Climate Dynamics
https://doi.org/10.1007/s00382-022-06277-2

On deep learning-based bias correction and downscaling of multiple climate models simulations

Fang Wang[1] · **Di Tian**[1]

Received: 13 October 2021 / Accepted: 27 March 2022

Abstract
Bias correcting and downscaling climate model simulations requires reconstructing spatial and intervariable dependences of the observations. However, the existing univariate bias correction methods often fail to account for such dependences. While the multivariate bias correction methods have been developed to address this issue, they do not consistently outperform the univariate methods due to various assumptions. In this study, using 20 state-of-the-art coupled general circulation models (GCMs) daily mean, maximum and minimum temperature (T_{mean}, T_{max} and T_{min}) from the Coupled Model Intercomparison Project phase 6 (CMIP6), we comprehensively evaluated the Super Resolution Deep Residual Network (SRDRN) deep learning model for climate downscaling and bias correction. The SRDRN model sequentially stacked 20 GCMs with single or multiple input-output channels, so that the biases can be efficiently removed based on the relative relations among different GCMs against observations, and the intervariable dependences can be retained for multivariate bias correction. It corrected biases in spatial dependences by deeply extracting spatial features and making adjustments for daily simulations according to observations. For univariate SRDRN, it considerably reduced larger biases of T_{mean} in space, time, as well as extremes compared to the quantile delta mapping (QDM) approach. For multivariate SRDRN, it performed better than the dynamic Optimal Transport Correction (dOTC) method and reduced greater biases of T_{max} and T_{min} but also reproduced intervariable dependences of the observations, where QDM and dOTC showed unrealistic artifacts ($T_{max} < T_{min}$). Additional studies on the deep learning-based approach may bring climate model bias correction and downscaling to the next level.

Keywords Climate models · Bias correction · Downscaling · Deep learning · Spatial dependence · Multivariate dependence · Model evaluation

1 Introduction

Simulated variables from coupled general circulation models (GCMs) can exhibit large systematic biases relative to observational datasets and may have limited usefulness for climate impact assessments unless the biases are corrected (Cannon 2018; Mearns et al. 2012; Sillmann et al. 2013). For this reason, various bias correction methods have been developed to tackle this issue, including univariate bias correction methods such as the widely used quantile mapping approach [QM; e.g., Panofsky and Brier (1968), Thrasher et al. (2012), Wood et al. (2002)] as well as multivariate bias correction approaches which incorporate intervariable dependence with capability of correcting multiple variables simultaneously (Bürger et al. 2011; Cannon 2016; Chen et al. 2018; Mehrotra and Sharma 2012, 2019; Robin et al. 2019).

Systematic errors in GCM outputs include large biases in spatial and intervariable dependences (Bürger et al. 2011; Nahar et al. 2018). Correcting biases on spatial distribution of climate variables is very important for climate impact studies such as agriculture or water resources planning and management (Nahar et al. 2018). Ignoring the intervariable dependence structure between variables can result in obtaining corrected outputs with inappropriate physical laws (Agbazo and Grenier 2020; Thrasher et al. 2012) and, thereby, distorting the results of impact studies (François et al. 2020; Maraun et al. 2017; Zscheischler et al. 2020). However, most current bias correction approaches are applied at the grid point basis and do not consider spatial dependence across the domain. The existing multivariate bias correction methods simplify calculation process as

✉ Di Tian
tiandi@auburn.edu

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

Published online: 15 April 2022

linear programming problem (Robin et al. 2019) or assume predefined intervariable relationships including Pearson correlation (Bürger et al. 2011; Cannon 2016; Mehrotra and Sharma 2012), Spearman rank correlation (Cannon 2016), lag 1 autocorrelation for rank dependence (Mehrotra and Sharma 2019), and linear combinations of normally distributed variables (Cannon 2018; Chen et al. 2014). Due to these simplifications and assumptions, recent intercomparison studies have revealed that the current multivariate bias correction approaches may fail to compete with univariate methods (Chen et al. 2018; François et al. 2020; Guo et al. 2020; Meyer et al. 2019; Van de Velde et al. 2020). For example, Van de Velde et al. (2020) found that the multivariate methods often perform worse than the univariate methods especially for temperature and advantages of multivariate bias correction are weakened for most climate regimes in climate change condition and thus they recommended the simpler univariate bias correction methods for assessing climate change impact. The advantages of using current multivariate bias correction methods for impact modeling are also region dependent, which is not as profound in the validation period as in the calibration period (Guo et al. 2020).

Deep learning with convolutional neural network (CNN) types of approaches have achieved notable progress in modeling spatial context data in computer vision field (LeCun et al. 2015). Deep learning for climate sciences are still at early stage (Reichstein et al. 2019) but growing rapidly during recent years. Recent studies successfully applied deep convolution based architecture to objectively extract spatial features to define and classify extreme weather (for example, hurricanes, storms, and atmospheric rivers) in numerical weather prediction model output (Liu et al. 2016; Racah et al. 2016), forecast ENSO (Ham et al. 2019; Liang et al. 2021), and improve precipitation nowcasting (Ravuri et al. 2021). In particular, image to image translation using generative adversarial networks (GANs) or single generator based models with stacked convolutional layers has successfully learned the mapping between an input image and an output image using a training set of aligned paired or unpaired images, so that the deep learning models can systematically convert or correct one image according to another (Isola et al. 2017; Yang et al. 2018; Zhu et al. 2017). CNN types of models have also successfully been used to systematically remove different types of noises at different levels from photos to generate clean ones as well as mixed noises that follow different distributions (Tian et al. 2020). These progresses of deep learning on spatial feature extraction and image conversion or correction provide potentials for addressing the existing issues on bias correcting GCM outputs. Furthermore, the CNN-based model can be multivariate model by taking multiple input-output channels. Relationships among different channels are integrated into model training process, which enables the models to capture complex relationships among different variables beyond our prior knowledge, providing a great potential for improving multivariate bias corrections of GCMs.

Deep CNN-based models have been used for GCM bias correction and downscaling in recent years. For example, Liu et al. (2020) developed an image super resolution architecture called YNet to downscale monthly mean temperature/precipitation from 35/33 GCMs, demonstrating that the model outperformed a shallow plain architecture (Vandal et al. 2017) and a traditional statistical downscaling method. Rodrigues et al. (2018) developed a CNN-based deep learning architecture namely DeepDownscale to downscale daily precipitation from 4 GCMs into the local scale and indicated that the model performed better than a regional dynamic downscaling method. François et al. (2021) applied a simple CNN-based GANs architecture to bias correct GCM outputs and found that its performance is generally better than QM and other two multivariate bias correction methods. In François et al. (2021), performances of spatial adjustments from their CNN-based GANs were assessed through spatial correlations and energy distances. Pan et al. (2021) developed a sophisticated CNN-based GANs architecture namely RADA with other variables as dynamical constraints to bias correct daily precipitation and found that RADA performs mostly better than QM and a multivariate method (Cannon et al. 2018) particularly on inter-field correlation. Despite the advances achieved from these studies, they did not account for multivariate bias correction nor explicitly quantify model performance for correcting biases of spatial dependence.

In this study, using 20 state-of-the-art GCMs daily mean, maximum and minimum temperature outputs from the Coupled Model Intercomparison Project phase 6 (CMIP6), we comprehensively evaluate a new Super Resolution Deep Residual Network (SRDRN) model (Wang et al. 2021), which was developed based on an advanced deep CNN type architecture, in comparison with conventional bias correction approaches, for addressing the aforementioned challenges for climate model downscaling and bias correction. The structure of this paper is organized as follows: Sect. 2 introduces data and methodology, including the experimental design, the SRDRN model and the benchmark bias correction approaches; Sect. 3 presents results; discussions and conclusions are provided in Sects. 4 and 5, respectively.

2 Data and methodology

2.1 Data and area of study

The climate system is highly complex and it remains fundamentally impossible to consider everything in one single climate model (Tebaldi and Knutti 2007). Kumar and Ganguly (2018) have found that climate variability of multi-model

ensemble is larger than climate internal variability of single climate models for all projection time horizons and spatial resolutions for precipitation and temperature, and the latter is dominated only for the initial few decades. To reasonably account for uncertainties in climate simulations and projections, this study explores bias correction for multi-model ensemble.

For the climate simulations, daily mean (T_{mean}) surface air temperatures simulated by 20 commonly used GCMs were extracted from the newest Coupled Model Intercomparison Project phase 6 (CMIP6) (Eyring et al. 2016). Maximum (T_{max}) and minimum (T_{min}) surface air temperature are available in 18 out of the 20 GCMs. The 20 GCMs' were developed by different climate research centers all over the world and have different spatial resolution varying from 0.5° to as large as 2.8° (see Table 1). The 20 GCMs provide outputs of climate variables under the historical scenario for the reference period (1850–2014), but the period of 1979 to 2014 was used in this study in order to match with available climate reanalysis (used as a surrogate to observations). Most GCMs archived in CMIP6 include multiple ensemble members (named with r*i*n*p*n*f*n, where r represents realization, i represents initialization method, p represents physics, f represents forcing, and *n* can be different numbers) and we used single member for each model in this study for fair comparisons. The outputs of the 20 GCMs were extracted from the ensemble member of r1i1p1f1, except for CNRM-CM6-1 (from r1i1p1f2), CNRM-CM6-1-HR (from r1i1p1f2), BCC-CSM2-MR (from r2i1p1f1), HadGEM3-GC31-LL (from r1i1p1f3) and HadGEM3-GC31-MM (from r1i1p1f3). Prior to bias correction and downscaling, the GCM outputs were regridded into a common 1° × 1° resolution using bilinear interpolation.

Table 1 Basic information of 20 CMIP6 GCMs

NO.	GCMs	Resolution (lat × lon °)	Country
1	GFDL-ESM4	1.0×1.25	NOAA, USA
2	GFDL-CM4	1.0×1.25	NOAA, USA
3	CESM2-WACCM	0.94×1.25	NCAR, USA
4	CESM2	0.94×1.25	NCAR, USA
5	CanESM5	2.81×2.81	Canada
6	CNRM-CM6-1	1.41×1.41	France
7	CNRM-CM6-1-HR	0.5×0.5	France
8	EC-Earth3	0.70×0.70	Europe
9	ACCESS-ESM1-5	1.24×1.88	Australia
10	IPSL-CM6A-LR	1.26×2.5	France
11	MPI-ESM1-2-LR	1.88×1.88	Germany
12	MPI-ESM1-2-HR	0.94×0.94	Germany
13	NorESM2-MM	0.94×1.25	Norway
14	NorESM2-LM	1.88×2.5	Norway
15	MIROC6	1.41×1.41	Japan
16	MRI-ESM2-0	1.13×1.13	Japan
17	INM-CM5-0	1.5×2.0	Russia
18	BCC-CSM2-MR	1.13×1.13	China
19	HadGEM3-GC31-LL	1.25×1.88	UK
20	HadGEM3-GC31-MM	0.56×0.83	UK

The European Center for Medium-Range Weather Forecast's (ECMWF) ERA5 dataset was used as high-resolution observations (hereafter "observations"), which has 0.25° resolution covering the period from 1979 to 2014 (Hersbach et al. 2020). The bias correction and downscaling experiments were performed in the rectangle area covering the entire southeastern of the United States and Gulf of Mexico, ranging from 99° W to 75.25° W in longitude and from 25°N to 36.25°N in latitude. The study area falls into humid subtropical climate and is highly influenced by hot extreme events.

2.2 Methods

This section presents a brief description of the SRDRN deep learning model, a widely used univariate quantile delta mapping (QDM) bias correction method and a multivariate bias correction approach namely dynamic Optimal Transport Correction (dOTC, Robin et al. 2019). The univariate QDM and multivariate dOTC methods serve as benchmark approaches to measure the potential added values from the SRDRN downscaling and bias correction. We designed two experiments to perform and evaluate the bias correction methods. For the experiment 1, we bias corrected and downscaled daily mean temperature (T_{mean}) from 20 CMIP6 GCMs using univariate SRDRN with single input-output channel and QDM, respectively. For the experiment 2, as demonstrated in Fig. 1, we bias corrected and downscaled T_{max} and T_{min} simultaneously using multivariate SRDRN with multiple input-output channels, as well as separately using QDM. For both experiments, SRDRN sequentially stacked GCMs daily temperature data, which greatly augments the data size for training a robust model but also accounts for relative relations among different GCMs against observations. As indicated in previous studies, there are significant differences in bias among different GCMs (Nahar et al. 2017). The underlying hypothesis posed here is that the SRDRN can identify mixed biases from different GCMs, given that CNN-type based architectures has been successfully used to remove mixed noises that follows different probability distribution (Tian et al. 2020). The observations (i.e. ERA5) were replicated and stacked correspondingly to match with each set of GCMs for generating pairs of coarse and fine resolution 2-dimentional temperature data on each day. Furthermore, we executed the multivariate dOTC in a relatively complex topography area around the state of Tennesse within the study area. We did not run dOTC in the entire research area due to its limitation of correcting

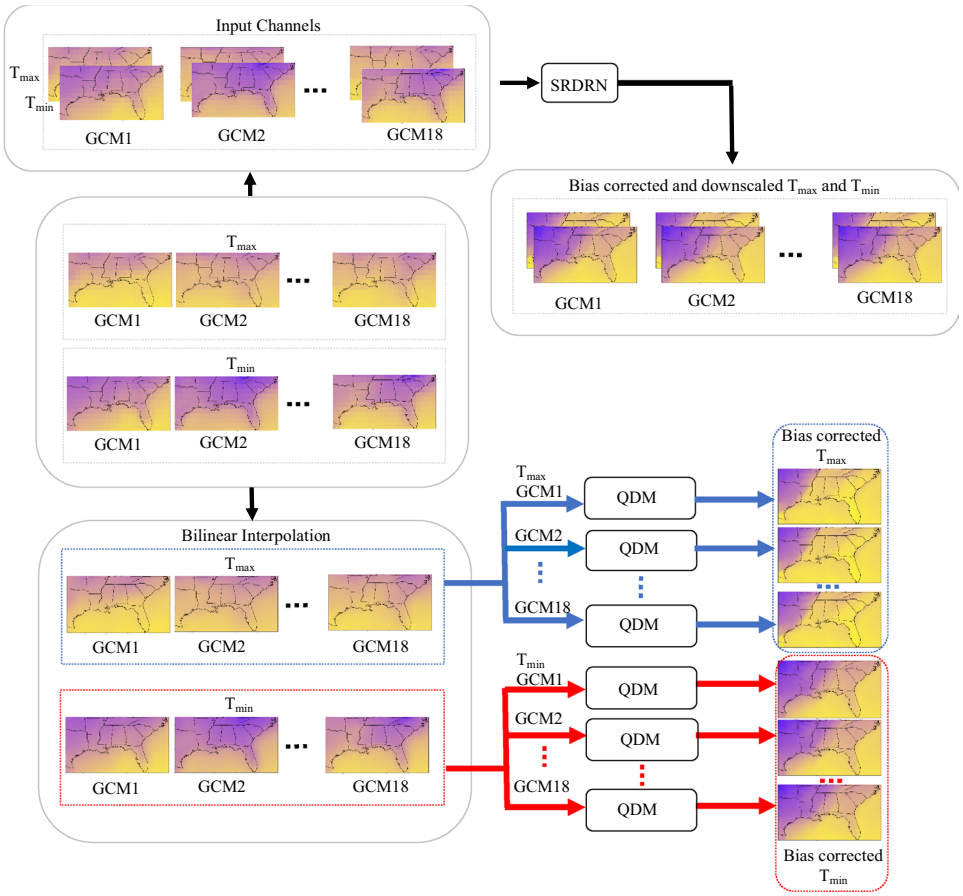

Fig. 1 Schematic of the experiment for bias correcting T_{max} and T_{min} using SRDRN (upper panel) and QDM (lower panel). Note: 18 out of 20 GCMs include both daily T_{max} and T_{min} data

variables in a large geographical area (François et al. 2020). We used the first 26 years daily data (1979 to 2004) as the training dataset, the middle 5 years (2005 to 2009) as the validation dataset to adjust hyperparameters and the last 5 years (2010 to 2014) as the testing dataset for a "future" climate, so that the model can be tested in the nonstationary context. QDM and dOTC bias corrected each GCM individually, with all the GCMs outputs spatially disaggregated from 1° into 0.25° resolution (Abatzoglou and Brown 2012; François et al. 2021) (same as the resolution of ERA5 observations) using bilinear interpolation before performing bias correction (Fig. 1).

2.2.1 SRDRN model

Here we provide a description of the SRDRN algorithm. For more details, the readers refer to Wang et al. (2021). The SRDRN algorithm was inspired by a novel super scaling deep learning approach in the computer vision field (Ledig et al. 2017), which is mainly comprised of residual blocks and upsampling blocks with CNN and batch normalization layers. To extract spatial patterns, the CNN layers apply filters to go through the input data to build local connection within nearby grids by computing the element-wise dot product between the filters and different patches of the input. The result is followed by a nonlinear activation function, here parametric ReLU (He et al. 2015) in this study. Batch normalization is a technique to standardize the inputs to a layer for each mini-batch and thus stabilize the learning process and accelerate the training of the model (Ioffe and Szegedy 2015).

The residual blocks equipped with CNN and batchnormalization layers are designed to extract fine spatial features and avoid degradation issue for the very deep neural network. For plain deep neural networks, model accuracy can be easily saturated and degrade rapidly with the increase of network depth (He et al. 2016). Residual blocks, however, can improve model performance even for extensively deep networks (Silver et al. 2017) because residual blocks execute residual mapping and include skipping connections (see Fig. 14 in "Appendix A"). The way that skipping connection skips layers and connects next layers for the SRDRN is through element-wise addition in this study. The total number of 16 residual blocks were used in the SRDRN architecture, which makes the network very deep and has the potentials to extract fine spatial features.

The upsampling blocks are used to increase data resolution from coarse to high for the downscaling purpose. The

upsampling process is executed directly on the feature maps from convolutional layers and each upsampling block consists of one convolutional layer and one upsampling layer followed by parametric ReLU activation function. The upsampling layer with nearest neighbor interpolation was chosen to increase spatial resolution. Each upsampling block sequentially and gradually increases the input low resolution feature maps by a factor of 2. In this study, the downscaling ratio (the ratio between coarse resolution and high resolution data) is 4 and thus we used 2 upsampling blocks.

Data normalization was executed as a data preprocessing step. Specifically, daily temperature was normalized by subtracting the mean (μ) and dividing by the standard deviation (σ). Here μ and and σ are scalar values that were calculated based on the flattened variable for the training dataset. During the testing period, the model prediction was inversely normalized with μ and σ calculated from the GCM statistics in order to preserve the trend in the test dataset.

In this study, the mean square error (MSE) was chosen as the loss function (Wang et al. 2021). For the model output statistics (MOS), Maraun et al. (2010) indicated that observations and model simulations are not synchronized in time. In this study, we synchronously paired coarse resolution data from GCMs and observations and used MSE loss function to search an optimal solution. The underlying assumption is that the SRDRN is capable of reproducing distributions of the observations if synchronized biases are well corrected. The optimization algorithm Adam was applied to train the network with a learning rate of 0.0001 and other parameters with defaulted values. Through a series of experiments, we found that the learning rate of 0.0001 worked well in this study and the mini-batch size of 64 was chosen. We output validation loss for the validation dataset for each epoch to choose the best model for prediction.

2.2.2 Benchmark approaches

We used quantile delta mapping (QDM) as a univariate benchmark approach. Compared to quantile mapping (QM) method (Panofsky and Brier 1968; Thrasher et al. 2012; Wood et al. 2002), QDM accounts for the difference between historical and future climate scenario data and thus is capable of preserving trend of the future climate (Cannon et al. 2015). QDM has been widely used to bias correct climate variables including daily temperature in recent studies, which indicated better performance compared to the other bias correction approaches (Cannon et al. 2015; Eum and Cannon 2017; Jose and Dwarakish 2021; Kim et al. 2021; Tegegne and Melesse 2021; Tong et al. 2020).

The basic equation of the QDM method comprises the bias corrected value term obtained from the observational dataset and the relative change term (delta) obtained from the GCM data, as defined in Eq. (1) below.

$$\hat{x}_{m,p}(t) = \hat{x}_{o:m,h:p}(t) + \Delta_m(t), \tag{1}$$

Where $\hat{x}_{o:m,h:p}(t) = F_{o,h}^{-1}\left[\tau_{m,p}(t)\right]$ and $\Delta_m(t) = \frac{x_{m,p}(t)}{F_{m,h}^{-1}\left[\tau_{m,p}(t)\right]}$. Here $\hat{x}_{m,p}(t)$ is the bias corrected value of the model data for the projection period. $\hat{x}_{o:m,h:p}(t)$ is the bias corrected value of the observed data for the historical period and $\Delta_m(t)$ is the relative change in the model data between the historical and future periods. Thus, the bias corrected future projection at time t is given by adding the relative change $\Delta_m(t)$ by the historical bias-corrected observed value. $\tau_{m,p}(t)$ is the percentile of $\hat{x}_{m,p}(t)$ in the empirical cumulative density function (F) over a time window around t. $F_{o,h}^{-1}$ represents inverse empirical cumulative density function or quantile for the observed data in the historical period and $F_{m,h}^{-1}$ for quantile for the model data in the historical period. The time window to construct the empirical cumulative density function around time t was set to be 45 days in order to preserve the seasonable cycle. In this study, the historical and projection periods correspond to the training and testing data periods, respectively. Details about the QDM method are referred to Cannon et al. (2015).

In addition to QDM, we also considered a widely used multivariate bias correction method dOTC (François et al. 2020, 2021; Robin et al. 2019; Van de Velde et al. 2020). Here we give brief description of this method. The dOTC corrects the marginal distributions and dependence structures altogether as the same time. Taking advantage of the optimal transport theory, dOTC constructs a multivariate transfer function, called a transport plan, for the adjustment of climate simulations with respect to references while minimizing an associated cost function. The coefficients subjected to constraints in the cost function are solved as a linear programming problem. Through this transfer function, multivariate distribution of a biased random variable (need to be estimated) and its correction are linked together. Any value of the variable to correct is associated with a conditional law linking the biased value and its correction. Corrections are drawn randomly from these conditional laws, which introduces some stochasticity into the bias correction procedure. Similar to QDM, dOTC also considers nonstationarity of the dependence structure between the calibration and the projection periods and, permits the evolution of the model (e.g., induced by climate change) to be considered in the bias correction procedure. Time series at each grid cell is considered as one dimension. We ran dOTC at a complex topography area around the state of Tennessee within the study area including 351 grid cells. With two variables (T_{max} and T_{min}), dOTC simultaneously bias corrected 702 time series and the setting is similar to François et al. (2021). The publicly available SBCK python code was used to execute dOTC and more details about dOTC are referred to Robin et al. (2019).

3 Results

The univariate SRDRN and the QDM bias correction results for T_{mean} are presented in the Sect. 3.1 to 3.4. The multivariate SRDRN, the QDM, and the dOTC bias correction results for T_{max} and T_{min} are shown in the Sect. 3.5. We also present bilinear interpolation results of GCMs raw temperature simulations without bias correction (named as Bilinear), and the differences between Bilinear and observations are considered as total biases originated from raw GCMs outputs.

3.1 Overall agreement

The overall agreement between the observed and bias corrected T_{mean} was quantified using root mean square error (RMSE) and correlation coefficient. These metrics were calculated on a daily basis over the entire testing period (2010 to 2014) for all the grid points over the study region (Fig. 2) by flattening the spatial and temporal axes as one-dimensional data. The results show that the SRDRN bias corrected and downscaled T_{mean} have much lower RMSE and higher correlation coefficient values for all the 20 GCMs than QDM, suggesting SRDRN downscaled data are closer to observations. While the overall RMSE for QDM is lower than Bilinear, the correlation coefficient is not consistently higher than the Bilinear method for several GCMs, due to the limitation of QDM, which adjusts quantiles on a grid-by-grid basis and is not able to correct the biases in spatial context. The SRDRN model, however, fully accounts for the spatial feature-based relationship between GCM outputs and observations on a day-to-day basis in the training process and systematically removes the biases when applying to the testing dataset. Based on the overall RMSE, SRDRN reduced 8 to 20% total biases while QDM only reduced 0.6 to 11% total biases depending on GCMs.

3.2 Spatial bias and dependence

To evaluate the performance on correcting spatial biases, we calculated annual average of daily map correlation between bias corrected GCMs and observations for T_{mean} (Fig. 3). Here we firstly calculated daily map correlation by flattening the two-dimensional spatial data as one-dimensional vector and then calculated the average over each year during the testing period. The results indicate that SRDRN shows greater map correlations (0.73 for SRDRN versus 0.68 for QDM) and a much smaller inter-model variability (standard deviation of 0.017 for SRDRN

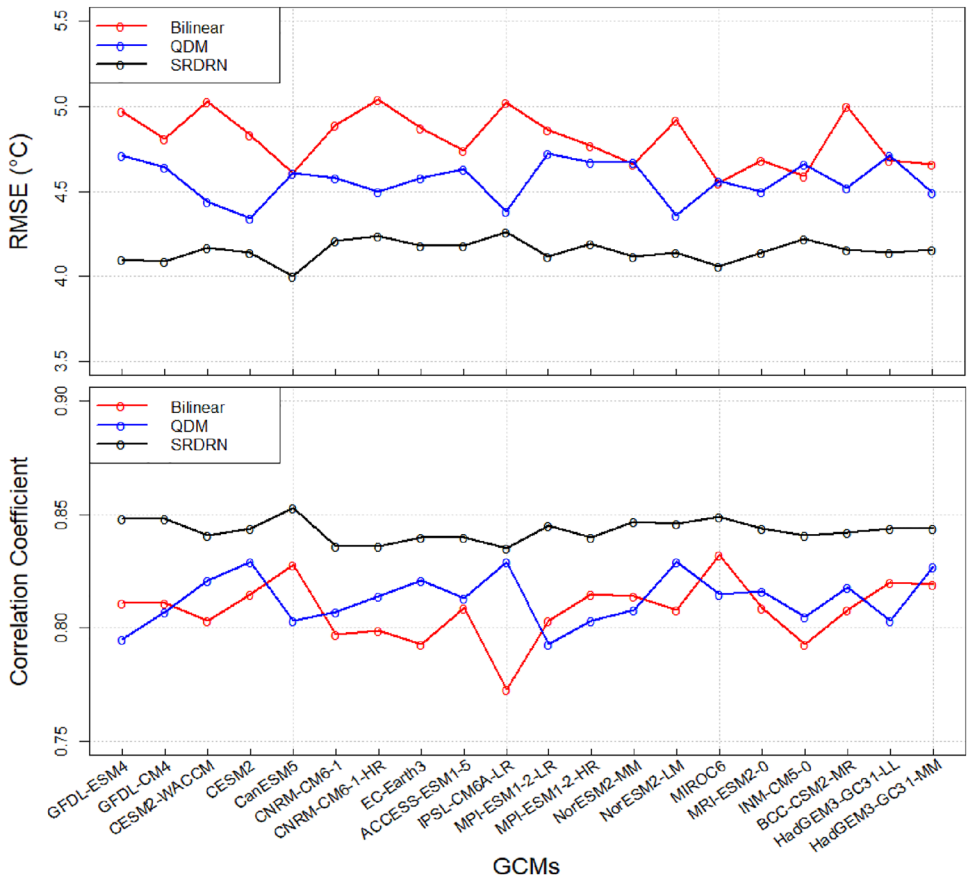

Fig. 2 Overall assessment for daily mean temperature (T_{mean}) from 20 CMIP6 GCMs

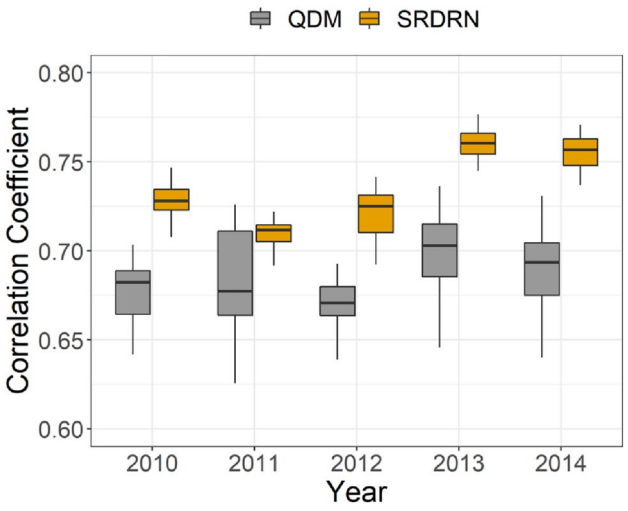

Fig. 3 Annual average of daily map correlation for SRDRN and QDM

versus 0.028 for QDM) compared to QDM, demonstrating the strength of SRDRN for correcting spatial biases.

We evaluated spatial dependence by comparing the spatial semivariance of the bias corrected and downscaled T_{mean} and the observed T_{mean}. The spatial semivariance can be calculated using the equation below:

$$\gamma(h) = \frac{1}{2}\langle [I(j+h) - I(j)]^2 \rangle \quad (2)$$

where $I(j)$ is the T_{mean} at location j and h is a displacement vector (spacing between grid cells). The R library "gstat" (Pebesma 2004) was used to calculate the spatial semivariance. Figure 4 shows the RMSE and correlation coefficient of spatial semivariance between observed and bias corrected/downscaled data by QDM and SRDRN as well as Bilinear for all the GCMs. These metrics were calculated daily over the entire testing period (2010 to 2014), which reveals how well each method can capture spatial dependence against the observations for each day. We can see that RMSE of SRDRN spatial semivariance is consistently lower than QDM and Bilinear, and correlation coefficient is much higher for all the GCMs, demonstrating that SRDRN corrected greater biases (as large as 40% total biases) than QDM (no greater than 30%) in terms of spatial dependences. We also evaluated spatial dependence for each month. Figure 5 shows spatial variograms (spatial semivariance versus distance) at a winter and a summer month for a randomly selected GCM (MPI-ESM1-2-HR). We can see there are large differences between the spatial variograms of Bilinear and OBS. QDM

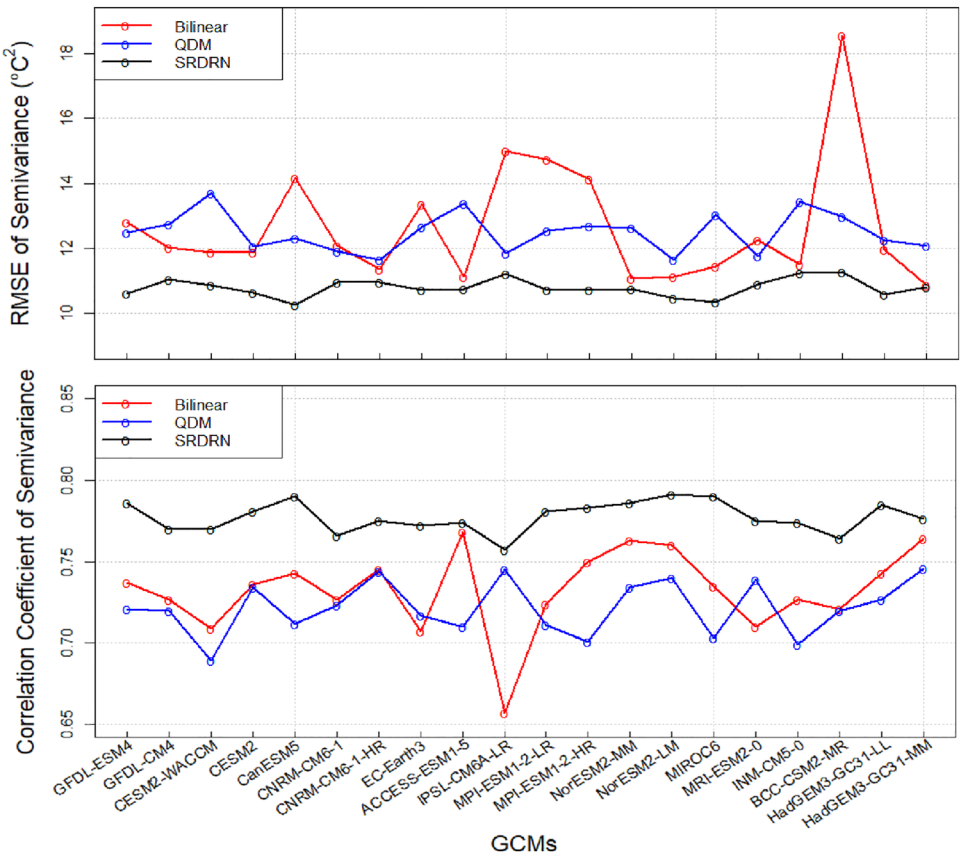

Fig. 4 RMSE and correlation coefficient of spatial semivariance for SRDRN, QDM, and Bilinear T_{mean} with observed T_{mean}

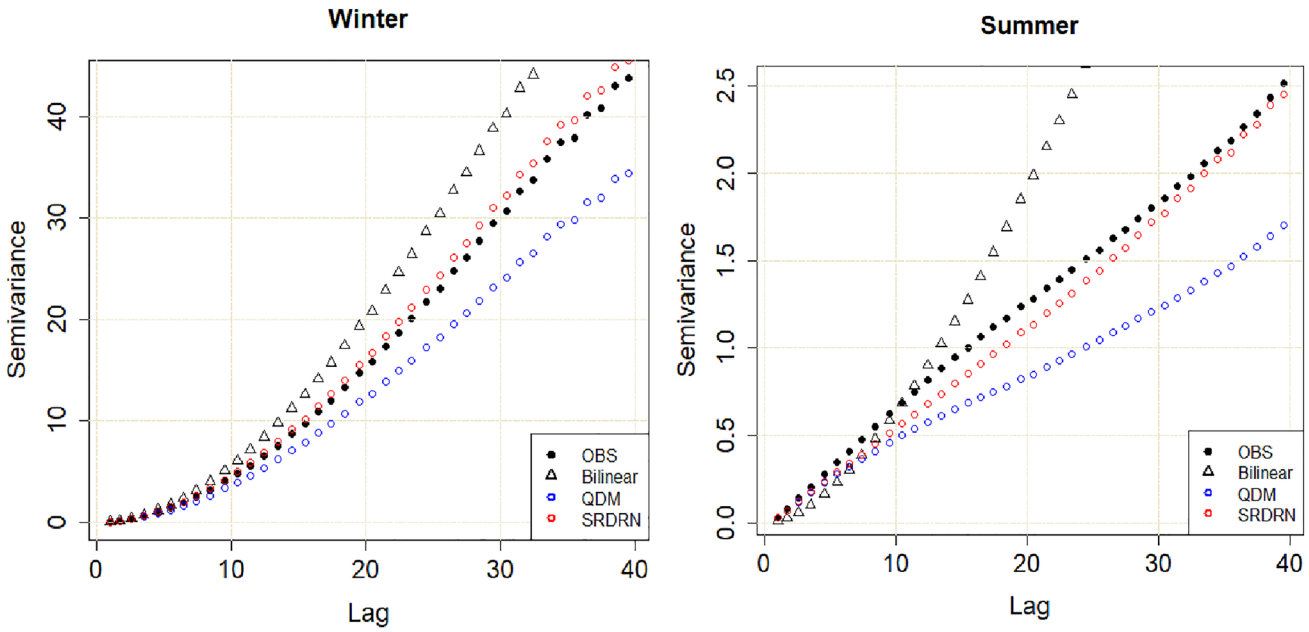

Fig. 5 Spatial variogram of SRDRN and QDM bias corrected and downscaled MPI-ESM1-2-HR T_{mean}, a randomly selected GCM, as well as Bilinear and observations (OBS) during a winter and summer months

highly underestimates the semivariances of the observations, the semivariances calculated from SRDRN, however, are very close to the ones calculated from the observations, suggesting that the trained SRDRN during 1979 to 2004 well adjusted spatial dependence of GCMs during the testing period (2010 to 2014).

3.3 Temporal bias

The SRDRN model treated daily spatial temperature data independently, which did not explicitly account for temporal dependence during bias correction. However, given daily spatial temperature data is well corrected and downscaled by SRDRN, we hypothesize that the SRDRN model could well reduce temporal biases. To test that, we evaluate both daily and climatology of downscaled and bias corrected T_{mean} against the observations in time. Figure 6 shows downscaled and bias corrected results versus observed daily series of T_{mean} averaged over the study area. While both QDM and SRDRN reduced the inter-model variability among 20 GCMs especially during summer seasons, SRDRN removed more biases for low temperature than QDM during winter seasons. To examine how well the bias corrected and downscaled T_{mean} reproduces the climatology of observations, Fig. 7 shows the monthly mean of QDM and SRDRN bias corrected daily T_{mean} as well as Bilinear for the 20 GCMs against the observations. Overall, both SRDRN and QDM greatly reduced monthly mean biases compared with Bilinear (removed about 4 °C biases for SRDRN and 3 °C biases for QDM). The monthly means of SRDRN bias corrected daily temperature well reproduced the observed climatology with small inter-model variability. In comparison, the monthly means of QDM bias corrected daily T_{mean} mostly underestimated the observed climatology during summer season and had relative larger inter-model variability in winter season compared to SRDRN.

3.4 Extremes

For T_{mean} extremes, we evaluated 98th percentile temperature and annual maximum warm spells of the time series at each grid cell, which are two commonly used temperature extreme indices (Baño-Medina et al. 2020; Hertig et al. 2019) and can have varying effects on agriculture, ecosystems, energy use, and human health (Nairn and Fawcett 2015; Nicholls et al. 2008; Pattenden et al. 2003; Pierce et al. 2014). Figure 8 shows that the RMSE for T_{mean} at the 98th percentile between SRDRN bias corrected results and observations are around 1 °C among different GCMs. While RMSEs of SRDRN and QDM bias corrected T_{mean} are much lower than Bilinear for most of the 20 GCMs, QDM shows much larger inter-model variability than SRDRN. Figure 9 presents the spatial distribution of 98th percentile T_{mean} of observations (OBS), the raw output, Bilinear, QDM and SRDRN of a randomly selected GCM (CESM2-WACCM) as well as their biases. We can see that GCM has very high temperature biases around the northwest region of the study area and the states border between Tennessee and North Carolina (the edge of the Appalachian Mountains). Both QDM and SRDRN substantially reduced biases in the

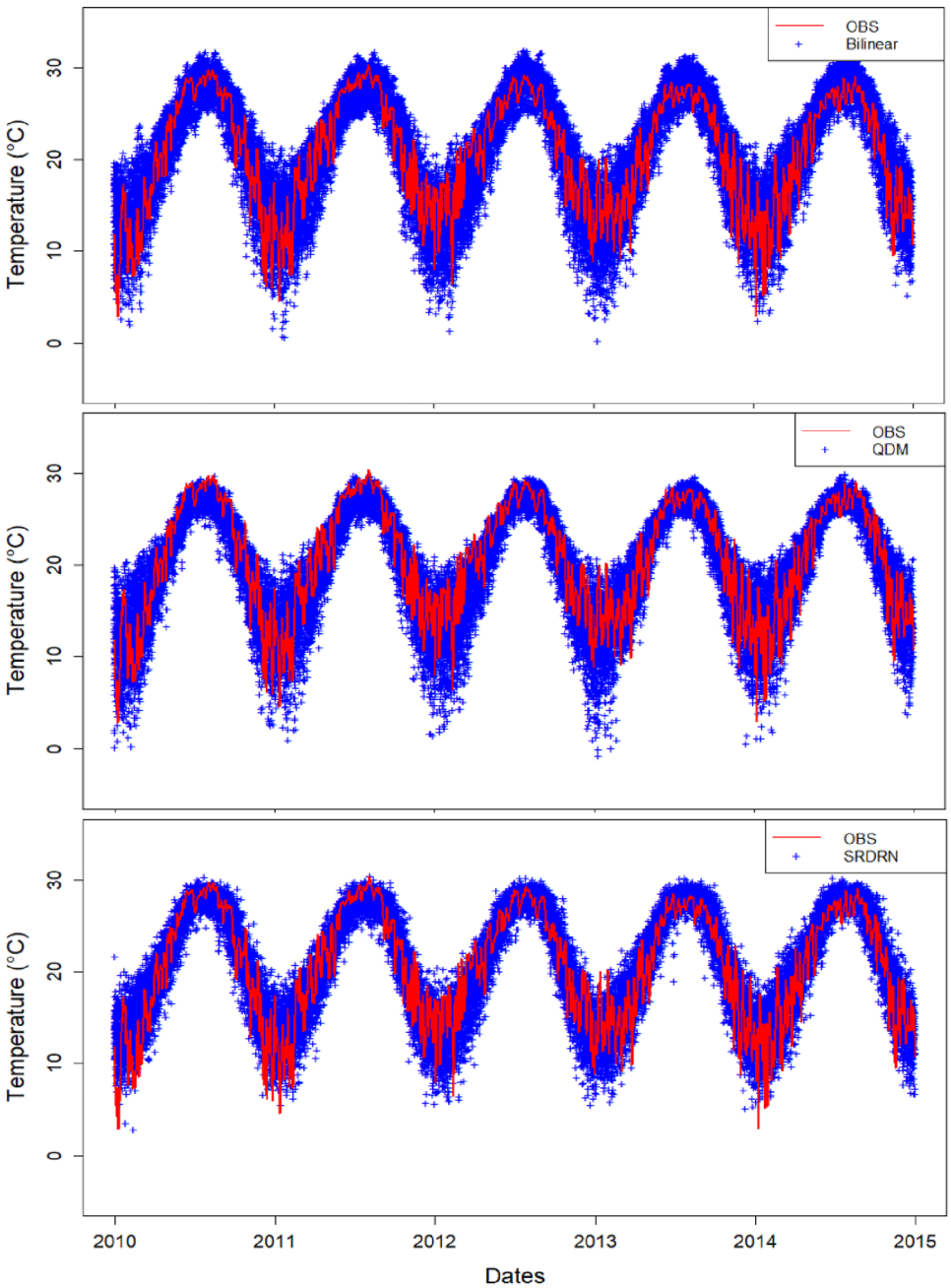

Fig. 6 Time series of daily mean temperature (T_{mean}) averaged over the study area. Blue cross symbols (+) represent T_{mean} from the Bilinear (the top plot) downscaled, and QDM (the middle plot) and SRDRN (the bottom plot) downscaled and bias corrected GCMs. Red lines represent ERA5 data (OBS)

areas with high biases. Compared to QDM, SRDRN reduced more biases over ocean areas. Since T_{mean} over the ocean surface is more homogeneous than over land, SRDRN's feature extraction ability likely treated T_{mean} over the ocean as background features, which has been found relatively easier to be captured than more detailed fine features (Sobral and Vacavant 2014). QDM, however, only adjusted temporal sequence of daily temperature at each individual grid point without accounting for homogeneity feature of the ocean.

Warm spell is defined as consecutive days when T_{mean} is greater than 90th percentile (Baño-Medina et al. 2020; Hertig et al. 2019). Figure 10 presents RMSE of the annual maximum warm spells derived from the SRDRN and QDM bias corrected T_{mean}. The results indicate that, overall, RMSE of annual maximum warm spell between SRDRN and observations is consistently lower than Bilinear and has small inter-model variability. The RMSE of QDM, however, has high inter-model variability, much higher than Bilinear for several GCMs (e.g., CanESM5 and ESM1-2-HR).

3.5 Multivariate bias correction

We investigated multivariate bias correction and downscaling of T_{max} and T_{min} using SRDRN, which took both T_{max}

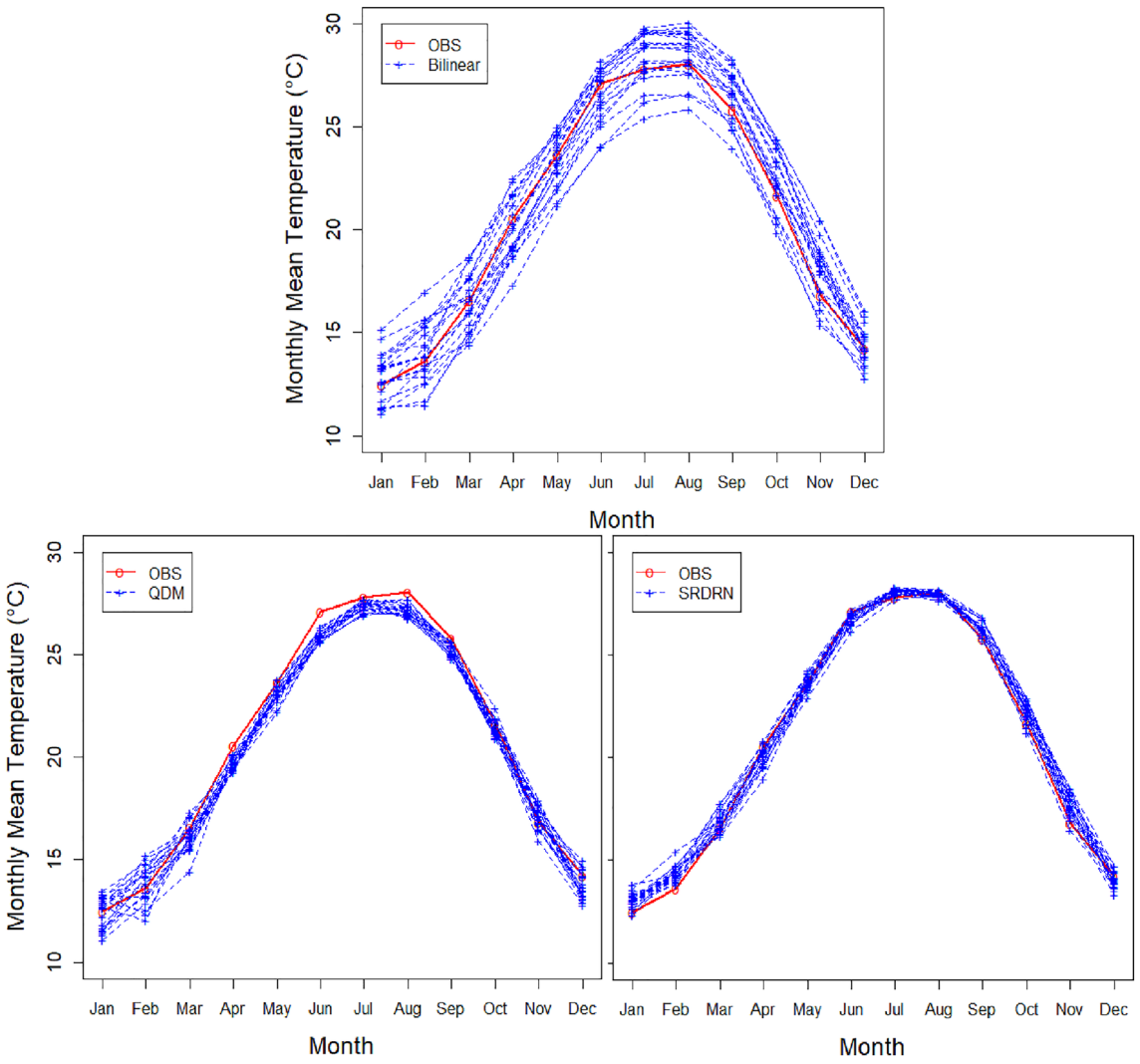

Fig. 7 Monthly mean of daily mean temperature T_{mean} from Bilinear, QDM and SRDRN methods compared to observations as averaged over the study area. The blue dash lines represent 20 Bilinear interpolated GCMs (the top plot), as well as QDM (the left plot) and SRDRN (the right plot) downscaled and bias corrected GCMs, and the red lines represent observations

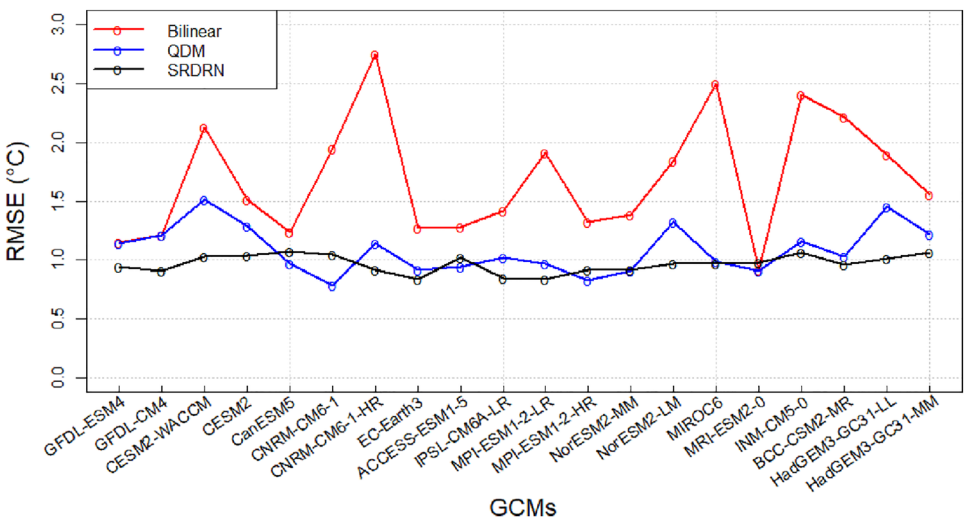

Fig. 8 RMSE of 98th percentile of daily mean temperature (T_{mean}) from Bilinear, QDM, and SRDRN

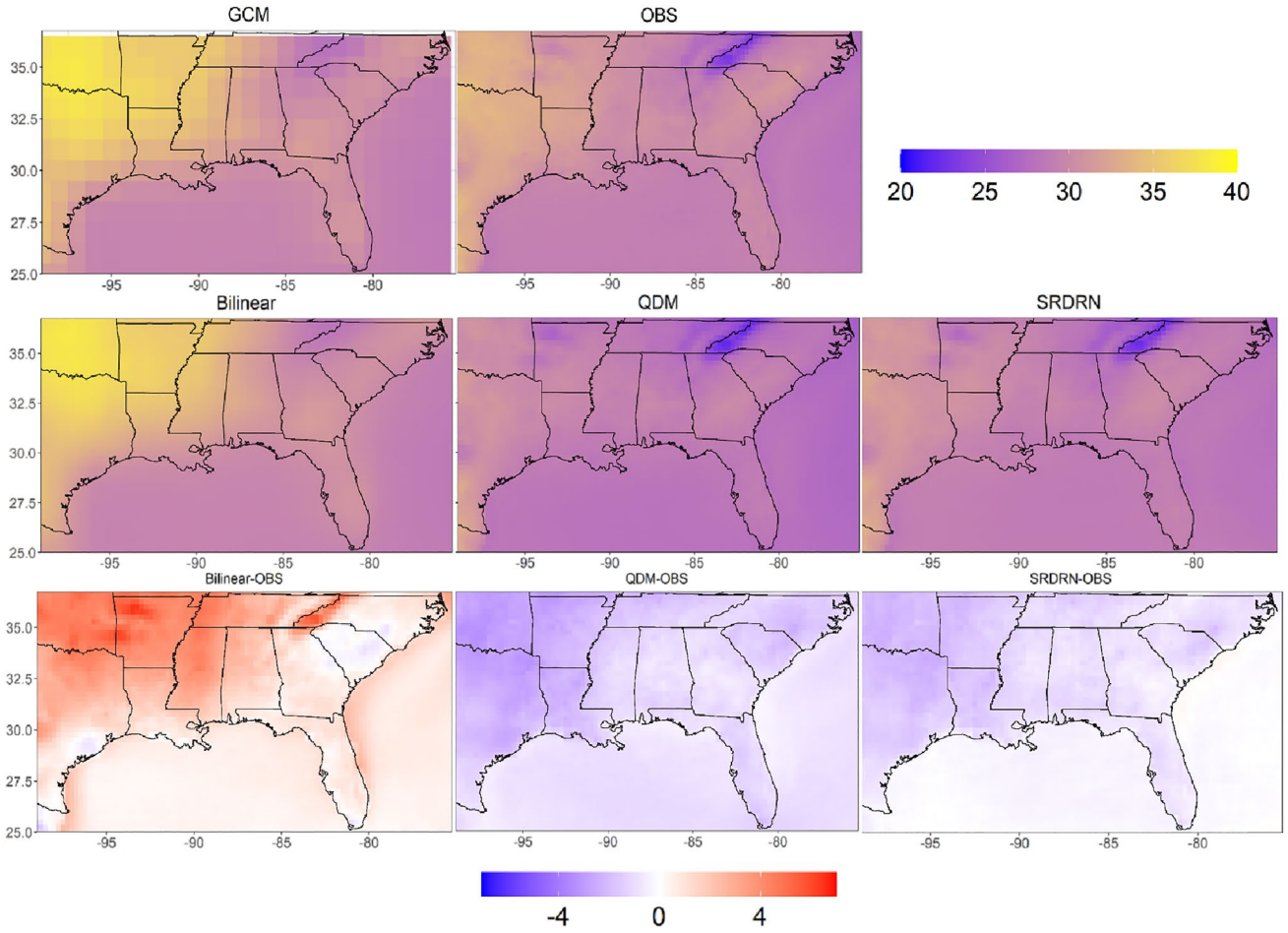

Fig. 9 Spatial distribution of T_{mean} at 98th percentile for a randomly selected GCM (CESM2-WACCM) and OBS (upper row), Bilinear, QDM, and SRDRN (middle row), and their differences with observations (bottom row)

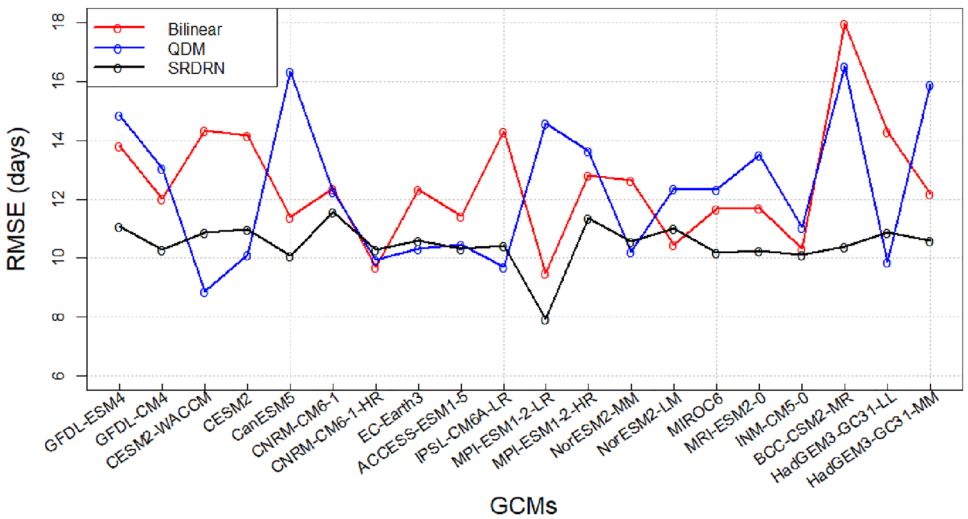

Fig. 10 RMSE of annual maximum warm spell for Bilinear, QDM, and SRDRN of 20 GCMs

and T_{min} as input-output channels (Fig. 1). In this case, filters with different parameters for the T_{max} and T_{min} input channels were simultaneously applied in the convolutional operation and these parameters were optimized during training period by decreasing the differences between the GCM models and observations (namely loss function). Thus, the

intervariable relationship between T_{max} and T_{min} was captured by the algorithm itself without predefining any relationships as the other multivariate bias correction methods usually do. In general, the RMSE for T_{max}, T_{min} and their diurnal range (i.e. T_{max}-T_{min}) between SRDRN bias corrected data and observations are much lower than Bilinear and QDM for all the 18 GCMs (Fig. 11). Considering RMSE of Bilinear as the total GCM bias, SRDRN reduced up to 24% biases for T_{max} and up to 22% for T_{min}, while QDM approach reduced only up to 14% for both T_{max} and T_{min}. Unlike other multivariate bias correction methods (Bürger et al. 2011; Cannon 2016; Chen et al. 2018; Mehrotra and Sharma 2012, 2019), the SRDRN model does not assume any relationships among different variables. SRDRN offers an "end-to-end" modeling workflow, which allows the model to learn auto-customized relationship rather than subject to prior knowledge and has improved bias correction skills.

Studies have demonstrated that quantile mapping (QM) bias correction method may generate physically unrealistic artifacts (Agbazo and Grenier 2020; Thrasher et al. 2012). To examine this issue, we compared the difference between bias corrected T_{max} and bias corrected T_{min}. The results showed that about 4% of QDM bias corrected T_{max} is smaller than QDM bias corrected T_{min}, while the multivariate SRDRN bias corrected T_{max} is always greater than or approximately equal to T_{min} (Fig. 12 and Fig. S1 in the

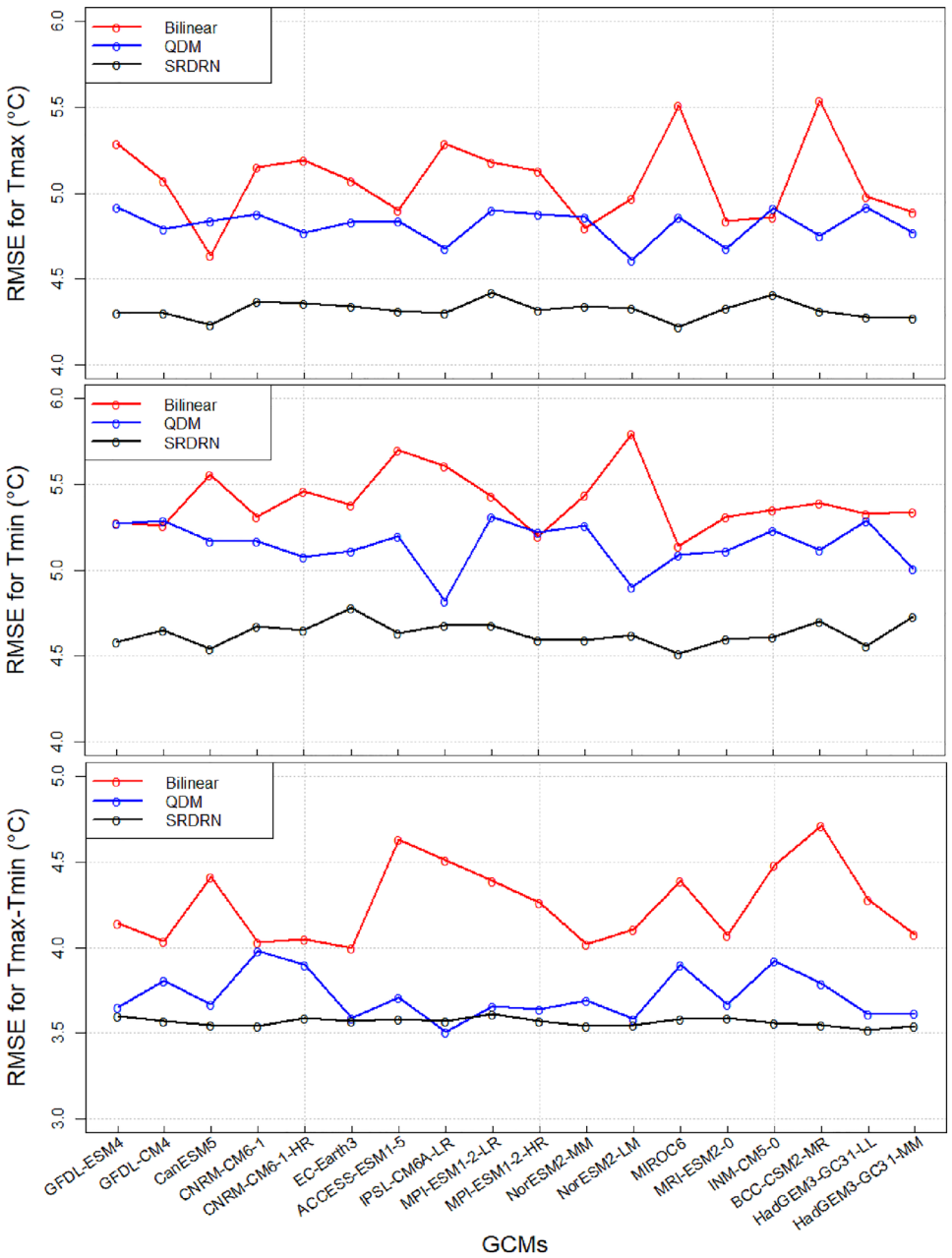

Fig. 11 RMSE for multi-channel SRDRN bias corrected T_{max}, T_{min} and T_{max}-T_{min}, comparing with Bilinear and QDM

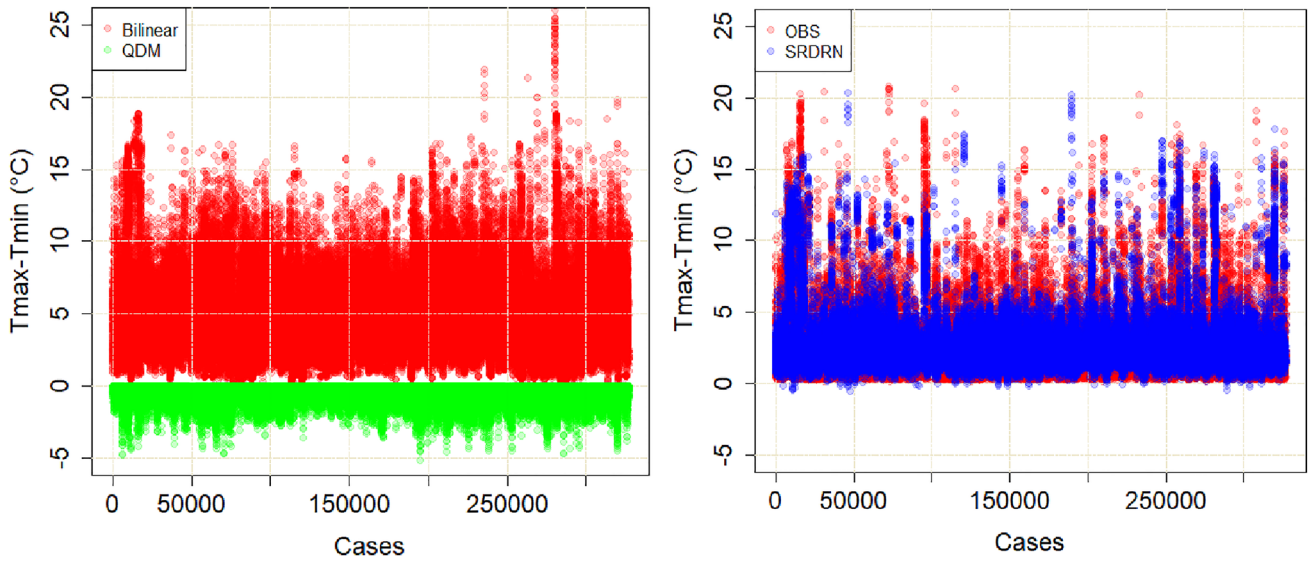

Fig. 12 Cases where $T_{min} > T_{max}$ in QDM bias correction (left panel) and corresponding cases generated by SRDRN bias correction (right panel)

Supplementary Information). This result suggests that the SRDRN model can capture the intervariable dependence between T_{max} and T_{min}, and thus address intervariable physical inconsistency.

We compared the performance of SRDRN with the dOTC in a relatively complex topography area around the state of Tennesse within the study area (see Fig. 13). Figure 13 shows the spatial distribution of mean on grids of the diurnal range (T_{max}-T_{min}) for a randomly selected GCM (HadGEM3-GC31-LL). It indicates that the spatial distribution of the mean diurnal range corrected by dOTC seems very close to observations (RMSE 0.23 °C), which is better than QDM (RMSE 0.31 °C), but is worse than SRDRN (RMSE 0.18 °C). The overall assessment of dOTC in terms of RMSE and correlation coefficient varies depending on different GCMs (see figures S2, S3 and S4 in the Supplementary Information) and is not consistently better than QDM, and is worse than SRDRN for all the GCMs. Regarding inter-variable dependence, dOTC generates even more negative values (2.7%) of the diurnal range (T_{max}-T_{min}) (see Fig. S5 in the Supplementary Information) than QDM (0.16%), SRDRN bias corrected results, however, do not have such issue. For the dOTC, empirical high dimensional multivariate distributions have to be esimated first and then multivariate coefficients with constraints were solved as a linear programming problem, which cannot gurrantee a optimum solution especially for very high dimensions and thus may cause its deterioration.

4 Discussion

4.1 On stacking multiple GCMs for bias correction

It is challenging to train a deep neural network that can perform well on bias correcting unseen data (testing data in this study, in a broader context, can be considered as "future climate") in the nonstationary context. Overfitting occurs when a neural network model learns the details and noise of training data to the extent that negatively impacts the performances on the testing data. The most robust way to avoid overfitting is to train the model with sufficiently large dataset in order to provide the deep learning algorithm more resources to learn the underlying mapping of inputs to outputs. Here we sequentially stacked 20 GCMs that greatly increased the amount of training dataset (Fig. 1), which resulted in a more robust model. Another way to improve model performance on the testing dataset is to use regularization techniques when training the model. Regularization is the process which discourages learning a more complex model by regularizing or shrinking certain parameters towards zero. The SRDRN algorithm includes batch normalization layers following each convolutional layer. Besides solving the internal covariate shift and speeding up the training process as mentioned in the methodology section, the batch normalization layers also served as an implicit regularization technique to avoid

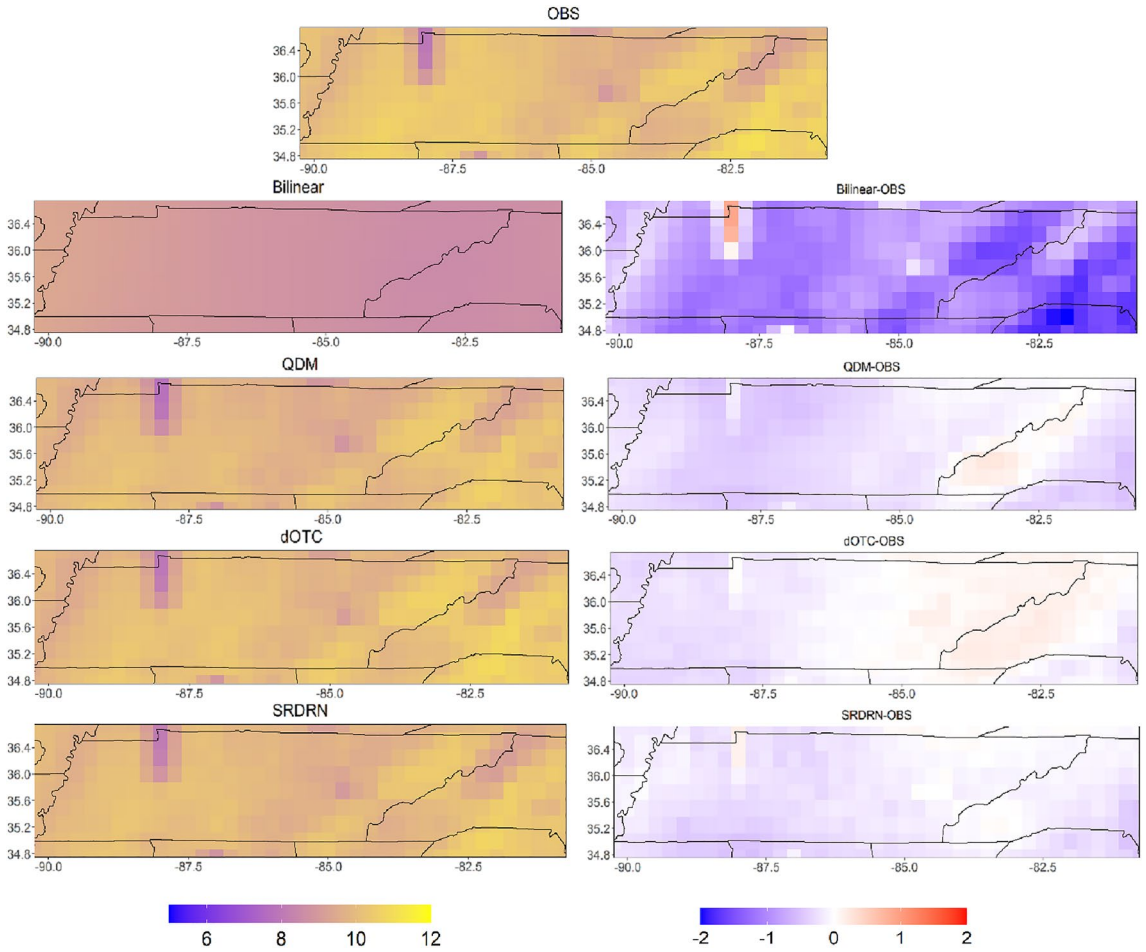

Fig. 13 Spatial distribution of mean of T_{max} -T_{min} at a complex topography area (around Tennessee State area) for a randomly selected GCM (HadGEM3-GC31-LL) and OBS (top row), Bilinear, QDM, dOTC and SRDRN (left column), and their differences with observations (right column)

overfitting and thus improve the model performance when facing completely new data (Luo et al. 2018).

Besides benefitting from large training data by stacking the 20 GCMs, SRDRN simultaneously bias corrected and downscaled multiple GCMs, which enabled the model to account for relative performances among different GCMs against observations and provided more resources to discover common noises among GCMs. The SRDRN algorithm updated weights per mini-batch size (64) by decreasing differences or loss between bias corrected products and observations through back propagation algorithm. Since 20 GCMs were stacked in sequence, each mini-batch size of data may include some data from relative well performed GCMs as well as relative poorly performed GCMs. Through the training process, the SRDRN algorithm exhaustively learned the relationships between different GCMs and observations, and therefore removed more noises for the relative poorly performed GCMs and converged towards observations as close as possible, which made the SRDRN bias corrected GCMs consistent and stable. The distributions for individual GCM were also well corrected by SRDRN (see Fig. S6 in the Supplementary Information) compared to QDM which explicitly adjusted quantiles individually for each GCM. We bias correct multi-model ensemble members together instead of each model member separately, for the sake of retaining uncertainties of climate simulations contributed from each model while correcting total biases. This is conceptually similar to the study of bias correcting large ensembles from a single climate model (Ayar et al. 2021), which used all members together to construct the cumulative density function (CDF) for bias correction in order to preserve internal variability while correcting biases. On the other hand, bias correcting each member separately may over adjust the model, which ignore the nature of climate simulation uncertainty represented by multi-model ensembles or single-model large ensembles.

4.2 On bias correcting spatial and multivariate dependences

The SRDRN deep learning model extracted and evaluated spatial features between GCMs outputs and observations, and then corrected the biases after reconstructing the spatial features. That means SRDRN can correct spatial dependence biases (lower RMSE and higher correlation coefficient of spatial semivariance for all the GCMs as shown in Fig. 4 as well as Fig. S4 in the Supplementary Information), which is critical for climate impact assessment considering the importance of spatial climate variability. In recent studies, Nahar et al. (2018) developed an independent component analysis (ICA) approach based on principal components to bias correct spatial biases for monthly GCM outputs (precipitation and temperature), but its improvement is limited for the testing dataset due to strictly stationary assumption. François et al. (2021) applied a deep learning architecture based on generative adversarial networks called CycleGAN to correct spatial biases after quantile mapping correction for daily GCM outputs. The results are comparable with quantile mapping and two other multivariate bias correction methods. As a proof of concept, François et al. (2021) used a simple architecture of neural networks with a small number of CNN layers (7), which may limit its capability on capturing more complex spatial relationships for the correction of climate simulations. The SRDRN, however, includes 37 CNN layers and can potentially capture more complex spatial relationships and correct fine spatial differences between model simulations and observations.

SRDRN considered multiple variables as different input-output channels. The intervariable dependence was captured during model training process and SRDRN can learn the real complex relationship from the climate "big data" beyond our prior knowledge. While this study explored how well the SRDRN can capture the intervariable relationship between maximum and minimum temperatures, it can be potentially applied to simultaneously downscale and bias correct more climate variables besides temperatures. Current multivariate bias correction approaches including dOTC, however, present differences in terms of assumptions and philosophical features (such as deterministic versus stochastic) (François et al. 2020). Consequently, the performance of the corrected outputs can vary largely from one method to another and mostly fail to compete with univariate bias correction methods particularly for testing dataset as shown in the figures S2-S5 in the Supplementary Information as well as in the previous studies (Cannon 2018; Chen et al. 2014; Chen et al. 2018; François et al. 2020; Guo et al. 2020). Furthermore, the existing univariate bias correction methods that ignore the intervariable relationship may generate unrealistic artefacts and thus it is necessary to consider multivariate bias correction, which is also important for large-scale modeling frameworks such as assessing risks from compound extreme climate events (Zscheischler et al. 2018).

4.3 Caveats and future study

The results indicate that the SRDRN deep learning model can reduce more biases in space, time, and extremes compared to conventional bias correction methods, as well as correct biases in spatial and intervariable dependence for 'future' climate in the nonstationary context. However, while SRDRN, as a convolutional type of deep learning approach, excels at learning spatial patterns, it did not consider sequential connections of daily temperature or temporal structure. Since temperature has high autocorrelation, incorporating time dependence between images by replacing 2-dimension convolutional layers with 3-dimension ones has the potentials to further improve bias correcting performance, which can be explored in the future study. Furthermore, the SRDRN was evaluated for bias correction and downscaling daily temperature, which is a continuous variable. We expect the model would also work for other continuous climate variables, such as solar radiation, humidity, etc. However, for highly skewed, non-continuous climate variables like precipitation, it requires additional research, for example, by incorporating distribution into loss function (Tao et al. 2016) or adding regularization term that penalizes deviations at the grid cell resolution between GCM outputs and observations (Ravuri et al. 2021). Another issue is that we synchronized observations and model simulations in time and used MSE as loss function, which means daily maps from the 20 GCMs are forced to resemble those observed, without considering the atmospheric state of the different climate models. As a result, dynamics of the climate models may be lost during the correction procedure, which may be a limitation of using the SRDRN corrections for impact analyses. This limitation may be addressed in the future by modifying loss functions to match distributions of climate models with observations (instead of day-to-day sequence) while bias correcting all the required climate variables simultaneously considered for impact studies.

The SRDRN algorithm is capable of learning finer and more intricate features especially for extreme events compared to shallow plain architectures (Wang et al. 2021). Systematically bias correcting intricate feature differences between GCM and observations is critical for impact assessments facing climate change. But one weakness of the SRDRN algorithm is that it does not explicitly account for physical processes in the bias correction procedure François et al. 2020; Ivanov et al. 2018; Maraun et al. 2017; Wang et al. 2021) and lacks interpretability to help the users gain understanding of the model (McGovern et al. 2019). Thus, additional studies, such as incorporating causality analysis (Liang et al. 2021), can be helpful to improve its

interpretability in the future. Furthermore, one limitation of deep learning is its computational cost. The training time for this study for each scenario was about 5 h using one graphic processing unit (GPU). For larger scale applications, more GPUs may be needed to reduce the computing time, which will accordingly increase the computational cost.

5 Conclusions

Climate models requires bias correction and downscaling for accurate impact analysis. Current univariate bias correction approaches are mostly applied to one variable, one location at a time, which may result in misrepresentation of the spatial and intervariable structures of the bias corrected outputs. On the other hand, current multivariate bias correction approaches assumed pre-defined relationships among variables and over simplifications for handling very high dimensional data, causing disparity between training and testing periods and loss of capability under nonstationary climate change condition. The deep learning approach based on convolutional neural network can systematically learn complex relationships among multivariate 2-dimensional data without pre-assuming any relationships and thus has great potentials to address the current issues faced by climate model downscaling and bias corrections. In this study, using 20 CMIP6 GCMs T_{mean}, T_{max} and T_{min} outputs, we comprehensively evaluated SRDRN, a deep convolutional neural network based model, comparing with one widely used univariate (QDM) and one multivariate (dOTC) bias correction approaches, for climate model downscaling and bias correction. We found that SRDRN considerably corrected more biases in space, time, and extremes compared to QDM and dOTC. It also well addressed the artefacts found in QDM and dOTC by well reproducing spatial and intervariable structures of the observations. It greatly reduced inter-model variability among downscaled and bias corrected GCMs by considering historical relations between different GCMs and observations, and potentially preserved inter-model uncertainty through combining all the GCMs together during training. Further studies on deep learning by additionally accounting for sequential relationships of climate variables and further evaluating and improving for highly skewed non-continuous climate variables may take the climate model downscaling and bias correction to the next level.

Appendix A

It is difficult to estimate a mapping function from x to $H(x)$ as plain architectures do (Fig. 14). Rather than expect stacked layers (Layer 1, Layer 2 to Layer n in Fig. 14) to approximate

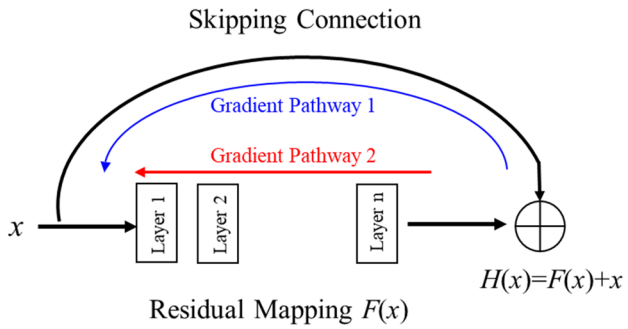

Fig. 14 Illustration of residual block with residual mapping and skipping connection

$H(x)$, residual mapping explicitly let these layers approximate a residual function $F(x)$ (i.e., $H(x)$-x) instead. It has been approved that learning residual mapping is relatively easier than directly estimating $H(x)$ and can mitigate degradation issue in very deep architectures. Skipping connections also provide an alternative path for backpropagation algorithm to update weights (see the blue line in Fig. 14) and avoid vanishing gradient issue.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00382-022-06277-2.

Acknowledgements This study is supported in part by the NOAA RESTORE program (No. NA19NOS4510194), by the Alabama Agricultural Experiment Station and the Hatch program of the USDA National Institute of Food and Agriculture (NIFA) (No. 1012578), by the USDA NIFA Agriculture and Food Research Initiative (AFRI) program (No. 1019690) and by the NASA EPSCoR R3 program support (No. AL-80NSSC21M0138). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

Funding This study was funded by National Oceanic and Atmospheric Administration, National Institute of Food and Agriculture, and National Aeronautics and Space Administration.

Data and Code Availability Statement The surface air temperature dataset from the 20 GCMs as part of the CMIP6 climate model simulations can be downloaded through the Earth System Grib 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/). dOTC code is publicly available at https://github.com/yrobink/SBCK (Robin et al. 2019).

Declarations

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

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

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

Ayar PV, Vrac M, Mailhot A (2021) Ensemble bias correction of climate simulations: preserving internal variability. Sci Rep 11(1):1–9

Baño-Medina J, Manzanas R, Gutiérrez JM (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. Geosci Model Dev 13:2109–2124

Bürger G, Schulla J, Werner A (2011) Estimates of future flow, including extremes, of the Columbia River headwaters. Water Resourc Res 47

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

Chen J, Zhang XJ, Brissette FP (2014) Assessing scale effects for statistically downscaling precipitation with GPCC model. Int J Climatol 34:708–727

Chen J, Li C, Brissette FP, Chen H, Wang M, Essou GR (2018) Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. J Hydrol 560:326–341

Deser C, Phillips A, Bourdette V, Teng H (2012) Uncertainty in climate change projections: the role of internal variability. Clim Dyn 38(3):527–546

Eum HI, Cannon AJ (2017) Intercomparison of projected changes in climate extremes for South Korea: application of trend preserving statistical downscaling methods to the CMIP5 ensemble. Int J Climatol 37:3381–3397

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

François B, Vrac M, Cannon AJ, Robin Y, Allard D (2020) Multivariate bias corrections of climate simulations: which benefits for which losses? Earth Sys Dyn 11:537–562

François B, Thao S, Vrac M (2021) Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. Clim Dyn 1–31

Guo Q, Chen J, Zhang XJ, Xu CY, Chen H (2020) Impacts of using state-of-the-art multivariate bias correction methods on hydrological modeling over North America. Water Resour Res 56:e2019WR026659

Ham Y-G, Kim J-H, Luo J-J (2019) Deep learning for multi-year ENSO forecasts. Nature 573:568–572

He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, 1026–1034

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 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

Hertig E et al (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

Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, PMLR, 448–456

Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 1125–1134

Ivanov MA, Luterbacher J, Kotlarski S (2018) Climate model biases and modification of the climate change signal by intensity-dependent bias correction. J Clim 31:6591–6610

Jose DM, Dwarakish GS (2021) Bias correction and trend analysis of temperature data by a high-resolution CMIP6 model over a tropical River Basin. Asia-Pac J Atmos Sci,1–19

Kim S, Joo K, Kim H, Shin J-Y, Heo J-H (2021) Regional quantile delta mapping method using regional frequency analysis for regional climate model precipitation. J Hydrol 596:125685

Kumar D, Ganguly AR (2018) Intercomparison of model response and internal variability across climate model ensembles. Clim Dyn 51(1):207–219

LeCun Y, Bengio Y, Hinton G (2015) Deep Learn Nat 521:436–444

Ledig C et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 4681–4690

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(1):1–14

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, Ganguly AR, Dy J (2020) Climate Downscaling Using YNet: A Deep Convolutional Network with Skip Connections and Fusion. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3145–3153

Luo P, Wang X, Shao W, Peng Z (2018) Towards understanding regularization in batch normalization. arXiv preprint arXiv:1809.00846

Maraun D et al (2017) Towards process-informed bias correction of climate change simulations. Nat Clim Change 7:764–773

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys, 48(3)

McGovern A, Lagerquist R, Gagne DJ, Jergensen GE, Elmore KL, Homeyer CR, Smith T (2019) Making the black box more transparent: Understanding the physical implications of machine learning. Bull Am Meteorol Soc 100:2175–2199

Mearns LO et al (2012) The North American regional climate change assessment program: overview of phase I results. Bull Am Meteorol Soc 93:1337–1362

Mehrotra R, Sharma A (2012) An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations. Water Resourc Res, 48

Mehrotra R, Sharma A (2019) A resampling approach for correcting systematic spatiotemporal biases for multiple variables in a changing climate. Water Resour Res 55:754–770

Meyer J, Kohn I, Stahl K, Hakala K, Seibert J, Cannon AJ (2019) Effects of univariate and multivariate bias correction on

hydrological impact projections in alpine catchments. Hydrol Earth Syst Sci 23:1339–1354
Nahar J, Johnson F, Sharma A (2017) Assessing the extent of non-stationary biases in GCMs. J Hydrol 549:148–162
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
Nairn JR, Fawcett RJ (2015) The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. Int J Environ Res Public Health 12:227–253
Nicholls N, Skinner C, Loughnan M, Tapper N (2008) A simple heat alert system for Melbourne, Australia. Int J Biometeorol 52:375–384
Pan B, Anderson GJ, Goncalves A, Lucas DD, Bonfils CJ, Lee J, Ma HY (2021) Learning to correct climate projection biases. J Adv Model Earth Syst 13(10): e2021MS002509
Panofsky HA, Brier GW (1968) Some applications of statistics to meteorology, The Pennsylvania State University, University Park, PA, USA, 224
Pattenden S, Nikiforov B, Armstrong B (2003) Mortality and temperature in Sofia and London. J Epidemiol Community Health 57:628–633
Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30:683–691
Pierce DW, Cayan DR, Thrasher BL (2014) Statistical downscaling using localized constructed analogs (LOCA). J Hydrometeorol 15:2558–2585
Racah E, Beckham C, Maharaj T, Kahou SE, Pal C (2016) ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. arXiv preprint arXiv:1612.02095
Ravuri S et al (2021) Skillful Precipitation Nowcasting using Deep Generative Models of Radar. arXiv preprint arXiv:2104.00954
Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566:195–204
Robin Y, Vrac M, Naveau P, Yiou P (2019) Multivariate stochastic bias corrections with optimal transport. Hydrol Earth Syst Sci 23:773–786. https://doi.org/10.5194/hess-23-773-2019
Rodrigues ER, Oliveira I, Cunha R, Netto M (2018) DeepDownscale: a deep learning strategy for high-resolution weather forecast. In: 2018 IEEE 14th International Conference on e-Science (e-Science), IEEE, 415–422
Sillmann J, Kharin V, Zhang X, Zwiers F, Bronaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. J Geophys Res Atmos 118:1716–1733
Silver D et al (2017) Mastering the game of go without human knowledge. Nature 550:354–359
Sobral A, Vacavant A (2014) A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput Vis Image Underst 122:4–21
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, 1349–1355
Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans R Soc A Math Phys Eng Sci 365(1857):2053–2075
Tegegne G, Melesse AM (2021) Comparison of Trend Preserving Statistical Downscaling Algorithms Toward an Improved Precipitation Extremes Projection in the Headwaters of Blue Nile River in Ethiopia. Environ Process 8:59–75
Thrasher B, Maurer EP, McKellar C, Duffy PB (2012) Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. Hydrol Earth Syst Sci 16:3309–3314
Tian C, Fei L, Zheng W, Xu Y, Zuo W, Lin C-W (2020) Deep learning on image denoising: an overview. Neural Networks
Tong Y, Gao X, Han Z, Xu Y, Xu Y, Giorgi F (2020) Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. Clim Dyn 1–19
Van de Velde J, Demuzere M, De Baets B, Verhoest NE (2020) Impact of bias nonstationarity on the performance of uni-and multivariate bias-adjusting methods. Hydrol Earth Syst Sci Discuss 1–47
Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR (2017) Deepsd: Generating high resolution climate change projections through single image super-resolution. In: Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining, 1663–1672
Wang J, Chen Y, Tett SF, Yan Z, Zhai P, Feng J, Xia J (2020) Anthropogenically-driven increases in the risks of summertime compound hot extremes. Nat Commun 11:1–11
Wang F, Tian D, Lowe L, Kalin L, Lehrter J (2021) Deep Learning for Daily Precipitation and Temperature Downscaling. Water Resour Res 57:e2020WR029308
Wood AW, Maurer EP, Kumar A, Lettenmaier DP (2002) Long-range experimental hydrologic forecasting for the eastern United States. J Geophys Res Atmos 107:ACL 6-1-ACL 6–15
Xu Y, Noy A, Lin M, Qian Q, Li H, Jin R (2020) WeMix: How to Better Utilize Data Augmentation. arXiv preprint arXiv:2010.01267
Yang C, Kim T, Wang R, Peng H, Kuo C-CJ (2018) : ESTHER: extremely simple image translation through self-regularization. BMVC, 110
Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, 2223–2232
Zscheischler J et al (2018) Future climate risk from compound events. Nat Clim Change 8:469–477
Zscheischler J et al (2020) A typology of compound weather and climate events. Nat reviews earth Environ 1:333–347