Title: Statistically Downscaled Precipitation Sensitivity to Gridded Observation Data and Downscaling Technique

Short Title: Downscaled Precipitation Sensitivity to Observations and Downscaling

Authors:

Adrienne M. Wootten[1]*, Keith W. Dixon[2], Dennis Adams-Smith[2,3], Renee A. McPherson[1]

[1]South Central Climate Adaptation Science Center, University of Oklahoma

[2]National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory

[3]University Corporation for Atmospheric Research Cooperative Program for the Advancement of Earth System Science

***Corresponding Author**

Address: 201 Stephenson Parkway, Suite 2100, Norman, OK, 73019

Email: amwootte@ou.edu

Phone: 405-325-0685

Fax: 405-325-1122

Research Sponsors: United States Geological Survey (Award Number G15PG00088), the Oklahoma EPSCoR project through the Oklahoma State Regents for Higher Education and the National Science Foundation under NSF Grant OIA-1301789.

Keywords: downscaling, climate projections, impact assessments, climate model evaluation

Statistically Downscaled Precipitation Sensitivity to Gridded Observation Data and Downscaling Technique

A.M. Wootten, K. W. Dixon, D. Adams-Smith, R. A. McPherson

Abstract

Future climate projections illuminate our understanding of the climate system and generate data products often used in climate impact assessments. Statistical downscaling (SD) is commonly used to address biases in global climate models (GCM) and to translate large-scale projected changes to the higher spatial resolutions desired for regional and local scale studies. However, downscaled climate projections are sensitive to method configuration and input data source choices made during the downscaling process that can affect a projection's ultimate suitability for particular impact assessments. Quantifying how changes in inputs or parameters affect SD-generated projections of precipitation is critical for improving these datasets and their use by impacts researchers.

Through analysis of a systematically designed set of 18 statistically downscaled future daily precipitation projections for the south-central United States, this study aims to improve the guidance available to impacts researchers. Two statistical processing techniques are examined: a ratio delta downscaling technique and an equi-ratio quantile mapping method. The projections are generated using as input results from three GCMs forced with representative concentration pathway (RCP) 8.5 and three gridded

observation-based data products. Sensitivity analyses identify differences in the values of precipitation variables among the projections and the underlying reasons for the differences.

Results indicate that differences in how observational station data are converted to gridded daily observational products can markedly affect statistically downscaled future projections of wet-day frequency, intensity of precipitation extremes, and the length of multi-day wet and dry periods. The choice of downscaling technique also can affect the climate change signal for variables of interest, in some cases causing change signals to reverse sign. Hence, this study provides illustrations and explanations for some downscaled precipitation projection differences that users may encounter, as well as evidence of symptoms that can affect user decisions.

1. Introduction

Global climate modeling is an important tool for researching the large-scale climate system (Weart, 2010) and developing climate projections that help guide decision making related to anthropogenic climate change (Smith and Stern, 2011). Regional and local-scale climates often are not well represented in global climate models (GCMs; Rummukainen, 2010). Downscaling techniques can be used to translate a GCM-simulated climate response to smaller spatial scales, reducing GCM biases and providing added information scaled for decision-makers (Rummukainen, 2016; Tabari et al., 2016). Because statistical downscaling and bias correction methods (hereafter SD methods) are relatively flexible and computationally efficient, SD-generated climate projections are frequently used as input for multiple types of impacts models (e.g. Gergel et al., 2017; Basso et al., 2015) as part of impact assessment studies and adaptation planning.

The generation of SD-processed future climate projections involve three main ingredients: 1) observation-based data for climate variables of interest from a historical period, 2) GCM output for both historical and future time periods, and 3) a statistical technique that uses 1) and 2) as inputs to determine statistical relationships used to produce refined future projections as output. This statistical processing aims to add value to the raw GCM output by using information gleaned from observations, rendering output more suitable for direct use in research applications. SD assumes that large-scale climate conditions can be linked to local-scale effects and that model biases during the historical period are applicable to future climate conditions (Wilby and Wigley, 1997).

The climate services community has encouraged stakeholders to use ensembles of emissions scenarios, GCMs, and downscaling techniques (e.g. Markus et al., 2018) to capture a range of possible future climates, with differences among the projections (i.e., the spread) resulting from multiple sources of uncertainty (Wootten et al., 2017). Unfortunately, simply capturing the uncertainty range does not provide adequate insight to researchers or adaptation experts regarding the sources of this uncertainty or, more practically, how choices made during the process of downscaling may influence the resulting climate variable. This study examines the sensitivity of precipitation variables projections to the choice of SD method and the gridded observations used to train the methods. Herein, the term “sensitivity” refers to the change in precipitation variables resulting from choices made in the downscaling process.

Numerous studies have examined the sensitivity of dynamical climate model-simulated climate changes to model formulation and factors affecting radiative forcings (e.g., the Coupled Model Intercomparison Project, CMIP [Taylor et al., 2012]; the Coordinated Regional Downscaling Experiment [Whitehall et al., 2012]). Fewer studies have examined the sensitivity of statistically refined climate change signals to downscaling technique choices (e.g. Action ES1102 (VALUE) of the European Cooperation in Science and Technology [Maraun et al., 2014]). A limited set of studies have examined the influence of different training datasets and SD techniques on projections of different precipitation metrics or their regional effects on ecosystem or hydrological modeling applications (e.g., Pourmoktharian et al., 2016, Werner and Cannon, 2016), the accuracy of gridded observations and GCMs (Behnke et al., 2016, Sillmann et al., 2013), and the effect of gridded observations on hydrologic model calibration (Elsner et al., 2014). Our study examines the influence of different gridded observations and SD techniques on projections of precipitation metrics important to climate impacts researchers in the south-central United States.

Regional and local scale precipitation remains challenging to simulate well in large scale dynamical climate models (e.g. Cesana et al., 2017; Tapiador, 2019). Accordingly, impacts researchers benefit from using SD-generated precipitation projections as input to their work, as long as the sensitivities of precipitation-related variables connected to their needs are better documented. For example, (1) What aspects of the training data can drive the sensitivity of downscaled precipitation projections? (2) Which precipitation-related variables are more sensitive to SD choices than others? (3) What training data or SD techniques are better suited to refine climate projection data for use in particular types of impacts models and their applications?

We compare high-resolution climate projections for 2070-2099 created with two SD techniques trained separately with three gridded observation-based datasets and forced by representative concentration pathway (RCP) 8.5 (van Vuuren et al., 2011; Riahi et al., 2007) and three CMIP5 GCMs (Taylor et al., 2012). Our analysis focuses on the differences in the change signal present in this set of 18 SD-generated precipitation projections and those in the three GCMs' pre-downscaled data. We then assess the influence of different observation data and downscaling techniques on the resulting projections of several precipitation metrics of interest to impacts researchers and decision makers. Though not comprehensive in scope (e.g., not all SD methods nor all observational products are tested), these results illustrate often under-appreciated aspects of SD-generated precipitation data products. Similarly, though this study focuses on analyses of precipitation metrics that are of interest to climate impacts researchers in the topographically complex south-central United States (i.e., region of our funding source), our results highlight sensitivities present in SD-generated projections covering other regions.

2. Study Region, Data, and Methods

2.1. Study Region

The study region was the south-central United States, from about 26°N 108.5°W to 40°N 91°W, excluding Mexico and the Gulf of Mexico. The region has varied topography with the Mississippi River Valley and the Ozark Mountains in the east (elevations of 200-800 m), the Rocky Mountains in the west (1,500-4,400 m), and the coast of the Gulf of Mexico in the southeast (near sea level). This region has a sharp gradient of mean annual precipitation from east to west, with some southeastern locations receiving eight times more precipitation than the driest western locations (Figure 1). Strong gradients of both mean annual number of days with rain and annual daily maximum rainfall also exist. Snow dominates precipitation across the mountainous terrain during winter; tropical systems can provide substantial portions of any given year's precipitation along the Gulf coast. The strong climatological differences in precipitation type and amounts provide an opportunity to assess the sensitivity of SD-generated precipitation projections to the choice of downscaling techniques and training data.

2.2. Observation-based data

Though some downscaling efforts used weather stations for training or calibration purposes, many publicly available downscaled projections are created using gridded observation-based data (hereafter gridded observations) for training. Gridded observations are processed data products largely based on station data (but can include remotely sensed observations) that are adjusted and interpolated to a grid in a manner that attempts to account for missing station data, as well as temporal and spatial incoherence. The absence of a standard practice to create gridded observations results in the production of numerous gridded observations from multiple sources. Three gridded observations over the 48 conterminous United States (CONUS) serve as training data for downscaling in this study (Table 1): Daymet Version 2.1 (hereafter Daymet, Thornton et al., 2017), Livneh Version 1.2 (hereafter Livneh, Livneh et al., 2013), and PRISM Version AN81d (hereafter PRISM, Daly et al., 2008; Daly et al., 2013). These datasets reflect different methods for creating gridded observations (Table 1) and are used in numerous prior downscaling efforts, including the Multivariate Adaptive Constructed Analogs (MACA, Abatzoglou et al., 2012) and the Localized Constructed Analogs (LOCA, Pierce et al., 2014). In particular, these gridded observations were created using different methods to adjust the daily precipitation time series from station data to ensure temporal and spatial coherence. Differing methods for adjusting the daily precipitation time series have known impacts on precipitation variables, including the frequency of occurrence and intensity (Oyler and Nicholas, 2018). Therefore, using daily precipitation from these three gridded observations allowed us to assess if such differences affect the output of SD experiments.

2.3. GCM data

We used one ensemble member from three GCMs from CMIP5 (Taylor et al., 2012): the Community Climate System Model version 4 (CCSM4, Gent et al., 2011), the

Model for Interdisciplinary Research on Climate version 5 (MIROC5, Watanabe et al., 2010), and the Max-Planck-Institute Earth System Model running on a low-resolution grid (MPI-ESM-LR, Giorgetta et al., 2013). These GCMs produced reasonable historical climate simulations for global models of this class based on the bias for temperature and precipitation across the south-central U.S. compared to other GCMs (Sheffield et al., 2013). The three GCMs also reflected the range of simulated future change signals in the study region, capturing what is referred to as "model uncertainty" (Hawkins and Sutton, 2011). Here, we focused on GCM simulations driven by RCP 8.5, a scenario with a large radiative forcing increase by 2100 (8.5 W m^{-2}, van Vuuren et al., 2011; Riahi et al., 2007), as it was more likely to produce readily detectable change signals by the end of the 21st century and allowed us to analyze greater possible differences between the projections in the sensitivity analysis.

2.4 Analysis Metrics and Periods

Our analyses focus on years 2070-2099 and variables that reflected means, extremes, and occurrences of precipitation. These variables were a subset of the CLIMDEX indices (Zhang et al., 2011, Bronaugh, 2014) calculated using the "climdex.pcic" package in the R software language (https://CRAN.R-project.org/package=climdex.pcic). The chosen CLIMDEX variables were: the total precipitation (prcptot), the number of days with precipitation (r1mm), the maximum dry spell length (cdd), the maximum wet spell length (cwd), and the maximum 1-day precipitation (rx1day). These variables (calculated per year) were those requested by stakeholders for impacts assessments across the study region.

The historical period used for downscaling was 1981-2005. Thirty-year averaging periods are a standard for climatological normals, as they encompass a range of climate variability; however, the PRISM gridded observations were only available from 1981 onward, so the other gridded observations and GCM historical data were trimmed to match PRISM's 25-year period (Table 1).

3. Methods

To investigate the impact of training data and SD techniques on the resulting climate projections, the publicly available GCM and gridded observations were first interpolated to a common grid with bilinear interpolation applied at a daily timescale using the NCAR Command Language (NCAR et al., 2018) Earth System Modeling Framework regridding functions (NCAR, 2019). These functions interpolated each GCM and observation dataset to a 0.1° latitude by 0.1° longitude rectilinear grid, resulting in >19,000 grid points within a study domain which included the Upper Rio Grande, Red, and Canadian River basins (Figure 1). We used a conservative interpolation for coastal grid points where there was inadequate information in gridded observations for a bilinear interpolation. The resulting interpolated GCMs and gridded observations were the inputs for two SD techniques (detailed below). The interpolation is a key component of the downscaling procedure, but the same interpolation is used to allow the differences between SD techniques (also referred to as bias correction techniques) to be the focus of this study.

Conventions used herein are that *MF* (model future) refers to the GCM precipitation during the future time period, and *MH* (model historical) and *OH* (observed historical) refer to the GCM and observation-based precipitation during the corresponding historical time window. Downscaled output (*DS*) represents refined future precipitation.

3.1. Ratio Delta Method (DeltaSD)

Variants of a simple technique commonly referred to as the delta method (Maraun et al., 2010; Fowler et al., 2007; Räty et al., 2014) combine information from observations and GCMs to produce climate projections. In general, delta methods (including the DeltaSD version used here) operate by first determining a GCM-simulated time-mean change signal over some specified time range. Next, delta methods apply that change signal to the observations to yield an adjusted future projection. Therefore, the downscaled output time series produced using DeltaSD exhibits daily weather sequences that resemble those of the observations used in the statistical processing (Teutschbein and Seibert, 2012).

How the time-mean signal is computed and applied can differ across applications of the delta approach. In this study, we implemented a multiplicative scaling (ratio) approach for precipitation adjustments. Accordingly, for each time period, DeltaSD calculated the ratio (Δ) at each grid point based upon the GCM being used:

$$\Delta = \frac{mean(MF)}{mean(MH)} \quad (1)$$

Ratios were computed for the future period (2070-2099) relative to the 1981-2005 historical period. In our implementation of the DeltaSD method, the seasonal cycle was captured by using 12 time windows of three months each for which the means of *MH* and *MF* are calculated over the multi-decadal period. For training (calibration) purposes, we used three-month time windows to obtain an adequate sample size of wet days in arid regions of the domain. These *MH* and *MF* means were used to calculate the change ratio (Δ) that is applied to the daily observations of the center month of each time window. The appropriate *OH* time series was multiplied by the ratio (Δ) to calculate the final downscaled precipitation values (*DS*):

$$DS = OH * \Delta \quad (2)$$

Following the multiplicative adjustments performed by DeltaSD, any projection values less than the standard U.S. daily trace amount (0.01 inches day^{-1} = 0.254 mm day^{-1}) were set as dry days (i.e., 0.00 inches) in the output.

For example, to generate the adjusted daily April precipitation values for the late 21st century, the ratio (Δ) was calculated using the three-month means (March, April, and May) during 1981-2005 (*MH*) and 2070-2099 (*MF*), then applied to the series of historical observations (*OH*) for all Aprils during the historical period to calculate *DS*. Therefore, if a

GCM's simulation exhibited a 21 percent increase in mean precipitation amounts for March-April-May of 2070-2099 relative to its simulation for March-April-May of 1981-2005, then the April daily observations were multiplied by 1.21 to generate the downscaled future time series. Because the GCM future simulations span 94 years (2006 through 2099, inclusive) and the observational time series cover only 25 years, DeltaSD output was generated by recycling the observational sequence starting with 1 Jan 1981 being the basis for 1 Jan 2006. Accordingly, 2070-2080 DeltaSD outputs are based on scaled versions of the 1995-2005 observational sequence, and 2081-2099 outputs are based on scaled versions of the 1981-1999 observations. Variants of the ratio delta approach downscale daily precipitation to ensure values of zero or greater; an additive delta approach is more common for variables such as daily temperatures.

3.2. Equi-ratio Quantile Mapping (ERQM)

Equi-ratio quantile mapping (ERQM) was designed to retain the relative change signal at all quantiles of the precipitation distribution while performing bias correction (Li et al., 2010; Wang and Chen, 2014; Cannon et al., 2015). This approach was distinct from that of DeltaSD, which only retained the change signal of monthly mean precipitation. Our implementation of ERQM used the QDM (quantile delta mapping) code of Cannon and colleagues from the R-language MBC (multivariate bias correction) package (https://CRAN.R-project.org/package=MBC). However, configuration choices here differed enough from those used by Cannon et al. (2015) to warrant our adoption of a different acronym to avoid confusion. Many of these choices aligned with earlier pilot studies across the Arkansas-Red River basin (Bertrand and McPherson, 2018, 2019), facilitating comparisons. Like other SD methods that employ a bias correction approach, the weather sequences present in ERQM future projections resemble those of the future GCM used as input.

Prior to execution of ERQM, we applied a trace adjustment (similar to Pierce et al. [2015]) to correct the wet-day fraction of the *MH* precipitation data to match that of the *OH* training data. Then we applied a cube-root transformation on daily precipitation values from all GCM inputs to yield a more Gaussian distribution.

First, the ERQM method calculated τ_{MF}, the non-exceedance probability of the *MF* value (x_{MF}) at day *t*:

$$\tau_{MF}(t) = F^{(t)}_{MF}\left[x_{MF}(t)\right] \tag{3}$$

The inverse cumulative distribution functions (CDFs) of the *MH* and *MF* precipitation were used then to define a relative change factor at the quantile corresponding to the GCM future precipitation value at day *t*.

$$\Delta_M(t) = \frac{F^{-1}_{MF}\left[\tau_{MF}(t)\right]}{F^{-1}_{MH}\left[\tau_{MF}(t)\right]} \tag{4}$$

This equation defined a relative change at the quantile of the GCM value. ERQM also used the inverse CDF of *OH* used for training to bias correct the *MF* for day *t* as follows:

$$x^{\wedge}_{O:M,H:F}(t) = F^{-1}_{OH}\left[\tau_{MF}(t)\right], \tag{5}$$

where $x^{\wedge}_{O:M,H:F}$ was the bias-corrected value of the GCM future precipitation. The calculated value from ERQM for a future GCM value at day *t* is

$$x_{DS}(t) = x^{\wedge}_{O:M,H:F}(t)\Delta_{M}(t), \tag{6}$$

where x_{DS} is the *DS* value for day *t*.

We patterned our time-window approach for ERQM calibration training and output-generation steps after that of the aforementioned Arkansas-Red River work, including using the same training period (1981-2005). It differs from that used with the DeltaSD herein and that used for QDM in Cannon et al. (2015). In our ERQM processing, the seasonal cycle was represented by calculating $\Delta_{M}(t)$ for each of four non-overlapping three-month seasonal periods. For example, $\Delta_{M}(t)$ for March-April-May was calculated using the inverse CDFs of March-April-May precipitation from *MH* and *MF*. The $\Delta_{M}(t)$ value for March-April-May was then multiplied by the bias-corrected GCM precipitation in

March-April-May to produce the final result. We applied the trained ERQM to the entire future period (2006-2099), rather than using non-overlapping 30-year periods (as in DeltaSD) or overlapping 30-year future periods (as in Cannon et al. 2015).

3.3. Technique differences

DeltaSD and ERQM are relatively simple techniques that have one common feature – they use the relative change between GCM historical and future simulations to refine future projections in a manner that is informed by historical observations. However, there are two noteworthy differences between these downscaling techniques. First, while both techniques use ratios representing relative change, DeltaSD preserves only the relative change in the monthly mean precipitation, while ERQM preserves the relative change at the quantiles of the precipitation distribution. This methodological difference impacts the variability of daily precipitation that results from ERQM and DeltaSD.

Consider Figure 2, an example of October precipitation using MPI-ESM-LR and PRISM. Using the DeltaSD approach, the mean of the three-month, GCM-simulated precipitation totals is 21 percent greater in the future period than in the historical period. That is, the relative change from the historical period to the future period is 1.21 at all quantiles for the DeltaSD approach. In contrast, ERQM calculates the relative change at each quantile; here, the GCM-simulated climate change-induced relative increases are greater at the right-hand tail of the distribution (e.g., $\Delta = 1.97$ at the 95th percentile). Thus, although ERQM and DeltaSD yield similar values at the 50th percentile (10 mm/day and 8 mm/day, respectively), there is roughly a 60 percent difference in precipitation at the 95th percentile using ERQM (78 mm/day) versus DeltaSD (48 mm/day). While this example is not necessarily representative for all seasons and locations, it demonstrates that ERQM can produce substantially different projected changes for extreme values of precipitation compared to DeltaSD.

Second, the DeltaSD method is designed such that the output is time synchronous with the training data, while ERQM's output is time synchronous with the GCM-simulated future daily weather variations. Thus, the two methods likely will represent the time sequence of daily precipitation and the length of dry and wet spells differently. Using the *OH* time series as the basis for the future time sequence also constrains the output of DeltaSD to reflect a similar frequency of wet days between the historical and future time periods. That is, DeltaSD ignores GCM-simulated dynamical changes that influence weather sequences, while ERQM incorporates those dynamic changes. Some users of SD-generated daily precipitation projections prefer using downscaled data products that have weather sequences based on observations (as produced by DeltaSD and other "change factor" methods), especially if a GCM's historical weather sequence characteristics (e.g., wet or dry spells) differ markedly from observations. Other practitioners may prefer to use bias correction types of SD methods that, like ERQM, produce future projections based on GCM weather sequences, and hence can represent climate change-induced changes in spell lengths not represented by methods such as DeltaSD (Teutschbein and Seibert, 2012).

Though the techniques used here do not encompass all downscaling methods used in climate studies, ERQM and DeltaSD are two techniques that provide insight into sensitivities of the downscaled output to differing objectives for representing the change signal in climate projections. We expect that the differing objectives will affect the projected change signal for precipitation spells and extremes and be compounded by the differing observational choices available for training data. Also, by using three different gridded observations for training, the experimental design allows quantitative exploration of additional factors that influence the results of the two downscaling methods.

Though not very sophisticated, methods like DeltaSD are used by some impacts researchers in studies that support decision-making, such as climate scenario development for the Netherlands and Switzerland (KNMI, 2014; Kotlarski et al., 2018), ecological modeling of climate impacts (Bucklin et al., 2013), and addressing data availability limitations (Walsh et al., 2018). Given the many users of DeltaSD and similar methods, it is important to compare it to more sophisticated methods, such as ERQM, if only to remind users of its limitations.

4. Results and Discussion

We used this suite of downscaled projections (created with two downscaling techniques, three GCMs and three sets of gridded observations) to address several questions. We focused on precipitation and several derived variables to answer these questions because many stakeholders in the south-central U.S. are keenly interested in future changes of precipitation. As SD-generated output is affected by both the gridded observations and the GCMs, we set the context of how these inputs affect the downscaled projections, detailed in section 4.1. In section 4.2, we discuss our findings related to questions 1 and 2. Finally, section 4.3 more broadly examines our last question.

4.1. Input Data Context

Among the gridded observations used for training purposes, Daymet and PRISM used daily precipitation amounts recorded at any time during the day to represent the daily total for the period from midnight to midnight local time. However, Livneh apportioned observations of daily precipitation based on their time of observation, prorating the precipitation total by the number of hours overlapping the date of the gridded observation (Livneh et al., 2013). For example, if the daily total were recorded at 8AM local time, representing the past 24 hours, the Livneh method apportioned two-thirds of the total precipitation to the previous day and the remaining third to the recorded date of observation.

Livneh's apportioning approach affected each daily precipitation amount, but not the annual total precipitation (prcptot). Figure 3a displays averages of prcptot across the south-central U.S. for the historical period 1981–2005 (see also Figure 4). Visual inspection revealed that differences between the observations are similar to those between observations and GCMs and among the three GCMs. The areal-average for Daymet, Livneh, and PRISM was 782.7 mm, 753.4 mm, and 653 mm, respectively. The

GCM-calculated values (areal averages of 814.4 mm for CCSM4, 701.5 mm for MIROC5, and 696.1 mm for MPI-ESM-LR) and spatial patterns across this region also were similar among GCMs during the historical period (Figure 4), including the known wet bias in the western portion of the domain (e.g., Mejia et al., 2018). The range (maximum minus minimum grid point values) of prcptot across the region, however, was consistently smaller for the GCMs than for the observations. The known tendency for GCMs to underestimate precipitation (Pendergrass and Hartmann, 2014; Stephens et al., 2010) results in fewer extreme precipitation values and prompts the need for bias correction.

Unlike annual total precipitation, the gridded observations showed marked differences in the annual number of days with precipitation (r1mm). The Livneh gridded observations have about 60 percent more days with precipitation (mean = 99.7 days) compared to Daymet and PRISM (mean = 62.1 days and 62.5 days, respectively, Figure 3b). In contrast, the GCMs tended to overestimate the observed number of days with precipitation, with CCSM4 having the largest r1mm (mean = 143.4 days) and MPI-ESM-LR having the smallest r1mm (mean = 104.1 days). The known tendency for GCMs to overestimate the number of days with light precipitation (Stephens et al. 2010; Pendergrass and Hartmann, 2014) also caused the GCM climatology for the total number of precipitation events to be larger than observed. This overestimation is another issue that SD attempts to correct, while retaining the differing future change signals from the GCMs.

Although Livneh's apportioning process improved the temporal alignment and spatial coherence, this adjustment also increased the frequency and decreased the intensity of events (Oyler and Nicholas, 2018). For large, single-day events (rx1day), this apportioning process divided precipitation across two days, decreasing the amounts on a single day (Figure 5). This effect was mitigated for multi-day events, as any single day receives prorated precipitation amounts from two days. The known tendency for GCMs to underestimate variance for daily precipitation also caused GCM estimates of precipitation extremes to be smaller than observed. Therefore, the OH training data likely will influence the downscaled number of wet and dry days, along with rainfall amounts for single-day extreme events.

Livneh's precipitation adjustment (apportioning) also affected the length of dry and wet spells in its gridded observations. For example, when a single day's precipitation was split over two days and both days had values larger than 1 mm, then two wet days were counted for Livneh rather than one for Daymet and PRISM. Since the apportioning separated rain across two consecutive days, the average length of the longest wet spell in a year (cwd) was larger in Livneh (mean 8.2 days) than Daymet (mean 5.6 days) or PRISM (mean 5.6 days; Figure 6a). Dry spells (consecutive days with precipitation less than 1 mm) also decreased in length for a similar reason. Although the apportioning effect is limited, the average length of the longest dry spell in a year (cdd) in Livneh was still smaller (mean 37.7 days) than Daymet (mean 42.6 days) and PRISM (mean 41.7 days; Figure 6b). The GCM's bias toward overestimating the number of rain days also caused the wet (dry) spells represented by most GCMs to be larger (smaller) than observations (Figure 6), though the climate change signal was slightly different among GCMs. This result suggested that while the GCM change signal had a strong influence on the downscaled output, the training data also had some influence on the downscaled projections of cdd and cwd.

Figures 3, 5, and 6 display the undownscaled future projection (2070-2099) from each GCM for each variable. The simulated climate change signals range from minimal (CCSM4) to large (MPI-ESM-LR) model-simulated changes for all variables. The future GCM values had less spatial variability and overestimate the frequency of rain days as compared to the gridded observations. This result stems in part from the coarse resolution of the GCMs that reduced precipitation variability across complex topography or coastal boundaries. SD attempted to correct these aspects based on the training data; yet, these gridded observations exhibited marked differences resulting in part from the choice of precipitation adjustment used to create them. Therefore, it is likely that the choice of training data, GCM, and downscaling technique influenced the projected precipitation output.

4.2. Sensitivities in the Downscaled Projections

The GCMs and training data were two inputs for SD, but the downscaling technique also played a critical role in the resulting high-resolution climate projections. For the temporal subsamples used in a given SD training step, DeltaSD applied a single ratio to translate relative change from each GCM to local scales across the entire precipitation distribution of the training data, while ERQM used quantile-specific ratios to translate change across all quantiles and maintain the GCM weather sequence.

From Figure 7 it is apparent that, for the annual number of precipitation days (r1mm), the change signal was highly sensitive to the GCM, training data, and downscaling technique. When downscaled using Daymet as training data, the DeltaSD-downscaled MPI-ESM-LR indicated an increase in r1mm (~+1 day), whereas ERQM-downscaled MPI-ESM-LR (also trained with Daymet) and the raw MPI-ESM-LR indicated a decrease in r1mm (~-7 and ~-16 days respectively; Figure 7, left). The difference in climate signal sign resulted from the different relative change calculations for the two methods. For DeltaSD, the ratios were generally > 1, causing future precipitation values less than 1 mm to sometimes become greater than 1 mm, slightly increasing r1mm. For ERQM, the ratios at low quantiles were generally < 1, causing the r1mm to decrease. This effect was not limited to the MPI-ESM-LR-forced results, but also occurred for the other two GCMs (Figure 7, right). Thus, ERQM better preserved the r1mm change signal at low quantiles provided by the trace-adjusted GCM (see Supporting Information) than the DeltaSD method. The resulting ERQM change signal was more than double that of the DeltaSD change signal for 75.3% of the domain on average (Table 2). In addition, switching from DeltaSD to ERQM caused the r1mm change signal to switch from *increasing* to *decreasing* for 25.5% of the domain on average.

The choice of gridded observations for training data also influenced the r1mm climate change signal. Recall that, following the CLIMDEX conventions, “wet days” in our analyses were those with more than 1 mm day^{-1} (~0.04 inches day^{-1}). For example, using MPI-ESM-LR and ERQM, downscaled output trained with Daymet, Livneh, and PRISM projected a mean decrease of 7.5 days, 14.4 days, and 8.6 days, respectively, for r1mm. All of the statistically downscaled projections created using Livneh projected a *decrease* in the number of precipitation days that was up to twice as large as those created with the other training data. Using Livneh in place of Daymet or PRISM for training caused the r1mm climate change signal to be twice as large (or larger) for 41.5% of the domain on average (Table 2). Again, as the Livneh apportioning caused the single-day precipitation amounts to become smaller (i.e., closer to the trace value), ratios < 1 caused more rain days to be converted to dry days and a larger decrease in r1mm compared to PRISM- and Daymet-trained projections. Table 3 shows the mean and interquartile range for the projected changes across the study domain for all simulations. We examine the sensitivity of the results to each GCM, downscaling technique, or gridded dataset by comparing the difference between the maximum and minimum values of the mean projected change, or the range of mean projected change (RMC). For example, the sensitivity of r1mm to the choice in GCMs is represented by the RMC for each GCM across all downscaling techniques and training data, or -0.9 days (CCSM4 average) – (-5.8 days, MPI-ESM-LR average) = 4.9 days. The RMC is calculated using the mean values in Table 3. Overall, for the SD-generated output, the projected number of days with precipitation was sensitive to the GCM (RMC = 4.9 days), downscaling technique (RMC = 5 days), and training data (RMC = 3.5 days) used (Table 3).

As one may suspect, the different ratios used by DeltaSD and ERQM also affect precipitation extremes. For example, although the raw MPI-ESM-LR output projected an area-averaged mean *increase* of 8 mm in 1-day maximum precipitation (rx1day) by the end of the century, the DeltaSD downscaled MPI-ESM-LR trained with Daymet projected a mean *decrease* of 1 mm (Figure 8, left). This difference was broadly consistent with ERQM applying larger relative change to the right tail of the GCM's precipitation distribution than for the middle of the distribution, due to the previously mentioned tendency for GCMs to underestimate daily precipitation amount variances (Figure 8). The sign of the change signal switched for rx1day based on the downscaling technique used in 58.7% of the domain on average (Table 2). On the other hand, ERQM trained with Daymet had a slightly higher positive change (mean change of +11 mm) compared to the GCM (Table 3).

The training data also appeared to have less influence on rx1day than the downscaling technique and GCM. The effect of apportioning on precipitation extremes is mitigated in the raw output by prorating over two days and further diminished in the change signal. For example, using the MPI-ESM-LR downscaled with ERQM, the mean change in rx1day was +11.2 mm, +9.3 mm, and +11.4 mm for Daymet, Livneh, and PRISM, respectively. In our example with MPI-ESM-LR, the Livneh based projected change was 80% of the Daymet or PRISM based changes. For 45.9% of the domain on average, the Livneh-based projected change to rx1day was 80% or less of the Daymet- or PRISM-based changes (Table 2). Though the training data used influenced the future projections, we found that the downscaling technique (RMC = 9.9 mm) and GCMs (RMC = 2.2 mm) had the largest influence on the change signal for rx1day (Table 3).

Similarly, the differences between DeltaSD and ERQM influenced the annual longest dry and wet spells (cdd and cwd). For cwd (Figure 9), the GCMs projected a mean change of -0.85 to 0.21 days, DeltaSD projected a mean change of -0.39 to 0 days, and ERQM projected a mean change of -0.66 to 0.54 days (Table 3). For cdd (Figure 10), the three GCMs projected a mean change of one to three days and their downscaled projections using DeltaSD and ERQM projected a mean change of zero to two days, and two to six days, respectively (Table 3). For cdd and cwd, the ERQM based projected change was twice as large (or larger) than the DeltaSD based changes for 73.6% and 71.6% of the domain on average respectively (Table 2). The ERQM result exhibits a larger range across the domain in the downscaled output than DeltaSD for both metrics (Figures 9 and 10).

For the cdd and cwd metrics, the Livneh-based projected change was twice as large (or larger) than Daymet- or PRISM-based changes for 20.2% and 59.2% of the domain on average, respectively (Table 2). That the DeltaSD future spell length metrics cdd and cwd were not much different from those of the observations' was expected because the DeltaSD future weather sequence was simply based on the observed OH time series multiplied by ratio scale factors (Table 3). For DeltaSD, spell length change signals arose largely from when the ratio multiplication caused the output value to cross the trace-value threshold. In contrast, the ERQM future values were scaled versions of the GCM's future time sequences, thus changes in spell lengths simulated by the dynamical models were reflected in the ERQM downscaled future projections (also reflected by the interquartile range, Table 3). We found that downscaled projections of cdd were most sensitive to the downscaling technique (RMC = 2.3 days) and GCM (RMC = 1.8 days, Table 3). Also, downscaled projections of cwd were most sensitive to the GCM (RMC = 0.5 days) and training data (RMC = 0.2 days, Table 3).

For the annual total precipitation (prcptot), the mean change from the GCMs was -68 to -19 mm, the mean change from DeltaSD across all GCMs and training data was -68 to -6 mm, and the mean change across all GCMs and training data from ERQM was -72 to 6 mm (Figure 11). In addition, regardless of the training data or downscaling technique used, the projections for prcptot were nearly identical for 48-56% of the domain on average when the GCM is held fixed (Table 2). We find that projections of prcptot are less sensitive to the training data (RMC = 10.4 mm) and downscaling technique (RMC = 7.5 mm) and far more sensitive to the GCM used (RMC = 63.0 mm, Table 3).

4.3. Broader Issues

Although limited in geographic region and downscaling methods examined, this study illustrates the types of uncertainties present in SD-generated precipitation projections produced by a broader range of statistical methods for various regions using different input data sources. Our analyses of downscaled precipitation climate projections for the south-central United States are broadly consistent with studies that focused on other regions (e.g., Bürger et al. [2013]). Accordingly, our results can help guide and inform users of downscaled precipitation projections, especially those who use output from the projections as input to their climate impacts-related model of interest.

First, analyses demonstrate that both SD techniques examined here are sensitive to the training data used. Second, variables such as the occurrence and extremes of precipitation are more sensitive to the choice of training data or downscaling technique than the GCM used. These results can be especially relevant for users (e.g., water managers, agricultural producers) who care about number of precipitation days and the daily maximum amount.

Third, the analyses suggest that developers of gridded observations should be strategic in selecting how they treat once-a-day precipitation station data (e.g., Section 4.1), as that choice can drive the sensitivity for several downscaled precipitation variables. The apportioning used to create Livneh data product results in more rain days with simultaneously smaller values. The effect of this apportioning increases the number of rain days and decreased the intensity of single day extreme events. However, for the number of rain days, we also find that the ERQM and DeltaSD were similar with respect to how sensitive they are to the training data. Using Livneh in place of Daymet or PRISM caused the change signal to decrease by 3.28 days and 2.24 days on average in ERQM and DeltaSD, respectively.

Finally, although the DeltaSD and ERQM approaches generally were similar in the middle of the precipitation distribution, they could be significantly different at both high and low quantiles. The differences in the SD-generated climate change signal at high and low quantiles is related to how each downscaling approach incorporated the GCM projected change. For the south-central U.S., ERQM projected a future increase in precipitation at the high quantiles. In contrast, DeltaSD projected a decrease in large, daily precipitation amounts. This difference at high quantiles resulted from the use of one change factor at all quantiles (as in DeltaSD) versus a unique, and typically varying, change factor at each quantile (as in ERQM).

These findings return us to a broader research question: With increasing demand for projections in climate impact assessments and adaptation planning, are certain training data or SD techniques better suited than others to refine climate projection data for use in impacts modeling? Based on our analyses, we answer with a qualified "yes," noting that it is important to carefully consider how projected precipitation variables are affected by the modeling choices made to produce a set of downscaled projections. The qualification arises in part from recognizing that different climate impacts applications are sensitive to different precipitation-related characteristics. For example, an application driven by changes in annual mean precipitation would be less sensitive to different statistically downscaled data products than one sensitive to changes in the length of wet or dry spells. Also, our results illustrate that practitioners should be cautious about using a single statistically downscaled data product without considering how sensitive their particular application is to the SD technique's characteristics and limitations. The degree to which any of these downscaling choices (GCM, downscaling technique, and training data) influences the projection depends on the variable in question. Here, we found that the annual number of days with precipitation was the most sensitive, while the annual total precipitation was the least sensitive to these downscaling choices. One must account for the varying sensitivity when using these variables either to make decisions or as input to impacts models, such as those commonly used to project future streamflow, crop yield, or species distribution. If an adaptation decision or impacts model requires daily precipitation values (where the number of days with rain, precipitation intensity, and dry- or wet-spell length would be important) or exceedance thresholds, we have demonstrated that certain training data or downscaling techniques are more appropriate than others.

To illustrate this, let us consider a practitioner and impacts researcher interested in changes to stream discharge under a changing climate. Calculating streamflow and discharge generally requires daily precipitation as an input to a hydrology model (Devi et al., 2015). To assess changes to stream discharge or streamflow in a changing climate requires the use of downscaled (or bias-corrected) projections of precipitation (e.g. Mizukami et al., 2016; Sunde et al., 2017). The question is, what is important to the practitioner? If the practitioner were only concerned with annual total discharge, then this study indicates that a sufficient number of GCMs (to account for model uncertainty) could be employed with any combination of downscaling technique and gridded observations. If, however, the practitioner were concerned with peak daily discharge, then they will plan for changes to precipitation extremes. This study indicates then that using downscaled projections created with DeltaSD or trained with the Livneh observations can dampen the projected intensity of single-day events, reducing the likelihood the impacts researcher will develop a simulation that adequately guides the practitioner. Generally, an impacts researcher or practitioner should ask themselves: “Do heavy precipitation events, precipitation occurrence, the length of dry and wet spells, or seasonality matter for my problem?" If so, then it is important to select climate projections that represent each of these variables in the historical and future periods.

More broadly, there are several recommendations based on this study. First, when using downscaled projections for impact assessments, one should carefully consider how the training data were created. In many cases, it is advisable to consider using downscaled projections based on more than one training dataset so that the uncertainty associated with the different options for training data is included for a more complete assessment of risk. We recognize that many times the choices of training data and downscaling technique are meant to meet specific needs (such as capturing the effect of complex topography or reflecting an important time period); however, at a minimum, users need to acknowledge the uncertainty associated with these choices.

Second, our study focuses on the precipitation adjustment used to create the gridded observations, but that adjustment is only one component in the creation of gridded observations that could be translated into downscaled projections. Methods to account for wind-induced undercatch, wetting, or evaporation losses in station measurements of precipitation (Yang et al., 1998) also have been used to create gridded observations. There are also numerous interpolation techniques and elevation corrections that have been applied with differing station networks to create gridded observation-based datasets. We have not analyzed all these corrections or interpolations here, and it was beyond the scope to completely isolate and determine which components of the gridded observations affected the precipitation projections. To our knowledge, current literature does not document any sensitivities in statistically downscaled climate projections that are caused by the many components of generating gridded observations. Therefore, we recommend that future studies examine the sensitivity of SD techniques to these other choices made during the creation of gridded observations.

Finally, DeltaSD and ERQM represent two simple downscaling techniques, but they do not represent the breadth of publicly available downscaled projections. More recent datasets have been created with complex techniques such as the Multivariate Adaptive Constructed Analogs (MACA, Abatzoglou et al., 2012) and the Localized Constructed Analogs (LOCA, Pierce et al., 2014). We recommend examining the effect of downscaling choices with these more complex downscaling techniques and impacts-specific variables, given that small changes may result in effects to the projected change signal not shown using simpler SD techniques.

The results from this study disagree somewhat with those of Pourmoktharian et al. (2016) that found that the training data had more influence on the change signal than the downscaling technique. The results from this study agree with Alder and Hostetler (2019) with respect to downscaling technique and training data as well as with Timmermans et al. (2019) with respect to precipitation extremes in gridded observations. However, Pourmoktharian et al. (2016) and Alder and Hostetler (2019) focused on the results projected by an ecosystem model and a hydrology model respectively using downscaled projections. It is plausible that the non-linear physical relationships between climate and impacts modeling could magnify the effect of the training data over the choice of downscaling technique. Therefore, a question raised by this study, is how far the effect of training data is conveyed from downscaling through impact modeling.

Pourmoktharian et al. (2016) and Alder and Hostetler (2019) are both natural extensions of this study, focusing on the effect of modeling choices in the climate projections on impact modeling. The work of Timmermans et al. (2019) focused on the gridded observations with respect to precipitation extremes. While this study went beyond precipitation extremes, we agree with the note of Timmermans et al. (2019), that one should be particularly cautious of the gridded observations used. The gridded observations used in this study are “standard” products used for SD and other research activities. The influence of the gridded observations upon downscaled projections documented in this study alongside the work of other recent studies implies the need to continuing probing the uncertainty involved and emphasize caution when using gridded observations for SD of precipitation.

5. Conclusions

This study examines the effects of the choices that climate scientists make in developing projections of multiple precipitation variables, in this case, using two simple downscaling techniques (DeltaSD and ERQM) and three gridded observations (Livneh, PRISM, and Daymet) for training. The domain of interest was the south-central United States. For simplicity, we use three GCMs (CCSM4, MIROC5, MPI-ESM-LR) that exhibited sufficient spread in their future projections while also generating a representative simulation of this region's historical climatology.

This study finds that downscaled projections of precipitation variables can be sensitive to the training data used in SD. This effect is apparent in results generated by both SD techniques. Specifically, the precipitation adjustment (apportioning) used to create the Livneh observations leads to an increased frequency of days with measurable precipitation and decreased intensity of daily events, matching the results of Oyler and Nicholas (2018). The apportioning also causes Livneh to have shorter dry spells, longer wet spells, and less intense precipitation extremes then Daymet or PRISM. These differences between gridded observations are translated into projections that use SD, causing the effects present in the Livneh observations to be present in the historical output, future output, and change signal for both downscaling techniques regardless of the GCM used. This finding is important for those who apply results from the Fourth National Climate Assessment (Easterling et al., 2017), as the LOCA-based downscaled climate projections were trained using Livneh gridded observations.

The ERQM and DeltaSD methods show marked differences for projected changes of the number of days with rain, longest dry spells, longest wet spells, and precipitation extremes. This result is caused by the different approaches to translating relative change with each downscaling technique. Although ERQM preserves the change signal at all quantiles of a distribution, DeltaSD only preserves the change in the mean. Therefore, other methods which preserve the change signal at all quantiles will likely produce change signals different from a ratio delta method for many of the variables in this study.

Although the choice of GCM, downscaling technique, and gridded observations used for training all affect projections of precipitation, we find that there are varying degrees of influence depending on the variable. Annual total precipitation is the least sensitive and the annual number of days with precipitation is the most sensitive variable. Annual total precipitation is sensitive primarily to the GCM, while the annual number of days with precipitation is sensitive to the GCM, downscaling technique, and training data.

Given that we find some projected variables are more sensitive than others, certain training data or downscaling techniques appear to be better suited than others to use in specific types of impacts modeling and adaptation planning efforts. As the use of climate projections grows across multiple applications, impacts researchers should carefully consider the potential effects of training data and downscaling technique when selecting downscaled datasets for impacts modeling in agriculture, ecology, and other fields. Such additional studies will allow the climate science community to identify potential issues where overconfidence may exist, expand the breadth of climate model evaluation, and provide users of climate information with more robust guidance for impact assessments and adaptation planning.

6. Acknowledgements

This work was funded by the United States Geological Survey under award G15PG00088, the Oklahoma EPSCoR project through the Oklahoma State Regents for Higher Education and the National Science Foundation under NSF Grant OIA-1301789, and the Office of the Vice President for Research of the University of Oklahoma through the South Central Climate Adaptation Science Center. The authors thank the Geophysical Fluid Dynamics Laboratory of the National Oceanic and Atmospheric Administration for in-kind contributions of research support and computational time, including the efforts of Mary Jo Nath and Carolyn Whitlock. Finally, the authors thank the anonymous reviewers of this article for their critiques to strengthen this article.

7. Supporting Information

The separate supporting information a description of the trace adjustment procedure used with ERQM (Appendix S1) which describes the trace adjustment procedure used with ERQM.

8. References

Abatzoglou, JT and TJ Brown. 2012. A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology* **32**: 772-780. DOI: 10.1002/joc.2312

Alder, J and S Hostetler. 2019. The dependence of hydroclimate projections in snow-dominated regions of the western U.S. on the choice of statistical downscaled climate data. *Water Resources Research*. DOI: 10.1029/2018WR023458

Basso B, Hyndman DW, Kendall AD, Grace PR, and GP Roberston. 2015. Can impacts of climate change agricultural adaptation strategies be accurately quantified if crop models are annually re-initialized? *PLoS ONE* **10**: DOI: 10.1371/journal.pone.0127333

Behnke R, Vavrus S, Allstadt A, Albright T, Thogmartin WE, and VC Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecological Applications* **26**: 1338-1351. DOI: 10.1002/15-1061

Bertrand D and RA McPherson. 2018. Future hydrologic extremes of the Red River Basin. *Journal of Applied Meteorology and Climatology* **57**: 1321-1336. DOI: 10.1175/JAMC-D-17-0346.1

Bertrand D and RA McPherson. 2019: Development of downscaled climate projections: A case study of the Red River Basin, South-Central U.S. *Advances in Meteorology*, Volume 2019, Article ID 4702139. DOI: 10.1155/2019/4702139.

Bronaugh D. 2014. R package climdex pcic: PCIC implementation of Climdex routines. version 1.1-6. Victoria, British Columbia, Canada: Pacific Climate Impact Consortium. Available from: https://CRAN.R-project.org/package=climdex.pcic

Bucklin DN, Watling JI, Speroterra C, Brandt LA, Mazzotti FJ, and SS Romanach, 2013. Climate downscaling effects on predictive ecological models: a case study for threatened and endangered vertebrates in the southeastern United States. *Regional Environmental Change* **13**: 57-68. DOI: 10.1007/s10113-012-0389-z

Bürger G, Sobie SR, Cannon AJ, Werner AT, and TQ Murdock. 2013. Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate. *Journal of Climate* **26**: 3429-3449. DOI: 10.1175/JCLI-D-12-00249.1

Cannon AJ, Sobie SR, and TQ Murdock. 2015. Bias Correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate* **28**: 6938-6959. DOI: 10.1175/JCLI-D-14-00754.1

Cesana G, Suselj K, and F Brient. 2017. On the Dependence of Cloud Feedbacks on Physical Parameterizations in WRF Aquaplanet Simulations. *Geophysical Research Letters* **44**: 10,762-10,771. DOI: 10.1002/2017GL074820

Daly C, Halblieb M, Smith JI, Gibson WP, Dogget MK, Taylor GH, Curtis J, and PP Pasteries. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Journal of Climate* **28**: 2031-2064. DOI: 10.1002/joc.1688

Daly C and K Bryant. 2013. The PRISM Climate and Weather System – An Introduction. Available from: http://prism.oregonstate.edu/documents/PRISM_history_jun2013.pdf

Devi GK, Ganasri BP, and GS Dwarakish. 2015. A review on hydrological models. *Aquatica Procedia* **4**: 1001-1007. DOI: 10.1016/j.aqpro.2015.02.126

Easterling DR, Kunkel KE, Arnold JR, Knutson T, LeGrande AN, Leung LR, Vose RS, Waliser DE, and MF Wehner. 2017. Precipitation change in the Unites States. In: Climate Science Special Report: Fourth National Climate Assessment Volume I [Wuebbles DJ, Fahey DW, Hibbard KA, Dokken DJ, Stewart BC, and TK Maycock (eds.)]. U.S. Global Change Research Program, Washington DC USA, pp. 207-230, DOI: 10.7930/J0H993CC.

Elsner MM, Gangopadhyay S, Pruitt T, Brekke LD, Mizkami N, and MP Clark, 2014. How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations? **15**: 1384-1403. DOI: 10.1175/jhm-d-13-083.1

Fowler HJ, Blenkinsop S, Tebaldi C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* **27**(12): 1547–1578. DOI: 10.1002/joc.1556.

Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley, PH, Yang Z-L, and M Zhang. 2011. The Community Climate System Model Version 4. *Journal of Climate* **24**: DOI: 10.1175/2011JCLI4083.1

Gergel DR, Nijssen B, Abatzoglou JT, Lettenmaier DP, and MR Stumbaugh. 2017. Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change* **141**: DOI: 10.1007/s10584-017-1899-y

Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Bottinger M, Brokin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segschneider J, Six KD, Stockhause M, Timmreck C, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, and B Stevens. 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project Phase 5. *Journal of Advances in Modeling Earth Systems* **5**: DOI: 10.1002/jame.20038

Hawkins E and R Sutton. 2011. The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics.* **37**: 407-418.

KNMI, 2014. KNMI'14: Climate Change scenarios for the 21st Century – A Netherlands perspective; by Bart van den Hurk, Peter Siegmund, Albert Klein Tank (Eds), Jisk Attema, Alexander Bakker, Jules Beersma, Janette Bessembinder, Reinout Boers, Theo Brandsma, Henk van den Brink, Sybren Drijfhout, Henk Eskes, Rein Haarsma, Wilco Hazeleger, Rudmer Jilderda, Caroline Katsman, Geert Lenderink, Jessica Loriaux, Erik van Meijgaard, Twan van Noije, Geert Jan van Oldenborgh, Frank Selten, Pier Siebesma, Andreas Sterl, Hylke de Vries, Michiel van Weele, Renske de Winter and Gerd-Jan van Zadelhoff. Scientific Report WR2014-01, KNMI, De Bilt, The Netherlands. www.climatescenarios.nl

Kotlarski S, Rajczak J, Casaneuva A, Feigenwinter I, and M Liniger, 2018. Localized Projections. CH2018 – Climate Scenarios for Switzerland Technical Report. https://www.nccs.admin.ch/dam/nccs/de/dokumente/website/klima/CH2018_Technical_Report.pdf.download.pdf/CH2018_Technical_Report.pdf

Li H, Sheffield J, and EF Wood. 2010. Bias correction of monthly precipitation and temperature fields Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research Atmospheres* **115**: DOI: 10.1029/2009JD012882

Livneh B, Rosenberg EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, and DP Lettenmaier. 2013. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate* **26**: 9384–9392. DOI: 10.1175/JCLI-D-12-00508.1

Maraun DF, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeβl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, and I Thiele-Eich. 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*. **48**: DOI: 10.1029/2009RG000314

Maraun D and Coauthors. 2014. VALUE: COST Action ES1102 (2012-2015). European Cooperation in Science and Technology, accessed March 26, 2019. Available from: http://www.value-cost.eu/

Markus M, Angel J, Byard G, McConkey S, Zhang C, Cai X, Notaro M, and M Asfhaq. 2018. Communicating the Impacts of Projected Climate Change on Heavy Rainfall Using a

Weighted Ensemble Approach. *Journal of Hydrologic Engineering* **23**: DOI: 10.1061/(ASCE)HE.1943-5584.0001614.

Maurer E, Wood AW, Adam JC, Lettenmaier DP, and B Nijssen, 2002. A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. *Journal of Climate* **15**, DOI: https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

Mejia J, Koracin D, and E Wilcox. 2018. Effect of coupled global climate models sea surface temperature biases on simulated climate of the western United States. *International Journal of Climatology* **38**: 5386-5404. DOI: 10.1002/joc.5817

Mizukami N, Clark MP, Gutmann ED, Mendoza PA, Newman AJ, Nijssen B, Livneh B, Hay LE, Arnold JR, and LD Brekke. 2016. Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled Forcing Data and Hydrologic Models. *Journal of Hydrometeorology* **17**: 73-98. DOI: 10.1175/JHM-D-14-0187.1

NCAR. 2019. Regridding using NCL with Earth System Modeling Framework (ESMF) software. Accessed January 9, 2019. Available from https://www.ncl.ucar.edu/Applications/ESMF.shtml

NCAR, UCAR, CISL, TDD. 2018. The NCAR Command Language (Version 6.5.0) [Software]. Boulder, Colorado, DOI: http://dx.doi.org/10.5065/D6WD3XH5

Oyler JW and RE Nicholas. 2018. Time of observation adjustments to daily station precipitation may introduce undesired statistical issues. *International Journal of Climatology* **38**: 364-377. DOI: 10.1002/joc.5377

Pendergrass AG and DL Hartmann. 2014. Two Modes of Change of the Distribution of Rain. *Journal of Climate* **27**: 8357-8371. DOI: 10.1175/JCLI-D-14-00182.1

Pierce DW, Cayan DR, and BL Thrasher. 2014. Statistical downscaling using localized constructed analogs (LOCA). *Journal of Hydrometeorology* **15**: 2558–2585.DOI:10.1175/JHM-D-14-0082.1.

Pierce DW, Cayan DR, Maurer EP, Abatzoglou JT, and KC Hegewisch. 2015. Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology* **16**: 2421-2442. DOI: 10.1175/JHM-D-14-0236.1

Pourmokhtarian, A, Driscoll CT, Campbell JL, Hayhoe K, and AMK. Stoner. 2016. The Effects of Climate Downscaling Technique and Observational Dataset on Modeled

Ecological Responses. *Ecological Applications* **26**: 1321-1337. DOI: https://doi.org/10.1890/15-0745

Räty O, Räisänen J, and JS Ylhäisi. 2014. Evaluation of delta change and bias correction methods for future daily precipitation: intermodal cross-validation using ENSEMBLES simulations. *Climate Dynamics* **42**: 2287-2303. DOI: 10.1007/s00382-014-2130-8

Riahi, K, Grubler A, and N Nakicenovic. 2007. Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change* **74**: DOI: 10.1016/j.techfore.2006.05.026

Rummukainen M. 2010. State-of-the-art with regional climate models. *WIREs Climate Change* **1**: DOI:10.1002/wcc.008.

Rummukainen M. 2016. Added value in regional climate modeling. *WIREs Climate Change* **7**: 145-159. DOI:10.1992/wcc378

Sheffield J, Barrett AP, Colle B, Fernando DN, Fu R, Geil KL, Hu Q, Kinter J, Kumar S, Langenbrunner B, Lombardo K, Long LN, Maloney E, Mariotti A, Meyerson JE, Mo KC, Neelin JD, Nigam S, Pan Z, Ren T, Ruiz-Barradas A, Serra YL, Seth A, Thibeault JM, Stroeve JC, Yang Z, and L Yin, 2013. North American Climate in CMIP5 Experiments.

Part 1: Evaluation of historical simulations of continental or regional climatology. **26(**23): 9209-9245. DOI: 10.1175/JCLI-D-12-00592.1

Sillmann J, Kharin VV, Zwiers FW, Zhang X, and D Bronaugh, 2013. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. **118**: 2473-2493. DOI: 10.1002/jgrd.50188

Smith L and N Stern. 2011. Uncertainty in science and its role in climate policy. *Philosophical Transactions of the Royal Society A* **369**: 1-24. doi:10.1098/rsta.2011.0149

Stephens GL, L'Ecuyer T, Forbes R, Gettleman A, Golaz J-C, Bodas-Salcedo A, Suzuki K, Gabriel P, and J Haynes. 2010. Dreary State of precipitation in global models. *Journal of Geophysical Research Atmospheres* **115**: 1-14. DOI: 10.1029/2010JD014532

Sunde MG, He HS, Hubbart JA, and MA Urban. 2017. Integrating downscaled CMIP5 data with a physically based hydrologic model to estimate potential climate change impacts on streamflow processes in a mixed-use watershed. *Hydrological Processes* **31**: 1790-1803. DOI: 10.1002/hyp.11150

Tabari H, DeTroch R, Giot O, Hamdi R, Termonia P, Saeed S, Brisson E, Van Lipzig N, and P Williams. 2016. Local impact analysis of climate change on precipitation extremes: are high-resolution climate models needed for realistic simulations? *Hydrology* and *Earth System Science*. 1-21. DOI:105194/hess-2016-106.

Tapiador FJ, Roca R, Genio AD, Dewitte B, Petersen W, and F Zhang. 2019. Is Precipitation a Good Metric for Model Performance? *Bulletin of the American Meteorological Society* **100**: 223-233. DOI: 10.1175/bams-d-17-0218.1

Taylor A, Gregory JM, Webb MJ, and KE Taylor. 2012. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters* **39**: DOI: 10.1029/2012GL051607

Teutschbein C, Seibert J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456–457: 12–29. DOI: 10.1016/j.jhydrol.2012.05.052.

Thornton, P.E., S.W. Running, and M.A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* **190**: 214 - 251. DOI:http://dx.doi.org/10.1016/S0022-1694(96)03128-9

Thornton PE, Thornton MM, Mayer BW, Wei Y, Devarakonda R, Vose RS, and RB Cook. 2017. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. DOI: 10.3334/ORNLDAAC/1328

Timmermans B, Wehner M, Cooley D, O'Brien T, and H Krishnan. 2019. An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dynamics* **52**(11): 6651-6670. DOI: 10.1007/s00382-018-4537-0

Van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Mainshausen M, Nakicenovic N, Smith SJ, and SK Rose. 2011. The representative concentration pathways: an overview. *Climatic Change* **109**(5): DOI: 10.1007/s10584-011-0148-z

Walsh JE, Bhatt US, Little JS, Leonawicz M, Lindgren M, Kurkowski TA, Bieniek PA, Thoman R, Gray S, and TS Rupp, 2018. Downscaling of climate model output for Alaskan stakeholders **110**: 38-51. DOI: 10.1016/j.envsoft.2018.03.021

Wang L and W Chen. 2014. Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmospheric Science Letters* **15**: 1-6. DOI: 10.1002/asl2.454

Watanabe M, Suzuki T, O'ishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T, Hasumi H, Tatebe H, and M Kimoto, 2010: Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate* **23**: DOI: 10.1175/2010JCLI3679.1

Weart S. 2010. The development of general circulation models of climate. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics* **41**: 208-217. DOI: 10.1016/j.shpsb.2010.06.002

Werner AT and AJ Cannon. 2016. Hydrologic extremes - an intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences* **20**: 1483-1508. DOI: 10.5194/hess-20-1483-2016

Whitehall K. and Coauthors. 2012. Building Model Evaluation and Decision Support Capacity for CORDEX. World Meteorological Organization, accessed March 26, 2019, Available from: https://public.wmo.int/en/resources/bulletin/building-model-evaluation-and-decision-support-capacity-cordex

Wilby RL and TML Wigley. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography: Earth and Environment* 21(4): 530–548. DOI: 10.1177/030913339702100403.

Wootten A, Terando A, Reich BJ, Boyles RP, and F Semazzi. 2017. Characterizing Sources of Uncertainty from Global Climate Models and Downscaling Techniques. *Journal of Applied Meteorology and Climatology* **56**: 3245-3262. DOI: 10.1175/JAMC-D-17-0087.1

Yang D, Goodison BE, and S Ishida. 1998. Adjustment of daily precipitation data at 10 climate stations in Alaska: Application of World Meteorological Organization intercomparison results. *Water Resources Research* **34**: DOI: 10.1029/97WR02681

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW. 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIRES Climate Change* **2**(6): 851–870. DOI: 10.1002/wcc.147.

Figure Captions

Figure 1. Study domain overlaid with annual average precipitation (mm) from Daymet v. 2.1 (Thornton et al. 1997; Thornton et al. 2017).

Figure 2. Daily precipitation cumulative distribution functions (CDFs) from PRISM (thick black line), MPI-ESM-LR historical values (thin, dark blue line) and future projections (red line with open circles), and the ERQM (light blue line with diamonds) and DeltaSD (purple line with triangles) downscaled results. The historical period is 1981-2005 and the future period is 2070-2099 with RCP8.5. The horizontal dashed lines correspond to the 50^{th} and 95^{th} percentiles.

Figure 3. Boxplots of the (a) climatology of the annual total precipitation (prcptot, in mm) and (b) climatology of the annual number of days with precipitation (r1mm) for the gridded observations during the historical period (1981-2005; left set), the GCMs used during the historical period (center set), and the GCMs during the future period (2070-2099; right set). Open circles indicate the minimum and maximum for each boxplot, and horizontal lines reflect the 5^{th}, 25^{th}, 50^{th} (median; thickest line), 75^{th}, and 95^{th} percentiles, from bottom to top of each box plot. The domain average is the asterisk overlaid on each boxplot.

Figure 4. Average annual total precipitation (prcptot, in mm) during the historical period (1981–2005) for the three GCMs (top row) and three gridded observation datasets (bottom row). From left to right on the top row: CCSM4, MIROC5, and MPI-ESM-LR. From left to right on the bottom row: Daymet, Livneh, and PRISM. Domain-wide maximum, mean, and minimum are plotted in the lower left of each map. Note the similarities within GCMs and within gridded observations, but the large differences between the GCMs and gridded observations.

Figure 5. As in Figure 3 except for the annual 1-day maximum precipitation (rx1day).

Figure 6. Boxplots of the (a) climatology of the annual longest wet spell (cwd; left half) and (b) climatology of the annual longest dry spell (cdd; right half) for the gridded observations during the historical period (1981-2005; left set for each variable), the GCMs used during the historical period (center set for each variable), and the GCMs during the future period (2070-2099; right set for each variable). Open circles indicate the minimum and maximum for each boxplot, and horizontal lines reflect the 5th, 25th, 50th (median; thickest line), 75th, and 95th percentiles, from bottom to top of each box plot. The domain average is the asterisk overlaid on each boxplot. Note the different scales for cwd and cdd.

Figure 7. Left, the projected change (2070-2099 values minus 1981-2005 values) of the annual average of the number of days with precipitation (r1mm, in days) for MPI-ESM-LR

(top; not downscaled), MPI-ESM-LR downscaled with DeltaSD using Daymet for training (middle), and MPI-ESM-LR downscaled with ERQM using Daymet for training (bottom). Positive (negative) values indicate more (fewer) rain days in the future. Domain-wide maximum, mean, and minimum are plotted in the lower left of each map. Right, boxplots of projected change of r1mm for all GCMs and downscaling experiments. Open circles indicate the minimum and maximum for each boxplot, and horizontal lines reflect the 5th, 25th, 50th (median; thickest line), 75th, and 95th percentiles, from bottom to top of each box plot. The domain average is the asterisk overlaid on each boxplot.

Figure 8. As in Figure 7 except for the annual average of the maximum 1-day precipitation (rx1day, in mm). Positive (negative) values indicate higher (lower) values for rx1day in the future.

Figure 9. As in Figure 7 except for the annual average of the longest wet spell (cwd, in days). Positive (negative) values indicate longer (shorter) duration of the longest wet spell in the future.

Figure 10. As in Figure 7 except for the annual average of the longest dry spell (cdd, in days). Positive (negative) values indicate longer (shorter) duration of the longest dry spell in the future.

Figure 11. As in Figure 7 except for the annual total precipitation (prcptot). Positive (negative) values indicate more (less) precipitation in the future.

Tables

Table 1. Gridded observation-based datasets used in this study.

Dataset (Citation)	Interpolation or gridding method	Precipitation Adjustments	Native resolution	Time Period Available
Daymet version 2.1 (Thornton et al. 1997; Thornton et al. 2017)	Geographically weighted regression	No adjustment	1 km^2	1980 -
Livneh version 1.2 (Maurer et al. 2002; Livneh et al. 2013)	Synergraphic mapping system (SYMAP), precipitation scaled to match PRISM climatology	Uniform adjustment similar to that tested by Oyler and Nicholas (2018)	1/16 degree (~6 km^2)	1950 - 2013
PRISM version AN81d (Daly et al. 2008; Daly et al. 2013).	Geographically and elevation-weighted regression, station weighting by topography, distance to coast, atmospheric factors.	No adjustment	2.5 min (~4 km^2)	1981-

Table 2. Mean percentage of study domain where switching between training data or downscaling techniques causes the magnitude of the change signal to more than double, become 80% or less, remains approximately the same, or causes the sign of the change signal to reverse. Where the training dataset is changed, the value is the mean percentage across GCMs and downscaling technique. Where the downscaling technique is changed, the value is the mean percentage across GCMs and training datasets.

	Variable of Interest	Changing training data to Livneh from Daymet or PRISM	Changing downscaling technique to ERQM from DeltaSD
Mean percentage of the domain where the change signal more than doubles in magnitude	r1mm	41.5	75.3
	rx1day	8.2	55.8
	cdd	20.2	73.6
	cwd	59.2	71.6
	prcptot	8.0	11.4
Mean percentage of the domain where the change signal becomes 80% or less in magnitude	r1mm	12.2	7.7
	rx1day	45.9	19.8
	cdd	50.3	11.0
	cwd	16.8	11.9
	prcptot	18.4	28.6
Mean percentage of the domain where the change signal within 25% of the same.	r1mm	15.7	6.3
	rx1day	44.0	11.8
	cdd	19.6	6.8
	cwd	10.5	7.3
	prcptot	56.0	48.0
Mean percentage of the domain where the sign of the change signal reverses	r1mm	15.7	25.5
	rx1day	7.2	58.7
	cdd	21.4	32.7
	cwd	31.6	45.4
	prcptot	5.0	8.2

Table 3. Means and Interquartile ranges (IQR) for projected changes of r1mm (days), rx1day (mm), cdd (days), cwd (days), and prcptot (mm) across the study domain for each GCM and downscaled projection in this study.

DS	GCM	Training Data	r1mm		rx1day		cdd		cwd		prcptot	
			mean	IQR	mean	IQR	mean	IQR	mean	IQR	mean	IQR
N/A	CCSM4	N/A	-5.1	10.4	1.2	5.4	1.2	3.3	0.2	2.2	-20.2	67.8
	MIROC5		-10.8	5.2	6.8	5.9	3.0	4.0	-0.2	1.3	-69.1	54.2
	MPI-ESM-LR		-16.3	9.3	7.7	4.6	2.9	7.3	-0.9	1.2	-51.1	99.3
DeltaSD	CCSM4	Daymet	0.6	1.5	0.8	3.0	0.1	1.4	0.0	0.2	-7.0	47.1
		Livneh	-1.8	2.7	0.5	2.5	1.4	2.1	-0.1	0.3	-8.2	45.0
		PRISM	-1.1	1.4	0.6	2.9	1.3	1.9	-0.1	0.2	-7.5	38.0
	MIROC5	Daymet	0.0	1.2	-4.2	6.7	0.5	1.3	0.0	0.1	-68.4	79.7
		Livneh	-3.1	2.3	-3.4	4.9	2.1	2.5	-0.1	0.4	-66.3	77.1
		PRISM	-1.7	1.4	-4.9	6.5	1.8	2.1	-0.1	0.2	-62.5	73.0
	MPI-ESM-LR	Daymet	0.7	1.8	-1.4	8.7	-0.1	1.5	0.0	0.1	-35.0	98.3
		Livneh	-3.5	5.8	-1.6	7.4	1.6	2.6	-0.4	0.6	-38.3	103.4
		PRISM	-1.7	2.3	-1.1	8.3	1.6	2.4	-0.1	0.2	-28.4	74.8
ERQM	CCSM4	Daymet	-0.2	4.3	4.6	7.5	2.0	6.0	0.4	0.8	1.3	50.7
		Livneh	-2.0	5.7	3.8	6.2	2.2	5.2	0.5	1.2	-1.4	45.2
		PRISM	-1.0	4.0	4.6	7.6	2.5	6.0	0.3	0.8	6.0	44.1
	MIROC5	Daymet	-6.7	6.2	10.4	11.8	5.3	7.4	-0.1	0.6	-73.2	98.7
		Livneh	-9.1	5.7	8.7	9.8	4.5	5.7	-0.2	0.9	-65.7	75.9
		PRISM	-7.5	5.8	10.5	12.8	5.9	7.7	-0.1	0.6	-58.7	83.0
	MPI-ESM-LR	Daymet	-7.5	4.9	11.2	13.1	2.8	10.5	-0.3	0.7	-27.8	104.1

Livneh	-14.4	7.3	9.3	10.8	2.8	8.0	-0.7	1.0	-36.4	95.8
PRISM	-8.7	4.7	11.4	13.6	3.4	10.6	-0.3	0.7	-11.4	82.0