



Multivariate bias correction and downscaling of climate models with trend-preserving deep learning

Fang Wang¹ · Di Tian¹

Received: 4 January 2024 / Accepted: 16 August 2024 / Published online: 4 September 2024
© The Author(s) 2024

Abstract

Global climate models (GCMs) and Earth system models (ESMs) exhibit biases, with resolutions too coarse to capture local variability for fine-scale, reliable drought and climate impact assessment. However, conventional bias correction approaches may cause implausible climate change signals due to unrealistic representations of spatial and intervariable dependences. While purely data-driven deep learning has achieved significant progress in improving climate and earth system simulations and predictions, they cannot reliably learn the circumstances (e.g., extremes) that are largely unseen in historical climate but likely becoming more frequent in the future climate (i.e., climate non-stationarity). This study shows an integrated trend-preserving deep learning approach that can address the spatial and intervariable dependences and climate non-stationarity issues for downscaling and bias correcting GCMs/ESMs. Here we combine the super-resolution deep residual network (SRDRN) with the trend-preserving quantile delta mapping (QDM) to downscale and bias correct six primary climate variables at once (including daily precipitation, maximum temperature, minimum temperature, relative humidity, solar radiation, and wind speed) from five state-of-the-art GCMs/ESMs in the Coupled Model Intercomparison Project Phase 6 (CMIP6). We found that the SRDRN-QDM approach greatly reduced GCMs/ESMs biases in spatial and intervariable dependences while significantly better-reducing biases in extremes compared to deep learning. The estimated drought based on the six bias-corrected and downscaled variables captured the observed drought intensity and frequency, which outperformed state-of-the-art multivariate bias correction approaches, demonstrating its capability for correcting GCMs/ESMs biases in spatial and multivariable dependences and extremes.

Keywords Multivariate bias correction · Downscaling · Trend preserving · Deep learning · Climate model · Drought assessment

1 Introduction

Global climate models (GCMs) and Earth system models (ESMs), which are based on our physical understanding of the climate and earth system and its evolutions in the future, play a key role in climate impacts, adaptation, and resilience studies. However, GCM/ESM outputs suffer from systematic biases, and their spatial resolutions are too coarse to capture local variability. Therefore, it needs to reduce, postprocess, or cope with GCMs/ESMs model biases and scale issues for subsequent climate impact modeling and analysis.

Many statistical bias correction and downscaling methods have been developed to correct the discrepancies between GCM simulations and observed references. Most of these methods, such as the widely used quantile mapping (QM), are designed to correct variables individually at each location [QM; e.g., Panofsky and Brier (1968); Thrasher et al. (2012); Wood et al. (2002)] without accounting for biases of spatial and intervariable dependences (Bürger et al. 2011; Nahar et al. 2018). Spatial and intervariable dependences in downscaled and bias-corrected GCMs/ESMs are particularly important in subsequent impact modeling and analysis that depend on multiple variables such as droughts (Van Loon et al. 2016; Zscheischler et al. 2018). Neglecting those biases can cause generated outputs not to obey physical laws (Agbazo and Grenier 2020; Thrasher et al. 2012; Wang and Tian 2022), distorting the results of impact studies (Zscheischler et al. 2018), or causing implausible climate

✉ Di Tian
tiandi@auburn.edu

¹ Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, USA

change signals (Maraun 2016; Maraun et al. 2017). These issues have led to the recent development of multivariate bias correction methods based on different statistical techniques. These techniques either directly adjust multivariate distributions over a region (e.g., MBCn; Cannon 2018; Lange 2019) or assume predefined intervariable relationships, including Pearson correlation (Bürger et al. 2011; Cannon 2016; Mehrotra and Sharma 2012), Spearman rank correlation (Cannon 2016) and lag one autocorrelation for rank dependence (Mehrotra and Sharma 2012). Due to the difficulties of adjusting multivariate distributions at very high dimensions (i.e., multiple variables at many point locations over a large area) and inherent assumptions, the current multivariate bias correction methods have limited capability of bias correcting spatial and intervariable dependences for a large geographical area (François et al. 2020). For example, Multivariate Bias Correction with N-dimensional probability density function transform (MBCn) maps a multivariate source distribution to a same-dimensional target distribution through orthogonal matrix rotation operations iteratively, which has been shown competitive performance compared to either univariate (e.g., QM) or other multivariate bias correction approaches for bias correcting climate data at a relatively small number of point locations (Cannon 2018; François et al. 2020). However, when it was applied to bias-correct gridded precipitation over a large geographic region (i.e., the contiguous United States (CONUS) with a spatial resolution of 1°), it performed much worse compared to the QM method, potentially due to its difficulties in handling a large number of grid points (Pan et al. 2021).

Deep learning techniques have emerged as a promising approach for simulating highly nonlinear and complex relationships between different variables. In particular, deep learning models with convolutional neural networks [CNNs, LeCun et al. (2015)] can capture complex spatial structures showing great performances. Originally developed for computer vision problems, CNNs for climate and Earth system sciences have been growing rapidly during recent years (Reichstein et al. 2019) including detection of extreme weather events (Liu et al. 2016; Rcah et al. 2017) and weather and climate forecasting (Chapman et al. 2019; Ham et al. 2019; Liang et al. 2021; Ravuri et al. 2021; Scher and Messori 2019; Shi et al. 2015). Besides capturing complex spatial features, recent study has shown deep learning with deep convolutional layers and upsampling layers can process multiple input–output spatial variables at one time, accounting for their physical relations, suggesting a great potential for bias correcting and downscaling dependences of multiple variables (Wang and Tian 2022). Since the intervariable relationships are integrated into model training with no prior assumptions, deep learning models with convolutional layers taking multiple physically associated variables at once have opportunities to capture complex relationships

beyond our prior knowledge and provide potentials for improving bias corrections of both spatial and intervariable dependences in GCMs/ESMs.

Deep learning with convolutional layers models have been applied to downscale and bias correcting climate variables and show superiority over traditional methods (Baño-Medina et al. 2022; François et al. 2021; Fulton et al. 2023; Hess et al. 2023; Hess et al. 2022; Liu et al. 2020; Pan et al. 2021; Quesada-Chacón et al. 2022; Quesada-Chacón et al. 2023; Rodrigues et al. 2018; Vandal et al. 2017; Wang and Tian 2022). However, most of these studies focus on bias correcting individual climate variables, neglecting their physical relationships, while only several studies accounted for the intervariable relationships among multiple climate variables. Wang and Tian (2022) used a Super Resolution Deep Residual Network (SRDRN) deep learning model to simultaneously bias correct and downscale minimum temperature (T_{\min}) and maximum temperature (T_{\max}) of 20 CMIP6 GCMs, showing that the intervariable relationship between T_{\min} and T_{\max} is well captured, while statistical multivariate bias correction methods experienced unrealistic artifacts (e.g., showing $T_{\min} > T_{\max}$ in the downscaled and bias-corrected fields). Fulton et al. (2023) used a generative adversarial network (GAN) based model, namely the Unsupervised Image-to-Image Translation (UNIT) neural network architecture, to simultaneously bias correct five climate variables. The results indicate that the UNIT model creates fewer extreme values than the target distribution and applying the quantile mapping (QM) method to postprocess the UNIT model outputs show the optimum outcome. This method, however, has a strict stationary assumption between model training (past) and testing (future) periods and may face great challenges when a distinct trend exists between past and future periods, considering climate non-stationarity, particularly when applying to the GCMs/ESMs projections. This, in fact, is a common issue of using deep learning for climate change projections due to its inability to extrapolate beyond their training data to unseen circumstances where system has considerably changed (de Silva et al. 2020; Read et al. 2019; Wi and Steinschneider 2022).

This study shows an integrated trend-preserving deep learning approach can address the spatial and intervariable dependences and climate non-stationarity issues for downscaling and bias correcting GCMs/ESMs. Here we combine the SRDRN model (Wang et al. 2021) with a trend-preserving quantile delta mapping approach (QDM; Cannon 2018) for bias correcting and downscaling six primary climate variables at once, including daily precipitation, maximum temperature, minimum temperature, relative humidity, solar radiation, and wind speed, from five state-of-the-art GCMs/ESMs in the Coupled Model Intercomparison Project phase 6 (CMIP6) over the CONUS. This trend preserving deep learning approach, hereafter referred to as SRDRN-QDM,

accounts for spatial and intervariable relations and non-stationarity. We show SRDRN-QDM greatly improve state-of-the-art GCMs/ESMs by significantly reducing biases of individual variables, and better reducing biases in spatial and multivariate dependences compared to the current multivariate bias correction approach, and better reducing biases in extreme events compared to the current deep learning downscaling and bias correction approach. We performed comprehensive evaluations of the SRDRN-QDM downscaled and bias-corrected variables, as well as drought characteristics derived from the six downscaled and bias-corrected variables, in comparison with the current approaches. The structure of this paper is organized as follows: Sect. 2 introduced data and methodology, including the SRDRN model, QDM method to preserve trend, and calculation of drought index. Section 3 presents results; discussion and conclusions are provided in Sect. 4.

2 Data and methodology

2.1 Data and study area

In this study, we consider historical simulations from five state-of-the-art CMIP6 GCMs/ESMs (Eyring et al. 2016) and six climate variables, including daily accumulative precipitation, minimum temperature at 2-m, maximum temperature at 2-m, relative humidity at 2-m, solar radiation, and wind speed at 10-m. The five GCMs/ESMs were developed by major climate centers all over the world and have different spatial resolutions varying from 0.7° to 2.5° (see Table 1). These selected variables are commonly used for climate impact studies such as drought (e.g., Ahmadalipour et al. 2017; Haile et al. 2020; Lee et al. 2019), wildfire (e.g., Bedia et al. 2013; Brown et al. 2023; Grose et al. 2014) and crop failures (e.g., Goulart et al. 2023, 2021; Schillerberg and Tian 2023). While most of the five CMIP6 GCMs/ESMs include multiple ensemble members (named with *r_{in}p_nf_n*, where *r* represents realization, *i* represents initialization method, *p* represents physics, *f* represents forcing, and *n* can be different numbers), only one single member (*r1i1p1f1*) for each model was used in this study for fair comparisons. Prior to bias correction and downscaling, the GCM/ESM

outputs were re-gridded into a common 1° spatial resolution with bilinear interpolation.

The European Center for Medium-Range Weather Forecast's (ECMWF) ERA5 dataset was used as high-resolution observational references (Hersbach et al. 2020), which has a spatial resolution of 0.25° . The overlapped period from 1979 to 2014 was used for both GCMs and ERA5 reanalysis data. The multivariate bias correction and downscaling experiments were performed over the CONUS, where a wide range of climatic zones exist, including the temperate to continental climate in the Northeast, subtropical climate in the Southeast, and Mountains and Great Plains climate in the West as well as oceanic climate at the western coast from the Pacific Ocean and Gulf of Mexico in the Southeast.

2.2 Improving SRDRN for multivariate bias correction and downscaling

The SRDRN model is a deep learning architecture inspired by a novel single-image super-resolution study in the computer vision field (Ledig et al. 2017). This model has been tested for downscaling daily precipitation and temperature individually through synthetic experiments (Wang et al. 2021), bias correcting minimum temperature and maximum temperature at once for GCM/ESM outputs (Wang and Tian 2022), and customized for bias correcting and downscaling hourly precipitation from reanalysis data using radar observations (Wang et al. 2023). These previous studies have demonstrated its superiority over conventional deep learning approaches. Compared with the widely used U-Net architecture (e.g., Sha et al. 2020; Sun and Tang 2020), the SRDRN model directly extracts features at the coarse resolution input; therefore, the coarse resolution input data do not need to be firstly interpolated into higher resolution as observational references, resulting in the decrease of computational and memory complexity.

The model is mainly comprised of residual blocks and upsampling blocks with CNN layers, parametric ReLU activation function, and batch normalization layers. The residual blocks enable a very deep architecture that has notable advantages for extracting fine spatial features without degradation issues (He et al. 2016). The upsampling blocks equipped with upsampling 2D layers enable the model to

Table 1 Selected GCM information

No	GCMs	Horizontal resolution	Institutions
1	EC-Earth3	$0.7^\circ \times 0.7^\circ$	European EC-Earth Consortium (Europe)
2	GFDL-ESM4	$1.0^\circ \times 0.8^\circ$	Geophysical Fluid Dynamics Laboratory (USA)
3	IPSL-CM6A-LR	$1.26^\circ \times 2.5^\circ$	Institute Pierre Simon Laplace (France)
4	MPI-ESM1-2-HR	$0.94^\circ \times 0.94^\circ$	Max Planck Institute for Meteorology (Germany)
5	MRI-ESM2-0	$1.125^\circ \times 1.125^\circ$	Meteorological Research Institute, Japan Meteorological Agency (Japan)

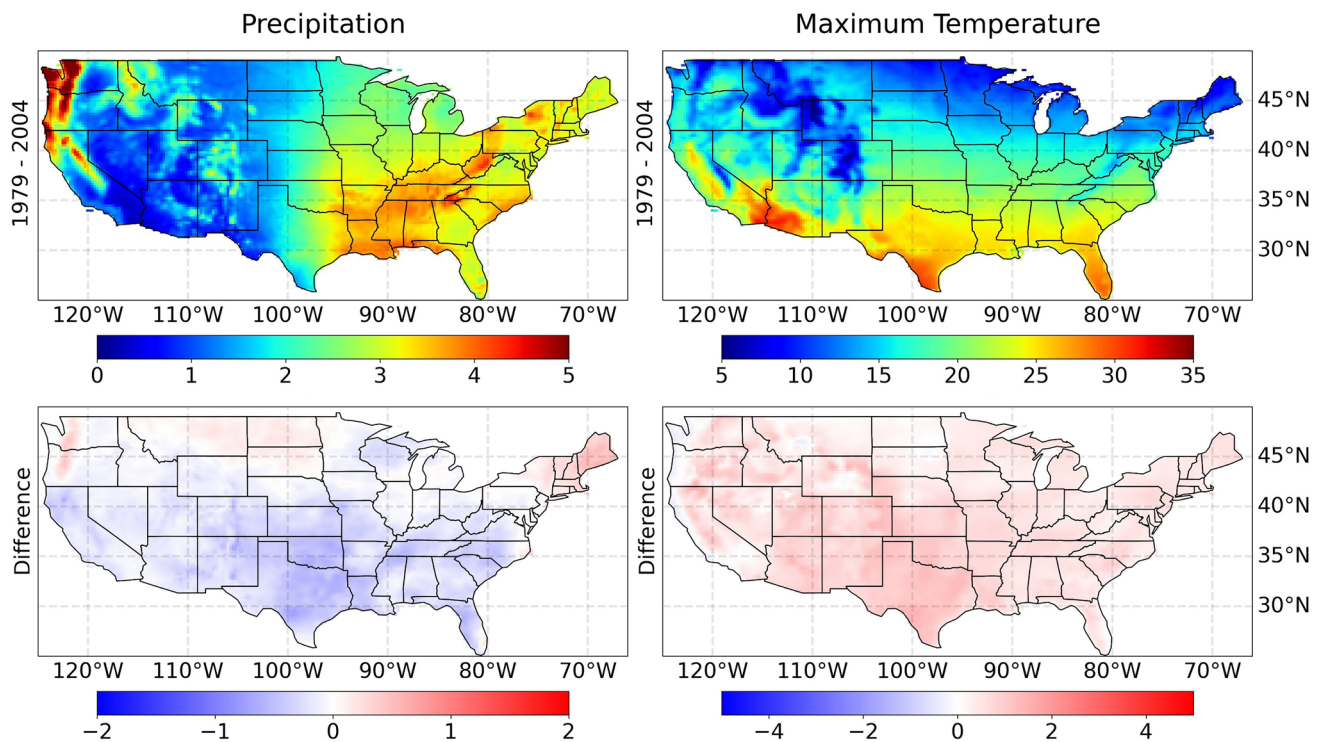


Fig. 1 Climatology of daily precipitation (1st column) and maximum temperature (2nd column) for ERA5 from 1979 to 2004 (1st row) and differences (2nd row) between the period of 2005 to 2014 and the

period of 1979 to 2004. Units for the color bars are mm for precipitation and °C for maximum temperature

have downscaling capability for generating high-resolution data. Each upsampling block sequentially and gradually increases the input coarse resolution feature maps by a factor of 2 or 3. In this study, the downscaling ratio (the ratio of spatial resolutions between GCMs/ESMs and ERA5) is 4, and thus, we used two upsampling blocks, with each increasing spatial resolution by a factor of 2. There are 37 CNN layers in this architecture including one CNN layer in the first layer, 33 CNN layers in the residual blocks, two CNN layers in the upsampling blocks and one CNN layer for reconstruction at the end. For more details about the original model architecture, the readers refer to Wang et al. (2021). Through a series of additional tests, we made modifications to the SRDRN architecture to achieve optimum results for multivariate bias correction and downscaling. Specifically, the number of filters for the CNN layers in the upsampling blocks increased from 256 to 512, and a spatial 2D channel-wise dropout layer was added at the end of the second upsampling block based on suggestions from Kong et al. (2022). These adjustments increase model complexity for handling six variables at once but also increase generalization ability to battle overfitting issues (Srivastava et al. 2014).

Data normalization was executed as a data preprocessing step. Specifically, each variable was normalized by

subtracting the grid mean and dividing by the grid standard deviation at each grid point, which is different from our previous studies (Wang and Tian 2022; Wang et al. 2023; Wang et al. 2021) in which variables are normalized with mean and standard deviation calculated based on the flattened variable of all the grid points (i.e., the flattened vectors). Our tests indicate that normalization with grid mean and grid standard deviation better captures spatial variability at the continental scale potentially because these grid-based parameters accurately attain the long term climatological features at local scale, which is very important for retaining spatial variability at the large scale. Note that precipitation data was firstly logarithmically transformed to reduce skewness (Sha et al. 2020b; Wang et al. 2023) before being normalized.

2.3 SRDRN-QDM trend-preserving deep learning

Global warming continues to increase under different scenarios and modeled pathways, causing increased extreme events (IPCC 2023, Lee et al. 2023). Figure 1 shows the climatology differences between the period of 1979 to 2004 and the period of 2005 to 2014 for daily precipitation and maximum temperature over the CONUS based on the observational reference dataset (i.e., ERA5). We can see that precipitation climatology decreased as much as 20% in the south of CONUS,

while maximum temperature climatology increased in a similar area as large as 1 °C. These differences between historical and future climate will likely be greater, especially under the higher emission scenarios (IPCC 2023, Lee et al. 2023). However, pure data-driven deep learning models have experienced difficulties in inferring testing dataset that have different distributions with the training dataset (e.g., non-stationary data such as climate) (Arik et al. 2022; Li et al. 2022), and have difficulties capturing extreme events that are never or rarely seen in the training dataset (e.g., Wilson et al. 2022). This is a common issue of using deep learning for climate change projections due to its inability to extrapolate beyond their training data to unseen circumstances where system has considerably changed (de Silva et al. 2020; Read et al. 2019; Wi and Steinschneider 2022).

In order to address this issue, we combined the adjusted trend-preserving quantile delta mapping approach (QDM; Cannon et al. 2015) with SRDRN deep learning approach for better capturing climate trends and extremes (denoted as SRDRN-QDM). Here we take the precipitation variable as an example. Firstly, we calculate the non-exceedance probability associated with the value in the projection period at time t for the SRDRN model output, $\tau_{SRDRN,p}(t)$,

$$\tau_{SRDRN,p}(t) = F_{SRDRN,p}^{(t)}[x_{SRDRN,p}(t)] \quad (1)$$

where $F_{SRDRN,p}^{(t)}$ is the estimated empirical cumulative density function (CDF) over a time window around t in the projection period for SRDRN model output. $x_{SRDRN,p}(t)$ is the precipitation value from the SRDRN model at time t in the projection period. Secondly, we calculate the relative change of precipitation value at the GCM CDFs between the projection and historical periods based on GCM model data, $\Delta_{GCM}(t)$,

$$\Delta_{GCM}(t) = \frac{F_{GCM,p}^{(t)-1}[\tau_{SRDRN,p}(t)]}{F_{GCM,h}^{(t)-1}[\tau_{SRDRN,p}(t)]} \quad (2)$$

where $F_{GCM,p}^{(t)-1}$ ($F_{GCM,h}^{(t)-1}$) is the inverse CDF of GCM outputs in the projection (historical) period. $F_{GCM,p}^{(t)-1}[\tau_{SRDRN,p}(t)]$ is the precipitation value in the projection period corresponding to $\tau_{SRDRN,p}(t)$, while $F_{GCM,h}^{(t)-1}[\tau_{SRDRN,p}(t)]$ is the precipitation value in the historical period corresponding to the same probability $\tau_{SRDRN,p}(t)$. Similar to other trend preserving studies (Cannon 2016, 2018; Lange 2019), here we assume the trends from the GCM outputs ($\Delta_{GCM}(t)$) are realistic and no trend biases exist. Thirdly, the SRDRN model output is also bias-corrected with quantile mapping based on the observed data.

$$\hat{x}_{o:SRDRN,h:p}(t) = F_{o,h}^{-1}[\tau_{SRDRN,p}(t)] \quad (3)$$

where $\hat{x}_{o:SRDRN,h:p}(t)$ is quantile mapping bias-corrected precipitation for SRDRN model output. $F_{o,h}^{-1}$ is the inverse CDF for the observed data in the historical period. Finally, the QDM bias-corrected result, $\hat{x}_{SRDRN,p}(t)$, is calculated by multiplying the relative change $\Delta_{GCM}(t)$,

$$\hat{x}_{SRDRN,p}(t) = \hat{x}_{o:SRDRN,h:p}(t) \cdot \Delta_{GCM}(t) \quad (4)$$

The time window to construct the empirical CDF around time t was set to be 45 days to preserve the seasonality. Taking 20 years of historical period and 10 years of projection period as an example, the total number of days for constructing the empirical CDF will be 1820 days $[(45 + 45 + 1) \times 20]$ for the historical period and 910 $[(45 + 45 + 1) \times 10]$ days for the projection period. Since the 45 days are moving with time t moving and the days within the 45 days are also used to construct empirical CDF at other times around time t , any changes within the 45 days would not be neglected. In this study, we consider precipitation, relative humidity, and wind speed as relative changes in quantiles in Eqs. (2 and 4). To preserve absolute changes in quantiles, the Eqs. (2 and 4) can simply be applied additively rather than multiplicatively. The variables of minimum temperature, maximum temperature, and solar radiation are considered absolute changes in this study.

2.4 Model training

We first performed SRDRN bias correction and downscaling using the first 26 years (1979 to 2004) as the training dataset and the remaining ten years (2005 to 2014) as the testing dataset. We used grid mean and grid standard deviation calculated from the training dataset to normalize the training data and used the same ones for denormalization during the inference period. The parameters of grid mean and standard deviation are obtained from all year round training dataset without season separations. Similar to Wang and Tian (2022), we stacked the five GCMs daily data with six channels, which greatly augments the data size and also allows the model to consider inter-model variability. The referenced data of ERA5 were replicated and stacked to match each set of GCMs. While GCM outputs are not synchronized in time with ERA5, we synchronously paired coarse resolution data from GCMs and observations and assumed that the SRDRN has the capability of reproducing distributions of the observations if synchronized biases are well corrected. The GCM GFDL-ESM4 (see Table 1) used the 365-day calendar, while ERA5 and the other four GCMs used Gregorian calendar. In order to synchronize GCMs and ERA5 for model training, we removed the data on days of Feb. 29 so that all the GCMs and ERA5 to have 365 days each year. The mean absolute error (MAE) was chosen as the loss function.

For the channel of precipitation, weighted MAE was used according to Wang et al. (2023) to better balance the precipitation data and weights w were calculated following,

$$w = \begin{cases} 0.1 & y'_{true} \leq 0.1 \\ y'_{true} & 0.1 < y'_{true} < 1.0 \\ 1.0 & y'_{true} \geq 1.0 \end{cases} \quad (5)$$

where y'_{true} is the natural log transformed ERA5 precipitation scaled by dividing the maximum value of natural log transformed ERA5 precipitation. For other channels, MAE loss is used. The Adam optimization algorithm was used to train the network with a learning rate of 0.0001, and default values for other parameters were used during model training. The mini-batch size of 64 was used, and the number of epochs was set to 160. We applied QDM as introduced above to each output variable from SRDRN to better preserve trends and extremes. The historical and projection periods mentioned in the previous section correspond to the training and testing periods in this study, respectively. The model was trained with approximately 1.2×10^5 iterations and was executed using NVIDIA V100 GPU provided by the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS), formerly known as the Extreme Science and Engineering Discovery Environment (XSEDE) (Towns et al., 2014).

2.5 Model evaluations

We evaluated the SRDRN-QDM model performance for each variable as well as spatial and intervariable relationships against ERA5 observations during the testing period. The agreements between the modeled and observed values were quantified by the root mean squared error (RMSE), Kling-Gupta efficiency (KGE) and Kolmogorov–Smirnov (KSS) statistics. The KGE is an overall performance metric combining correlation, bias, and variability (Kling et al. 2012), representing perfect agreement when it equals 1. The KSS statistic is used to test whether the modeled sample came from the same distribution as the referenced data, which has been used in climate downscaling and bias correction (e.g., Quesada-Chacón et al. 2023). Besides evaluating each individual variable as well as spatial and intervariable relationships, we also take drought as an example to evaluate the SRDRN-QDM performance for capturing climate extreme events that are characterized by multiple variables. We used a multivariate drought index, the standardized precipitation evapotranspiration index (SPEI; developed by Vicente-Serrano et al. 2010), to examine the inter-variable physical coherence of involved key essential variables, i.e., the six variables, including precipitation, maximum temperature, minimum temperature, relative humidity, solar

radiation, and wind speed from the SRDRN-QDM outputs. The procedure of SPEI calculation involves a climatic water balance, and it considers both the role of precipitation and evaporation in drought assessment. SPEI is based on variations in the deficit of precipitation and potential evapotranspiration (P-PET). Various methods have been proposed for calculating PET, and it has been shown that the Penman–Monteith (PM) approach provides more accurate results due to a more physically-based formulation of atmospheric evaporative demand (Donohue et al. 2010). Therefore, our PET is calculated based on the FAO-56 PM equation (FAO 56 PM; Allen et al. 1998), which is recommended by the World Meteorological Organization (WMO) as the standard method for estimating PET. The FAO-56 PM equation requires five variables: minimum temperature, maximum temperature, solar radiation, relative humidity, and wind speed. Therefore, we calculated daily PET according to the FAO-56 PM equation with the five bias-corrected and downscaled variables. Based on Vicente-Serrano et al. (2010), monthly precipitation and PET are used to calculate the climatic water deficits. Thus, we aggregated daily precipitation and daily PET into the monthly timescale. It is worth noting that the calculated climatic water deficits at the monthly timescale can be aggregated at different time scales. In this study, we focus on the monthly timescale for short- or long-term drought analysis (Ansari et al. 2023). After calculating monthly climatic water deficits, normalization is performed based on a log-logistic probability distribution to obtain the SPEI series. The log-logistic distribution is used and recommended by many researchers (e.g., Ansari et al. 2023; Vicente-Serrano et al. 2010). The R package ‘SPEI’ was used to calculate SPEI in this study (Beguiría and Vicente-Serrano 2017). As a summary, Fig. 2 outlines the overall deep learning-based framework of multivariate bias correction and downscaling for drought assessment.

3 Results

We first present the effects of QDM on adjusting distributions for the outputs of SRDRN and statistics at 1st, 33rd, 66th, and 99th percentiles for each variable. Then, we show the SRDRN-QDM model performance for individual variables on climatological statistics. Bias reduction is quantified by comparing the SRDRN-QDM bias-corrected and downscaled results with the bilinear interpolation of raw GCMs without bias correction (named Bilinear). We also include two state-of-the-art trend preserving multivariate bias correction methods, including MBCn (Cannon 2018) and Inter-Sectoral Impact Model Intercomparison Project version 3 (ISIMIP3) (Lange 2019). Furthermore, the SRDRN-QDM model performance on reducing biases of spatial and intervariable dependences is presented and compared with

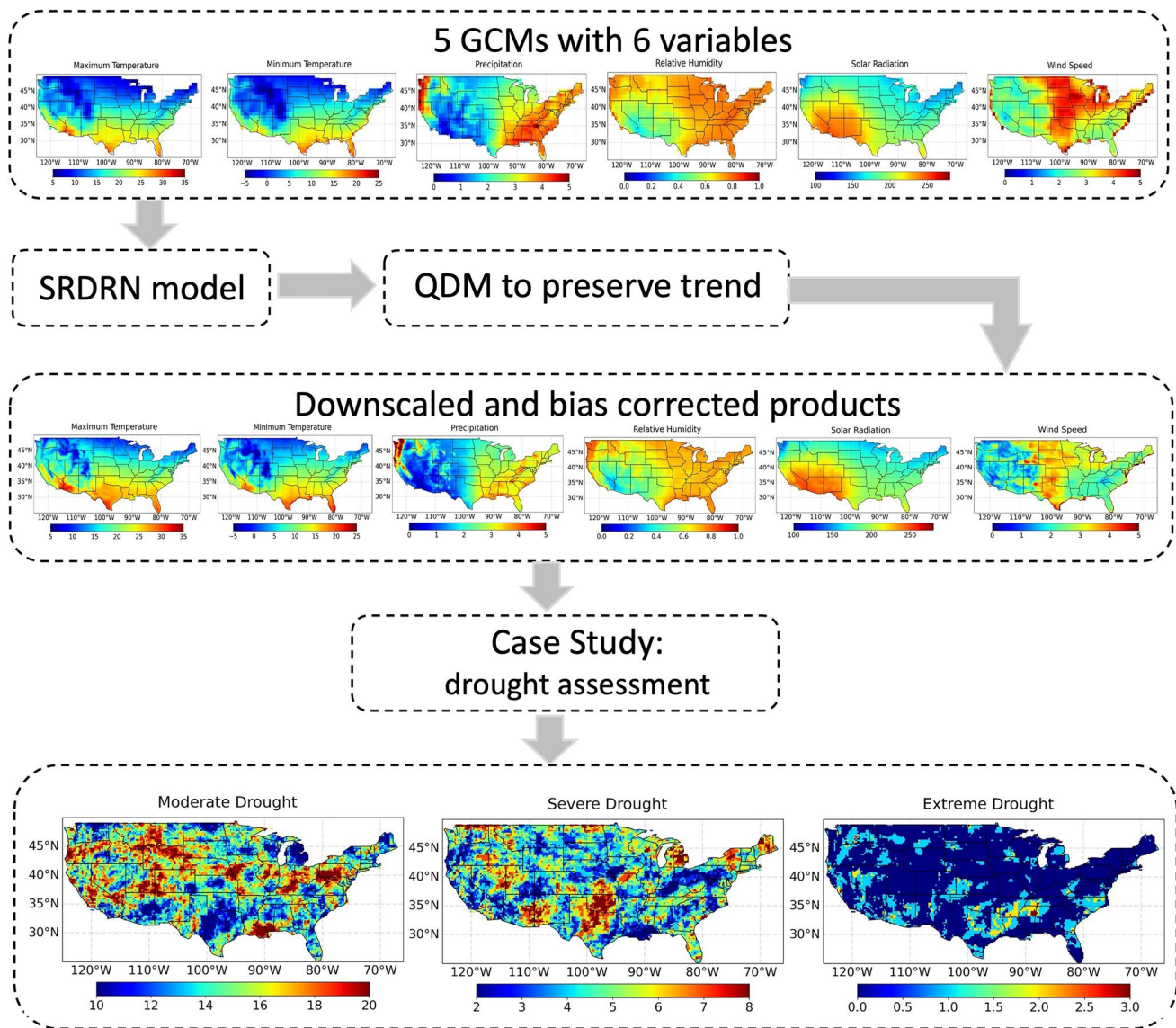


Fig. 2 Schematic of the experiment for downscaling and bias correcting six variables using SRDRN-QDM model for five GCMs/ESMs

MBCn and ISIMIP3. Finally, we present a drought assessment based on the SPEI index as a case study.

3.1 Effects of QDM

Figure 3 shows the probability distributions of all the six variables flattened in the spatial and temporal dimensions from ERA5, five raw GCMs, bias-corrected by SRDRN and SRDRN-QDM. There are large biases between the ERA5 and raw GCMs at the extreme percentiles particularly for precipitation and wind speed variables. The SRDRN model reduces the occurrence of both high and low extremes and shifts more of the distributions toward ERA5's central peak for all the variables. The

SRDRN-QDM approach, however, greatly improved the distributions and well-matched distributions from ERA5 for all the six variables, particularly at the extreme percentiles (less than 1st percentile and greater than 99th percentile). QDM is explicitly designed to match one-dimensional distributions at grid point, while the SRDRN deep learning model matches the distributions only as an emergent feature of optimizing its loss function and tends to neglect the small portion of data examples that occurred extremely infrequent (e.g., < 1%), making it especially challenging to correct large biases at extremes between the coarse resolution GCMs/ESMs and fine resolution ERA5 observational reference.

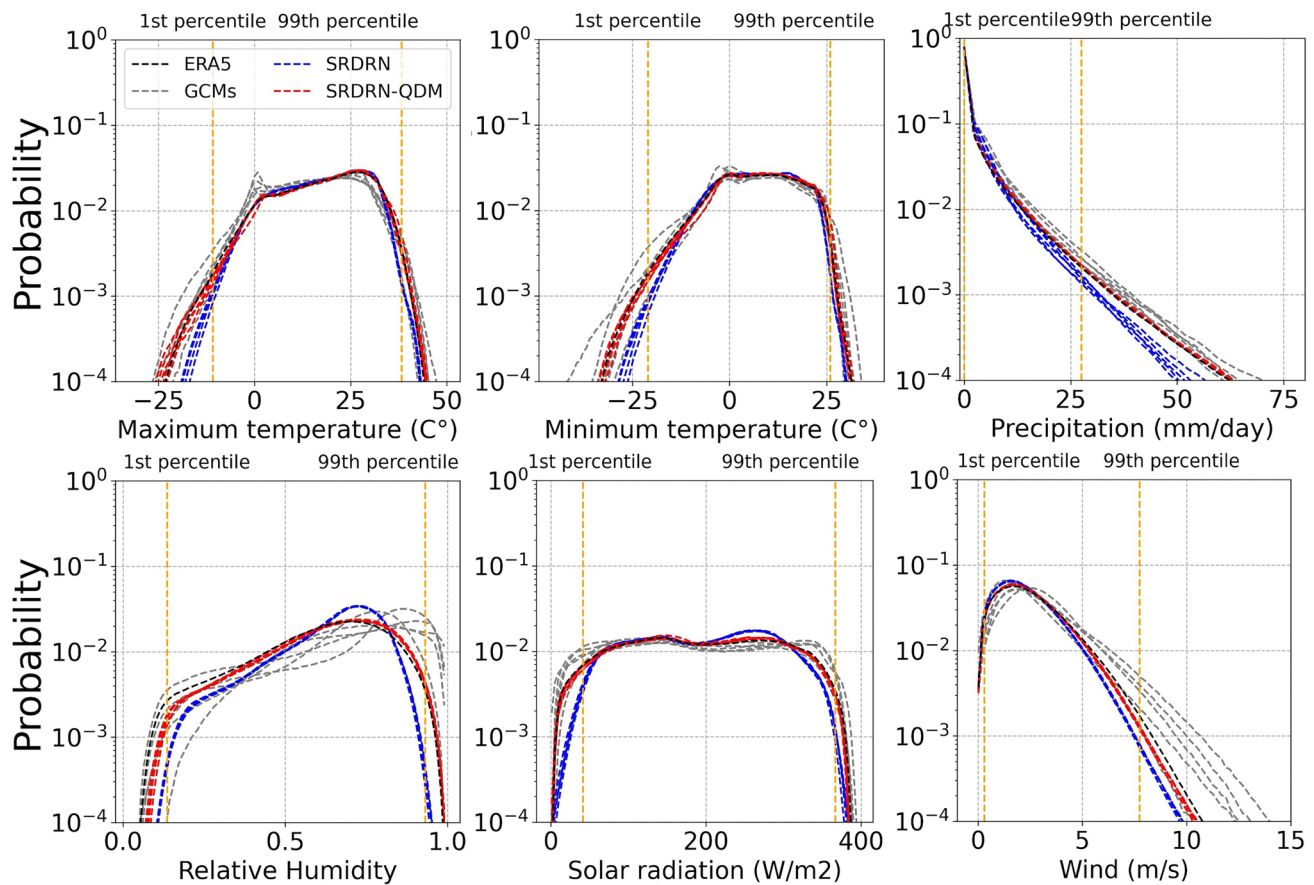


Fig. 3 Probability distributions of maximum temperature, minimum temperature, precipitation, relative humidity, solar radiation and wind speed for 5 raw GCMs and bias-corrected models (SRDRN and

SRDRN-QDM) as well as the referenced ERA5. Note that the normalized histogram in the y axis is a log scale so that the differences in the distribution can be better seen

3.2 Overall performance

We further evaluated the performance of each variable at four different percentiles (1st, 33rd, 66th, and 99th percentiles) from SRDRN-QDM, compared to bilinear interpolation of GCMs/ESMs without bias correction (Bilinear). Table 2 shows the RMSE and KGE values between the models (bias-corrected products of the ensemble mean of the five GCMs/ESMs) and referenced data at the four percentiles. As shown in the table, the SRDRN-QDM model greatly reduced RMSE and increased KGE values at all four percentiles, indicating that the model captures the distributions of each individual variable at each grid point. Taking RMSE at the extreme 99th percentile as an example, the SRDRN-QDM model reduced biases of 82.2% for relative humidity, 75.2% for solar radiation, 70.5% for maximum temperature, 80.7% for minimum temperature, 85.0% for wind speed and 54.6% for precipitation, respectively. The increases in KGE values between SRDRN-QDM and Bilinear at the extreme percentile are substantial, particularly for the variables of relative humidity and wind speed.

Besides Bilinear, we also ran two state-of-the-art multivariate bias correction methods at the grid basis including MBCn and ISIMIP3. Previous studies have demonstrated the competitive performance of the MBCn approach compared to univariate (e.g., QDM) or other multivariate bias correction approaches with limited dimensional data (Cannon 2018; François et al. 2020). Given the deterioration issue of MBCn on handling very high dimensional data (over 6×10^4 dimensions in this study), we applied MBCn at each grid point to bias correct intervariable dependences among the six variables on a grid point basis. The method ISIMIP3 was designed at grid point basis, which firstly applied parametric quantile mapping for bias correction at coarse resolution and then used MBCn for bias correcting the bilinear interpolated variables at the fine resolution (Lange 2019). Table 2 indicates that the KGE for MBCn, ISIMIP3 and SRDRN-QDM are all close to each other for all the six variables at the four percentiles, while ISIMIP3 have relatively higher RMSE compared to SRDRN-QDM and MBCn for all the six variables at most of the four percentiles. The KSS statistics (Table S1

Table 2 RMSE and KGE at 1st, 33rd, 66th, and 99th percentiles

Variables	Model	1st percentile		33rd percentile		66th percentile		99th percentile	
		RMSE	KGE	RMSE	KGE	RMSE	KGE	RMSE	KGE
Relative humidity	Bilinear	0.064	0.85	0.099	0.71	0.12	0.36	0.077	-0.026
	MBCn	0.036	0.88	0.029	0.90	0.0305	0.85	0.015	0.91
	ISIMIP3	0.037	0.91	0.038	0.91	0.0361	0.87	0.0219	0.89
	SRDRN-QDM	0.033	0.89	0.031	0.89	0.027	0.89	0.014	0.92
Solar radiation	Bilinear	16	0.34	11	0.93	11	0.92	9.6	0.57
	MBCn	3.4	0.94	3.4	0.95	3.2	0.95	2.3	0.91
	ISIMIP3	4.7	0.91	3.7	0.97	4.1	0.96	2.7	0.92
	SRDRN-QDM	3.1	0.97	3.5	0.91	3.1	0.96	2.4	0.93
Maximum temperature	Bilinear	2.5	0.41	3.2	0.67	2.6	0.88	2.2	0.81
	MBCn	0.84	0.98	0.61	0.95	0.57	0.92	0.75	0.96
	ISIMIP3	1.2	0.97	0.79	0.95	0.98	0.93	1.3	0.93
	SRDRN-QDM	0.81	0.94	0.59	0.96	0.58	0.93	0.65	0.97
Minimum temperature	Bilinear	2.6	0.85	1.4	0.70	1.5	0.95	2.3	0.88
	MBCn	0.95	0.99	0.53	0.76	0.50	0.94	0.52	0.97
	ISIMIP3	1.27	0.99	0.89	0.75	0.83	0.95	0.87	0.97
	SRDRN-QDM	0.95	0.96	0.66	0.81	0.50	0.93	0.45	0.99
Wind speed	Bilinear	0.12	0.56	0.42	0.47	0.63	0.47	2.2	0.50
	MBCn	0.041	0.93	0.070	0.97	0.12	0.96	0.29	0.94
	ISIMIP3	0.15	0.93	0.20	0.92	0.31	0.93	0.70	0.91
	SRDRN-QDM	0.034	0.95	0.064	0.99	0.14	0.97	0.33	0.98
Precipitation	Bilinear	0	–	0.082	–0.23	0.80	0.28	4.9	0.75
	MBCn	0	–	0.015	0.56	0.19	0.79	2.4	0.95
	ISIMIP3	0	–	0.0289	–2.3	0.26	0.85	2.8	0.94
	SRDRN-QDM	0	–	0.014	0.60	0.18	0.80	2.2	0.96

*RMSE units for relative humidity are unitless, solar radiation is W/m², maximum and minimum temperatures are °C, wind speed is m/s, and precipitation is mm/day

in the Supplementary Information) indicate that most of the distributions at grids from Bilinear are very different from the ones of the referenced ERA5 and SRDRN-QDM greatly increased the percentages of grids that match the distributions of the referenced ERA5. Compared to MBCn and ISIMIP3, SRDRN-QDM has an overall better performance.

Figure 4 shows the box plots of the differences of the model outputs (Bilinear, MBCn, ISIMIP3 and SRDRN-QDM) compared to the observational reference data at the 33rd and 99th percentiles for all six variables from each GCM/ESM simulation, respectively. The differences of the SRDRN-QDM bias-corrected products compared to the referenced data are around 0, with a small spread for each variable and GCM/ESM, indicating that the bias-corrected results from the SRDRN-QDM model well match the distribution of the referenced data, much better compared to Bilinear. Comparing MBCn and SRDRN-QDM, ISIMIP3 has relatively larger spread for most of the GCMs particularly at 99th percentile.

3.3 Climatology

We evaluated the long-term mean (i.e., climatology) during the testing period for all six variables with multi-model (five GCMs/ESMs) ensemble mean compared with Bilinear. Figure 5 shows the climatology mean for each daily variable at each grid point over the CONUS, including the differences between models and reference data (ERA5) during the testing period. We can see that the SRDRN-QDM model greatly reduced biases for all six variables over the CONUS. The effects are more obvious over the regions with complex topographies (Great Plains and Mountains region in the West) in which there are large biases from bilinear interpolated GCMs/ESMs, while the ISIMIP3 model has relatively large climatology mean differences from ERA5 in those areas for all the variables. We further present the root mean squared error (RMSE) and Kling-Gupta efficiency (KGE) statistics for the climatology mean in Table 3 including Bilinear, MBCn, ISIMIP3 and SRDRN-QDM. The results indicate that the SRDRN-QDM model reduced RMSE and

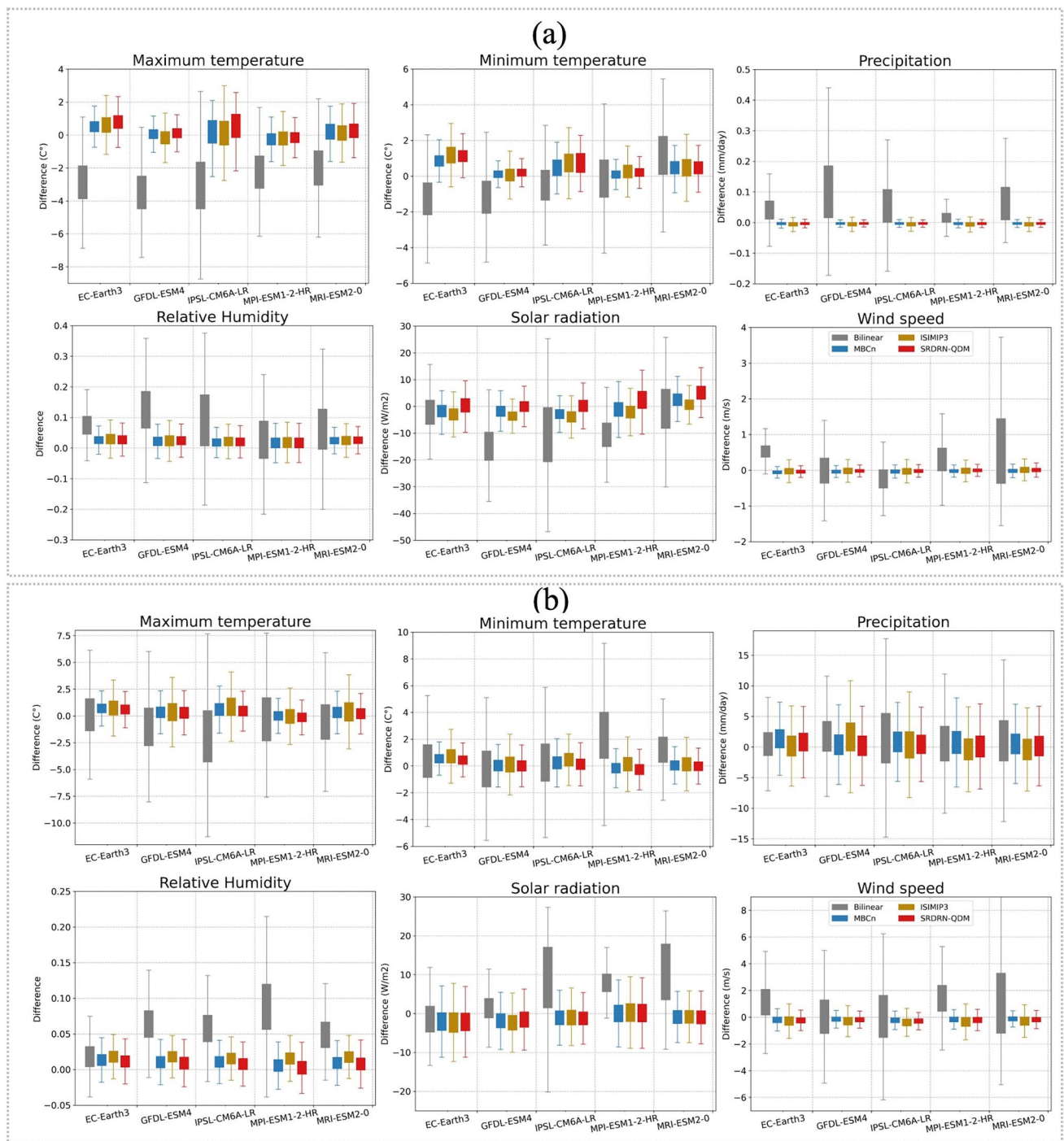


Fig. 4 Box plots for differences between models (Bilinear, MBCn, ISIMIP3 and SRDRN-QDM) and referenced ERA5 at the 33rd percentile **a** and 99th percentile **b** for maximum temperature, minimum

temperature, precipitation, relative humidity, solar radiation, and wind speed for each GCM

increased KGE values of all six variables. SRDRN-QDM reduced RMSE by 71.6% for relative humidity, 77.7% for solar radiation, 83.8% for maximum temperature, 64.4% for minimum temperature, 82.4% for wind speed, and 70.0% for precipitation, respectively. In particular, the increase of KGE

values from the SRDRN-QDM model is tremendous for the variables of relative humidity, wind speed, and precipitation since these three variables are difficult to simulate and have larger biases compared to the other three variables even at the monthly timescale (Xuan et al. 2017).

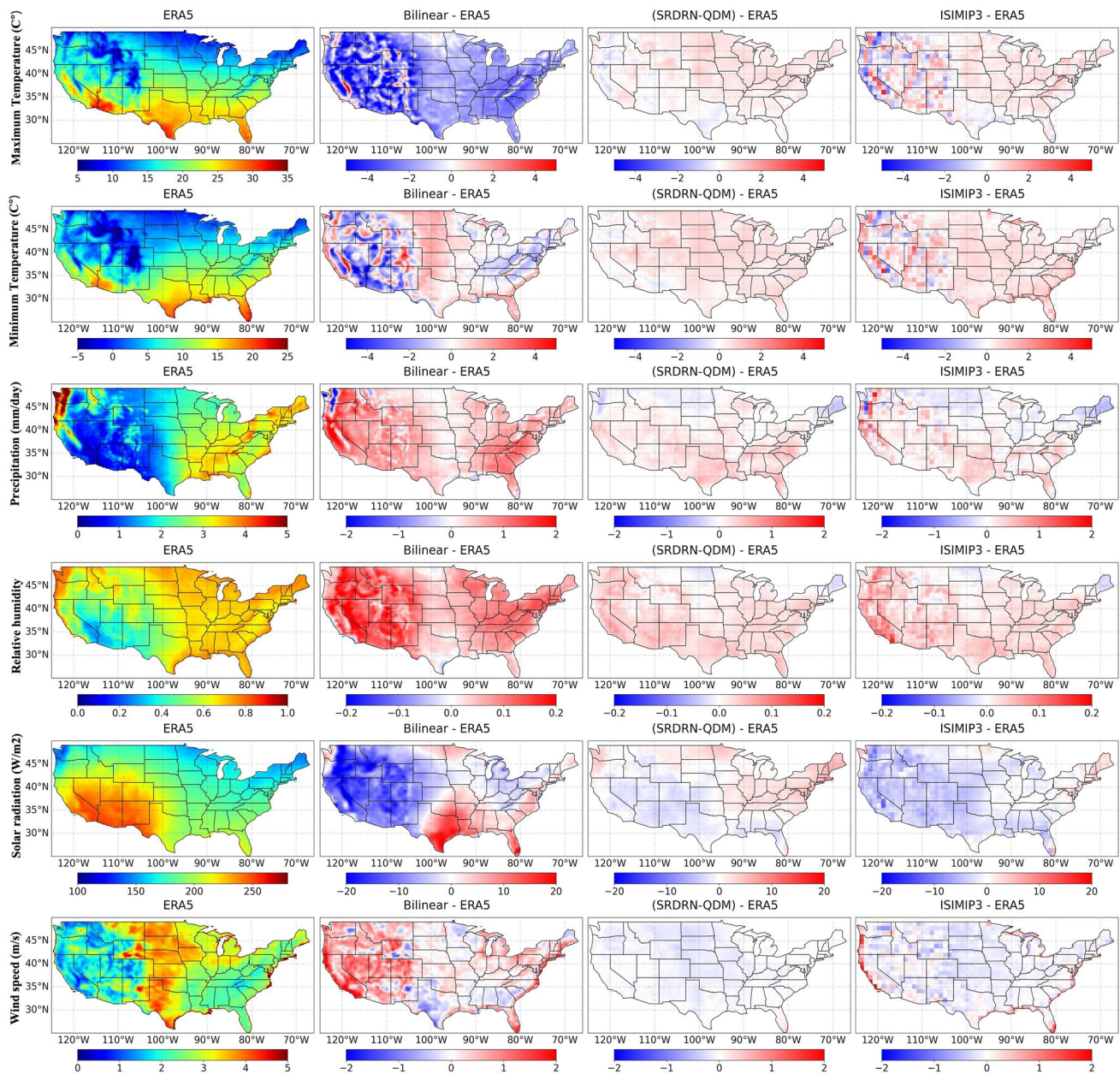


Fig. 5 Climatology means for maximum temperature (1st row), minimum temperature (2nd row), precipitation (3rd row), relative humidity (4th row), solar radiation (5th row), and wind speed (6th row) at each grid point over the CONUS. 1st column is the climatology mean for ERA5, 2nd column is the climatology mean difference between Bilinear and ERA5, 3rd column is the climatology mean difference

between the SRDRN-QDM model and ERA5, 4th column is the climatology mean difference between the ISIMIP3 model and ERA5. Units for the color bars are °C for maximum and minimum temperature, mm/day for precipitation, no unit for relative humidity, W/m² for solar radiation, and m/s for wind speed

Besides the climatology mean, we also evaluated the standard deviation of all six variables in the testing period. Figure 6 indicates that the SRDRN-QDM model greatly reduced biases of standard deviations for all six variables, while the ISIMIP3 model has relatively large standard deviation differences from ERA5, particularly for the variables of maximum temperature, minimum temperature and wind

speed around the boundary areas between the land and the ocean. Table 3 also shows that the SRDRN-QDM model has much lower RMSE and higher KGE values for standard deviation compared to bilinear interpolation over all the grid points in the CONUS. Specifically, SRDRN-QDM has reduced the standard deviation RMSE by 76.0% for relative humidity, 73.3% for solar radiation, 69.8% for maximum

Table 3 RMSE and KGE for climatology mean and standard deviation

Variables	Model	Climatology mean		standard deviation	
		RMSE	KGE	RMSE	KGE
Relative humidity	Bilinear	0.096	0.59	0.036	0.75
	MBCn	0.027	0.89	0.0071	0.98
	ISIMIP3	0.034	0.90	0.0091	0.96
	SRDRN-QDM	0.027	0.90	0.0087	0.96
Solar radiation	Bilinear	8.9	0.91	6.8	0.62
	MBCn	2.2	0.95	1.7	0.96
	ISIMIP3	3.2	0.96	2.0	0.95
	SRDRN-QDM	2.0	0.94	1.8	0.98
Maximum temperature	Bilinear	2.5	0.85	0.81	0.80
	MBCn	0.40	0.96	0.25	0.98
	ISIMIP3	0.78	0.98	0.37	0.92
	SRDRN-QDM	0.41	0.96	0.24	0.97
Minimum temperature	Bilinear	1.4	0.91	0.62	0.94
	MBCn	0.48	0.92	0.28	0.99
	ISIMIP3	0.82	0.92	0.30	0.94
	SRDRN-QDM	0.50	0.96	0.27	0.98
Wind speed	Bilinear	0.58	0.47	0.42	0.51
	MBCn	0.10	0.97	0.09	0.92
	ISIMIP3	0.27	0.92	0.15	0.92
	SRDRN-QDM	0.10	0.97	0.09	0.93
Precipitation	Bilinear	0.66	0.61	1.0	0.71
	MBCn	0.24	0.91	0.46	0.94
	ISIMIP3	0.26	0.94	0.54	0.96
	SRDRN-QDM	0.20	0.93	0.42	0.95

*RMSE units for relative humidity are unitless, solar radiation is W/m², maximum and minimum temperatures are °C, wind speed is m/s, and precipitation is mm/day

temperature, 56.1% for minimum temperature, 77.2% for wind speed, and 58.1% for precipitation, suggesting great improvements. Similar to Table 2, Table 3 shows that KGE values are close to each other for climatology mean and standard deviation for MBCn, ISIMIP3 and SRDRN-QDM, while ISIMIP3 has relative higher RMSE than MBCn and SRDRN-QDM.

3.4 Intervariable dependence

SRDRN-QDM took multiple variables at once as input–output channels (six input variables and six output variables in this study; see Fig. 2), making the model learn very complex relationships beyond our prior knowledge, so that the intervariable dependences can be captured during the training process. In particular, the intervariable correlation between temperature and precipitation has been extensively explored in previous studies (e.g., Guo et al. 2020, 2019; Li et al. 2014), since the biases at intervariable dependences

greatly affect GCM/ESM simulated processes such as snow-melt, evapotranspiration, and runoff generation (Buishand and Brandsma 2001; Immerzeel et al. 2014; Maurer et al. 2010; Mueller and Seneviratne 2014) and further affect climate change on crop yields (Lobell and Field 2007). Figure 7 shows the Spearman correlation coefficients between anomalies of precipitation and maximum temperature at daily time scale in winter months (December to February, DJF) and summer months (June to August, JJA) for the ERA5 observational reference, and differences between model bias-corrected results from the multi-model ensemble mean for Bilinear, MBCn, ISIMIP3 and SRDRN-QDM. We can see that the observed Spearman correlations from the referenced data vary regionally and seasonally. In the winter months, the relationship is positive in the mountain areas in the west and eastern CONUS, while negative around the west coast and middle to high altitudes areas. In the summer months, negative relationship is dominant at southern coastal area, while positive relationship is located at the Southwestern arid area (i.e., South California, Arizona and New Mexico states) and Northeast area around the Great Lakes. The Bilinear product from the 5-model ensemble mostly overestimated the correlations in winter months (RMSE of 0.139) and underestimated the correlation in south central and the arid areas in the west (RMSE of 0.109). MBCn appears reducing more biases for the winter months with lower RMSE of 0.115, while MBCn slightly increased the correlation biases in the summer months with a RMSE of 0.120 particularly at the arid area around Arizona state. ISIMIP3 performs worse than MBCn with a RMSE of 0.151 in winter months and 0.124 in summer months. In contrast, SRDRN-QDM reduces biases in the simulated precipitation-temperature correlation fields in both winter and summer months with a RMSE of 0.107 in winter months and 0.113 in summer months. We also aggregated daily time scale to monthly (see Figure S1 in the Supplementary Information). Most noticeably in the monthly time scale, the positive high correlations (i.e., hot spots) in the middle-to-high altitudes in the winter months and in the Southwestern arid area in the summer months are well enhanced. However, we noticed that SRDRN-QDM still overestimated the correlation in the west coastal areas in both winter and summer months and overestimated the correlation around Florida state in the summer months at both daily and monthly time scale, highlighting the challenges for capturing intervariable relationships at once across various climate regions over the CONUS.

3.5 Spatial dependence

To evaluate the performance of SRDRN-QDM on correcting spatial biases, we calculated map correlation (also called spatial correlation) of day of the year average between

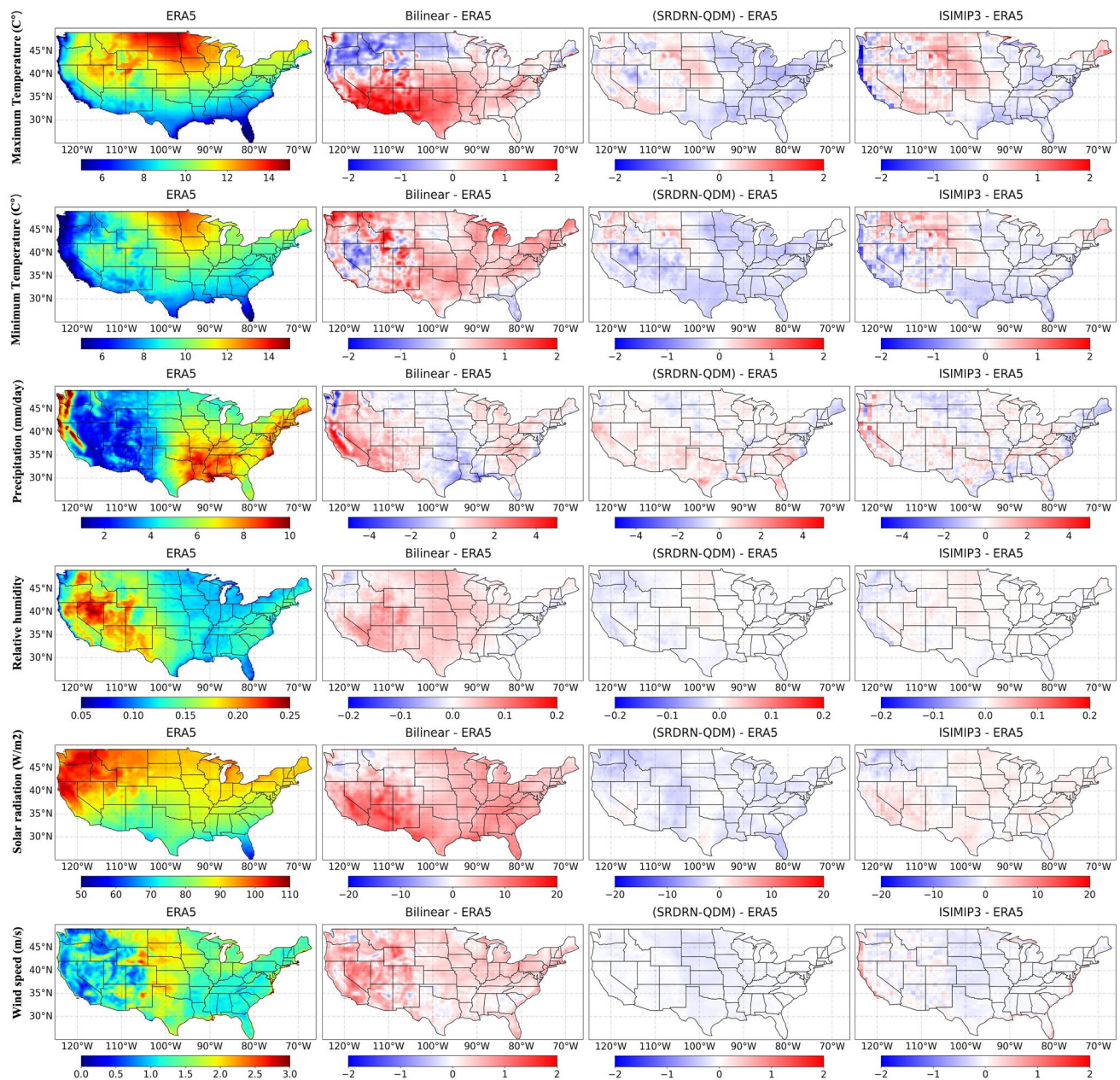


Fig. 6 The standard deviation for maximum temperature (1st row), minimum temperature (2nd row), precipitation (3rd row), relative humidity (4th row), solar radiation (5th row), and wind speed (6th row) at each grid point over the CONUS. 1st column is the standard deviation for ERA5, 2nd column is the standard deviation difference between Bilinear and ERA5, 3rd column is the standard deviation difference between the SRDRN-QDM model and ERA5, and 4th column is the standard deviation difference between the ISIMIP3 model and ERA5. Units for the color bars are °C for maximum and minimum temperature, mm/day for precipitation, no unit for relative humidity, W/m² for solar radiation, and m/s for wind speed

models (Bilinear, MBCn, ISIMIP3 or SRDRN-QDM) and referenced data for each GCM/ESM and each variable (Fig. 8). Figure 8 shows that correlation coefficients for all the six variables are mostly higher than 0.9, while SRDRN-QDM model further increased map correlations for all of the six variables with relatively narrower spread and much better than either MBCn or ISIMIP3. The map correlations

from MBCn are mostly lower than Bilinear, indicating grid-based bias correction process may increase spatial biases. ISIMIP3 performs much worst for precipitation variable, which potentially caused by assuming precipitation distribution follows gamma distribution (Lange 2019). This result suggests that SRDRN-QDM mostly reduced biases on spatial dependence, showing better performance than other two

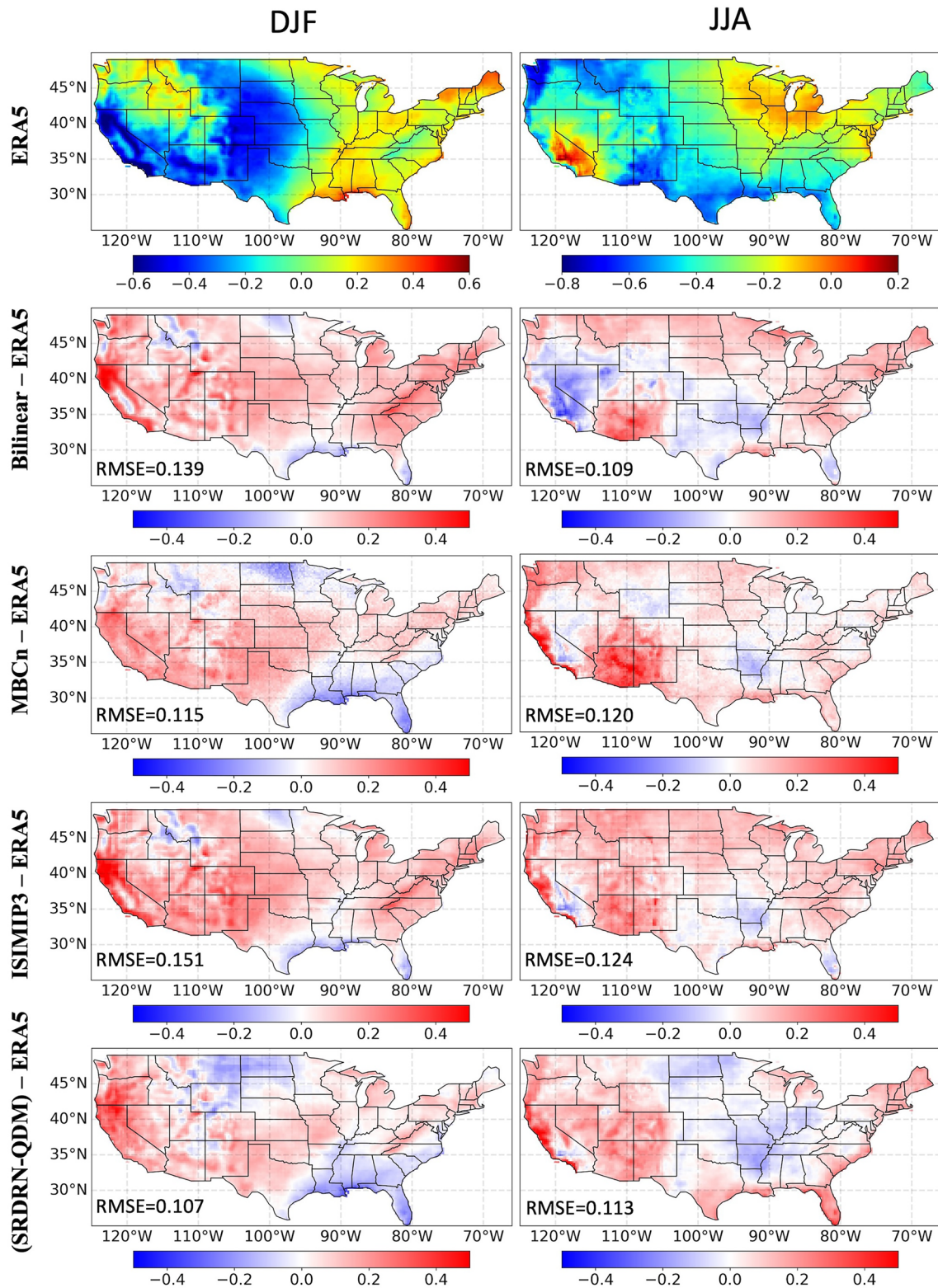


Fig. 7 Spatial distribution of Spearman correlation coefficient between precipitation and maximum temperature for referenced ERA5 (1st row), and difference between bilinear interpolation (2nd

row), bias-corrected by MBCn (3rd row), ISIMIP3 (4th row) and SRDRN-QDM (5th row) with ERA5 in winter days (DJF, 1st column) and summer days (JJA, 2nd column)

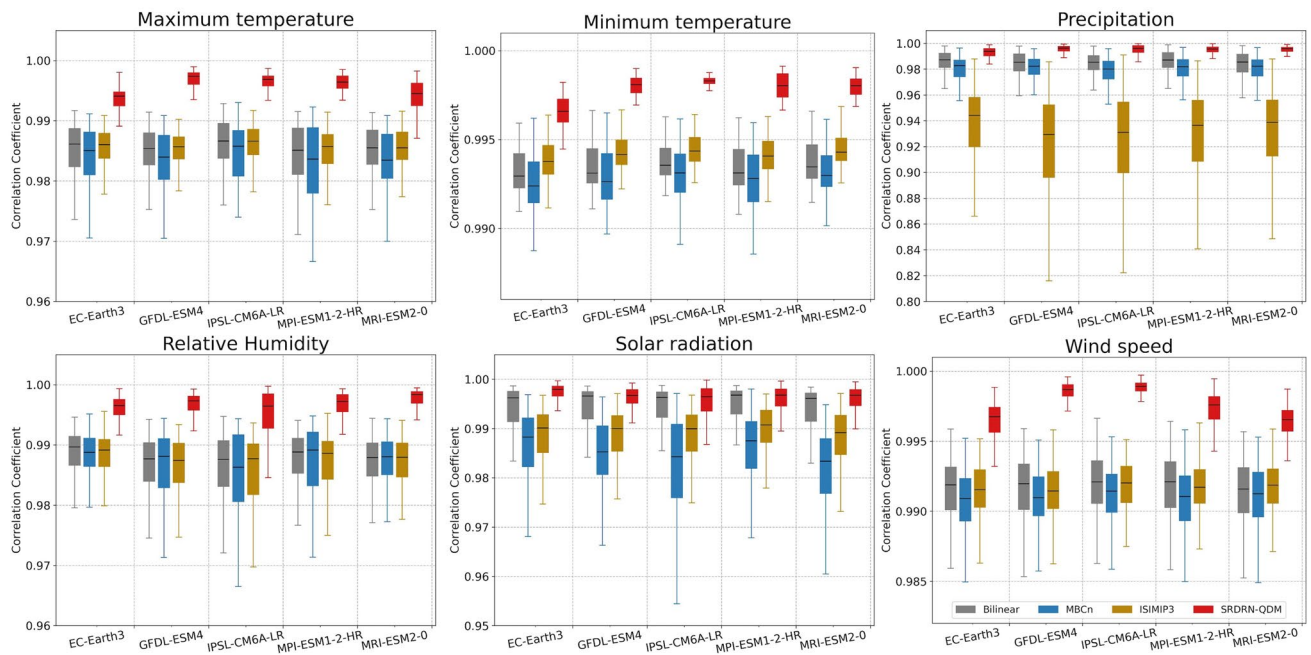


Fig. 8 Box plots for day of the year average map correlation between models (Bilinear and bias-corrected products by SRDRN-QDM and MBCn for each GCM/ESM) and referenced ERA5 for maximum tem-

perature, minimum temperature, precipitation, relative humidity, solar radiation, and wind speed

bias correction methods. This is likely because both MBCn and ISIMIP3 were executed on the grid basis (due to its limitations in handling high dimensional data), which did not account for spatial dependences when performing bias corrections, while SRDRN-QDM includes convolutional layers accounting for spatial patterns between inputs and outputs when learning their relationships.

3.6 Assessing drought

We defined a drought event as a negative SPEI ($\text{SPEI} \leq -1$) lasting for at least one consecutive month. Based on this definition, the duration of most drought events is one month based on the reference data (ERA5) during the testing period. Figure 9 shows the average drought intensity for all the drought events for ERA5 and model outputs (Bilinear, MBCn, ISIMIP3 and SRDRN-QDM) for each GCM/ESM. As shown in the figure, drought intensity varies across the CONUS, and high intensity presents in the southwestern region, the central south, and part of the northeast region. For the same GCM/ESM, the spatial patterns of drought intensity from MBCn and ISIMIP3 models appear to be similar to Bilinear, but very different from ERA5, indicating grid-based bias correction methods are not able to bias correct spatial patterns of drought intensity. Among different GCMs/ESMs, there are dramatic spatial pattern differences of drought intensity for the Bilinear, MBCn and ISIMIP3 models, causing the multi-model ensemble mean

results greatly underestimated the drought intensity over the regions where the observed absolute drought intensity is relatively high while overestimated the drought intensity at locations where the observed absolute drought intensity is relatively low (Figure S4 in the Supplementary Information). The SRDRN-QDM model, however, corrected the spatial patterns of drought intensity from original GCMs, while not exactly matching the ones from the referenced ERA5, but appear close to it, causing the multi-model ensemble mean from the SRDRN-QDM model roughly captured the hot spot areas with high absolute drought intensity (Figure S4 in the Supplementary Information). Nevertheless, we must admit that SRDRN-QDM still experiences difficulties to capture exact spatial patterns and exact locations of high drought intensity from ERA5.

We further classified droughts into three categories based on different thresholds (McKee et al. 1993), and calculated their frequencies for each category. SPEI index of -1 to -1.49 , -1.5 to -1.99 , and less than -2 corresponds to moderate, severe, and extreme drought, respectively. Figure 10 shows the spatial distribution of frequency for the three classified drought conditions from the ERA5 observational reference and model products from a GCM/ESM (EC-Earth3). We find large frequencies of moderate drought over the northern regions (states of Montana, North Dakota, and Minnesota), west coastal states (California and Arizona), western Appalachians, and southeastern regions. The frequencies of severe drought

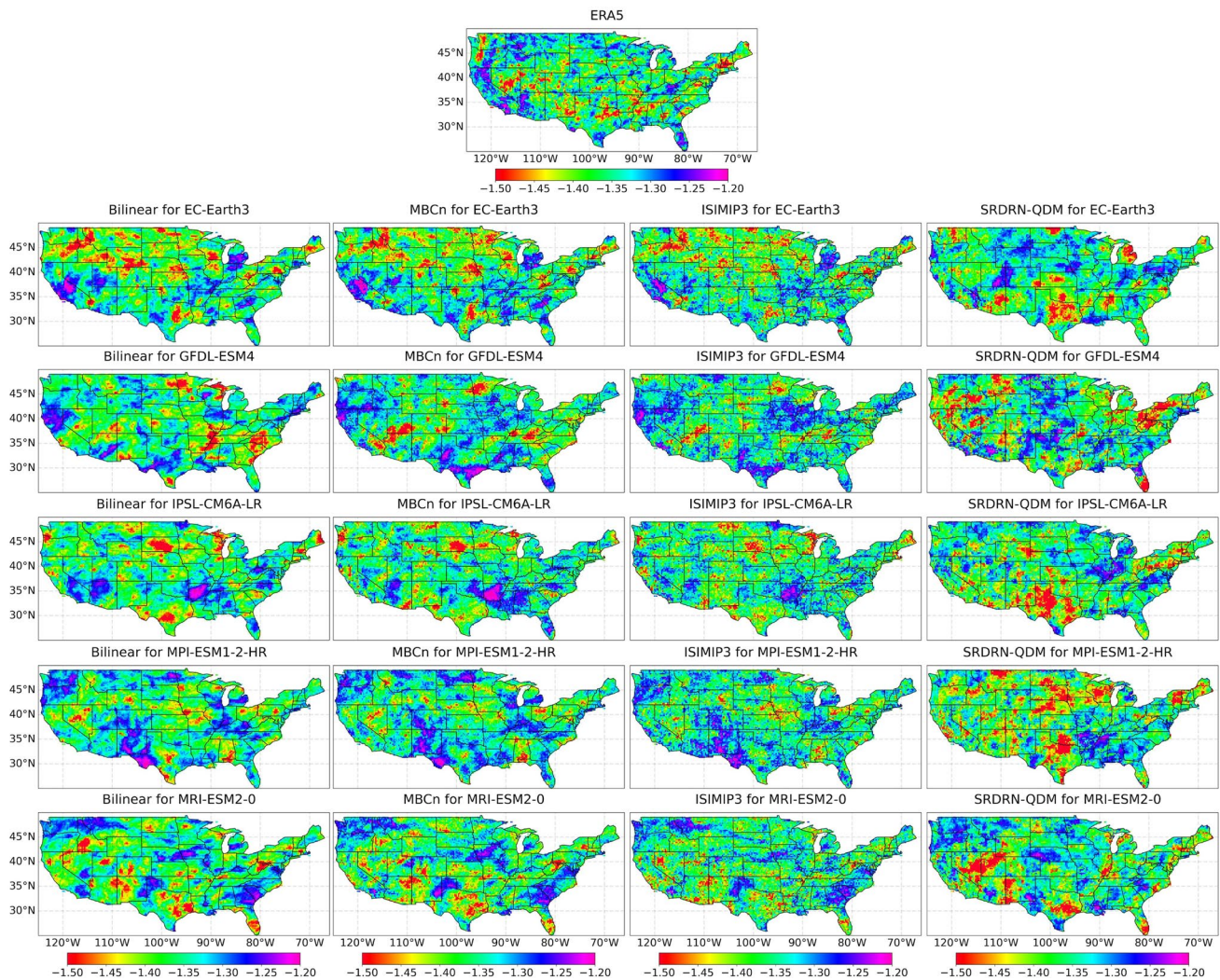


Fig.9 Spatial distribution of average drought intensity for drought events (SPEI index < -1) for referenced ERA5 (top plot), bilinear interpolation (Bilinear, 1st column), bias-corrected by MBCn (2nd column), ISIMIP3 (3rd column), and SRDRN-QDM (4th column)

from each GCM/ESM, including EC-Earth3 (1st row), GFDL-ESM4 (2nd row), IPSL-CM6A-LR (3rd row), MPI-ESM1-2-HR (4th row) and MRI-ESM2-0 (5th row)

appear to be different from the moderate drought conditions, with high frequencies in the central CONUS and around the Appalachians. The high frequencies of extreme drought are scattered around the western and central south regions. The simulations from Bilinear greatly underestimated observed frequencies over the north regions for moderate drought and overestimated observed frequencies over these regions for severe drought. The scattered high frequency spatial pattern for extreme drought from Bilinear appears very different from the observed one. The performance of drought frequencies from the SRDRN-QDM shows improvements over the north regions and western coastal area for the moderate drought, central south region for severe drought, and spatial patterns for the extreme drought appear roughly match with the observations. By

contrast, the spatial patterns of drought frequencies from either MBCn or ISIMIP3 appears to be very close to Bilinear for all three drought categories, suggesting little effects from bias corrections.

Figure 11 shows the distribution of frequency difference between modeled products from the five GCMs/ESMs and ERA5 for moderate and severe drought (the one for the extreme drought category was not plotted due to limited nonzero frequencies). As shown in the figure, frequency differences between modeled products and ERA5 are around zero, indicating simulations and bias-corrected products all captured observed mean frequencies for all five GCMs. Overall, the frequency differences of the SRDRN-QDM have relatively narrower spread around zero compared to MBCn and ISIMIP3 for most GCMs/ESMs for moderate

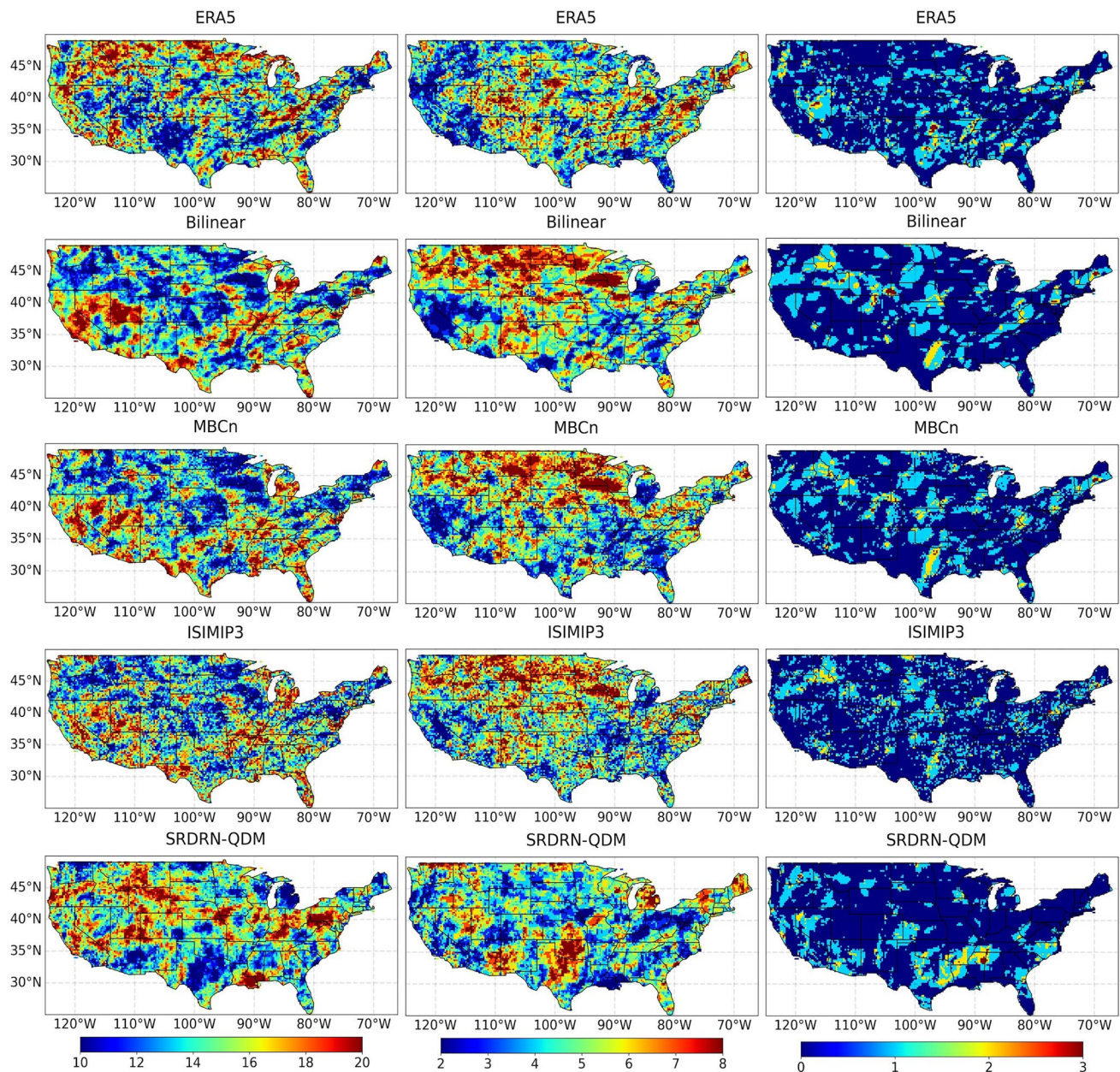


Fig. 10 Frequency spatial distribution of moderate drought (1st column), severe drought (2nd column), and extreme drought (3rd column) for ERA5 (1st row), bilinear interpolation (2nd row), bias-

corrected by MBCn (3rd row), ISIMIP3 (4th row) and SRDRN-QDM (5th row) from a single GCM/ESM (EC-Earth3) with the unit of month for the color bars

drought and for all five GCMs for severe drought, suggesting that SRDRN-QDM reduced more biases of all six variables, resulting in lower biases in drought characteristics. Notably, ISIMIP3 shows greater spread for the severe drought frequency differences.

4 Discussion and conclusions

The study provides a trend-preserving deep learning framework for downscaling and bias correcting multiple variables from GCMs/ESMs at once, accounting for complex spatial and intervariable relations and climate non-stationarity. We presented and evaluated the SRDRN-QDM trend-preserving deep learning for multivariate bias correcting and downscaling daily precipitation, maximum temperature, minimum temperature, solar radiation, relative humidity, and wind

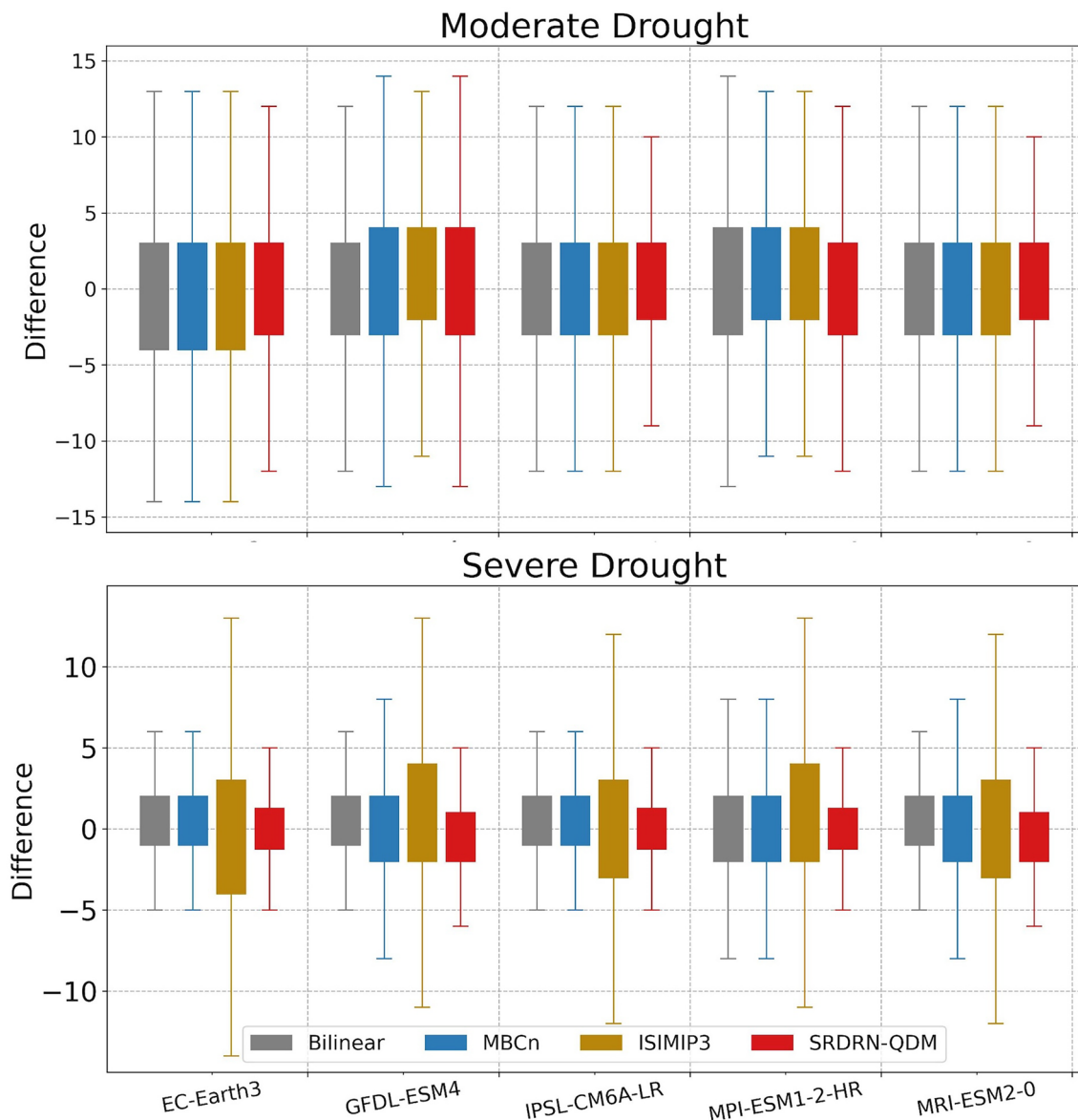


Fig. 11 Box plots of frequency difference (unit: month) between models (Bilinear, MBCn, ISIMIP3, and SRDRN-QDM) and ERA5 for moderate and severe drought categories

speed from five GCMs/ESMs at once over the CONUS. This approach applied the trend preservation approach, quantile delta mapping (QDM) to the SRDRN to adjust distributions at extremes and preserve climate trends. The performance of the six SRDRN-QDM bias-corrected and downscaled variables were comprehensively evaluated for assessing climatology, extremes, spatial dependences, intervariable dependences, and droughts, in comparison with state-of-the-art methods.

The SRDRN-QDM model greatly reduced discrepancies between the model and observations compared to SRDRN only, particularly at the extremes (Fig. 3), and spatial distributions of statistics at the 1st, 33rd, 66th, and 99th

percentiles also match well with the observations (Fig. 4 and Table 2). The model greatly reduced biases in terms of climatology statistics (mean and standard deviation, Figs. 5 and 6, and Table 3). It is worth noting that Quesada-Chacón et al. (2023) explored downscaling and bias correcting seven variables both individually and multivariately using a deep learning model in a small region in Germany, while the evaluations only focused on individual variables. They noted that multivariately trained models tend to focus more on certain variables resulting in better performance on these variables, while performing poorly in the others. However, multivariate bias correction and downscaling is needed, since bias correcting and downscaling individual variables may lose

the physically coherent or intervariable dependences, which is critical for most of the impact studies that need multiple variables at the same time (Cannon 2018; Guo et al. 2020; Zscheischler et al. 2019) or assessing compound events (Zscheischler et al. 2018). To further improve multivariate bias correction and downscaling approach for addressing biases in individual variables and interdependencies, one potential avenue is to bring physical constraints among variables through mass and energy balance into the loss function or customized layers as discussed by Harder et al. (2022).

The SRDRN-QDM model also reduces biases on inter-variable dependencies (e.g., the relationship between precipitation and temperature in Fig. 7) and spatial dependencies for most of the variables (e.g., increased correlation coefficients of map correlation in Fig. 8). The intervariable dependencies are learned during model training process without defining any pre-established functional relationships. The SRDRN-QDM model includes 37 CNN layers and has the potential to capture more complex spatial relationships and correct fine spatial feature differences between model simulations and observations. The SRDRN-QDM model has difficulties to fully handle the complex intervariable dependencies over the Florida peninsular in the summer months at both daily and monthly time scale. This is likely because climate conditions in the Florida peninsular are very different with other CONUS regions. These limitations may be improved by performing the SRDRN-QDM locally in individual climate regions with relatively homogeneous conditions.

Taking drought assessment as an example, the SRDRN-QDM model to some extent reduced biases of the SPEI drought index in terms of both intensities (Fig. 9) and frequencies under moderate, severe, and extreme drought categories (Figs. 10 and 11). In a previous study, the added value of multivariate bias correction methods for the SPEI index was explored by Ansari et al. (2023). The authors found comparable performance for different multivariate bias correction methods, including MBCn, in terms of reduced biases for the SPEI index. In this study, we showed the improved performance of SRDRN-QDM compared to the MBCn and ISIMIP3 multivariate bias correction methods. The spatial patterns of drought intensity and frequency from the SRDRN-QDM model (Figs. 9, 10 and S4 in the Supplementary Information) generally match with the observations (ERA5), while still not precisely capturing the exact hot spot locations. This is potentially due to the challenges for the model to learn across high spatial heterogeneity of climate conditions over the CONUS, and the model has difficulties in reducing biases for all the regions at one time. Training the SRDRN-QDM locally at each climate region has the potential to improve performance on the SPEI index while at the expense of increasing computing cost.

Combining SRDRN with QDM particularly improved the model performance at extremes by considering climate trends simulated by physics based GCMs/ESMs models. SRDRN-QDM model greatly reduced biases at extreme percentiles (less than 1st percentile and greater than 99th percentile) for each variable compared with the SRDRN model output (Fig. 3). While recent studies found that combining GAN-based model with quantile mapping (QM) led to overall improvement (Fulton et al. 2023; Hess et al. 2023), they did not account for climate trends or non-stationarity in these approaches. The importance of GCM trend preservation becomes more significant for GCM projections under a strong anthropogenic signature on the climate (IPCC 2023, Lee et al. 2023). Previous studies have found that impact studies are sensitive to non-stationary biases, and bias correction approaches performed worse in the testing period due to model stationary assumptions (e.g., Chen et al. 2021; Guo et al. 2020). SRDRN-QDM tackles this issue and can be used to bias correct and downscale climate projections from GCMs/ESMs accounting for climate trends or non-stationarity.

As mentioned in the methodology section, we synchronized the referenced ERA5 and model simulations in time, which means daily maps from the 5 GCMs are forced to match with those observed and used MAE as loss function without considering the atmospheric state of the different climate models. As a result, while the results from this work showed notably improved performance for correcting spatial and intervariable dependences, the representation of climate model dynamics at the daily timescale may be affected during the correction procedure, which may restrict the usefulness of the output data for assessing compound extreme events at the daily timescale. This issue may be potentially addressed in the future by modifying loss function to match distributions of climate models with observed distributions instead of day-to-day matches (Tao et al. 2016).

The SRDRN-QDM model treated daily spatial data independently and did not explicitly account for temporal dependence during bias correction. The SRDRN-QDM model is capable of capturing seasonality for all the six variables (Figure S2 in the Supplementary Information), but the SRDRN-QDM slightly underestimated the lag-1 autocorrelations for most of the variables (see Figure S3 in the Supplementary Information). Incorporating time dependence between sequence images by replacing 2-dimension convolutional layers with 3-dimension ones has the potentials to further improve model performance on temporal dependence, which can be explored in the future study.

This study evaluated the SRDRN-QDM model performance on the joint effects of the six variables in terms of drought index SPEI at monthly time scale. However, even at the monthly time scale, notable biases still exist after bias corrections (see Figs. 9 and 10). We also evaluated the

SPEI at the daily scale. However, the results indicate that the SRDRN-QDM greatly overestimated drought intensity in the eastern CONUS and drought duration in the western CONUS, much worse compared to the monthly scale (not shown), which is likely due to larger noises/biases for the joint variability of the six variables at daily time scale. Thus, further work is still needed to improve multivariate aspects of the model performance at the daily time scale.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-024-07406-9>.

Acknowledgements This project was supported in part by the NSF CAREER award (NSF-EAR-2144293), the NOAA RESTORE program (NOAA-NA19NOS4510194), and with federal funding for the Alabama Center of Excellence from the Department of the Treasury under the Resources and Ecosystems Sustainability, Tourist Opportunities, and Revived Economies of the Gulf Coast States Act of 2012 (RESTORE Act) in cooperation with the State of Alabama Department of Conservation and Natural Resources under the Alabama Center of Excellence Program at the MESCDauphin Island Sea Lab. The authors acknowledge ACCESS and NCAR CISL for providing high performance computing resources and support.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Fang Wang and Di Tian. The first draft of the manuscript was written by Fang Wang and was reviewed and edited by Di Tian. All authors read and approved the final manuscript.

Funding This project was supported in part by the NSF CAREER award (NSF-EAR-2144293), the NOAA RESTORE program (NOAA-NA19NOS4510194), and with federal funding for the Alabama Center of Excellence from the Department of the Treasury under the Resources and Ecosystems Sustainability, Tourist Opportunities, and Revived Economies of the Gulf Coast States Act of 2012 (RESTORE Act) in cooperation with the State of Alabama Department of Conservation and Natural Resources under the Alabama Center of Excellence Program at the MESCDauphin Island Sea Lab.

Data availability The SRDRN-QDM code is shared at <https://osf.io/bw8js/> (DOI: 117605/OSF.IO/BW8JS). The GCM/ESM dataset from the five GCMs as part of the CMIP6 climate model simulations can be downloaded through the Earth System Grid Federation (ESGF) portals (e.g., <https://esgf-data.dkrz.de/projects/cmip6-dkrz/>). The ERA5 reanalysis dataset is produced and provided by the European Centre for Medium-range Weather Forecasts (ECMWF) and can be acquired via the ECMWF data portal (<https://apps.ecmwf.int/data-catalogues/era5/>). MBC R code is publicly available at <https://rdrr.io/cran/MBC/man/MBC-package.html> (Cannon 2018) and SPEI R code is publicly available at <https://github.com/sbegueria/SPEI> (Begueria and Vicente-Serrano 2017). ISIMIP3 python code is publicly available at <https://zenodo.org/records/7151476> (Lange 2019).

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agbazo MN, Grenier P (2020) Characterizing and avoiding physical inconsistency generated by the application of univariate quantile mapping on daily minimum and maximum temperatures over Hudson Bay. *Int J Climatol* 40:3868–3884
- Ahmadalipour A, Moradkhani H, Demirel MC (2017) A comparative assessment of projected meteorological and hydrological droughts: elucidating the role of temperature. *J Hydrol* 553:785–797
- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56. Fao, Rome 300:D05109
- Ansari R, Casanueva A, Liaqat MU, Grossi G (2023) Evaluation of bias correction methods for a multivariate drought index: case study of the Upper Jhelum Basin. *Geoscientific Model Development* 16:2055–2076
- Arik SO, NC Yoder, and T Pfister (2022) Self-adaptive forecasting for improved deep learning on non-stationary time-series. *arXiv preprint arXiv:2202.02403*
- Baño-Medina J, Manzanar R, Cimadevilla E, Fernández J, González-Abad J, Cofiño AS, Gutiérrez JM (2022) Downscaling multi-model climate projection ensembles with deep learning (Deep-ESD): contribution to CORDEX EUR-44. *Geoscientific Model Development* 15:6747–6758
- Bedia J, Herrera S, Martín DS, Koutsias N, Gutiérrez JM (2013) Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. *Clim Change* 120:229–247
- Begueria S, and SM Vicente-Serrano (2017) SPEI: calculation of the standardised precipitation-evapotranspiration index. R package version, 1
- Brown PT et al. (2023) Climate warming increases extreme daily wildfire growth risk in California. *Nature* 621(7980):760–766
- Buishand TA, Brandsma T (2001) Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. *Water Resour Res* 37:2761–2776
- Bürger G, J Schulla, Werner A (2011) Estimates of future flow, including extremes, of the Columbia River headwaters. *Water Res Res* 47(10). <https://doi.org/10.1029/2010WR009716>
- Cannon AJ (2016) Multivariate bias correction of climate model output: matching marginal distributions and intervariable dependence structure. *J Clim* 29:7045–7064
- Cannon AJ (2018) Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim Dyn* 50:31–49
- 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 Clim* 28:6938–6959
- Chapman W, Subramanian A, Delle Monache L, Xie S, Ralph F (2019) Improving atmospheric river forecasts with machine learning. *Geophys Res Lett* 46:10627–10635
- Chen J, R Arsenault, FP Brissette, and S Zhang (2021) Climate change impact studies: should we bias correct climate model outputs or post-process impact model outputs? *Water Res Res* 57: e2020WR028638

- de Silva BM, Higdon DM, Brunton SL, Kutz JN (2020) Discovery of physics from data: universal laws and discrepancies. *Front Artif Intellig* 3:25
- Donohue RJ, McVicar TR, Roderick ML (2010) Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *J Hydrol* 386:186–197
- 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 Develop* 9:1937–1958
- François B, Vrac M, Cannon AJ, Robin Y, Allard D (2020) Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth System Dynamics* 11:537–562
- François B, Thao S, Vrac M (2021) Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. *Clim Dyn* 57:3323–3353
- Fulton DJ, Clarke BJ, Hegerl GC (2023) Bias correcting climate model simulations using unpaired image-to-image translation networks. *Artif Intell Earth Syst* 2:e220031
- Goulart HMD, van der Wiel K, Folberth C, Balkovic J, van den Hurk B (2021) Storylines of weather-induced crop failure events under climate change. *Earth Syst Dynam* 12:1503–1527
- Goulart HM, K van der Wiel, C Folberth, E Boere, and B van den Hurk (2023) Increase of simultaneous soybean failures due to climate change. *Earth's Future*, 11: e2022EF003106
- Grose MR, Fox-Hughes P, Harris RM, Bindoff NL (2014) Changes to the drivers of fire weather with a warming climate—a case study of southeast Tasmania. *Clim Change* 124:255–269
- Guo Q, Chen J, Zhang X, Shen M, Chen H, Guo S (2019) A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Clim Dyn* 53:3603–3623
- Guo Q, J Chen, XJ Zhang, CY Xu, and H Chen (2020) Impacts of using state-of-the-art multivariate bias correction methods on hydrological modeling over North America. *Water Res Res* 56, e2019WR026659
- Haile GG et al. (2020) Projected impacts of climate change on drought patterns over East Africa. *Earth's Future* 8: e2020EF001502
- Ham Y-G, Kim J-H, Luo J-J (2019) Deep learning for multi-year ENSO forecasts. *Nature* 573:568–572
- Harder P et al. (2022) Generating physically-consistent high-resolution climate data with hard-constrained neural networks. *arXiv preprint arXiv:2208.05424*
- He K, X Zhang, S Ren, and J Sun (2016) Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778
- Hersbach H et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049
- Hess P, Drüke M, Petri S, Strnad FM, Boers N (2022) Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nat Mach Intellig* 4:828–839
- Hess P, S Lange, C Schötz, Boers N (2023) Deep learning for bias-correcting CMIP6-class Earth system models. *Earth's Future* 11(10):e2023EF004002
- Immerzeel WW, Petersen L, Ragettli S, Pellicciotti F (2014) The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. *Water Resour Res* 50:2212–2226
- Kling H, Fuchs M, Paulin M (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J Hydrol* 424:264–277
- Kong X, X Liu, J Gu, Y Qiao, and C Dong (2022) ReFlash dropout in image super-resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6002–6012
- Lange S (2019) Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1. 0). *Geosci Model Develop* 12:3055–3070
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Ledig C et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690
- Lee M-H, Im E-S, Bae D-H (2019) A comparative assessment of climate change impacts on drought over Korea based on multiple climate projections and multiple drought indices. *Clim Dyn* 53:389–404
- Lee H et al. (2023) AR6 Synthesis Report: Climate Change 2023. Summary for Policymakers
- Li C, Sinha E, Horton DE, Diffenbaugh NS, Michalak AM (2014) Joint bias correction of temperature and precipitation in climate model simulations. *J Geophys Res Atmosph* 119(13):153–113
- Li L, S Huang, Z Ouyang, and N Li (2022) A deep learning framework for non-stationary time series prediction. 2022 3rd international conference on computer vision, image and deep learning and international conference on computer engineering and applications (CVIDL & ICCEA), IEEE, 339–342
- Liang XS, Xu F, Rong Y, Zhang R, Tang X, Zhang F (2021) El Niño Modoki can be mostly predicted more than 10 years ahead of time. *Sci Rep* 11:17860
- Liu Y et al. (2016) Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*
- Liu Y, AR Ganguly and J Dy (2020) Climate downscaling using YNet: a deep convolutional network with skip connections and fusion. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3145–3153
- Lobell DB, Field CB (2007) Global scale climate–crop yield relationships and the impacts of recent warming. *Environ Res Lett* 2:014002
- Maraun D (2016) Bias correcting climate change simulations—a critical review. *Curr Climate Change Rep* 2:211–220
- Maraun D et al (2017) Towards process-informed bias correction of climate change simulations. *Nat Clim Chang* 7:764–773
- Maurer EP, Hidalgo HG, Das T, Dettinger MD, Cayan DR (2010) The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol Earth Syst Sci* 14:1125–1138
- McKee TB, NJ Doesken, and J Kleist (1993) The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, California, 179–183
- Mehrotra R, Sharma A (2012) An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations. *Water Resour. Res* 48(12). <https://doi.org/10.1029/2012WR012446>
- Mueller B, Seneviratne SI (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys Res Lett* 41:128–134
- Nahar J, Johnson F, Sharma A (2018) Addressing spatial dependence bias in climate model simulations—an independent component analysis approach. *Water Resour Res* 54:827–841
- Pan B et al. (2021) Learning to correct climate projection biases. *J Adv Model Earth Syst* 13: e2021MS002509
- Panofsky H, and G Brier (1968) Some applications of statistics to meteorology, 224 pp. Pa. State Univ., University Park, Pa
- Quesada-Chacón D, Barfus K, Bernhofer C (2022) Repeatable high-resolution statistical downscaling through deep learning. *Geosci Model Develop* 15:7353–7370
- Quesada-Chacón D, J Baño-Medina, K Barfus, and C Bernhofer (2023) Downscaling CORDEX through deep learning to daily 1

- km multivariate ensemble in complex terrain. *Earth's Future*, 11: e2023EF003531
- Racah E, C Beckham, T Maharaj, S Ebrahimi Kahou, M Prabhat, and C Pal (2017) Extremeweather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Adv Neur Informat Process Syst* 30
- Ravuri S et al (2021) Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597:672–677
- Read JS et al (2019) Process-guided deep learning predictions of lake water temperature. *Water Resour Res* 55:9173–9190
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat F (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566:195–204
- Rodrigues ER, I Oliveira, R Cunha, and M Netto (2018) DeepDownscale: a deep learning strategy for high-resolution weather forecast. 2018 IEEE 14th International Conference on e-Science (e-Science), IEEE, 415–422
- Scher S, Messori G (2019) Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development* 12:2797–2809
- Schillerberg T, Tian D (2023) Changes in crop failures and their predictions with agroclimatic conditions: analysis based on earth observations and machine learning over global croplands. *Agriculture for Meteorology* 340:109620
- Sha Y (2020b) Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: daily precipitation. *J Appl Meteorol Climatol* 59:2075–2092
- Sha Y, Gagne DJ II, West G, Stull R (2020a) Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: daily maximum and minimum 2-m temperature. *J Appl Meteorol Climatol* 59:2057–2073
- Shi X, Z Chen, H Wang, D-Y Yeung, W-K Wong, and W-C Woo (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv Neural Informat Process Syst* 28
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15:1929–1958
- Sun AY, Tang G (2020) Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets. *Front Water* 2:536743
- Tao Y, Gao X, Ihler A, Hsu K, Sorooshian S (2016) Deep neural networks for precipitation estimation from remotely sensed information. In 2016 IEEE congress on evolutionary computation (CEC). IEEE, pp 1349–1355
- Thrasher B, Maurer EP, McKellar C, Duffy PB (2012) Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol Earth Syst Sci* 16:3309–3314
- Van Loon AF et al (2016) Drought in the anthropocene. *Nat Geosci* 9:89–91
- Vandal T, E Kodra, S Ganguly, A Michaelis, R Nemani, and AR Ganguly (2017) DeepSD: generating high resolution climate change projections through single image super-resolution. *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1663–1672
- Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J Clim* 23:1696–1718
- Wang F, Tian D (2022) On deep learning-based bias correction and downscaling of multiple climate models simulations. *Clim Dyn* 59:3451–3468
- Wang F, Tian D, Carroll M (2023) Customized deep learning for precipitation bias correction and downscaling. *Geoscientific Model Development* 16:535–556
- Wang F, D Tian, L Lowe, L Kalin, and J Lehrter (2021) Deep learning for daily precipitation and temperature downscaling. *Water Res Res* 57, e2020WR029308
- Wi S, and S Steinschneider (2022) Assessing the physical realism of deep learning hydrologic model projections under climate change. *Water Res Res* 58, e2022WR032123
- Wilson T, P.N. Tan and L Luo (2022) DeepGPD: a deep learning approach for modeling geospatial-temporal extreme events. *Proceedings of the AAAI Conference on Artificial Intelligence*, 4245–4253
- Wood AW, EP Maurer, A Kumar, and DP Lettenmaier (2002) Long-range experimental hydrologic forecasting for the eastern United States. *J Geophys Res Atmos* 107, ACL 6–1–ACL 6–15
- Xuan W, Ma C, Kang L, Gu H, Pan S, Xu Y-P (2017) Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China. *Theoret Appl Climatol* 128:207–222
- Zscheischler J et al (2018) Future climate risk from compound events. *Nat Clim Chang* 8:469–477
- Zscheischler J, Fischer EM, Lange S (2019) The effect of univariate bias adjustment on multivariate hazard estimates. *Earth System Dynamics* 10:31–43

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