Evaluating the Credibility of Downscaling: Integrating Scale, Trend, Extreme, and Climate Event into a Diagnostic Framework

FENGYUN SUN

School of Ecological and Environmental Sciences, and Shanghai Key Laboratory for Urban Ecological Processes and Eco-Restoration, East China Normal University, and Institute of Eco-Chongming, Shanghai, China

ALFONSO MEJIA

Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, Pennsylvania

SANJIB SHARMA

Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania

PENG ZENG AND YUE CHE

School of Ecological and Environmental Sciences, and Shanghai Key Laboratory for Urban Ecological Processes and Eco-Restoration, East China Normal University, and Institute of Eco-Chongming, Shanghai, China

(Manuscript received 27 March 2020, in final form 1 August 2020)

ABSTRACT

Because downscaling tools are needed to support climate change mitigation and adaptation practices, the guarantee of their credibility is of vital importance. To evaluate downscaling results, one needs to select a set of effective and nonoverlapping indices that reflect key system attributes. However, this subject is still insufficiently researched. With this study, we propose a diagnostic framework that evaluates the credibility of precipitation downscaling using five different attributes: spatial, temporal, trend, extreme, and climate event. A daily variant of the bias-corrected spatial downscaling approach is used to downscale daily precipitation from the GFDL-ESM2G climate model at 148 stations in the Yangtze River basin in China. Results prove that this framework is effective in systematically evaluating the performance of downscaling across the Yangtze River basin in the context of climate change and exacerbating climate extremes. Moreover, results also indicate that the downscaling approach adopted in this study yields good performance in correcting spatio-temporal bias, preserving trends, approximating extremes, and characterizing climate events across the Yangtze River basin. The proposed framework can be beneficial to planners and engineers facing issues relevant to climate change assessment.

1. Introduction

Mitigation and response plans that tackle real issues related to economic, societal, and environmental systems require an understanding of future conditions, particularly about climate because its impacts can cascade through these different systems (Knutti and Sedláček 2013; Ockenden et al. 2017; Sanderson et al. 2018). Credible and detailed information about future climate change is needed to modify existing or planned infrastructure designs (Fletcher et al. 2019; Martinich

and Crimmins 2019; Reidmiller et al. 2018). General circulation models (GCMs) are the main state-of-the-science source for information about future climate change. However, the outputs from GCMs tend to be spatially coarse and biased for application at local scales (Benestad 2010; Di Luca et al. 2012; Li et al. 2010; Mehrotra and Sharma 2010; Piani et al. 2010; Shiru et al. 2019). Downscaling techniques are used to improve the spatial resolution and correct systematic biases in climate projection data (Ali et al. 2019; Gudmundsson et al. 2012).

Climate downscaling can be broadly classified into dynamical and statistical downscaling categories

Corresponding author: Yue Che, yche@des.ecnu.edu.cn

DOI: 10.1175/JAMC-D-20-0078.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

(Boé et al. 2007; Murphy 1999; Schmidli et al. 2006; Wood et al. 2004). On the one hand, dynamical downscaling nests high-resolution regional models to simulate finescale physical processes that consistent with the large-scale weather evolution in a GCM (von Storch et al. 2000; Wood et al. 2004). A variety of Weather Research and Forecasting (WRF) Model versions (Lo et al. 2008; Powers et al. 2017; Skamarock and Klemp 2008) have been proposed for dynamical downscaling. On the other hand, statistical downscaling encompasses statistics-based techniques to build a statistical relationship between large-scale climate patterns and local climate responses to downscale climate outputs (Fowler et al. 2007; Kreienkamp et al. 2018; Maraun 2016). A wide range of techniques have been proposed for statistically downscaling over the past few decades, change factor (Chen et al. 2011; Diaz-Nieto and Wilby 2005), quantile mapping (Cannon et al. 2015; Maraun 2013), multiple linear regression (Jeong et al. 2012; Sachindra et al. 2013), artificial neural network (Khan et al. 2006; Wilby et al. 1998), support vector machines (Anandhi et al. 2008; Tripathi et al. 2006), and self-organized maps (Sinha et al. 2018), just to name a few. The downscaling methods differ in theoretical assumptions, calculation algorithms and processing procedures (Mudersbach et al. 2017). Effective evaluation should be adopted to inform end users about the strengths and weaknesses of downscaling approaches, and ultimately to aid the process of choosing a suitable downscaling method for use.

Studies are often carried out to evaluate the performance of downscaling methods. For example, Schmidli et al. (2007) used six indices, including mean precipitation, wet-day frequency, wet-day intensity, 90th quantile of precipitation on wet days, maximum number of consecutive dry days, and maximum 1-day/5-day precipitation total. Hessami et al. (2008) used five indices, including percentage of wet days, mean precipitation amount at wet days, maximum number of consecutive dry days, maximum 3-day precipitation total, and 90th percentile of rain day amount. Kreienkamp et al. (2018) adopted indicators of mean yearly bias, daily histograms, seasonal cycle, 99th percentile, two-sample Kolmogorov–Smirnov test value, lag autocorrelations, wet-day frequencies, and spatial pattern correlations. Hertig et al. (2019) adopted indicators of skewness, relative frequency of days with precipitation ≥ 10 mm, 98th percentile of wet days, total amount above 98th percentile of wet days, 20-yr return value, 90, 95, and 99th quantiles, median of the annual dry/wet spell maxima, and median of the annual wet. Ali et al. (2019) adopted indices of maximum number of consecutive

precipitation days, annual total precipitation in wet days, annual maximum 1-day/5-day precipitation, annual total 95th precipitation, and annual precipitation divided by the number of wet days. It is clear that the current selection of indicators, which communicates information of different system attributes (e.g., mean, frequency, and extreme), is highly depends on researchers' experience and interests.

Indeed, choosing right indices to describe the downscaling performance is nontrivial. Since existing indicators are diversified and the indicative meaning of many indicators are overlapping, it is impossible and unnecessary to include all the indices in a diagnosis. However, some key attributes should be identified and made sure to be comprehensively covered by selected indices in a systematic diagnosis. Currently, only a few diagnostic frameworks have been proposed to evaluate the skill of downscaling methods. In the framework proposed by Hayhoe (2010), three attributes are identified to evaluate the skill of four downscaling methods in simulating daily temperature at 20 stations across North America. The three attributes are (i) temporal values; (ii) thresholds, extremes, and quantities; and (iii) persistence. The National Climate Predictions and Projections platform teams proposed the Standardized Quantitative Evaluation Framework Prototype for evaluation of fine-scaled climate projections (Barsugli et al. 2013). This framework includes three cross-cutting groups: (i) statistical distribution and temporal characteristics, (ii) impact applications, and (iii) climate processes and phenomena. Another framework, known as "VALUE" [a European Cooperation in Science and Technology (COST) "action" more fully called Validating and Integrating Downscaling Methods for Climate Change Research] is used to validate downscaling approaches for climate change studies (Maraun et al. 2015). Five attributes are included in VALUE, that is, (i) marginal distribution, (ii) temporal dependence, (iii) spatial dependence, (iv) multivariate dependence, and (v) spatial climatological summaries. The attributes included in these frameworks are rather diverse and no general rules are applied for dividing attributes in these frameworks. Therefore, questions arise: What are the essential attributes that need to be covered in a systematic diagnostic framework? How effective are these attributes in evaluating the credibility of climate downscaling? To the best of our knowledge, no further studies have been carried out to discuss this subject in depth.

The attributes of spatial, temporal, trend, extreme, and climate event are potentially informative aspects to be integrated into climate downscaling evaluation. The spatial and temporal credibility are basic aspects and are included in all the aforementioned frameworks, which

give an idea of whether the downscaling method shows chronological and geographical disparity (Cannon 2018). Trend and climate extremes should also be treated as an indispensable part of the downscaling evaluation. For the reason that it is important to preserve the projected climate change signal in the downscaling process if realistic assessments are sought (Cannon et al. 2015). If trends are affected, variations in the projected mean and extremes, as well as any related factors, are likely to be misrepresented (Maraun 2013). Moreover, extreme values are more prone to errors after downscaling than are mean parameters (Themeßl et al. 2012). Given the downscaled products are often used for climate change assessments, such as droughts (Rhee and Cho 2016; Wang et al. 2014) and floods (Das et al. 2013; Roth et al. 2012), properly handling climate extremes during the downscaling process is necessary. Furthermore, questions exist regarding the effects that a downscaling method has on the projected signal of climate events. Specifically, the downscaling method may or may not deteriorate the signal of climate events in climate projections when making various corrections. Climate events are referred as occurrence of a value of a climate variable above (or below) a threshold value near the upper (or lower) ends of the variable range (IPCC 2012). Adopting climate event for diagnosis purposes could benefit the understanding of the compound effects of downscaling (Bürger et al. 2012). Although these attributes, that is, spatial, temporal, trend, extreme, and climate event, are frequently investigated and highlighted by separate studies, comprehensive evaluations that cover these attributes simultaneously are very limited.

The primary aim of this study is twofold. First, demonstrate the utility of the diagnostic framework for evaluating precipitation downscaling performance in the Yangtze River basin in China. The raw GCMs are biased in delineating the heterogeneity of precipitation at local scales (Li et al. 2010; Piani et al. 2010; Shiru et al. 2019). Five attributes are included: spatial, temporal, trend, extreme, and climate event credibility. Second, illustrate the performance of a daily variant of the bias-corrected spatial downscaling (BCSD) approach on precipitation correction. With increasing efforts on mitigating climate change impacts worldwide (Chai et al. 2019; Hansen et al. 2019), we believe the proposed framework can be beneficial to planners and engineers facing issues relevant to climate change assessment.

2. Study area and datasets

The credibility of precipitation downscaling is investigated across the Yangtze River basin in this study. The Yangtze River basin accounts for approximately 19% of

the total land surface of China, encompassing diverse climates and physiography (Sun et al. 2019). The historical observed daily precipitation data (1956–2005) at 148 stations in the Yangtze River basin were collected from the National Meteorological Information Center (<http://data.cma.cn/>). The historical simulated (1956–2005) and future projected (2021–2100) precipitation data were derived from the GFDL-ESM2G climate model (Dunne et al. 2012). This model is archived in phase 5 of the Coupled Model Intercomparison Project (CMIP5) database and is developed by the NOAA Geophysical Fluid Dynamics Laboratory. The spatial resolution of this model is 2° latitude $\times$ 2.5° longitude (Dunne et al. 2012). Using a single model instead of the CMIP5 ensemble average could allow clear observation of values changing between before and after downscaling.

To evaluate the downscaling performance for meteorological statistics, we compare raw precipitation data (uncorrected), downscaled precipitation data (corrected), and measured historical observations (observed) against each other during 1976–2005, with another 20 years during 1956–75 are used as the baseline training period. To evaluate the downscaling effects on climate events, we compare downscaled precipitation data (corrected) against the raw precipitation projections (uncorrected) during the future period of 2021–2100, with the historical period of 1956–2005 used as the baseline training period. We choose 2005 as the end year of the historical period as defined by the CMIP5.

3. Method

This section describes the diagnostic framework proposed in this study for evaluating the performance of precipitation downscaling. An overview of this framework is provided in Fig. 1. Spatial, temporal, trend, extreme, and climate event credibility are the main components of this framework. Users are free to choose indices to describe these main aspects. In this study, spatial credibility is evaluated using a geographical information system (GIS) mapping tool. Temporal credibility is evaluated using indicators of average total monthly precipitation (PCPMM), standard deviation for daily precipitation in the month (PCPSTD), probability of a wet day following a wet day in the month (PCPWW), and average number of days of precipitation in the month (PCPND) (detailed in Table 1). Trend credibility is evaluated using the Mann–Kendall test. Extreme credibility is assessed by comparing complementary cumulative distribution functions. Climate event effects are assessed using drought and flood events, as further characterized by standardized precipitation index (SPI) and generalized Pareto distribution (GPD)

FIG. 1. Schematic diagram illustrating the overall diagnostic framework. Spatial, temporal, trend, extreme, and climate event credibility are the main components of this framework. Users are free to choose indices to describe these main aspects. The detailed information of PCPMM, PCPSTD, PCPWW, and PCPND is in Table 1.

index, respectively. Datasets utilized are GCM future projected climates, GCM historical simulated climates, and historical observations.

a. The downscaling approach to be evaluated

To downscale precipitation, we adopt a daily variant of the BCSD approach (Girvetz et al. 2013). The BCSD approach designed by Wood et al. (2004) is widely used for climate downscaling across different spatial and

temporal scales (e.g., Barnett et al. 2008; Beyene et al. 2010). However, the BCSD only derives monthly time scale results, while the daily variant of the BCSD approach allows GCMs to be downscaled on a daily time scale (Zeng et al. 2020). The enhanced daily variant of the BCSD is also designed to preserve the projected climate change signal (Girvetz et al. 2013). To carry out the bias correction, the daily BCSD relies on the quantile mapping (QM) method (Maraun 2013; Thrasher

TABLE 1. Information of monthly precipitation indicators, adapted from Arnold et al. (2012).

Variable	Calculation	Definition
PCPMM: avg total monthly precipitation (mm)	$\bar{R}_{\text{mon}} = \frac{\sum_{d=1}^N R_{\text{day,mon}}}{\text{yrs}}$	$\bar{R}_{\text{mon}}$ is the mean monthly precipitation (mm), $R_{\text{day,mon}}$ is the daily precipitation for record d in month mon (mm), yrs is the no. of years, and N is the total number of records d in month mon (mm)
PCPSTD: std dev for daily precipitation in month (mm)	$\sigma_{\text{mon}} = \left[\frac{\sum_{d=1}^N (R_{\text{day,mon}} - \bar{R}_{\text{mon}})^2}{N - 1} \right]^{1/2}$	σ_{mon} is the std dev for daily precipitation in month mon (mm), $\bar{R}_{\text{mon}}$ is the avg precipitation for the month, and $R_{\text{day,mon}}$ and N are the same as above
PCPWW: probability of a wet day following a wet day in the month	$P_i(W/W) = \text{days}_{W/W,i} / \text{days}_{\text{wet},i}$	$P_i(W/W)$ is the probability of a wet day following a wet day in month i , $\text{days}_{W/W,i}$ is the no. of times that a wet day followed a wet day in month i for the entire period of record, and $\text{days}_{\text{wet},i}$ is the no. of wet days in month i during the entire period of record. A wet day is a day with > 0 mm of precipitation
PCPND: avg no. of days of precipitation in the month	$\bar{d}_{\text{wet},i} = \frac{\text{days}_{\text{wet},i}}{\text{yrs}}$	$\bar{d}_{\text{wet},i}$ is the avg no. of days of precipitation in month i , $\text{days}_{\text{wet},i}$ is the no. of wet days in month i during the entire period of record, and yrs is the no. of years of record

et al. 2012). Each projected data point $x_{s,p}$ is mapped to $F_{s,h}$, which is a cumulative distribution function (CDF) of the simulated climate variable in the historical period, and the resulting value is mapped to $F_{o,h}^{-1}$, which is the inverse CDF of the observed climate variable (Gudmundsson et al. 2012; Themeßl et al. 2011; Wang et al. 2014). This transformation is given by

$$\hat{x}_{s,p}(t) = F_{o,h}^{-1}\{F_{s,h}[x_{s,p}(t)]\}, \quad (1)$$

where $\hat{x}_{s,p}$ is the corrected data point and the subscripts o , s , h , and p denote the observed data, simulated data, historical period, and projected period, respectively. A moving window of ± 15 days (Thrasher et al. 2012) around the date of the data point in the observed baseline period is used to construct the CDFs. Empirical distributions are adopted to fit the CDFs due to their high skill in reducing precipitation biases (Gudmundsson et al. 2012). At least a data length of around 30 years is recommended for daily empirical CDF construction to mitigate sampling uncertainties. Compared to the widely used QM method that only uses the CDF of the entire period, this daily correction method explicitly accounts for the daily variations in distribution changes. Moreover, the BCSD method removes the trend before applying the QM method and adds the trend back afterward. The trend is qualified as the 9-yr running mean of monthly climate anomalies, which is calculated for each month in the period to be downscaled. A climate anomaly is defined as the difference between the monthly mean of the projected climate variable and the monthly mean of the observed climate variable (Girvetz et al. 2013).

Unfortunately, no specific indication on how to handle extreme values was given by Girvetz et al. (2013). If not properly treated, extreme values are more prone to errors after downscaling than are mean parameters (Themeßl et al. 2012). When handling the extreme values that fall outside the range of the underlying CDF distribution, extrapolation is required (Boé et al. 2007; Gudmundsson et al. 2012). Studies focus on extreme climate events commonly adopt the annual maximum (1/365, or $\sim 0.27\%$) or N -largest approaches to draw the extreme datasets from the overall samples (Sun et al. 2015; Villarini et al. 2011). Under this circumstance, values that may traditionally be treated as outliers deserve careful consideration. We, therefore, extend the BCSD method to allow consideration of extreme values. Specifically, on each day, if a value falls outside the constructed CDF, we correct this value using a CDF constructed for the entire period. If this value still falls outside the CDF of the entire period, we use a ratio to change the magnitude of the value. Boé et al. (2007) used a threshold ratio based on the largest 25% of data

points, but the study did not target extreme events. Therefore, for this threshold, we use 1% of the data, which roughly corresponds to an average of three extreme events each year and is a commonly adopted threshold in extreme value theory (Lang et al. 1999).

b. Indices evaluating temporal credibility

Four indices are included to describe the downscaling effects on seasonal precipitation characteristics (Table 1). These indicators are often used in hydrological software to generate a representative daily climate when measured data are missing (Arnold et al. 2012). Specifically, for each month, precipitation statistics include PCPMM, PCPSTD, PCPWW, and PCPND. In this study, calculations are performed monthly from January to December for 30 years during 1976–2005. This calculation is repeated for historical observation, uncorrected, and corrected CMIP5 simulations.

c. Indices evaluating climate event credibility

Questions exist with regard to the effects that a downscaling method has on climate events. Specifically, the downscaling method may or may not deteriorate the projected signal in climate models when making various corrections. To address this question, we present a comparison of drought and flood signals, which are among the climate events of most significant concern, before and after downscaling for the period of 2021–2100.

1) DROUGHT INDEX

To identify and estimate drought conditions, the SPI is used. We follow the approach by Sun et al. (2019) to calculate the SPI, which is based on a previously developed method by Farahmand and AghaKouchak (2015). To implement the approach, the monthly data at time t are denoted by $x(t)$. For a given n -month time scale (e.g., 3, 6, or 12 months), the accumulated variable $X_n(t)$ is calculated as follows:

$$X_n(t) = \sum_{t-n+1}^t x(t). \quad (2)$$

Second, for a particular month m , we subdivide the time series $X_n(t)$ into subseries $S_n(m)$ as follows:

$$S_n(m) = \{X_n(m), X_n(1 \times 12 + m), \dots, X_n[(N-1) \times 12 + m]\}, \quad (3)$$

where $m = 1, 2, \dots, 12$ are the calendar months and N represents the number of years considered. Third, the empirical Gringorten plotting position (Gringorten 1963) is used to compute the cumulative frequency of $S_n(m)$ as follows:

$$P_{nm}(i) = \frac{i_{nm} - 0.44}{N + 0.12}, \quad (4)$$

where i_{nm} is the rank of $S_n(m)$ from the smallest to the highest and $P_{nm}(i)$ denotes the estimate of the cumulative frequency of the i th term in the month m . Last, the empirical probabilities are transformed into the standard normal distribution function according to

$$\text{SPI}_{nm} = \Phi^{-1}(P_{nm}), \quad (5)$$

where SPI_{nm} denotes the standardized index for month m and time scale n and Φ denotes the standard normal distribution function (AghaKouchak 2014; Farahmand and AghaKouchak 2015).

2) FLOOD INDEX

To determine floods, we use extreme value theory (De Haan and Ferreira 2007), specifically, the GPD, which corresponds to the peaks-over-threshold (POT) approach (Mudersbach et al. 2017; Roth et al. 2012; Wang 1991). Note that unlike the long-lasting and cumulative nature of drought, floods develop rapidly and are highly related to precipitation peaks. First, we derive the POT sample from daily precipitation time series. The threshold value u is estimated from the daily precipitation series $x_1, \dots, x_n$, according to the criterion that on average three flood events are included for each year (Lang et al. 1999). The exceedances z fulfill the rule that $z > u$. Additionally, no two adjacent flood events should appear in a 14-day time window. Second, the GPD is fitted to describe the distributional behavior of z using

$$\text{GPD}(z) = 1 - \left[1 + \xi \left(\frac{z - u}{\sigma} \right) \right]^{-1/\xi}, \quad (6)$$

where σ is the location parameter, and ξ is the shape parameter (Coles et al. 2001). Stationary GPD fitting is adopted to fit the distribution. The return level is calculated by

$$z_N = \mu + \frac{\delta}{\xi} [(N n_y \zeta_\mu)^\xi - 1], \quad (7)$$

where Z_N is the N -year return level, ζ_μ is the probability of an individual observation exceeding the threshold u , and n_y is the total observations per year. In this study, return periods of 5, 10, and 20 years are selected.

4. Results

This section presents results of the diagnosis. Figure 2 shows the performance of downscaling on spatial bias correction, the top panel depicts the bias between the uncorrected and observed precipitation, and the bottom

FIG. 2. Investigation of the spatial credibility of downscaling. The (a) uncorrected bias measures the difference between the uncorrected (before downscaling) and observed precipitation. The (b) corrected bias is the difference between the corrected (after downscaling) and observed precipitation. The difference is calculated on the basis of daily precipitation means at 148 meteorological stations in the Yangtze River basin during the validation period of 1976–2005.

panel shows the bias between the corrected and observed precipitation. The bias is calculated as the mean value of daily precipitation difference over the 30 years validation period (1976–2005). It can be seen that the uncorrected precipitation generally overestimates the precipitation in the western Yangtze River basin and underestimates the precipitation in the eastern Yangtze River basin. The serious overestimation in parts of the western Yangtze River basin could largely be attributed to the steep terrain in these areas. Specifically, the elevation difference in the western part could be as large as 3000 m within a spatial area of 2° latitude $\times$ 2° longitude. The bias between the uncorrected and observed at the 148 stations ranges between -2.54 and 4.48 , with an average of 0.67 mm day^{-1} . After downscaling, the range of bias decreases to -0.74 – 0.64 , with an average of $-0.11 \text{ mm day}^{-1}$. In comparison with the precipitation correction ability of the downscaling method reported by Piani et al. (2010), this downscaling approach performs well in spatial bias correction.

Figure 3 shows the monthly precipitation statistics of the uncorrected (red), observed (blue) and corrected (yellow) climates. The values are averaged from 148 stations for 30 years during 1976–2005. Before downscaling, discrepancies exist for all four statistics. After

FIG. 3. Investigation of the temporal credibility of downscaling. The meaning of the acronyms (a) PCPMM, (b) PCPSTD, (c) PCPWW, and (d) PCPD is detailed in Table 1. The monthly uncorrected (red), observed (blue), and corrected (yellow) values are averaged from 148 stations for 30 years during 1976–2005 in the Yangtze River basin.

downscaling, significant improvements are made, as indicated by the similarity between observed and corrected statistics in Fig. 3. Nevertheless, the statistics in some months are not as satisfying, for example, PCPMM in June (-16.47 mm) and July (-18.00 mm) and PCPD in December ($+4.59$ days). We elaborate further on these findings in the discussion section.

Figure 4 presents the trends of uncorrected and corrected annual precipitation totals at 148 stations. The trend is calculated as the rate of increase in annual precipitation during the validation period. Mann–Kendall trend analysis is adopted to analyze the trend (Mann 1945). As shown, the climate change signals for these 148 stations all displayed increasing trends, with uncorrected precipitation trends of approximately 0.43 – 4.35 mm yr^{-1} and corrected precipitation trends of approximately -0.12 – 6.34 mm yr^{-1} . The data points are generally distributed along the 45° reference line. The trend at only one station reverts from positive (0.49) to negative (-0.12). The correlation between before and after downscaling is 0.47 (Pearson's R). These results indicate that the daily BCSD method effectively preserves the precipitation trend during the validation period.

Figure 5 evaluates the downscaling credibility of extreme values. The complementary CDF ($1 - \text{CDF}$) of

uncorrected (red), observed (blue) and corrected (yellow) precipitation at nine representative stations are plotted. These stations are representative meteorological stations of nine provinces/municipalities across the Yangtze River basin. Results are only shown for these nine stations for the sake of brevity and clarity. The y axis is shown using the log scale to highlight the upper tails of the distributions. The extremes of uncorrected precipitation are all below the corresponding observations, which can be seen from the shorter uncorrected x-axis ranges compared to the observed counterparts. The largest extreme values of the uncorrected, observed and corrected are 69.78 , 233.40 , and 206.96 mm day^{-1} , respectively. Moreover, the averaged maximum values, which are defined as the top 0.27% of uncorrected, observed, and corrected, are around 48.93 , 107.89 , and 116.25 mm day^{-1} , respectively. The downscaling process successfully transforms uncorrected extreme values that are much lower than the observed extremes to corrected extremes that are similar to observations. It is not surprising that the extremes of uncorrected precipitation are all below the corresponding observations because the grid value represents the average in a larger area. The daily BCSD method is able to generate values that approximate the observed extremes.

FIG. 4. Investigation of the trend credibility of downscaling. Uncorrected trend is plotted against the corrected trend for 148 stations in the Yangtze River basin. The nearer the absolute distance from a data point is to the 45° line, the more similar the trend is. The trend for each station is calculated using the Mann–Kendall test.

Figure 6 shows the future drought prediction of both level of 0.05 changed to 0.01 at two stations, the uncorrected (blue) and corrected (yellow) precipitation for nine representative stations. As shown, the uncorrected and corrected drought signals are highly correlated, as indicated by the Pearson's R , which range between 0.78 (Mianyang station, Fig. 6b) and 0.93 (Kunming station, Fig. 6a). Although the significance level of 0.05 changed to 0.01 at two stations, Mianyang (Fig. 6b) and Wuhan (Fig. 6f), the sign of the trends remains the same for all the stations. The differences between slope 1 (uncorrected) and slope 2 (corrected) vary between 0 and 0.04. These results suggest that the downscaled precipitation captures the projected drought signal reasonably well.

Unlike the long-lasting and cumulative nature of drought, floods rapidly develop and are highly related to precipitation peaks. In Fig. 7, we present the flood return levels for uncorrected (dashed lines) and corrected (solid lines) projections in the nine representative stations. The uncorrected and corrected return levels are computed based on the derived POT datasets of uncorrected (open black circles) and corrected (open green circles) projections, respectively. The historical flood return levels (solid circles on y axis) calculated from historical observed extreme events (not shown) are also computed as a baseline reference. The most prominent and relevant information from this figure is that the

uncorrected return levels in all the nine stations, regardless of return level, are much lower than the observed. While the corrected return levels are similar or higher than the observed, that is, the solid lines in Fig. 7 tend to be at the same level as the solid circles.

5. Discussion

a. Downscaling performance of the daily BCSD

To summarize, the daily BCSD approach exhibits satisfactory performance in precipitation downscaling as revealed by our diagnosis. Prior to downscaling, large discrepancies exist between the uncorrected and observed precipitation, such as station-dependent bias, unmatched monthly precipitation statistics, and underestimated extremes. After downscaling, the corrected precipitation statistics agree well with observations or expectations. The spatial bias is reduced, the monthly precipitation statistics are enhanced, the projected trend is preserved, the drought signal resembles the projection, and the projected return levels look more feasible. Notably, we advance the daily BCSD method implemented here by considering extreme values, which could be beneficial for studies on rare events. The handling of extremes is acceptable as validated by the flood projection. In theory, dynamical downscaling approaches based on physical laws rather than statistical relationships could better represent local conditions. However, they are computationally very demanding and could easily inherit errors from the global climate models.

We discuss next some of the poor performance conditions identified as part of our results. First, in June–August, the downscaled average monthly precipitation totals for the whole YRB are lower than the observed and projected, whereas the projected is similar to the observed in these same months. This could be related to the nature of quantile-dependent bias correction (Maraun 2013). Despite this, the amplitude and statistics are more reliable after downscaling compared to the uncorrected. Second, for drought prediction, the significance level changes from 0.05 to 0.01 for two of the stations. This is because the preserved trend (which is first removed and then added back after the QM) in the two stations are weakened by the QM process. Specifically, the QM changes the trend of the detrended data series (data with trend removed) to resemble historical observations, which exerts an adverse effect on the preserved trend. After this validation process, we conclude that the daily BCSD method performs well in climate downscaling and can be adopted by future studies.

It is important to note that the correction process is carried out on the simulated CMIP5 datasets but is compared with the real historical observations. Given

FIG. 5. Investigation of the extreme credibility of downscaling. Curves show the complementary CDF ($1 - \text{CDF}$) of uncorrected (red), observed (blue), and corrected (yellow) precipitation during the calibration period of 1976–2005. The nine representative meteorological stations are from nine provinces/municipalities across the Yangtze River basin. The y axis is shown using the logarithmic scale to highlight the upper tails of the distributions.

the current state of climate science, one would not expect the corrected climate simulation to reproduce the observed precipitation very accurately; rather, one expects the downscaled results to reasonably match the magnitude and pattern of observations. Not also that these results are trained with data from a 20-yr period and validated with data from a longer period of 30 years. By contrast, a longer period for training and a shorter period for validation could further improve the validation

performance. This choice of time periods is intended not only to maximize the testing ability for the downscaling method, but also to promote its application on multidecadal climate projections.

b. Significance and implications of downscaling diagnosis

Given the numerous transformations made during the downscaling process, uncertainties in downscaled products

FIG. 6. Investigation of the downscaling effect on droughts. Slope 1 and slope 2 are monthly changing rates, which are the slope values associated with linear regressions of the uncorrected and corrected SPI time series, respectively. Note that a single asterisk indicates decreasing or increasing trends and double asterisks indicate significant decreasing or increasing trends at the 0.05 confidence interval level.

can be large and originate from multiple sources (Kay et al. 2009). This study proposes a systematic diagnostic framework to help evaluate the performance of precipitation downscaling. As indicated by the evaluation process

conducted in this study, comprehensive evaluation is necessary as any individual source of uncertainty can be propagated through to bring uncertainty to the end result. Therefore, we argue that comprehensive downscaling

FIG. 7. Investigation of the downscaling effect on floods. Return levels of observed (dots on the vertical axis), uncorrected (dashed lines), and corrected (solid lines) floods are plotted together to inspect the effect of climate downscaling on flood signal projection. Return levels of observed floods represent historical conditions during 1976–2005, and the return levels of uncorrected and corrected floods are calculated for the future projected period of 2020–2100 in the Yangtze River basin. The derived POT datasets used to calculate uncorrected and corrected return levels are characterized by black and green open circles in the background, respectively. Return levels of 5 (red), 10 (blue), and 20 (yellow) years are selected.

evaluations should be carried out for recognizing the capability of downscaling approaches, especially when new downscaling methods are proposed. The significance of considering integrated attributes that can be beneficial for future studies are further discussed and highlighted below. First, a downscaling method should perform well spatially, which could guarantee effective spatial comparisons that are often carried out in climate-related studies. Second, ensuring a downscaling method performs well seasonally is also crucial to climate studies. Precipitation has strong seasonal dynamics. The seasonal indicators we adopted include not only two central moments but also wet/dry probabilities. Communicating information about rainfall-runoff processes, these indicators would benefit hydrometeorological studies. Third, the climate change signal is of paramount importance for long-term projection and must be properly preserved to inform long-term management (Themebl et al. 2012). For regional studies focused on future projection, if the trend is deteriorated in the downscaling process, how can we expect the final estimations are trustworthy? A downscaling approach that considers long-term trends could have profound implications for climate impact studies. Fourth, climate extremes are a major concern because of their potential for catastrophic impacts (Hertig et al. 2019). Hydrologic and hydraulic infrastructure designs often depend on the patterns of precipitation extremes (Roth et al. 2012). Wrongly projecting future precipitation peaks might lead to a considerable risk for infrastructure failure. These evaluations could not only inform end users about the performance of downscaling, but also benefit developers on the improvement of downscaling algorithms. Finally, in climate impact studies, downscaled precipitation is not an end in itself because the ultimate goal is to provide inputs for impact models (Chen et al. 2013; Olsson et al. 2009). Integrating climate event in the diagnosis tends to be much easier than previously as resources for quantifying climate impacts become more readily available. Therefore, it is necessary to develop a downscaling method that is applicable for the assessment of climate-sensitive events.

c. Limitations and future prospects

This study represents an initial attempt by proposing to include spatial, temporal, trend, extreme, and climate event into a systematic downscaling diagnosis. The disproportional development of downscaling evaluation standards compared to the fast-growing downscaling tools supports the need for focused research on downscaling diagnostics. Furthermore, a variety of indices can be adopted for each of the attributes considered. Currently, the indices describing each attribute are selected based on the resources at hand. However, some

indices may be more effective in describing these attributes than others, while discussion on the advantage and shortcomings of indices in describing downscaling performance are rarely presented. Thus, it is suggested that a unified framework with explicit indices, which adequately conveys how downscaling approaches perform, needs to be formulated in the future.

6. Conclusions

This study proposes a diagnostic framework for climate downscaling. The diagnosis covers five attributes (i.e., spatial, temporal, trend, extreme, and climate event). We demonstrate the credibility of this framework by evaluating the performance of a daily BCSD downscaling approach for the Yangtze River basin. Three datasets are involved in the downscaling process, including observed climates from 148 stations in the Yangtze River basin, projected climates, and simulated climates from the GFDL-ESM model. Notably, we further advance the daily BCSD method by considering extremes, which could be beneficial for studies focusing on rare events. As revealed by our diagnosis, our daily BCSD approach exhibits satisfactory performance in precipitation downscaling. Moreover, since downscaling is needed to bridge the gap between GCMs and regional studies, we argue that comprehensive downscaling evaluations should be carried out to better recognize the capability of various and emerging downscaling approaches. Finally, it is hoped that informative and non-overlapping indices can be established to interpret the performance of downscaling, which could explicitly reflect the credibility of downscaling approaches and further boost intercomparisons.

Acknowledgments. This study is supported by the China Postdoctoral Science Foundation (Grant 2019M661422) and the National Natural Science Foundation of China (Grant 41801314).

REFERENCES

- AghaKouchak, A., 2014: A baseline probabilistic drought forecasting framework using standardized soil moisture index: Application to the 2012 United States drought. *Hydrol. Earth Syst. Sci.*, **18**, 2485–2492, <https://doi.org/10.5194/hess-18-2485-2014>.
- Ali, S., and Coauthors, 2019: Assessment of climate extremes in future projections downscaled by multiple statistical downscaling methods over Pakistan. *Atmos. Res.*, **222**, 114–133, <https://doi.org/10.1016/j.atmosres.2019.02.009>.
- Anandhi, A., V. Srinivas, R. S. Nanjundiah, and D. Nagesh Kumar, 2008: Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *Int. J. Climatol.*, **28**, 401–420, <https://doi.org/10.1002/joc.1529>.

- Arnold, J. G., J. G. Kiniry, and R. Srinivasan, 2012: Soil and water assessment tool: Input/output documentation. *Trans. ASABE*, **7**, 168–170.
- Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083, <https://doi.org/10.1126/science.1152538>.
- Barsugli, J. J., and Coauthors, 2013: The practitioner's dilemma: How to assess the credibility of downscaled climate projections. *Eos, Trans. Amer. Geophys. Union*, **94**, 424–425, <https://doi.org/10.1002/2013EO460005>.
- Benestad, R. E., 2010: Downscaling precipitation extremes. *Theor. Appl. Climatol.*, **100**, 1–21, <https://doi.org/10.1007/s00704-009-0158-1>.
- Beyene, T., D. Lettenmaier, and P. Kabat, 2010: Hydrologic impacts of climate change on the Nile River basin: Implications of the 2007 IPCC scenarios. *Climatic Change*, **100**, 433–461, <https://doi.org/10.1007/s10584-009-9693-0>.
- Boé, J., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol.*, **27**, 1643–1655, <https://doi.org/10.1002/joc.1602>.
- Bürger, G., T. Q. Murdock, A. T. Werner, S. R. Sobie, and A. J. Cannon, 2012: Downscaling extremes—An intercomparison of multiple statistical methods for present climate. *J. Climate*, **25**, 4366–4388, <https://doi.org/10.1175/JCLI-D-11-00408.1>.
- Cannon, A. J., 2018: Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>.
- , S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>.
- Chai, Y., Y. Li, Y. Yang, B. Zhu, S. Li, C. Xu, and C. Liu, 2019: Influence of climate variability and reservoir operation on streamflow in the Yangtze River. *Sci. Rep.*, **9**, 5060, <https://doi.org/10.1038/s41598-019-41583-6>.
- Chen, J., F. P. Brissette, and R. Leconte, 2011: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J. Hydrol.*, **401**, 190–202, <https://doi.org/10.1016/j.jhydrol.2011.02.020>.
- , —, D. Chaumont, and M. Braun, 2013: Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *J. Hydrol.*, **479**, 200–214, <https://doi.org/10.1016/j.jhydrol.2012.11.062>.
- Coles, S., J. Bawa, L. Trenner, and P. Dorazio, 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer Science and Business Media, 208 pp.
- Das, T., E. P. Maurer, D. W. Pierce, M. D. Dettinger, and D. R. Cayan, 2013: Increases in flood magnitudes in California under warming climates. *J. Hydrol.*, **501**, 101–110, <https://doi.org/10.1016/j.jhydrol.2013.07.042>.
- De Haan, L., and A. Ferreira, 2007: *Extreme Value Theory: An Introduction*. Springer Science and Business Media, 418 pp.
- Diaz-Nieto, J., and R. L. Wilby, 2005: A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, **69**, 245–268, <https://doi.org/10.1007/s10584-005-1157-6>.
- Di Luca, A., R. de Elia, and R. Laprise, 2012: Potential for small scale added value of RCM's downscaled climate change signal. *Climate Dyn.*, **40**, 601–618, <https://doi.org/10.1007/s00382-012-1415-z>.
- Dunne, J. P., and Coauthors, 2012: GFDL's ESM2 global coupled climate-carbon Earth System Models. Part I: Physical formulation and baseline simulation characteristics. *J. Climate*, **25**, 6646–6665, <https://doi.org/10.1175/JCLI-D-11-00560.1>.
- Farahmand, A., and A. AghaKouchak, 2015: A generalized framework for deriving nonparametric standardized drought indicators. *Adv. Water Resour.*, **76**, 140–145, <https://doi.org/10.1016/j.advwatres.2014.11.012>.
- Fletcher, S., M. Lickley, and K. Strzepek, 2019: Learning about climate change uncertainty enables flexible water infrastructure planning. *Nat. Commun.*, **10**, 1782, <https://doi.org/10.1038/s41467-019-09677-x>.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Girvetz, E. H., E. P. Maurer, P. B. Duffy, A. Ruesch, B. Thrasher, and C. Zganjar, 2013: Making climate data relevant to decision making: The important details of spatial and temporal downscaling. World Bank Doc., 43 pp., <https://scholarcommons.scu.edu/ceng/13/>.
- Gringorten, I. I., 1963: A plotting rule for extreme probability paper. *J. Geophys. Res.*, **68**, 813–814, <https://doi.org/10.1029/JZ068i003p00813>.
- Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, <https://doi.org/10.5194/hess-16-3383-2012>.
- Hansen, B. B., and Coauthors, 2019: More frequent extreme climate events stabilize reindeer population dynamics. *Nat. Commun.*, **10**, 1616, <https://doi.org/10.1038/s41467-019-09332-5>.
- Hayhoe, K. A., 2010: A standardized framework for evaluating the skill of regional climate downscaling techniques. Ph.D. dissertation, University of Illinois at Urbana-Champaign, 158 pp., <http://hdl.handle.net/2142/16044>.
- Hertig, E., and Coauthors, 2019: Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation results from the perfect predictor experiment of the COST Action VALUE. *Int. J. Climatol.*, **39**, 3846–3867, <https://doi.org/10.1002/joc.5469>.
- Hessami, M., P. Gachon, T. Ouarda, and A. St-Hilaire, 2008: Automated regression-based statistical downscaling tool. *Environ. Modell. Software*, **23**, 813–834, <https://doi.org/10.1016/j.envsoft.2007.10.004>.
- IPCC, 2012: Glossary of terms. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 555–564.
- Jeong, D. I., A. St-Hilaire, T. B. Ouarda, and P. Gachon, 2012: Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator. *Climatic Change*, **114**, 567–591, <https://doi.org/10.1007/s10584-012-0451-3>.
- Kay, A. L., H. N. Davies, V. A. Bell, and R. G. Jones, 2009: Comparison of uncertainty sources for climate change impacts: Flood frequency in England. *Climatic Change*, **92**, 41–63, <https://doi.org/10.1007/s10584-008-9471-4>.
- Khan, M. S., P. Coulibaly, and Y. Dibikey, 2006: Uncertainty analysis of statistical downscaling methods. *J. Hydrol.*, **319**, 357–382, <https://doi.org/10.1016/j.jhydrol.2005.06.035>.
- Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Climate Change*, **3**, 369–373, <https://doi.org/10.1038/nclimate1716>.

- Kreienkamp, F., A. Paxian, B. Früh, P. Lorenz, and C. Matulla, 2018: Evaluation of the empirical-statistical downscaling method EPISODES. *Climate Dyn.*, **52**, 991–1026, <https://doi.org/10.1007/s00382-018-4276-2>.
- Lang, M., T. Ouarda, and B. Bobée, 1999: Towards operational guidelines for over-threshold modeling. *J. Hydrol.*, **225**, 103–117, [https://doi.org/10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5).
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, <https://doi.org/10.1029/2009JD012882>.
- Lo, J. C. F., Z. L. Yang, and R. A. Pielke, 2008: Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) Model. *J. Geophys. Res.*, **113**, D09112, <https://doi.org/10.1029/2007JD009216>.
- Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13**, 245–259, <https://doi.org/10.2307/1907187>.
- Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>.
- , 2016: Bias correcting climate change simulations - a critical review. *Curr. Climate Change Rep.*, **2**, 211–220, <https://doi.org/10.1007/s40641-016-0050-x>.
- , and Coauthors, 2015: VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, **3**, 1–14, <https://doi.org/10.1002/2014EF000259>.
- Martinich, J., and A. Crimmins, 2019: Climate damages and adaptation potential across diverse sectors of the United States. *Nat. Climate Change*, **9**, 397–404, <https://doi.org/10.1038/s41558-019-0444-6>.
- Mehrotra, R., and A. Sharma, 2010: Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. *Water Resour. Res.*, **46**, W07526, <https://doi.org/10.1029/2009WR008423>.
- Mudersbach, C., J. Bender, and F. Netzel, 2017: An analysis of changes in flood quantiles at the gauge Neu Darchau (Elbe River) from 1875 to 2013. *Stochastic Environ. Res. Risk Assess.*, **31**, 145–157, <https://doi.org/10.1007/s00477-015-1173-7>.
- Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate*, **12**, 2256–2284, [https://doi.org/10.1175/1520-0442\(1999\)012<2256:AEOSAD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2).
- Ockenden, M. C., and Coauthors, 2017: Major agricultural changes required to mitigate phosphorus losses under climate change. *Nat. Commun.*, **8**, 161, <https://doi.org/10.1038/s41467-017-00232-0>.
- Olsson, J., K. Berggren, M. Olofsson, and M. Viklander, 2009: Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden. *Atmos. Res.*, **92**, 364–375, <https://doi.org/10.1016/j.atmosres.2009.01.015>.
- Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, <https://doi.org/10.1175/BAMS-D-15-00308.1>.
- Reidmiller, D. R., C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock, and B. C. Stewart, Eds., 2018: *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment*. Vol. II, U.S. Global Change Research Program, 1515 pp., <https://doi.org/10.7930/NCA4.2018>.
- Rhee, J., and J. Cho, 2016: Future changes in drought characteristics: Regional analysis for South Korea under CMIP5 projections. *J. Hydrometeorol.*, **17**, 437–451, <https://doi.org/10.1175/JHM-D-15-0027.1>.
- Roth, M., T. Buishand, G. Jongbloed, A. Klein Tank, and J. Zanten, 2012: A regional peaks-over-threshold model in a nonstationary climate. *Water Resour. Res.*, **48**, W11533, <https://doi.org/10.1029/2012WR012214>.
- Sachindra, D., F. Huang, A. Barton, and B. Perera, 2013: Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *Int. J. Climatol.*, **33**, 1087–1106, <https://doi.org/10.1002/joc.3493>.
- Sanderson, B. M., K. W. Oleson, W. G. Strand, F. Lehner, and B. C. O'Neill, 2018: A new ensemble of GCM simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change*, **146**, 303–318, <https://doi.org/10.1007/s10584-015-1567-z>.
- Schmidli, J., C. Frei, and P. L. Vidale, 2006: Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods. *Int. J. Climatol.*, **26**, 679–689, <https://doi.org/10.1002/joc.1287>.
- , C. M. Goodess, C. Frei, M. R. Haylock, Y. Hündecha, J. Ribalaygua, and T. Schmith, 2007: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *J. Geophys. Res.*, **112**, D04105, <https://doi.org/10.1029/2005JD007026>.
- Shiru, M. S., S. Shahid, E.-S. Chung, N. Alias, and L. Scherer, 2019: A MCDM-based framework for selection of general circulation models and projection of spatio-temporal rainfall changes: A case study of Nigeria. *Atmos. Res.*, **225**, 1–16, <https://doi.org/10.1016/j.atmosres.2019.03.033>.
- Sinha, P., M. E. Mann, J. D. Fuentes, A. Mejia, L. Ning, W. Sun, T. He, and J. Obeysekera, 2018: Downscaled rainfall projections in south Florida using self-organizing maps. *Sci. Total Environ.*, **635**, 1110–1123, <https://doi.org/10.1016/j.scitotenv.2018.04.144>.
- Skamarock, W. C., and J. B. Klemp, 2008: A time-split non-hydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- Sun, F., A. Mejia, P. Zeng, and Y. Che, 2019: Projecting meteorological, hydrological and agricultural droughts for the Yangtze River basin. *Sci. Total Environ.*, **696**, 134076, <https://doi.org/10.1016/j.scitotenv.2019.134076>.
- Sun, X., U. Lall, B. Merz, and D. Nguyen Viet, 2015: Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany. *Water Resour. Res.*, **51**, 6586–6601, <https://doi.org/10.1002/2015WR017117>.
- Themeßl, M. J., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, **31**, 1530–1544, <https://doi.org/10.1002/joc.2168>.
- , —, and G. Heinrich, 2012: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, **112**, 449–468, <https://doi.org/10.1007/s10584-011-0224-4>.
- Thrasher, B., E. P. Maurer, C. McKellar, and P. B. Duffy, 2012: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol.*

- Earth Syst. Sci.*, **16**, 3309–3314, <https://doi.org/10.5194/hess-16-3309-2012>.
- Tripathi, S., V. Srinivas, and R. S. Nanjundiah, 2006: Downscaling of precipitation for climate change scenarios: A support vector machine approach. *J. Hydrol.*, **330**, 621–640, <https://doi.org/10.1016/j.jhydrol.2006.04.030>.
- Villarini, G., J. A. Smith, A. A. Ntelekos, and U. Schwarz, 2011: Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria. *J. Geophys. Res.*, **116**, D05103, <https://doi.org/10.1029/2010JD015038>.
- von Storch, H., H. Langenberg, and F. Feser, 2000: A spectral nudging technique for dynamical downscaling purposes. *Mon. Wea. Rev.*, **128**, 3664–3673, [https://doi.org/10.1175/1520-0493\(2000\)128<3664:ASNTFD>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2).
- Wang, L., W. Chen, and W. Zhou, 2014: Assessment of future drought in southwest China based on CMIP5 multimodel projections. *Adv. Atmos. Sci.*, **31**, 1035–1050, <https://doi.org/10.1007/s00376-014-3223-3>.
- Wang, Q., 1991: The POT model described by the generalized Pareto distribution with Poisson arrival rate. *J. Hydrol.*, **129**, 263–280, [https://doi.org/10.1016/0022-1694\(91\)90054-L](https://doi.org/10.1016/0022-1694(91)90054-L).
- Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.*, **34**, 2995–3008, <https://doi.org/10.1029/98WR02577>.
- Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>.
- Zeng, P., F. Sun, Y. Liu, and Y. Che, 2020: Future river basin health assessment through reliability-resilience-vulnerability: Thresholds of multiple dryness conditions. *Sci. Total Environ.*, **741**, 140395, <https://doi.org/10.1016/j.scitotenv.2020.140395>.