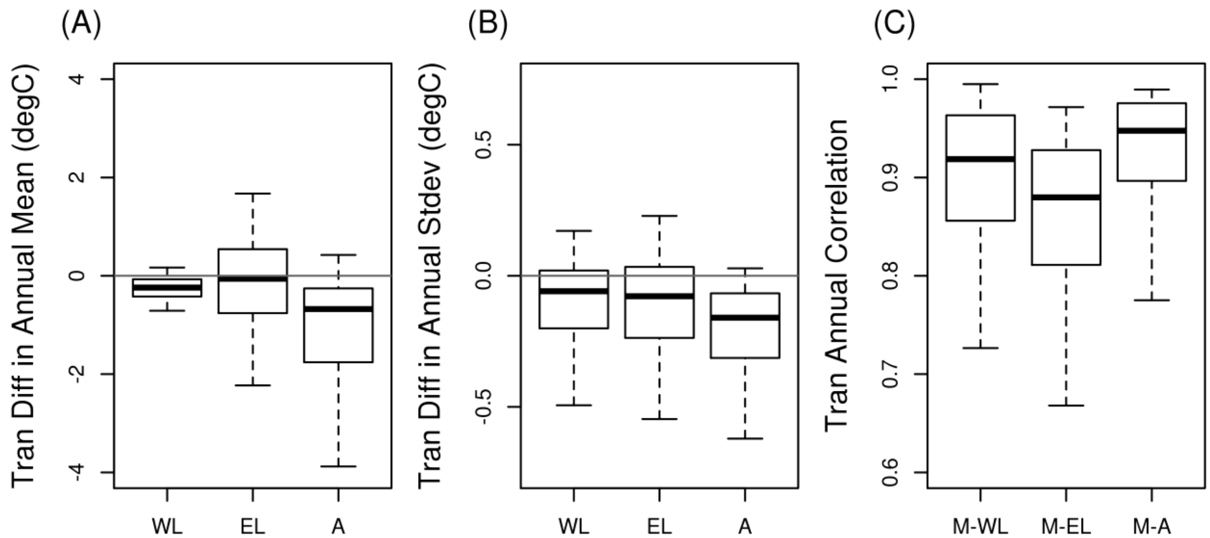


FIG 7a-b. Spatial comparison of difference (in degrees C) in monthly mean temperature (maximum [Tmax], minimum [Tmin], and diurnal range [Tran]) – January, top [A]; July, bottom [B] – comparing Wood-Lettenmaier, Elsner-Littell, and Abatzoglou datasets with respect to the Maurer dataset. Positive difference indicates high monthly temperature, while negative difference indicates lower monthly temperature.

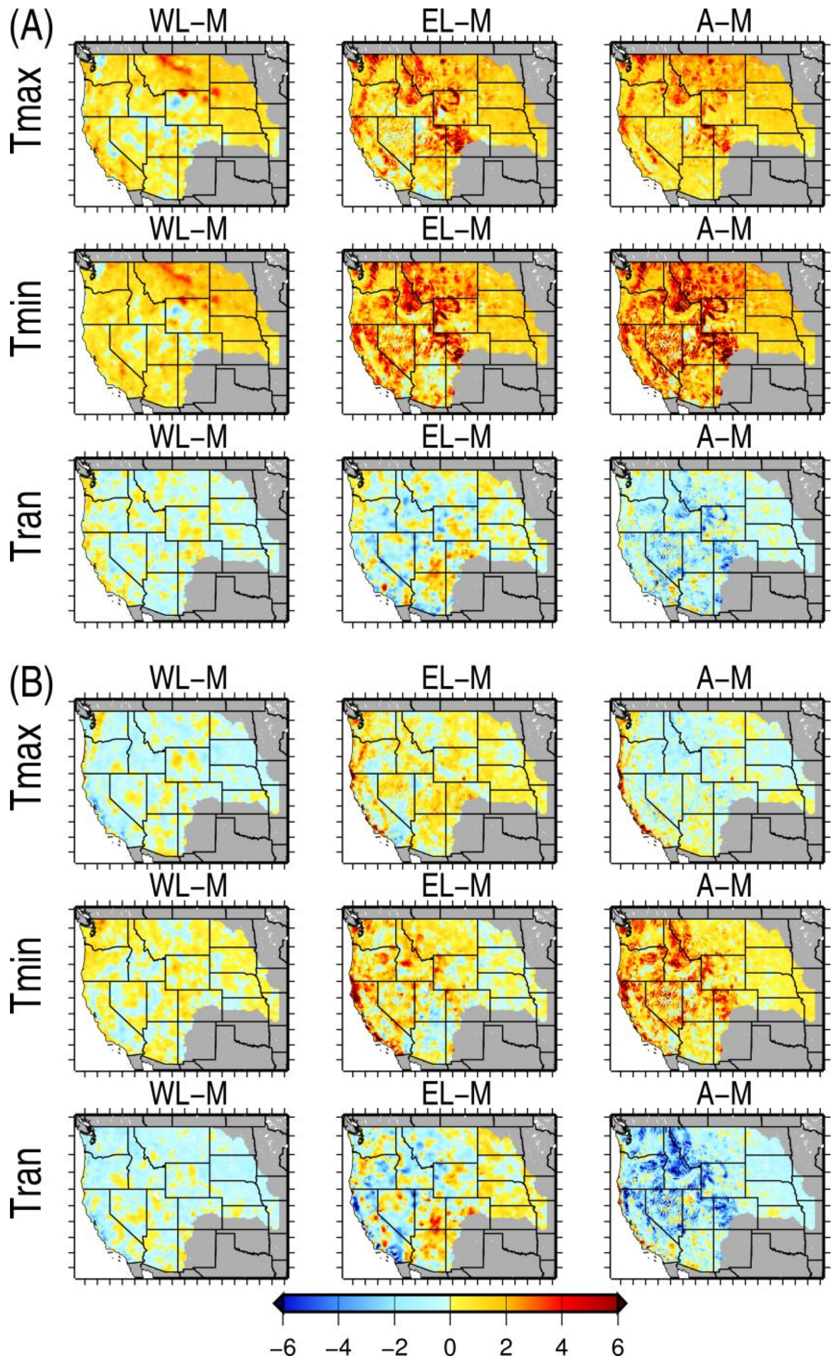


FIG 8. Summary of differences in mean annual precipitation and temperature (Tavg, Tmax, and Tmin) between Abatzoglou, Elsner-Littell, Wood-Lettenmaier and the reference Maurer dataset. Differences are shown over the seven case study watersheds and over 3 simulation periods: full simulation – 1980-1999 water years, calibration period, and validation period.

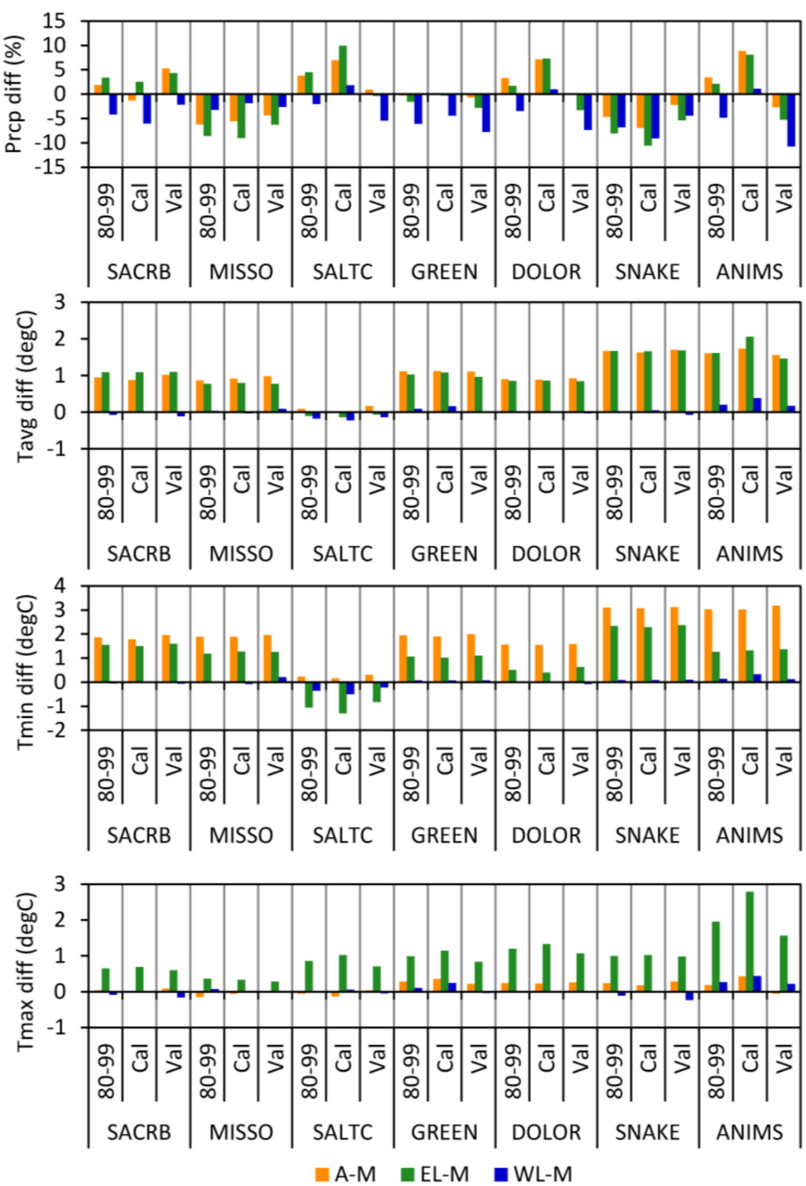
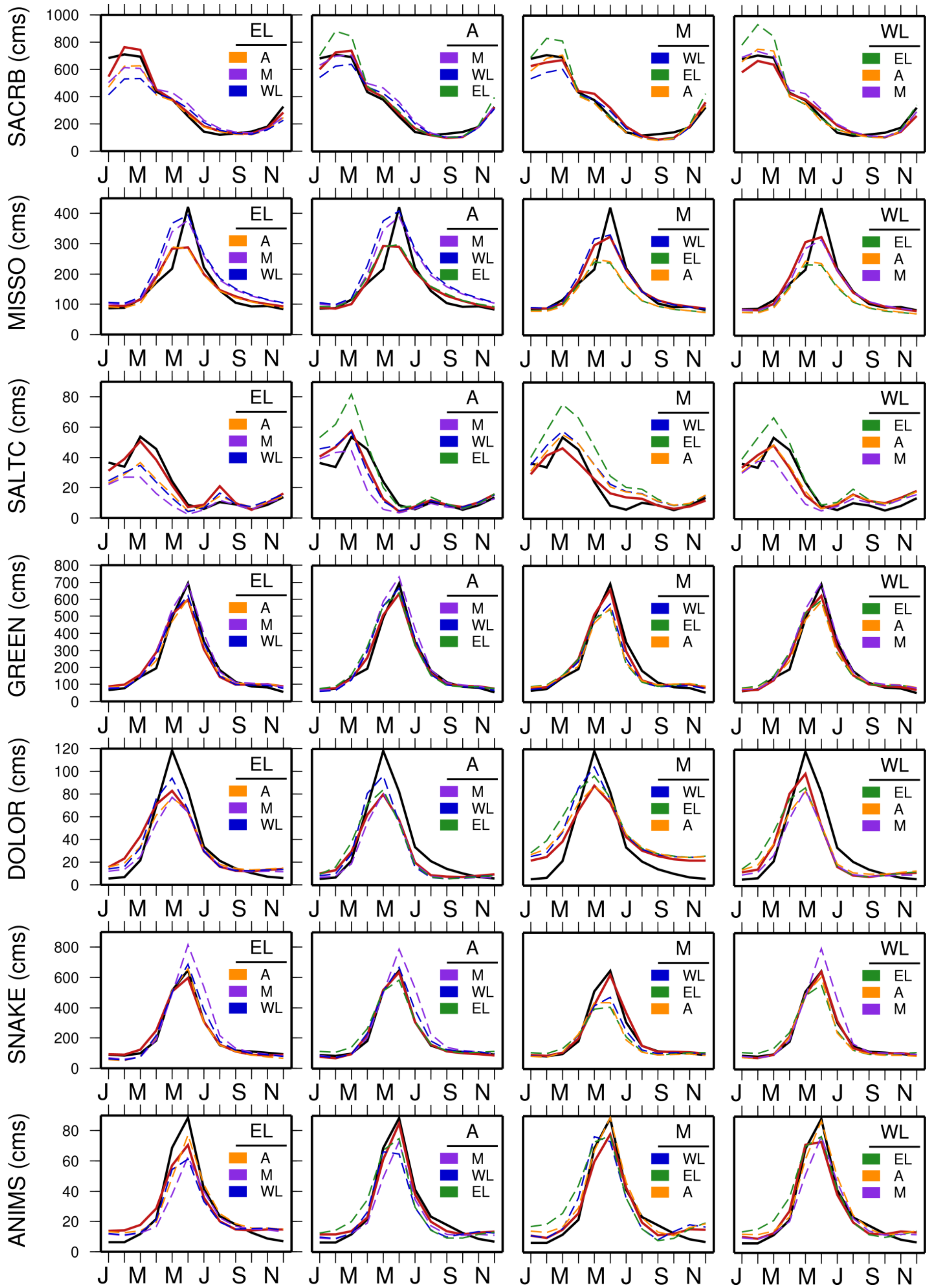


FIG 9. Summary of simulated flows based on calibrated models for seven case study watersheds (to each of the four meteorological forcing datasets) forced with alternate forcing datasets. In each figure panel, EL, A, M, and WL in the legend title (i.e. top row of legend above the line) indicate the base meteorological dataset used for model calibration. The black line represents mean monthly reconstructed natural streamflow at the watershed outlet. The red line represents resulting mean monthly streamflow from "base" calibrated simulations, having corresponding dataset and calibration parameters. The colored dashed lines represent mean monthly streamflow from simulations using calibrated parameters from the base simulation along with alternate meteorological datasets.



SACRB (cms)
MISSO (cms)
SALTC (cms)
GREEN (cms)
DOLOR (cms)
SNAKE (cms)
ANIMS (cms)
EL
A
M
WL
J M M J S N

FIG 10. Change in mean annual precipitation (Prcp) vs. change in mean annual runoff (RO), computed between the calibration period and selected warm-dry years (circles) and cool-wet years (diamonds) in the validation period. Size of shapes represents the relative magnitude (absolute value) of corresponding change in mean annual temperature.