# Identifying and correcting biases in localized downscaling estimates of daily precipitation return values

Mark D. Risser[1] · Daniel R. Feldman[1] · Michael F. Wehner[1] · David W. Pierce[2] · Jeffrey R. Arnold[3]

## Abstract

Extreme precipitation events are a major cause of economic damage and disruption, and need to be addressed for increasing resilience to a changing climate, particularly at the local scale. Practitioners typically want to understand local changes at spatial scales much smaller than the native resolution of most Global Climate Models, for which downscaling techniques are used to translate planetary-to-regional scale change information to local scales. However, users of statistically downscaled outputs should be aware that how the observational data used to train the statistical models is constructed determines key properties of the downscaled solutions. Specifically for one such downscaling approach, when considering seasonal return values of extreme daily precipitation, we find that the Localized Constructed Analogs (LOCA) method produces a significant low bias in return values due to choices made in building the observational data set used to train LOCA. The LOCA low biases in daily extremes are consistent across event extremity, but do not degrade the overall performance of LOCA-derived changes in extreme daily precipitation. We show that the low (negative) bias in daily extremes is a function of a time-of-day adjustment applied to the training data and the manner of gridding daily precipitation data. The effects of these choices are likely to affect other downscaling methods trained with observations made in the same way. The results developed here show that efforts to improve resilience at the local level using extreme precipitation projections can benefit from using products specifically created to properly capture the statistics of extreme daily precipitation events.

**Keywords** Extreme precipitation · Local impacts · Localized constructed analogs · Scaling factor

---

✉ Mark D. Risser
  mdrisser@lbl.gov

[1] Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2] Scripps Institution of Oceanography, La Jolla, CA, USA

[3] US Army Corps of Engineers, Seattle, WA, USA

Springer

# 1 Introduction

Large infrastructure decisions have in the past been based on average temperature and precipitation, their variability, and various measures of extreme events. Such decisions can require many years of lead time, often using compilations of temperature and precipitation statistics from long-duration observations and assuming stationarity in the distributions of environmental variables like temperature and precipitation (Klemes et al. 1982). Recognition of the effects of climate change has introduced new approaches using descriptions of the nonstationarity of those distributions, which are among the most important state variables for understanding climate change threats and impacts (Milly et al. 2008; Hallegatte 2009; Engle 2011). The climate science community has engaged in major international efforts through the Coupled Model Intercomparison Project (CMIP) exercises (e.g., Taylor et al. 2012; Eyring et al. 2016) to assess and understand changes in the planetary-to-regional-scale distributions of temperature and precipitation, and the associated analyses from that work are a key part of most applications of global-scale climate information for understanding and preparing for local impacts. These Global Climate and Earth System Models (GCMs and ESMs) use numerical representations of physical processes in Earth's climate system to project, among other effects, a range of responses of Earth's hydrological cycle as it adjusts to the warming resulting from specified concentration or emissions pathways.

However, GCMs and ESMs and the experiments they run for CMIP are not designed to depict local change. Computational limitations still require that most of the production-class GCMs and ESMs used in the CMIP experiments use 100 km (or larger) geospatial grids, which can be much larger than the scale of the decisions that practitioners must make concerning local impacts. One widely used approach to make native-scale GCM output more useful at finer spatial scales is to downscale the GCM outputs. Downscaling methods range across a continuum from empirical-statistical (e.g., Wood et al. 2004; Abatzoglou and Brown 2012; Stoner et al. 2013; Pierce et al. 2014), to fully dynamical depending on the extent to which physical processes are represented in the downscaling approach (e.g., Giorgi and Gutowski 2015; ), with hybrid methods between those two (e.g., Gutmann et al. 2016). However, the wide range of downscaling methods often produces a range of model results for historical and projected future conditions, which can be difficult to adapt to some long-term planning workflows (Brown et al. 2012; Ntegeka et al. 2014; Moss 2017; Feldman et al. 2021).

Practitioners using downscaled projections to describe and understand atmospheric and land-surface impacts at finer scales than those produced by the parent model (e.g., GCM) should consider multiple questions in tandem (e.g., Barsugli et al. 2013; Vano et al. 2018), including their tolerance for parent model spread across variables of interest, the level of confidence in, and fidelity of, the parent model to known and relevant processes at the parent model scale, and whether the downscaling solution introduces additional process information to fine-scale projections that is absent from the parent model but which materially improves the downscaled projection (e.g., Walton et al. 2020; Hall et al. 2020, which highlights that errors in under-resolved precipitation and snow albedo processes in parent model projections for California can be addressed, at least in part, through dynamical downscaling). These questions arise irrespective of whether practitioners directly apply downscaled fields to specific locations or indirectly apply downscaled outputs through approaches such as decision-scaling, in which a stochastic assessment of risk is combined with insights derived from climate projections at some scale to determine the likelihood of changes in events of particular interest at a given location (e.g., Brown et al. 2012). When the environmental variable involves daily precipitation extremes, directly using native

GCM or ESM outputs with $\geq 100$ km spatial resolution can often be excluded at the start of many practitioners' studies since the scales of their decisions are so much finer. However, downscaling extremes in daily precipitation is one of the key challenges for statistical downscaling methods (Chen and Zhang 2021), particularly since the gridded daily precipitation data products used for training are well-known to diminish variability and mute extremes (King et al. 2013; Risser et al. 2019c). The important point is that any data set or technique used should be able to capture extremes, which in principle can be accomplished by a variety of methods. One such example that can be tailored to reproduce and analyze extremal behavior is stochastic weather generators (SWGs; Vrac et al. 2007; Ayar et al. 2020), although coherency across space and variables is still a challenge for SWGs. Unfortunately, many of these approaches (including SWGs) require domain-specific scientific and statistical expertise that can be infeasible for practitioners to implement. It is far more practical for practitioners to analyze existing downscaled products than to, e.g., implement and run stochastic weather generators, and they can benefit from understanding when and where such products are appropriate for a particular problem of interest. To that end, we explore here how known biases in gridded daily precipitation products translate into corresponding biases in downscaling solutions.

In this paper, we demonstrate how the process of constructing observational data sets for training the empirical-statistical downscaling methods has a direct influence on the downscaling solutions, particularly for the problem of long-period return values of daily precipitation extremes. Specifically, we explore in detail the influence of training data on the Localized Constructed Analogs (LOCA) statistical downscaling method (Pierce et al. 2014), which used the daily gridded precipitation data set (Livneh et al. 2015a; 2015b) (henceforth L15) as training data, and show how using a product specifically constructed to accurately capture pointwise extreme precipitation statistics can give a more useful depiction of such extremes than the standard gridded daily precipitation product used to train the statistical models.

Before proceeding, an important piece of background has to do with the interpretation of gridded precipitation data sets. Generally speaking, gridded precipitation can either be interpreted as an areal average over a grid box or as a point estimate (Chen and Knutson 2008). As described in Gervais et al. (2014), precipitation arises from small-scale structures that are discontinuous in nature, which can lead to large differences between the areal average and point estimate interpretations. Such interpretational differences can furthermore lead to significant differences in the resulting statistics of precipitation, particularly for extremes (see, e.g., Gervais et al. 2014; Wehner et al. 2021). GCM-simulated precipitation (and, more generally, precipitation from any dynamical model) is best understood as an areal average regardless of how small the grid boxes are since a physical model can accumulate fluxes through the boundaries of its grid boxes. Measurements of precipitation from weather stations, on the other hand, follow the point estimate interpretation since they represent precipitation collected at an (essentially) infinitesimally small geospatial location. As is convincingly demonstrated in Gervais et al. (2014), any gridding method based on interpolation of daily weather station measurements (e.g., bilinear interpolation or inverse distance weighting) will yield a daily gridded precipitation product that should be interpreted as a point estimate. In Section 2, we describe how and why each of the gridded precipitation data products considered in this paper should be interpreted as point estimates and explicitly not areal averages.

The paper proceeds as follows: in Section 2, we describe the data sets (observational and modeled) and methods for comparing estimates of seasonal return values for daily precipitation. In Section 3, we present our results for an illustrative set of specific locations,

investigate two primary reasons for biases in the gridded daily product used for training the statistical downscaling solutions, derive a new data product of scaling factors that allows users of LOCA products derived from CMIP5 model outputs to adjust return value estimates, and apply the scaling factor to our illustrative locations. Finally, in Section 4, we conclude with the implications of these findings and suggestions for practitioners focused on location-specific analyses.

## 2 Data and methods

### 2.1 Observational data products

**Livneh daily gridded product** The Livneh et al. (2015a, b) daily gridded product provides estimates of daily precipitation at 1/16th degree spatial resolution over much of North America for 1950–2013. L15 is a widely used data product, and is relevant here because it is used as training data for the LOCA downscaling method. The L15 data product uses *in situ* measurements of daily total precipitation (over CONUS, the input data are from the Global Historical Climatology Network; Menne et al. 2012) to create a daily gridded product in two steps. First, for each day, the station measurements are interpolated to a $1/16°$ or $\sim 6$ km high-resolution grid using the SYMAP algorithm (Shepard 1968, 1984), which is an inverse distance weighting approach. Second, the interpolated data are multiplied by a monthly scaling factor determined by the ratio of its mean monthly baseline climatology (1981–2010) and the mean monthly climatology from the same period of the topographically aware PRISM data product (Daly et al. 1994, 2008). The L15 data product is largely an extension of an earlier version of the data set (Livneh et al. 2013, henceforth L13); all time-of-day splitting issues (see Section 3.2) were first raised in L13 and then inherited by L15.

As described in Section 1, the L15 product should be interpreted as a point estimate (as opposed to a grid box area average) because it is based on an interpolation of the underlying station data (Gervais et al. 2014) and does not preserve fluxes at the grid box boundaries. The SYMAP algorithm uses a weighted average of the four stations nearest to the grid cell (where the weights are based on the inverse distance from the grid cell centroid); hence, if a cell contains a station, then the interpolated value of that grid cell closely follows the observed value at that station since the station inside the grid cell is weighted more than stations outside the grid cell. However, in principle, if there are many observing stations that fall within a grid cell, the inverse distance weighting approach could be thought of as an area average when the grid box is sufficiently small (although the arguments in Gervais et al. 2014, would suggest that the point estimate interpretation is still correct even in this case). How often do two or more stations fall within a single 6 km Livneh grid cell? Taking into consideration the dynamic nature of stations falling in and out of the data set (due to missing data measurements and possible changes in station locations over the historical record), across all grid box-days, there is more than one station within 6 km of the grid cell just 0.32% of the time. All of these arguments verify that L15 is not an area-averaged product and instead should be considered a gridded representation of point (station) measurements.

**Risser climatological gridded product** Gridded daily products like L15 have been identified recently (King et al. 2013; Gervais et al. 2014; Timmermans et al. 2019) as inappropriate for characterizing extreme local precipitation since daily precipitation is known to exhibit fractal scaling (see, e.g., Yevjevich 1972; Lovejoy et al. 2008; Maskey et al.

2016, and numerous references therein). For this reason, any spatial averaging during the gridding process will diminish variability and extreme values in the original observations. As an alternative, Risser et al. (2019c) develop a "probabilistic" gridded data product that is appropriate for climatological analyses and is specifically designed to characterize local measurements of extreme precipitation. The Risser et al. (2019a) data product (henceforth R19) preserves the extreme statistics of weather station measurements by first calculating extreme statistics at each station and then interpolating these statistics using a data-driven technique to a 1/4° or ~25 km grid. Furthermore, R19 contains uncertainty quantification in the form of a bootstrap or resampling-based ensemble of estimates. Note that as an interpolation-based method the R19 product should be interpreted as a point measurement (Gervais et al. 2014). The benefits of the R19 data set (as documented in Risser et al. 2019c) include (1) the spatial aspect of the statistical analysis borrows strength from neighboring stations when estimating the extreme statistics at each station and (2) this borrowing of strength reduces the uncertainty such that the signal-to-noise ratio of the underlying spatial signal is increased.

There are several reasons why the R19 product is used in this paper as the "ground truth" for point estimates of the statistics of extreme precipitation over the CONUS. First, compared to a single-station analysis that estimates extreme statistics of daily precipitation at each station independently of all others, the R19 product has small root mean squared error in 20-year return values (ranging from 5.5mm in DJF to 11.2mm in JJA as a CONUS average; much of this "error" is smoothing over observational uncertainty) and significantly smaller bootstrap standard errors (ranging from a reduction of 40% in DJF to more than 50% in JJA). More importantly, Molter et al. (2021) determine that the statistics of seasonal extreme rainfall across CONUS vary only minimally on spatial scales smaller than 100 km (especially east of the Rocky Mountains), verifying that the R19 product can effectively describe the spatial variability of extreme precipitation since the average spacing of the underlying GHCN station data is ∼ 27 km. Furthermore, Molter et al. (2021) find that R19 accurately represents corresponding statistics of extreme precipitation estimated from the NEXRAD Stage IV radar-based daily precipitation product, which has a horizontal resolution of 4 km. As a result, the R19 product can safely be considered the "ground truth" for point estimates of the statistics of extreme precipitation, particularly in the eastern United States.

Finally, it is important to emphasize that R19 does not attempt to grid daily precipitation, and hence only provides climatological summaries (i.e., return values) with measures of uncertainty and so cannot be used to train traditional statistical downscaling methods in place of gridded products. The implication, however, is that daily extremes in precipitation as represented in R19 are not "smeared" over space, which is a fundamental property of any data product that attempts to interpolate daily precipitation without explicitly accounting for fractal scaling in precipitation (including L15). This includes other ensemble data products (e.g., Newman et al. 2015), which provide measures for estimating uncertainty but still diminish extreme values and variability by gridding daily precipitation data.

**Statistically downscaled data products** As previously mentioned, the L13 and L15 data sets have many applications, including as a training data set for the Localized Constructed Analogs (LOCA) statistical downscaling method (Pierce et al. 2014). LOCA has been used to downscale climate model outputs including those from the CMIP5 archive, and LOCA-downscaled CMIP5 model products for historical and future emissions scenarios have been widely used, including as the basis for projections of change in temperature and precipitation across the Conterminous United States in the 4th National Climate Assessment

(Wuebbles et al. 2017). Those LOCA outputs were downscaled from 32 GCMs and have a daily timestep for the period 1950–2100. Here we use GCM outputs downscaled from the CMIP5 historical period (1950–2005), for the 31 CMIP5 models shown in Supplemental Table 1.

LOCA is a constructed analog method for downscaling. In the constructed analog approach, fine-scale observations are first coarsened to the GCM grid, then the 30 coarsened observed days that best match the specific GCM day being downscaled are identified. These are termed the analog days. Traditional constructed analog downscaling then calculates the weighted average of the 30 analog days that best reproduces the specific GCM day being downscaled. The 30 weights are then applied to the 30 original high-resolution observational days, and the resulting weighted average is the high-resolution downscaled solution. However this averaging has the drawback that extremes are muted and (for precipitation) drizzle is generated in the downscaled field. LOCA avoids this by starting with the 30 analog days, but then using only the single best-matching analog day of the 30 in a $1° \times 1°$ latitude-longitude box around the point being downscaled to. With less averaging, extreme values as represented in the training data are better preserved in the downscaled field and extraneous drizzle is not generated. Details are given in Pierce et al. (2014). Note that, since the LOCA solutions are a statistical method trained on L15, the interpretation of LOCA is the same as the interpretation of L15: its gridded precipitation values should be interpreted as a point estimate.

## 2.2 Return value assessment

Our key objective with this paper is to compare estimates of daily precipitation return values at specific locations from LOCA and R19 in order to demonstrate how confidence in estimates of extreme daily precipitation return values can be enhanced by consulting a product that is specifically constructed on the basis of extreme value analysis. Return values (sometimes referred to as a return level) for daily precipitation define a threshold for what is considered an extreme or severe daily precipitation total at a given location. As such, return values are an important quantity for assessing local risk and designing infrastructure through quantifying the probability of rare events, including estimates for events that have not occurred in a particular observational period. Return values are defined for a particular "return period" $r$, where often $r = 10, 20, 50$, or 100 years, which specifies the event rarity: for example, in any given year, the probability of exceeding the threshold defined by the 20-year return value is $1/20 = 0.05$. In order to compare return value estimates between R19 and LOCA, we select the nearest R19 grid cell (25 km resolution) and LOCA grid cell (6 km resolution) for each of a set of locations across CONUS, focusing on US Department of Defense facilities where practitioners are planning for hydroclimatological extremes by considering downscaling solutions (Moss 2017). We evaluate the influence of the different spatial scales of these data products in the Supplemental Materials.

The R19 data product directly provides return value estimates for daily precipitation at each grid cell in each of the four three-month seasons (DJF, MAM, JJA, and SON), and hence contains the extreme statistics of interest. To calculate the corresponding seasonal quantities for the LOCA solutions, we apply an extreme value analysis to the time series of daily precipitation at each matched grid cell. First, for each grid cell in each season/year we calculate the maximum daily precipitation total (mm), commonly referred to as Rx1Day. Next, we characterize the extreme climatology for each grid cell using the Generalized Extreme Value (GEV) family of distributions, which is a statistical modeling framework for the maxima of a process over a specified time interval or "block"; here, the three-month

seasons. Coles (2001) (Theorem 3.1.1, page 48) shows that when the number of measurements per block is large enough, the cumulative distribution function (CDF) of the seasonal Rx1Day in year $t$ and grid cell $\mathbf{g}$, denoted $Y_t(\mathbf{g})$, can be approximated by a member of the GEV family

$$G_{\mathbf{g},t}(y) \equiv \mathbb{P}(Y_t(\mathbf{g}) \leq y) = \exp\left\{-\left[1 + \xi_t(\mathbf{g})\left(\frac{y - \mu_t(\mathbf{g})}{\sigma_t(\mathbf{g})}\right)\right]^{-1/\xi_t(\mathbf{g})}\right\}, \qquad (1)$$

defined for $\{y : 1 + \xi_t(\mathbf{g})(y - \mu_t(\mathbf{g}))/\sigma_t(\mathbf{g}) > 0\}$. The GEV family of distributions (1) is characterized by three statistical parameters: the location parameter $\mu_t(\mathbf{g}) \in \mathcal{R}$, which describes the center of the distribution; the scale parameter $\sigma_t(\mathbf{g}) > 0$, which describes the spread of the distribution; and the shape parameter $\xi_t(\mathbf{g}) \in \mathcal{R}$. The shape parameter $\xi_t(\mathbf{g})$ is the most important for determining the qualitative behavior of the distribution of daily rainfall at a given location. If $\xi_t(\mathbf{g}) < 0$, the distribution has a finite upper bound; if $\xi_t(\mathbf{g}) > 0$, the distribution has no upper limit; and if $\xi_t(\mathbf{g}) = 0$, the distribution is again unbounded and the CDF (1) is interpreted as the limit as the shape parameter $\xi_t(\mathbf{g})$ approaches zero (Coles 2001).

To account for changes over time in the distribution of extreme precipitation and temperature we use a simple trend model wherein the location parameter varies linearly in time, i.e., $\mu_t(\mathbf{g}) = \mu_0(\mathbf{g}) + \mu_1(\mathbf{g})t$, while the shape and scale parameters are constant over time, i.e., $\sigma_t(\mathbf{g}) \equiv \sigma(\mathbf{g})$ and $\xi_t(\mathbf{g}) \equiv \xi(\mathbf{g})$. Using just four parameters to model changes over time in Eq. 1 at each grid cell is a simplistic representation of the temporal evolution of extreme precipitation, since in reality the distribution may be shifting nonlinearly (e.g., due to nonlinear changes in external forcing) with possibly nonconstant variability. However, Risser et al. (2019b) explored both (a) including nonlinear trends in the location parameter and (b) allowing the shape and scale parameters to change over time, and found that the simpler model described by a linear trend in location with constant scale and shape performed as well (in a statistical sense) as either of these alternative characterizations. Hence, we argue that this statistical model is an appropriate way to characterize changes in extreme precipitation over CONUS. We use the climextRemes package for R (Paciorek 2016) to obtain maximum likelihood estimates (MLEs) of the GEV parameters $\{\widehat{\mu}_0(\mathbf{g}), \widehat{\mu}_1(\mathbf{g}), \widehat{\sigma}(\mathbf{g}), \widehat{\xi}(\mathbf{g})\}$, independently for each grid cell of interest and quantify uncertainty via the block bootstrap (see, e.g., Risser et al. 2019c). The MLEs and the bootstrap estimates can be used to derive estimates of the seasonal climatological $r$-year return value in year $t$, denoted $\widehat{\phi}_{t,r}(\mathbf{g})$, which can be calculated as:

$$\widehat{\phi}_{t,r}(\mathbf{g}) = \begin{cases} \widehat{\mu}_0(\mathbf{g}) + \widehat{\mu}_1(\mathbf{g})t - \frac{\widehat{\sigma}(\mathbf{g})}{\widehat{\xi}(\mathbf{g})}\left[1 - \{-\log(1 - 1/r)\}^{-\widehat{\xi}(\mathbf{g})}\right], & \widehat{\xi}(\mathbf{g}) \neq 0 \\ \widehat{\mu}_0(\mathbf{g}) + \widehat{\mu}_1(\mathbf{g})t - \widehat{\sigma}(\mathbf{g})\log\{-\log(1 - 1/r)\}, & \widehat{\xi}(\mathbf{g}) = 0 \end{cases} \qquad (2)$$

Coles (2001). In summary, after conducting the above analysis, we can calculate MLEs of the climatological Rx1Day $r$-year return value in year $t$ $\widehat{\phi}_{t,r}(\mathbf{g})$ and block bootstrap estimates $\{\widehat{\phi}_{t,r,b}(\mathbf{g}) : b = 1, \ldots, 250\}$ in each season, and for each of the LOCA-downscaled models. Again, the R19 data product yields these return value MLEs and bootstrap estimates directly. The bootstrap estimates are used to calculate a 95% confidence interval using the "basic bootstrap" method (see, e.g., Paciorek et al. 2018).

In order to compare the LOCA solutions versus R19, in addition to the return values we consider three quantities for each geospatial location: (1) the relative difference in return value (LOCA minus R19, divided by R19); (2) the 65-year trend (1951–2005) in return values, as determined by $\widehat{\mu}_1(\mathbf{g})$; and (3) the linear trend in the relative difference between LOCA and R19 as a function of the return period $r$. For each of these quantities, at a specific

location we first tally the number of LOCA solutions that differ from R19 (as determined by whether the 95% confidence intervals overlap) and, if there appear to be meaningful differences between the LOCA solutions and R19, we state differences between the LOCA average and R19.

# 3 Results

## 3.1 Return value comparison

To assess the performance of statistical downscaling solutions for location-specific analysis of extreme precipitation, we use, as a case study, the subset of United States Department of Defense (DoD) installations within the CONUS that serve as foci for environmental research supported by the Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP). We specifically focus on a total of 163 installations that are affiliated with the Air Force, Army, Coast Guard, Marine Corps, National Guard, or the Navy, and have a unique ZIP code; the geographic distribution of these sites is shown in Fig. 1. These specific locations are chosen for our analysis because practitioners who are assessing the risks at each site from climate-changed temperature and precipitation patterns would benefit from high-confidence, localized projections of statistical distributions of these quantities (Moss 2017).

As an illustration of our methodology, we first choose seven facilities that represent a variety of climatological regions in the CONUS; again see Fig. 1. For each facility, we compare estimates of wintertime (DJF) return values for daily maximum precipitation (Rx1Day). Return values are shown for a single year (2005); however, note that this represents an estimated return value based on the linear trend in Eq. 2 and can be considered a temporally smoothed or "climatological" quantity. Recall that the $r$-year return value summarizes the event (precipitation threshold) that is expected to be exceeded once every $r$ years (on average), or, equivalently the event that has a probability of being exceeded in any year
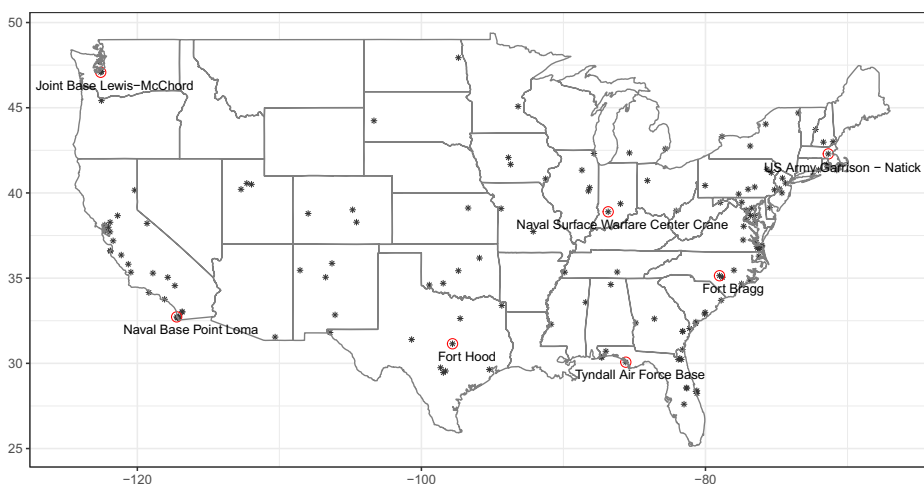


**Fig. 1** Geographic distribution of the 163 DoD facilities considered in this study (geocoded based on their ZIP code) with the seven facilities used for the case study in Section 3

of 1/r. Figure 2(a) shows the return values and 95% confidence band for R19 and the average of the 31 LOCA-downscaled CMIP5 models. For reference, the corresponding return values from L15 are also shown, as are return values estimated from the GHCN station closest to each facility (the "nearest-neighbor" GHCN station). Across the board, the LOCA average is significantly lower than R19; also note that LOCA generally agrees quite well with L15 and the R19 estimates compare favorably with the nearest-neighbor GHCN estimates. To specifically quantify the differences between LOCA and R19, Fig. 2(b) shows the relative difference between the LOCA average and R19: these differences are all negative (again highlighting the negative, or as we will use subsequently, "low" bias) and range between
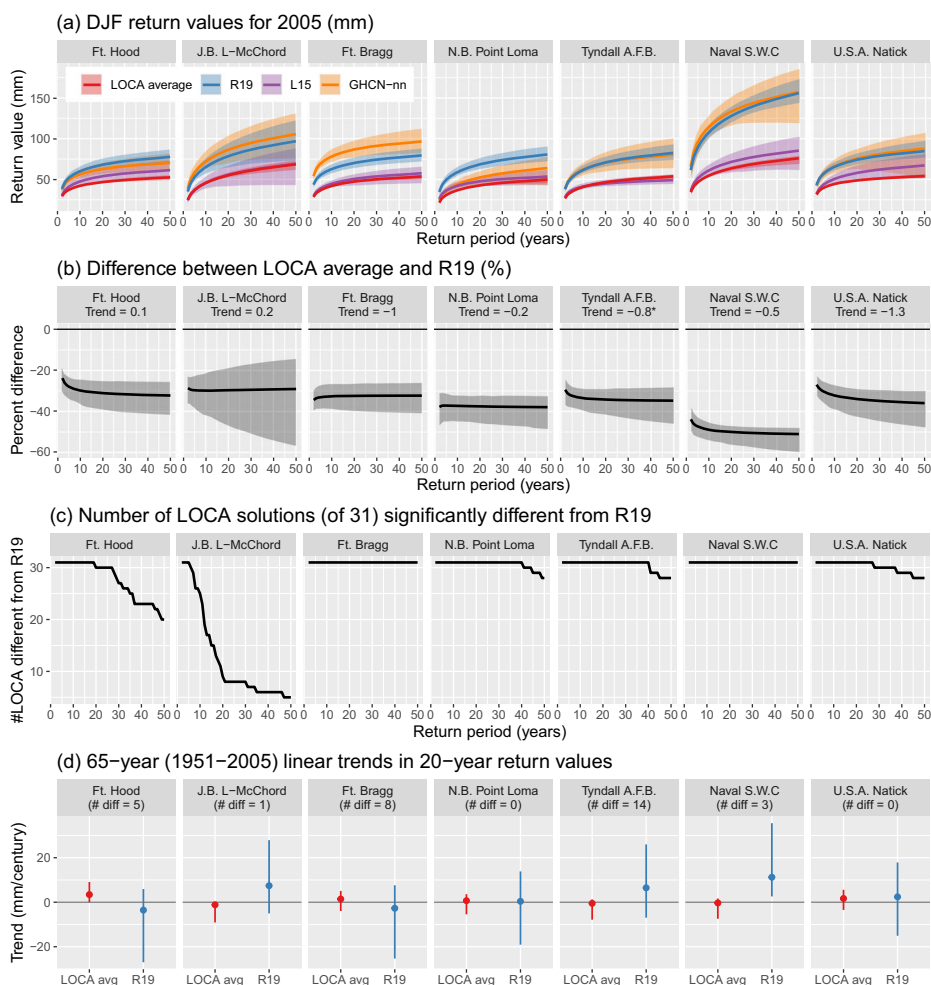


**Fig. 2** Best estimates and 95% basic bootstrap confidence intervals of the $r$-year return values of Rx1Day (panel **a**) and percent differences (panel **b**) for each of seven example facilities, comparing the Risser et al. (2019c) product (R19) and the average of the 31 LOCA-downscaled CMIP5 models. Panel **c** shows the number of individual LOCA solutions (31 total) that differ significantly from R19 for each return period. Panel **a** also shows return value estimates from L15 and the GHCN station that is closest to each facility (the nearest-neighbor GHCN station, labeled "GHCN-nn"): note that LOCA generally agrees with L15 and R19 agrees quite well with the GHCN-nn estimates

−30% and −50%, with the 95% confidence band showing that the differences are significant. This remains the case when considering the LOCA outputs derived from individual models: Fig. 2(c) shows the number of individual LOCA solutions that are significantly different from R19. For less rare events (up to the 20-year event), every LOCA solution differs from R19 except at Fort Hood. For more extreme events, the uncertainty is large enough that fewer of the LOCA solutions differ from R19, but a majority still show significant differences. In spite of the large differences between R19 and the LOCA average, the percent differences are relatively consistent across return period $r$. In other words, the relative errors (percent) in semi-rare events, e.g., the 1-in-5 year event, are approximately the same (after accounting for uncertainty) as much rarer events, e.g., the 1-in-50 year event. The strip text in Fig. 2(b) shows the trend in the percent difference as a function of return period $r$ (percent per 10-year increase in return period); only Tyndall Air Force Base has a statistically significant trend in the difference, although the trend is very small (less than 1% per 10-year increase in return period).

Having explored how the various data products compare in terms of the climatology of extreme precipitation (i.e., return values), we next move on to assess how the products compare in terms of how they quantify changes in extremes over time, specifically with respect to linear trends over 1951–2005. Recall that our GEV analysis quantifies changes over time in terms of a linear trend in the location parameter. Focusing on 20-year return values, we convert this trend into mm/century and again compute a 95% basic bootstrap confidence interval. Figure 2(d) compares the trend estimates at each facility for the LOCA average and R19: the uncertainty in the trends themselves are generally quite large, such that in most cases the trends are non-significant across data source; this agrees with significance statements regarding observed trends in GHCN return values from Risser et al. (2019b). Consequently, the differences in 1951–2005 trends across the LOCA average and R19 are generally non-significant, in that the confidence intervals overlap. This is again true for both the LOCA average as well as the individual models, where at most eight of the solutions have significantly different trends from R19 (with Tyndall Air Force Base being the lone exceptions; see the strip text of Fig. 2(d)).

Taking a step back, we next explore how these considerations play out across the other facilities shown in Fig. 1 and in all seasons. First, across all locations and seasons, the percent differences in return values between the LOCA average and R19 are approximately constant across event rarity: over 70% of locations have non-significant trends in the percent difference as a function of return period, and the other 30% (which are technically significant) only change by ≈ −2% per 10-year increase in return period. As a result, we again focus on the 20-year return value. Figure 3(a) shows the number of LOCA solutions with significantly different 20-year return values, where it is clear that across locations and seasons almost all of the solutions (approximately 25–27) are significantly different than R19. The magnitude of this difference is summarized in Fig. 3(b), which shows the percent difference between R19 and the LOCA average with black circles to identify locations for which the difference is significant. As with the case study locations, these differences are significant nearly everywhere, ranging between − 25% and −50%. Finally, Fig. 3(c) shows the number of LOCA solutions for which the 65-year (1951–2005) linear trend differs from R19. Again, as we saw in the case study, very few solutions have trends that differ significantly from R19, on average four across each season.

In summary, using these facilities as a case study, in all seasons the LOCA solutions have significant low biases in daily extreme precipitation return value estimates relative to R19, although the low biases are roughly consistent across return period. Estimates of trends
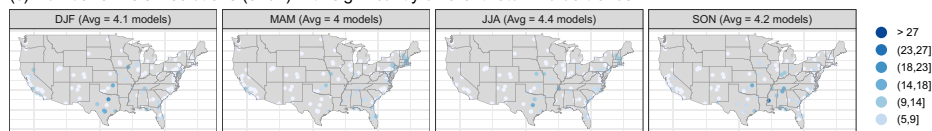
(a) Number of LOCA solutions (of 31) with significantly different 20−year return values
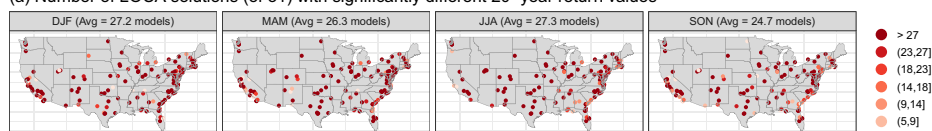


(b) Percent difference in 20−year return values, LOCA average vs. R19
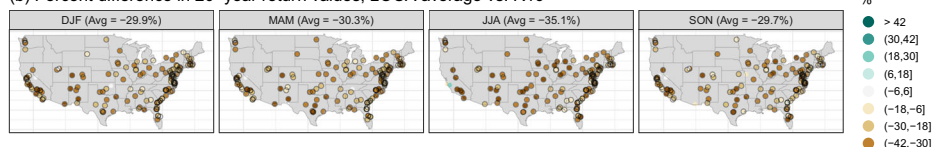


(c) Number of LOCA solutions (of 31) with significantly different return value trends



**Fig. 3** The number of LOCA solutions with significantly different 20-year return value estimates versus R19 (panel **a**; the average number of models is shown in the strip text), as well as the percent difference between the LOCA-average 20-year return value and the corresponding quantity from R19 (panel **b**). Locations with a significant difference between R19 and the LOCA average are plotted with a black circle. Panel **c** shows the number of LOCA solutions for which the 65-year (1951–2005) linear trend in 20-year return value differs from the R19 trend (with the seasonal average number of solutions shown in the strip text)

in extremes from both R19 and LOCA have very large uncertainties and do not appear to significantly differ from one another.

## 3.2 Reasons for large discrepancies in LOCA return values

In Section 3.1, we found that daily precipitation return value estimates from the LOCA-downscaled models at the locations of interest are approximately 25% to 50% too small relative to R19. As shown in Pierce et al. (2021), similar biases exist for the L15 training data relative to R19. Moreover, the LOCA solutions match the training data with respect to return value estimates and trends in extremes (see Section 1 of the Supplemental Materials). This latter finding is not surprising, since the L15 product was used to train the LOCA downscaling procedure and that procedure is dominated by values from the observations set. However, the differences between, specifically, L15 and R19 are concerning since the low biases in estimates of extreme precipitation from L15 relative to R19 have been propagated through to the downscaled solutions for return periods if not for space- and time-averaged distributions of extremes. The natural follow-up question is: why? We now show that there are two primary reasons for these differences: (1) a time-of-day correction in L13/L15, and (2) the application of gridding to fields of daily precipitation. We now examine each of these reasons in turn. In light of the fact that horizontal resolution is not the dominant driver of the L15 low bias in daily precipitation extremes (see Section 2 of the Supplemental Materials), in order to compare L15 and R19 in we regrid the L15 return values to the 1/4° R19 grid via a nearest-neighbor scheme that averages the four L15 nearest-neighbors to each R19 centroid.

**Time-of-day correction in L15** As described in Pierce et al. (2021), the weather station data used to create L15 are adjusted in order to "standardize" the daily data to all have the same measurement time. This adjustment is conducted prior to gridding, and is done as

follows: define $P^h(t)$ to be a measurement of precipitation representing one day $t$, taken at hour $h = 0, \ldots, 23$. In L15, these values are time-adjusted such that all stations have a measurement time of midnight ($h = 23$) as follows:

$$P^{23}(t) = P^h(t)\Delta + P^h(t+1)(1-\Delta),$$

where $\Delta = (h+1)/24$ (for further details, see Pierce et al. 2021). Pierce et al. (2021) show that the most common gauge observation times are between 6 and 8 AM local time; applying the time adjustment means that for such measurements, the precipitation for day $t$ is mostly added to the measurement on day $t + 1$. More importantly, Pierce et al. (2021) show that the time adjustment systematically reduces seasonal maxima (except in very rare situations), while also increasing the fraction of wet days and decreasing the average precipitation on wet days (these latter two issues are less relevant for the study at hand but are nonetheless important). Pierce et al. (2021) show that, averaged over CONUS, L15 underestimates 20-year return values by approximately 26% in DJF and 32% in JJA (similar to what we find in this paper). In light of this known feature of the L15 data processing algorithm, Pierce et al. (2021) develop a new version of the L15 data that does not include the time adjustment, henceforth "L15-unadj". Applying the same GEV analysis to L15-unadj significantly reduces the low biases in 20-year return values, to just 4% in DJF and 12% in JJA.

However, taking a closer look at Figure 5 in Pierce et al. (2021), it appears as though both of these CONUS-average low biases (for L15 and L15-unadj) are likely larger than they should be due to the inclusion of areas of complex topography in the western United States. In these areas, the L15 and L15-unadj return values are actually significantly larger than R19 (by as much as +30%). This is almost certainly due to the different ways each data set accounts for elevation: in R19, a simple linear elevation scaling is applied to the GEV parameter fields, while (as previously mentioned) L15 and L15-unadj use a PRISM-based adjustment (Daly et al. 1994) that incorporates other topographic variables. In light of these important differences and the difficulties of modeling precipitation in areas of complex topography (which are beyond the scope of this paper), we limit our comparison to CONUS grid cells with an elevation of less than 1500 m above sea level. We note that these problems may be influenced more by elevation gradients as opposed to absolute elevation, but using an absolute threshold is a simple way to navigate the challenges of modeling precipitation in places with complex topography.

**Gridding of daily data**   As was previously mentioned, a number of recent papers have documented how gridded daily products of precipitation (such as L15) are an inappropriate data source for characterizing extremes locally since daily precipitation exhibits fractal scaling and any spatial averaging will reduce extremes (Lovejoy et al. 2008; Maskey et al. 2016). To explicitly evaluate the magnitude of this reduction in extremes, we take the same input data as R19 (daily measurements of precipitation from 5,202 GHCN stations over 1950–2017) and apply bilinear interpolation (like the SYMAP algorithm used by L15) to convert the spatially irregular measurements to a regular 0.25° grid—the same grid as the one used in R19. Then, at each grid cell, we extract the seasonal maximum daily precipitation in each year and fit the GEV distribution defined in (1) with a linear trend in the location parameter; we then calculate the maximum likelihood estimate of the 20-year return value at each grid cell. While this is an admittedly simple interpolation scheme that does not make adjustments for complex topography, it allows us to isolate the specific influence of applying a standard gridding scheme to a fractal field. And, as mentioned in the previous paragraph, the facility locations under consideration in this paper are not in areas of complex topography.

Note that this procedure can be seen as reversing the order of operations in R19: in R19, one first calculates extreme statistics from the station data and then applies gridding, while here (as with L15) we instead first apply gridding to the daily station data and then calculate extreme statistics from the gridded result. As such, we denote this approach as "GHCN-grid-then-fit" since it uses GHCN station data as input but grids the daily data before fitting an extreme value distribution. And, importantly, this yields an apples-to-apples comparison since it uses the same input data as R19 and simply reverses the order of operations.

**Decomposition of low extreme daily precipitation bias in L15** Considering all 0.25° grid cells in CONUS with an elevation of less than 1500m (to avoid the influence of the PRISM adjustment in L15 and L15-unadj), we calculate the percent biases for L15, L15-unadj, and GHCN-grid-then-fit versus the R19 ground truth return values. Using these biases, we can explicitly quantify the contribution of the time-of-day (TOD) adjustment and the gridding of daily data to the total bias in L15. First, the contribution of the TOD adjustment can be calculated by taking the percent bias from L15 minus the percent bias in L15-unadj divided by the L15 bias, i.e.,

$$\text{Fraction of bias due to TOD adjustment} = \frac{\text{Bias(L15)} - \text{Bias(L15-unadj)}}{\text{Bias(L15)}},$$

where Bias($\cdot$) is the average percent difference in 20-year return values of the indicated data set relative to R19. Next, the contribution of gridding daily data can be calculated by directly comparing the two biases

$$\text{Fraction of bias due to gridding daily data} = \frac{\text{Bias(GHCN-grid-then-fit)}}{\text{Bias(L15)}},$$

since the primary difference between the GHCN-grid-then-fit and R19 data sets is the order of operations. The total percent bias in 20-year return values for L15 relative to R19 averaged over all 0.25° CONUS grid cells with an elevation less than 1500 m is shown in Table 1. Now that we have removed grid cells with complex topography (albeit rather crudely), we can see that the overall bias in L15 is slightly worse than was found in Pierce et al. (2021): the L15 return values are approximately 30% too small for DJF, MAM, and SON, and >35% too small in JJA. The largest fraction of this bias is due to the time-of-day adjustment in L15, which accounts for 53% of the bias in JJA and up to 74% of the bias in DJF. The gridding of daily data has a lesser but nonetheless significant influence on the return values, accounting for about 35% of the bias in DJF to 50% of the bias in JJA. Clearly, the two rightmost columns of Table 1 do not add up to 1, which indicates that there are some

**Table 1** Total percent bias in 20-year return values for L15 relative to R19 (second column from the left) averaged over all 0.25° CONUS grid cells with an elevation less than 1500 m

| Season | Total bias (L15) | Frac. of bias due to TOD adj. | Frac. of bias due to gridding daily data |
|---|---|---|---|
| DJF | −29.4% | 0.74 | 0.37 |
| MAM | −30.9% | 0.63 | 0.43 |
| JJA | −435.7% | 0.53 | 0.51 |
| SON | −30.4% | 0.63 | 0.41 |

The fraction of the bias due to both the time-of-day (TOD) adjustment and the gridding of daily data are also shown in the two rightmost columns

overlapping contributions to the bias from the TOD adjustment and gridding of daily data. Nonetheless, the fact that the sum of these two fractional contributions is close to 1 (actually 1.04 to 1.11) provides confidence that we are partitioning the bias into these two causes approximately correctly, especially since the return values from the different data sources (L15, L15-unadj, and GHCN-grid-then-fit) were calculated independently of one another.

## 3.3 Scaling factors for adjusting LOCA biases

When a user is interested in using the LOCA solutions for estimating return values, the results presented in this paper can be used to "bias correct" LOCA return values by way of a simple scaling factor. Given that 1951–2005 linear trends in return values for LOCA are generally indistinguishable from corresponding trends in R19 and differences are consistent across return period (see Section 3.1), a time- and return period-invariant scaling factor can be derived by calculating the ratio between L15 and R19 return values as

$$\text{sf}_{\mathbf{g}} = \frac{\widehat{\phi}^{\text{R19}}_{2005,20}(\mathbf{g})}{\widehat{\phi}^{\text{L15}}_{2005,20}(\mathbf{g})} \tag{3}$$

(of course, the scaling factors depend on season as well), where $\mathbf{g}$ corresponds to the $1/16°$ grid used in L15 and LOCA and we arbitrarily choose the year 2005 and the 20-year return period. The idea here is that if one estimates grid cell $r$-year return values for the LOCA solutions in exactly the same way as described in Section 2.2, one could multiply the return value by $\text{sf}_{\mathbf{g}}$ to yield return values that have the same climatology as R19. Furthermore, the scaling factors are based on L15 since biases present in the predictant of a statistical downscaling solution will also be present in the output (Maraun et al. 2010). Such a correction provides a path forward for users who require the daily data and range of solutions provided by the LOCA-downscaled CMIP5 GCMs but still wish to estimate pointwise return values. The seasonal scaling factors based on the percent differences are provided as a publicly available data product; see Risser et al. (2021).

As a simple example of how this correction works, we apply the scaling factor to the wintertime LOCA return values at two sample sites, Tyndall Air Force Base and Fort Hood (see Fig. 1). Recall from Fig. 2 that the return values from LOCA are too small relative to R19 for both sites, but much more so at Tyndall Air Force Base. The scaling factors for Tyndall and Fort Hood are 1.67 and 1.33, reflecting the larger (low) bias at Tyndall. The adjusted return values for these two sites are shown in Fig. 4, where we also show estimates from R19 and L15 for reference. The plotted lines for LOCA and LOCA-adjusted represent the multi-model mean and the confidence band shows the multi-model spread. At both sites, the adjusted return value curves agree much more closely with the R19 estimates, with the uncertainty bands overlapping in almost all cases. Our main point is that the adjusted LOCA return values could be used by practitioners. Given the well-documented difficulties in bias correcting downscaling projections of daily precipitation (Turco et al. 2017), these scaling factors could furthermore be applied to projected downscaling solutions, allowing practitioners to assess both relative and absolute changes in extreme precipitation return values under future emissions scenarios. However, an assessment of future projections is beyond the scope of this current work.
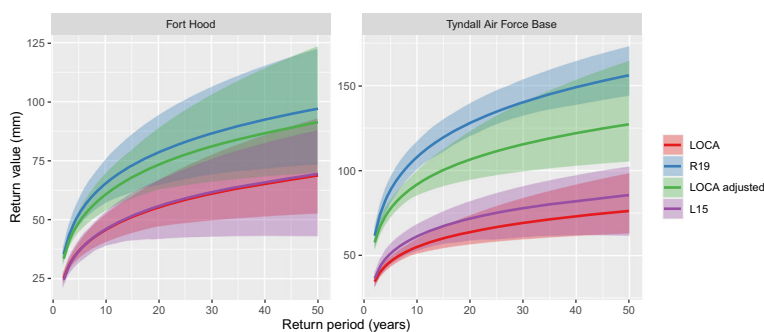
**Fig. 4** Return values for R19 and the LOCA solutions, before after applying the location-specific scaling factor to LOCA at the Tyndall Air Force Base (scaling factor = 1.67) and Fort Hood (scaling factor = 1.33) grid cells. The L15 return values are also shown for reference (recall that the scaling factors are calculated as R19/L15). For R19 and L15, the 95% basic bootstrap confidence band is shown; for the LOCA and LOCA-adjusted, the band represents the multi-model spread (also, the plotted line is the multi-model mean)

## 4 Discussion

Our findings in this paper show how construction of the training data set can influence key aspects of statistically downscaled solutions, and illustrate the utility of evaluating extreme statistics by using a product specifically designed for that purpose via extreme value analysis. These results echo downscaling intercomparison efforts that pointed to the central importance of training data (Gutmann et al. 2012, 2014). Other training data sets that do not incorporate the time-of-day correction may have smaller biases, but biases due to the gridding of daily data will persist in the absence of more sophisticated gridding techniques that do not over-smooth extreme values and variability.

To demonstrate the utility of our findings regarding representations of observed precipitation extremes, we provide a data product consisting of scaling factors that effectively correct the lower biases that we identify in LOCA return value estimates relative to a dedicated precipitation extremes data set. In this way we intend to support practitioners and other end-users of LOCA solutions if and as they use these data sets for location-specific analysis of extreme daily precipitation (Risser et al. 2021). These correction factors can also be applied to other downscaled products that are trained on L15, after accounting for the corresponding grid differences, and the findings can potentially serve as a guide to develop correction factors for other statistically downscaled products where training data sets are shown to exhibit biases in extreme behavior. We recognize that many downscaling approaches employ a range of bias-correction techniques out of necessity (Christensen et al. 2008; Mearns et al. 2012; Sillmann et al. 2013), though not without controversy (Ehret et al. 2012). The results presented indicate that an additional bias-correction for extreme precipitation is needed for historically downscaled products. We did not explore whether such a bias-correction would be needed for extremal values in future projections, but at least where statistical methods are used for downscaling future projections, such extremal values will need to be adjusted, and the approach and data set presented here can provide a starting point to inform such an analysis. We note that even though the difference in trends between LOCA and R19 is not statistically significant (at the 95% confidence level), the difference in percent changes in the future may still be different.

While the focus of this particular analysis is on L15 and LOCA, any statistical down-scaling method trained with L15 or using a similar time-of-observation correction scheme without correcting for this bias would be expected to show similar results. How observational training data sets are constructed can affect particular aspects of a downscaling solution, and using the alternative method of direct extreme value analysis of the variables of interest can add valuable information for local resilience planning. We specifically note that these low biases in daily extremes for L15 do not result in low (or dry) biases at longer timescales (e.g., multi-day or monthly). As described below, extreme daily precipitation values are diluted by being distributed across space (via gridding) and time (by a time-of-observation adjustment). Multi-day, seasonal, and annual precipitation totals are preserved by this "smearing" even as daily extremes are reduced.

The low bias in precipitation return values in commonly used gridded products of the historical record may be problematic for decisions regarding facility resource management. One common approach to using downscaled solutions, which recognizes the challenges with using the absolute values of those solutions, is to look at the relative change between the simulated historical record and a future projection in a given quantile (e.g., Cannon et al. 2015). We show that making decisions based on relative change in extreme precipitation at a given location will be problematic, since the relative change in extreme precipitation will propagate the low bias in extreme precipitation that we have shown can exist in downscaling solutions from the historical record.

Whenever multiple techniques can be considered for a common problem (here, an observational data set for assessing daily precipitation) a trade-off is typically involved. For example, the special-purpose precipitation extremes product R19 does not provide daily time series of precipitation and so cannot be used as a substitute for daily precipitation fields in all applications including ones to train statistical downscaling methods. Conversely, the standard daily gridded product and others like it cannot capture extremes as well as the dedicated product based on extreme value analysis. Depending on the application, using one or the other product (daily gridded or seasonal extreme measures) may be preferable, but we show that using aspects of both can produce more useful results than using either singly. We note that although we do not pursue the question in this study, we consider that gridded daily precipitation data products are appropriate for conducting GCM model evaluation (see, e.g., Chen and Knutson 2008; Gervais et al. 2014; Risser and Wehner 2020; Wehner et al. 2021); on the other hand, products such as Risser et al. (2019c, a) are the appropriate for considering pointwise applications, and are therefore used as the standard of comparison in this paper.

More broadly, this paper explores a thorny issue for downscaling: products that are developed for grid boxes at 6 km, which are very small relative to the parent model, may not be the most informative for analyses of extreme precipitation at other spatial scales. To be clear, this problem is based on the technique used to generate the training data set, and not due to the size of the downscaled grid box. Fundamentally, daily precipitation is a fractal field, and the gridding of that field produces artifacts. While the biases we identified here may decrease with the decreasing size of the grid boxes in the application domain, users of gridded observational data should exercise caution when using gridded data sets for analyzing point extremes of these environmental variables. Our results show the additional value that can be gained by considering results from a targeted extreme value analysis in addition to using statistics calculated from a daily gridded product. With respect to extreme precipitation, a correction is required to conventionally gridded products in order to incorporate unbiased estimates of extreme precipitation into location-specific analyses. While a correction factor may address gridding biases, the findings here show how attention must

be paid to the corruption of extrema information from gridding observational data. The bias induced from gridding in the manner explored here can be especially troublesome where changes in extreme precipitation need to be estimated. Baseline analyses of models and observational products over the historical period form a foundation for analyzing extreme precipitation in future climate. Regardless of whether location-specific analyses use future projections with statistical, dynamical, or hybrid methods to incorporate nonstationarity into their downscaled solutions, those analyses need to begin with information from historical data that is unbiased from gridding methods.

**Data availability** The LOCA solutions used in this paper are publicly available at ftp://gdo-dcp.ucllnl.org/pub/dcp/archive/cmip5/loca; the Livneh et al. (2015b) data are publicly available at ftp://ftp.nodc.noaa.gov/nodc/archive/arc0077/0129374/1.1/data/0-data/daily/. The Risser et al. (2019c) data product is publicly available at https://doi.org/10.7910/DVN/LULNUQ.

**Code availability** Not applicable

## Declarations

**Ethics approval** This report did not study humans and/or animals subjects.

**Consent to participate** This report did not study humans and/or animals subjects.

**Consent for publication** This report did not study humans and/or animals subjects.

**Conflict of interest** The authors declare no competing interests.

## References

Abatzoglou JT, Brown TJ (2012) A comparison of statistical downscaling methods suited for wildfire applications. Int J Climatol 32(5):772–780

Ayar PV, Blanchet J, Paquet E, Penot D (2020) Space-time simulation of precipitation based on weather pattern sub-sampling and meta-gaussian model. J Hydrol 581:124451

Barsugli JJ, Guentchev G, Horton RM, Wood A, Mearns LO, Liang XZ, Winkler JA, Dixon K, Hayhoe K, Rood RB et al (2013) The practitioner's dilemma: How to assess the credibility of downscaled climate projections. Eos, Trans Am Geophys Union 94(46):424–425

Brown C, Ghile Y, Laverty M, Li K (2012) Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resour Res 48(9)

Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? J Climate 28(17):6938–6959

Chen CT, Knutson T (2008) On the verification and comparison of extreme rainfall indices from climate models. J Clim 21(7):1605–1621

Chen J, Zhang XJ (2021) Challenges and potential solutions in statistical downscaling of precipitation. Clim Change 165(3):1–19

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys Res Lett 35(20)

Coles S (2001) An introduction to statistical modeling of extreme values. Lecture Notes in Control and Information Sciences, Springer, Berlin. https://books.google.com/books?id=2nugUEaKqFEC

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33(2):140–158

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. Int J Clim 28(15):2031–2064

Ehret U, Zehe E, Wulfmeyer V, Warrach-Sagi K, Liebert J (2012) Hess opinions should we apply bias correction to global and regional climate model data? Hydrol Earth Syst Sci Discuss 9(4)

Engle NL (2011) Adaptive capacity and its assessment. Global Environ Change 21(2):647–656

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9(5):1937–1958

Feldman DR, Tadić JM, Arnold W, Schwarz A (2021) Establishing a range of extreme precipitation estimates in california for planning in the face of climate change. J Water Resour Plan Manag 147(9):04021056

Gervais M, Tremblay LB, Gyakum JR, Atallah E (2014) Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods. J Clim 27(14):5201–5218

Giorgi F, Gutowski WJ Jr (2015) Regional dynamical downscaling and the CORDEX initiative. Annu Rev Environ Resour 40:467–490

Gutmann E, Pruitt T, Clark MP, Brekke L, Arnold JR, Raff DA, Rasmussen RM (2014) An intercomparison of statistical downscaling methods used for water resource assessments in the United States. Water Resour Res 50(9):7167–7186

Gutmann E, Barstad I, Clark M, Arnold J, Rasmussen R (2016) The intermediate complexity atmospheric research model (ICAR). J Hydrometeorol 17(3):957–973

Gutmann ED, Rasmussen RM, Liu C, Ikeda K, Gochis DJ, Clark MP, Dudhia J, Thompson G (2012) A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. J Clim 25(1):262–281

Hall A, Walton D, Berg N, Rahimi S, Grieco M, Lin YH, Cayan D, Pierce D, Maurer E (2020) Developing metrics to evaluate the skill and credibility of downscaling. Tech. rep., Univ of California. Los Angeles, CA (United States)

Hallegatte S (2009) Strategies to adapt to an uncertain climate change. Glob Environ Change 19(2):240–247

King AD, Alexander LV, Donat MG (2013) The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. Int J Climatol 33(10):2376–2387

Klemes V et al (1982) Empirical and causal models in hydrology. Scientific basis of water resource management, pp 95–104

Livneh B, Rosenberg EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP (2013) A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. J Climate 26(23):9384–9392

Livneh B, Bohn TJ, Pierce DW, Munoz-Arriola F, Nijssen B, Vose R, Cayan DR, Brekke L (2015a) A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. Sci Data 2(1):1–12

Livneh B, Bohn TJ, Pierce DW, Munoz-Arriola F, Nijssen B, Vose R, Cayan DR, Brekke L (2015b) A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada

(NCEI Accession 0129374). NOAA National Centers for Environmental Information Dataset (Daily precipitation) , accessed April 13, 2020. https://doi.org/https://doi.org/10.7289/v5x34vf6

Lovejoy S, Schertzer D, Allaire V (2008) The remarkable wide range spatial scaling of TRMM precipitation. Atmos Res 90(1):10–32. https://doi.org/10.1016/j.atmosres.2008.02.016. http://linkinghub.elsevier.com/retrieve/pii/S0169809508000562

Maraun D, Wetterhall F, Ireson A, Chandler R, Kendon E, Widmann M, Brienen S, Rust H, Sauter T, Themeßl M et al (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48(3)

Maskey ML, Puente CE, Sivakumar B, Cortis A (2016) Encoding daily rainfall records via adaptations of the fractal multifractal method. Stoch Env Res Risk A 30(7):1917–1931. https://doi.org/10.1007/s00477-015-1201-7

Mearns LO, Arritt R, Biner S, Bukovsky MS, McGinnis S, Sain S, Caya D, Correia JJr, Flory D, Gutowski W et al (2012) The North American regional climate change assessment program: overview of phase I results. Bull Am Meteorol Soc 93(9):1337–1362

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012) An overview of the Global Historical Climatology Network-Daily database. J Atmos Oceanic Tech 29(7):897–910

Milly P, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: Whither water management? Earth 4:20

Molter EM, Collins WD, Risser MD (2021) Quantitative precipitation estimation of extremes in conus with radar data, vol 48

Moss RH (2017) Nonstationary Weather Patterns and Extreme Events Informing Design and Planning for Long-Lived Infrastructure. Tech. rep., Workshop Report on Nonstationarity, ESTCP Project RC-201591. https://www.serdp-estcp.org/News-and-Events/Conferences-Workshops/Past-RC-Workshops/Nonstationary-Weather-Patterns-and-Extreme-Events-2017

Newman AJ, Clark MP, Craig J, Nijssen B, Wood A, Gutmann E, Mizukami N, Brekke L, Arnold JR (2015) Gridded ensemble precipitation and temperature estimates for the contiguous United States. J Hydrometeorol 16(6):2481–2500

Ntegeka V, Baguis P, Roulin E, Willems P (2014) Developing tailored climate change scenarios for hydrological impact assessments. J Hydrol 508:307–321

Paciorek C (2016) ClimextRemes: Tools for Analyzing Climate Extremes. https://CRAN.R-project.org/package=climextRemes, r package version 0.1.2

Paciorek C, Stone D, Wehner M (2018) Quantifying statistical uncertainty in the attribution of human influence on severe weather. Weather Clim Extremes 20:69–80. https://doi.org/10.1016/j.wace.2018.01.002

Pierce DW, Cayan DR, Thrasher BL (2014) Statistical downscaling using localized constructed analogs (LOCA). J Hydrometeorol 15(6):2558–2585

Pierce DW, Su L, Cayan DR, Risser MD, Livneh B, Lettenmaier DP (2021) An extreme-preserving long-term gridded daily precipitation data set for the conterminous United States. J Hydrometeorol

Risser M, Paciorek C, Wehner M, O'Brien T, Collins W (2019a) A probabilistic gridded product for daily precipitation extremes over the United States. Harvard Dataverse. https://doi.org/10.7910/DVN/LULNUQ

Risser M, Feldman D, Wehner M, Pierce D, Cayan D, Arnold J (2021) Scaling factors for improving resilience to extreme daily precipitation events for the L15 data product and LOCA solutions. Harvard Dataverse DOI pending. TBD

Risser MD, Wehner MF (2020) The effect of geographic sampling on evaluation of extreme precipitation in high-resolution climate models. Advances in Statistical Climatology. Meteorol Oceanogr 6(2):115–139

Risser MD, Paciorek CJ, O'Brien TA, Wehner MF, Collins WD (2019b) Detected changes in precipitation extremes at their native scales derived from in situ measurements. J Clim 32(23):8087–8109. https://doi.org/10.1175/JCLI-D-19-0077.1

Risser MD, Paciorek CJ, Wehner MF, O'Brien TA, Collins WD (2019c) A probabilistic gridded product for daily precipitation extremes over the United States. Clim Dyn 53(5):2517–2538. https://doi.org/10.1007/s00382-019-04636-0

Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference. pp 517–524

Shepard DS (1984) Computer mapping: The SYMAP interpolation algorithm. In: Spatial statistics and models. Springer, pp 133–145

Sillmann J, Kharin V, Zhang X, Zwiers F, Bronaugh D (2013) Climate extremes indices in the cmip5 multimodel ensemble: Part 1. model evaluation in the present climate. J Geophys Res Atmos 118(4):1716–1733

Stoner AM, Hayhoe K, Yang X, Wuebbles DJ (2013) An asynchronous regional regression model for statistical downscaling of daily climate variables. Int J Climatol 33(11):2473–2494

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485–498

Timmermans B, Wehner M, Cooley D, O'Brien T, Krishnan H (2019) An evaluation of the consistency of extremes in gridded precipitation data sets. Clim Dyn 52(11):6651–6670. https://doi.org/10.1007/s00382-018-4537-0

Turco M, Llasat MC, Herrera S, Gutiérrez JM (2017) Bias correction and downscaling of future rcm precipitation projections using a mos-analog technique. J Geophys Res Atmos 122(5):2631–2648

Vano JA, Arnold JR, Nijssen B, Clark MP, Wood AW, Gutmann ED, Addor N, Hamman J, Lehner F (2018) Dos and don'ts for using climate change information for water resource planning and management: guidelines for study design. Clim Serv 12:1–13

Vrac M, Stein M, Hayhoe K (2007) Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. Clim Res 34(3):169–184

Walton D, Berg N, Pierce D, Maurer E, Hall A, Lin YH, Rahimi S, Cayan D (2020) Understanding differences in california climate projections produced by dynamical and statistical downscaling. J Geophys Res Atmos 125(19):e2020JD032812

Wehner MF, Lee J, Risser MD, Ullrich P, Gleckler P, Collins WD (2021) Evaluation of extreme sub-daily precipitation in high-resolution global climate model simulations. Philos Trans R Soc :379. https://doi.org/10.1098/rsta.2019.0545

Wood AW, Leung LR, Sridhar V, Lettenmaier D (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim Change 62(1):189–216

Wuebbles DJ, Fahey DW, Hibbard KA (2017) Climate science special report: fourth national climate assessment, volume I

Yevjevich VM (1972) Structural analysis of hydrologic time series. PhD thesis, Colorado State University. Libraries

**Publisher's note**  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.