Customized Deep Learning for Precipitation Bias Correction and Downscaling

Fang Wang[1], Di Tian[1]*, and Mark Carroll[2]

[1]Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849, USA
[2]Computational and Information Science Technology Office, NASA Goddard Space Flight Center Greenbelt, MD 20771, USA

**Correspondence to*: Di Tian (tiandi@auburn.edu)

Abstract. Systematic biases and coarse resolutions are major limitations of current precipitation datasets. Many deep learning (DL) based studies have been conducted for precipitation bias correction and downscaling. However, it is still challenging for the current approaches to handling~~handle~~ complex features of hourly precipitation, resulting in the incapability of reproducing small-scale~~small scale~~ features, such as extreme events. This study developed a customized DL model by incorporating customized loss functions, multitask~~multitask~~ learning~~,~~ and physically relevant covariates to bias correct and downscale hourly precipitation data. We designed six scenarios to systematically evaluate the added values of weighted loss functions, multitask~~multi-task~~ learning, and atmospheric covariates compared to the regular DL and statistical approaches. The models ~~was~~ were trained and tested using the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA2) reanalysis and the Stage IV radar observations over the northern coastal region of the Gulf of Mexico at an hourly time scale. We found that all the scenarios with weighted loss functions performed notably better than the other scenarios with conventional loss functions and a quantile mapping-based approach at hourly, daily, and monthly time scales as well as extremes. Multitask~~Multitask~~ learning showed improved performance on capturing fine features of extreme events and ~~hourly precipitation climatology, aggregated precipitation at daily and monthly scales, and detailed features of extreme events, while the improvement is not as large as from weighted loss functions.~~accounting for atmospheric covariates, highly improved model performance at hourly and aggregated time scales~~-~~, while the improvement is not as large as from weighted loss functions. ~~Accounting for atmospheric covariates further improved the model performance for capturing extreme events.~~ We show that the customized DL model can better downscale and bias correct hourly precipitation datasets and provide improved precipitation estimates at fine spatial and temporal resolutions where regular DL and statistical methods ~~experiencing~~ experience challenges.

1 Introduction

Precipitation is a major component of the hydrological cycle and is fundamentally important for many applications, such as water resources planning and management, disaster risk management, and agriculture, amongst many others. Due to the limited coverage of ground-based rain gauges, numerous gridded precipitation datasets have been developed over the past decades, including gauge-based, satellite-based, reanalysis products, and merged products (Beck et al., 2019a; Sun et al., 2018). These datasets are different in terms of data sources, coverage, spatial and temporal resolution, and algorithms (see Sun et al., 2018 for a review), which provide a potential source of information to regions where conventional in situ precipitation measurements are lacking (Sun et al., 2018).

Gridded precipitation datasets have proven to be useful across a wide range of research fields, including climate trends~~trend~~ and extreme precipitation (Bhattacharyya et al., 2022; Degaetano et al., 2020; Fischer and Knutti, 2016; Kim et al., 2019; King et al., 2013), droughts and floods monitoring (Aadhar and Mishra, 2017; Peng et al., 2020; Suliman et al., 2020; Zhong et al., 2019), and driving hydrological models (Raimonet et al., 2017; Xu et al., 2016). However, many studies have identified that these gridded precipitation datasets include substantial biases in certain aspects compared to in situ observations (Aadhar and Mishra, 2017; Ashouri et al., 2016; Bitew and Gebremichael, 2011; Cavalcante et al., 2020; Jiang et al., 2021; Jury, 2009; Rivoire et al., 2021; Sun et al., 2018; Tong et al., 2014; Xu et al., 2016; Yilmaz et al., 2005). For example, Ashouri et al. (2016) evaluated the performance of NASA's~~NASA’s~~ Modern-Era Retrospective Analysis for Research and Applications (MERRA) precipitation reanalysis dataset and found that MERRA tends to overestimate the frequency at which the 99th percentile of precipitation is exceeded and underestimate the magnitude of extremes, especially over the Gulf Coast regions of the United States. Furthermore, spatial resolution for most of these gridded precipitation datasets is relatively coarse for local scale applications (mostly above 0.25°, Sun et al., 2018). Therefore, the gridded precipitation datasets require bias correction and downscaling (Duethmann et al., 2013; Emmanouil et al., 2021; Mamalakis et al., 2017; Seyyedi et al., 2014).

Bias correcting and downscaling gridded precipitation data is challenging due to its complex characteristics (e.g., highly skewed, unbalanced feature, and complex spatial-temporal structure). Various approaches have been developed to tackle this issue, including traditional quantile mapping (QM) based bias correction and downscaling methods (e.g., Cannon et al., 2015; Panofsky and Brier, 1968; Thrasher et al., 2012; Wood et al., 2002) and recent machine learning based approaches such as random forests (He et al., 2016b; Legasa et al., 2022; Long et al., 2019; Mei et al., 2020; Pour et al., 2016), support vector machines (Tripathi et al., 2006) and artificial neural networks (Schoof and Pryor, 2001; Vandal et al., 2019). Recently, advances in deep learning have made a significant impact on many fields and have been proven superior to traditional machine learning methods because of their powerful abilities to learn~~in learning~~ spatiotemporal feature representation in an end-to-end manner (Ham et al., 2019; Reichstein et al., 2019; Shen, 2018). In particular, deep learning (DL) with convolutional neural network (CNN) types of approaches have achieved notable progress in modeling spatial context data (Lecun et al., 2015) and have been used for bias correcting and downscaling low spatial resolution data (Kumar et al., 2021; Sha et al., 2020a, b; Vandal et al., 2018b; Wang et al., 2021; Xu et al., 2020), climate model outputs (François et al., 2021; Liu et al., 2020; Pan et al., 2021; Rodrigues et al., 2018; Wang and Tian, 2022), reanalysis products (Baño-Medina et al., 2020; Sun and Tang, 2020), and weather forecast model outputs (Harris et al., 2022; Li et al., 2022). While these studies have indicated many promising strengths and advantages over traditional downscaling and bias correction approaches, most of them have difficulties capturing~~to capture~~ local-, small-scale features such as extremes for an unseen dataset. For example, Baño-Medina et al. (2020) designed different DL configurations with a different number of plain CNN layers to bias correct and downscale daily ERA5-Interim reanalysis from 2° spatial resolution to 0.5°, and the overall performance is still marginal compared with simple generalized linear regression models and highly underestimated precipitation extremes. Harris et al. (2022) developed a generative adversarial networks (GANs) architecture to bias correct and downscale weather forecast outputs and found that it is more challenging to account for forecast error (or bias) in a spatially-coherent manner compared to the pure downscaling problem (Kumar et al., 2021; Sha et al., 2020a, b; Vandal et al., 2018b; Wang et al., 2021; Xu et al., 2020). The reason for that may be due to the sparsity of training data on extreme events. Deep learning (DL) models, however, need large training data in order to obtain a better regularization model for rare events in the unseen dataset.

Customized DL models have been proposed to generate physically consistent results and have better generalization ability for out-of-pocket~~out of pocket~~ datasets~~dataset~~ in the earth and environmental science field, which include incorporating customized loss functions (Kashinath et al., 2021), inputs from physically relevant auxiliary predictors (i.e., covariates) (Li et al., 2022; Rasp and Lerch, 2018), and customized multitask~~multitask~~ learning (Ruder, 2017). For example, Daw et al. (2017) indicated success in lake temperature modeling by incorporating a physics-based loss function in the DL objective compared to a regular loss function. Li et al. (2022) used a CNN-based approach to postprocess numerical weather prediction model output and found that the use of auxiliary predictors greatly improved model performance compared with raw precipitation data as the only predictor. ~~.~~A multitask~~multitask~~ model is trained to predict multiple tasks that are driven by the same underlying physical processes~~,~~ and thus has the potential to learn to better represent the shared physical process and better predict the variable of interest (Ruder, 2017). Multitask models have proven effective in several applications, including natural language processing (Chen et al., 2014; Seltzer and Droppo, 2013), computer vision (Girshick, 2015), as well as hydrology (Sadler et al., 2022). In addition, most of the previous bias correction and downscaling studies focused on the daily time scale (Baño-Medina et al., 2020; François et al., 2021; Harris et al., 2022; Kumar et al., 2021; Liu et al., 2020; Pan et al., 2021; Rodrigues et al., 2018; Sha et al., 2020a; Vandal et al., 2018b; Wang et al., 2021). However, the distribution of hourly precipitation data within a day is more important than daily or monthly aggregations for impacts and risks from warming-induced precipitation changes (Chen, 2020). Traditional DL loss functions have difficulties handling~~to handle~~ hourly precipitation data that are highly unbalanced with many zeros and highly positively~~positive~~ skewed for nonzero components. Therefore~~, therefore~~, customized DL with a weighted loss function to better balance nonzero components has the potential~~s~~ to improve the DL model performance. Besides the primary task of downscaling and bias correction task, adding a highly ~~correlated~~relevant classification task has ~~possibilities~~the potential to improve DL model performance on the primary task. ~~Including~~Incorporating covariates ~~that are highly correlated with~~selected based on precipitation formation theory (cloud mass movement and thermodynamics) also have the potential~~s~~ to improve ~~DL model performance on~~ precipitation downscaling and bias correction. ~~Customized deep learning, through incorporating customized loss functions, covariates, or customized multitask learning, have the potential to fundamentally improve hourly precipitation bias correction and downscaling.~~

Formatted: Normal, Line spacing: single

In this study, we will explore customized DL for precipitation bias correction and downscaling, aiming to take~~make~~ a step forward to address~~addressing~~ the current challenges described above. We designed a set of experiments to address this hypothesis using the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA2) reanalysis and the Stage IV radar precipitation data. The structure of this paper is organized as follows: Section 2 introduces~~introduced~~ data and ~~methodology~~study area; Section 3 introduces~~introduced~~ the methodology, ~~,~~ including the deep learning architecture and experimental designs for different scenarios, and a traditional bias correction approach as a benchmark; Section ~~3~~ 4 presents results; discussion and conclusions are provided in Section ~~4~~ 5 and ~~5~~6, respectively.

2 Data and ~~methodology~~Study Area

~~2.1 Data and study area~~

MERRA2 is a state-of-the-art global reanalysis product generated by the NASA Global Modeling and Assimilation Office (GMAO) using the Goddard Earth Observing System, version 5 (GEOS-5), and was introduced to replace and extend the original MERRA dataset (Reichle et al., 2017). It incorporates new satellite observations through data assimilation and benefits from advances in the GEOS-5 (Reichle et al., 2017). There are~~are~~ two hourly total precipitation (P) datasets available from the MERRA2 reanalysis product: the~~, the~~ model analysed~~analyzed~~~~analysed~~ precipitation computed from the atmospheric general circulation model and the observation-corrected P (Reichle et al., 2017). Both have a spatial resolution of 0.5° in latitude and 0.625° in longitude (~50km). MERRA2 observation-corrected precipitation has~~have~~ been used extensively in hydro-climatological analysis and modeling~~modelling~~ (Chen et al., 2021; Hamal et al., 2020; Xu et al., 2019; Xu et al., 2022). However, it still suffers from substantial biases (e.g., Hamal et al., 2020; Xu et al., 2019). This study will bias correct and downscale MERRA2 observation-corrected P using the Stage IV radar data (Lin and Mitchell, 2005) from the National Centers for Environmental Prediction (NCEP) as the observational reference. The Stage IV radar data has a 4 km spatial and hourly temporal resolution and covers the period from 2002 until the near present (2021 in this study). Stage IV radar was generated by merging data from 140 radars and about 5500 gauges over the continental United States (Lin and Mitchell, 2005; Nelson et al., 2016). ~~The~~ Stage IV provides highly accurate P estimates and has therefore been widely used as a reference for evaluating other P products (e.g., Aghakouchak et al., 2011; Aghakouchak et al., 2012; Beck et al., 2019b; Habib et al., 2009; Hong et al., 2006; Nelson et al., 2016; Zhang et al., 2018). The Stage IV dataset is a mosaic of regional analyses produced by 12~~by12~~ River Forecast Centers (RFCs) and is thus subject to the gauge correction and quality control performed at each individual RFC (Nelson et al., 2016).

The bias correction and downscaling experiments were performed in the rectangle coastal area of the Gulf of Mexico covering the entire states of Alabama, Mississippi, and Louisiana, and parts of neighbour states in the United States, ranging from -94.375° to -85.0° in longitude and from 29.0° to 35.0° in latitude. The study area falls into the humid subtropical climate and is highly influenced by extreme *P* events such as convective storms and hurricanes.

3 Methodology

Formatted: Normal, No bullets or numbering

~~2.2~~ 3.1 Customized DL approaches

Formatted: Font: Bold

~~2.2.1 Overview~~

This section first~~firstly~~ presents a brief description of a DL approach, namely, Super Resolution Deep Residual Network (SRDRN). Then, multitask~~multitask~~ learning, and customized loss functions are introduced based on the SRDRN architecture to construct customized DL approaches. Finally, we designed different modeling experiments, which include different combinations of multitask~~multitask~~ learning, customized loss functions, and *P* covariates as predictors, in order to evaluate the added values of each component of the customized DL approaches.

~~2.2.2~~3.1.1 SRDRN model

The SRDRN model is an advanced deep CNN-type~~CNN type~~ architecture and has been tested for downscaling daily *P* and temperature through synthetic experiments (Wang et al., 2021) and for bias-correcting~~bias correcting~~ near-surface~~near surface~~ temperature simulations from global climate models (Wang and Tian, 2022), considerably outperforming the conventional approaches. Furthermore, it has been proved that the SRDRN is capable of capturing much finer features than shallow plain CNN architecture (Wang et al., 2021). Compared~~Comparing~~ with the popular U-Net architecture (Sha et al., 2020a; Sun and Tang, 2020), the SRDRN directly extracts features~~feature~~ on the coarse resolution input~~,~~ and thus can potentially decrease computational and memory complexity.~~Furthermore, it has been proved that the SRDRN is capable of capturing much finer features than shallow plain CNN architecture (Wang et al., 2021). Comparing with the popular U-Net architecture~~ (Sha et al., 2020a; Sun and Tang, 2020)~~ (Sha et al. 2020b; Sun and Tang 2020), the SRDRN directly extracts feature on the coarse resolution input, and thus can potentially decrease computational and memory complexity~~

Here we provide a brief description of the SRDRN algorithm. For more details, the readers may refer to Wang et al. (2021). The SRDRN algorithm was developed based on a novel super-scaling~~super scaling~~ deep learning approach in the computer vision field (Ledig et al., 2017). Basically, the SRDRN algorithm is comprised of residual blocks and upsampling blocks with convolutional and batch normalization layers. For feature extraction, the convolutional layers apply filters to go through the input data to build a local connection within nearby grids by computing the element-wise dot product between the filters and different patches of the input. The outcome 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 so that the learning process can be stabilized and the training of the model can be accelerated (Ioffe and Szegedy, 2015).

With convolutional and batch normalization layers, the residual blocks are designed to extract fine spatial features while avoiding~~avoid~~ degradation issues~~issue~~ for the very deep neural network. Compared to plain CNN architectures, residual blocks can improve the performance of extensively deep networks (Silver et al., 2017) without suffering from model accuracy saturation and degradation (He et al., 2016a) because residual blocks execute residual mapping and include skipping connections. In this study, the way that skipping connection skips layers and connects the next layers is through element-wise addition. A~~The~~ total number of 16 residual blocks were used in the SRDRN architecture, which makes the network very deep and able to extract fine spatial features.

The upsampling blocks are applied to increase the spatial resolution for downscaling purposes~~purpose~~. The upsampling process is executed directly on the feature maps generated from the residual blocks, and each upsampling block is composed of one convolutional layer and one upsampling layer followed by a parametric ReLU activation function. The defaulted nearest neighbor interpolation was chosen in the upsampling layers to increase the spatial resolution, and the effects of different interpolation methods were not explored in this study. Each upsampling block sequentially and gradually increases the input low-resolution~~low resolution~~ feature maps by a factor of 2 or 3. In this study, the downscaling ratio (the ratio between coarse resolution and high-resolution data) is 12, and thus we used 3 upsampling blocks with two blocks having a factor of 2 and one block having a factor of 3.

~~2.2.3~~3.1.2 SRDRN model with multitask~~multitask~~ learning

We included an additional *P* classification task in the SRDRN model. Besides bias correcting and downscaling continuous hourly *P* values as a primary task, we added another task to bias correct hourly *P* categories. Studies have indicated that a multitask~~multitask~~ DL model could learn to better represent the shared physical processes and better predict the variable ~~that we are interested in~~of interest (e.g., Sadler et al., 2022). Since P categories and actual values are highly relevant, ~~Since *P* categories are generated based on different ranges of *P* values, it is~~adding ~~a expected that the~~ classification task can potentially improve the DL model ~~performance on~~for bias correcting and downscaling *P*.~~Since these two tasks are highly relevant to each other, it is expected that the classification task can improve the model performance on bias correcting and downscaling *P*.~~

Specifically, for the SRDRN with multitask~~multitask~~ learning, one convolutional layer (256 filters and 3x3 kernels) follows the last element-wise addition operation to summarize feature maps, then the architecture splits into two sections (Figure 1). The first section with two additional convolutional layers (the first one with 64 filters and the second with 4 filters) followed by the Softmax activation (Goodfellow et al., 2016) is ~~is~~ used for bias correcting *P* categories as a multiclass classification task, and the other section with upsampling blocks is used for the purpose of bias correcting and downscaling

Formatted: Font: Italic

hourly P. The classification task classifies the hourly P at each grid into four categories: 0-0.1mm/h as no rain, 0.1-2.5mm/h as light rain, 2.5-10mm/h as moderate rain, and >10mm/h as heavy rain (Tao et al., 2016). Due to radar sensors'~~sensors'~~ uncertainty in the very light rainfall, 0.1 mm/h is commonly used as a threshold to determine if there is rain (Tao et al., 2016). Since the classification task is executed on the feature maps at the coarse resolution, we aggregated Stage IV P (namely, coarsened Stage IV in this study) into the same spatial resolution as MERRA2 and classified the upscaled P data into the four groups as target labels.

[Insert Figure 1]

~~2.2.4~~3.1.3 Customized loss functions

Precipitation data is highly skewed and unbalanced, especially at an hourly time scale, which could cause the deep learning algorithm to focus more on no-rain~~no rain~~ events ~~and~~while ignoring~~ignore~~ heavy rain events ~~with~~if using regular loss functions. Wang et al. (2021); (Nelson et al., 2016; Ravuri et al., 2021)Here ~~We~~ we developed a weighted mean absolute error (MAE) loss function ($L_{MAE\_weighted}$) to balance precipitation data where weights change with precipitation values as shown below,

Formatted: Font: Not Bold

$$L_{MAE\_weighted} = \frac{\sum_{i=1}^{n} w_1 \cdot |y_{pred} - y_{true}|}{n} \quad (1)$$

where n is the total number of grids in a batch, $w_1$is the weight for each absolute error between the model predicted value y_{pred} and the true value y_{true}. The weight $w_1$changes with the actual true value y_{true},

$$w_1 = \begin{cases} MIN & y_{true} \le MIN \\ y_{true} & MIN < y_{true} < MAX \\ MAX & y_{true} \ge MAX \end{cases}$$

where MIN is the lowest threshold and MAX is the highest threshold for the weights. In other words, when the y_{true} value is below (above) MIN (MAX), $w_1$equals MIN (MAX), otherwise $w_1$equals y_{true} itself. Thus, the loss is weighted directly by the P value at the grid cell scale, which has been proven more effective than weighted by P bins (Ravuri et al., 2021; Shi et al., 2017). Note that all of the gridded P data, including Stage IV and MERRA-2, are logarithmically transformed [i.e., $y=log(x+1)$] in order to amplify the normality and reduce the skewness of P data (Sha et al., 2020a). In Equation 1, y_{true} and y_{pred} are transformed P values. MIN was set to $log(0.1+1)$ and MAX was set to $log(100+1)$, where maximum 100mm/h was chosen as the highest threshold before log transformation for robustness to spuriously large values in the Stage IV radar (Ravuri et al., 2021) and 0.1 mm/h is commonly used as a threshold to determine if there is rain for radar data (Tao et al., 2016).

For the four P categories, most data fall into the no rain category (over 88% in the coarsened Stage IV), and minority data fall into the heavy rain category (about 0.2% in the coarsened Stage IV). Thus, handling class imbalance~~class-imbalance~~ is of great importance in this situation, where the minority class for the heavy rain category is the class of most interest with respect to this learning task. The regular cross-entropy~~cross entropy~~ loss function for the classification task could result in the underestimation of the minority class (Fernando and Tsokos, 2021). Thus, we applied a weighted cross entropy as a loss function ($L_{weighted\ Cross-entropy}$) for the classification task in order to penalize more towards heavy rain category as follows,

$$L_{weighted\ Cross-entropy} = -\sum_{i=1}^{n}\sum_{j=1}^{k} w_{2,j} \cdot p\left(y_{i,j}\right) \cdot \log\left(q\left(y_{i,j}\right)\right) \tag{2}$$

where $w_{2,j}$ denotes the weight for the jth class, $p\left(y_{i,j}\right)$ represents the true distribution of the ith grid for the jth class, and $q\left(y_{i,j}\right)$ represents the predicted distribution. k is the number of classes (equals ~~to~~ 4 in this study). $w_{2,j}$ was set to 1, 5, 15, and 80 for no rain, light rain, moderate rain, and heavy rain classes, respectively, which is roughly based on the opposite percentage (i.e., 1, 5, 15, 80 are approximately from the percentages of heavy, moderate, light and no rain categories, respectively) for each category of the coarsened Stage IV. Since the weights for categories with rain are relatively larger than the no rain category, the loss $L_{weighted\ Cross-entropy}$ is relatively large when there are discrepancies between true and predicted categories with rain, resulting in guiding the training process towards ~~to~~ decreasing these differences with larger weights and thus better handling class-imbalance issues~~issue~~.

~~2.2.5~~3.1.4 Experiment Design

To comprehensively evaluate the added value of each component of customized DL models, including weighted loss function, multitask~~multitask~~ learning, and adding covariates, we designed six scenarios (Scenario1 to Scenario6 in Table 1). Scenario1 is based on the basic SRDRN architecture with hourly P from MERRA2 as coarse-resolution input, P from Stage IV as high-resolution labelled~~labelled~~ data, and regular MAE as loss function, which represents regular DL. Wang et al. (2021) used regular mean squared error (MSE) as a loss function, which works well for downscaling daily precipitation through synthetic experiments with no bias~~,~~ since the precipitation data was first coarsened and then downscaled into the original fine scale. However, in this study, the coarse resolution MERRA2 has substantial~~substantially~~ biases compared to Stage IV radar data, and Stage IV radar data also includes artefacts (e.g., large spurious~~spurious large~~ values) (Nelson et al., 2016). The previous~~Previous~~ study ~~have~~has shown that the MSE loss function is more sensitive to radar artefacts than the mean absolute error (MAE) loss function (Ravuri et al., 2021). Therefore, ~~and thus~~ we chose MAE as a regular loss function in this study. Scenario2 is the same as Scenario 1 except using weighted MAE loss function [Eqn. (1)]. The number of trainable parameters is the same for Scenario1 and Scenario2. Scenario3 includes the classification task, and the total loss is the combination of Eqn. (1) and Eqn. (2) with a weight λ [see Eqn. (3) below], where λ was set to 0.01 to ensure the two parts of the losses are in the same magnitude. The trainable parameters for Scenario3 increase~~increases~~ by 30% compared to Scenario1 and Scenario2.

$$L = L_{MAE\_weighted} + \lambda \cdot L_{weighted\ Cross-entropy} \qquad (3)$$

[Insert Table 1]

As described in Section 1, studies ~~(e.g., Baño-Medina et al., 2020; Li et al., 2022; Rasp and Lerch, 2018)~~have indicated that including atmospheric covariates is helpful for estimating precipitation (e.g., Baño-Medina et al., 2020; Li et al., 2022; Rasp and Lerch, 2018). The other three scenarios also consider atmospheric covariates of *P* from MERRA2 as predictors, which include geopotential height, specific humidity, air temperature, eastward wind, and northward wind at three different vertical levels (250, 500, 850 hPa) (e.g., Baño-Medina et al., 2020; Rasp and Lerch, 2018) as well as vertical wind (e.g., Trinh et al., 2021) at 500 hPa (OMEGA500), sea level pressure and 2-meter~~2 meter~~ air temperature in a single level (e.g., Panda et al., 2022; Rasp and Lerch, 2018) (see Table 2). We chose these variables based on precipitation formation theory (cloud mass movements and thermodynamics) ~~and other~~as well as findings from previous studies~~, on estimating precipitation as listed~~indicated above. ~~Similar~~Comparable to a classic multiple linear regression problem, covariates are multivariable predictors, and hourly precipitation is the only dependent variable. For each covariate listed in Table 2, data normalization was executed as a data preprocessing step. Specifically, each covariate was normalized by subtracting the mean (μ) and dividing by the standard deviation (σ). Here μ 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 made with the normalized testing dataset from MERRA2 with μ and σ calculated from the statistics of the coarse-resolution data during the testing period to preserve nonstationary. Scenario4 only included atmospheric covariates without using coarse resolution *P* as input and used Eqn. (1) as loss function to test whether only covariates are sufficient for estimating hourly *P*. The number of trainable parameters for Scenario4 is about 1% more compared to Scenario1 and Scenario2. Scenario5 is the same as Scenario4 except including *P* as a predictor besides atmospheric covariates, and the number of trainable parameters is very close to Scenario4. Scenario6 is the same as Scenario5 except including the classification task with Eqn. (3) as loss function and the number of trainable parameters is similar to Scenario3 (31% greater than scenarios with no multitask~~multitask~~ learning).

Field Code Changed

Field Code Changed

[Insert Table 2]

The Adam optimization algorithm was applied to train the six DL scenarios with a learning rate of 0.0001 and other default values. We found that the learning rate of 0.0001 worked stably in this study through a series of experiments. The batch size for each epoch was set to 64, and the number of epochs was set to 150 for each scenario listed in Table 1. Each scenario was trained ~~for~~with approximately 2.5x10^5 iterations. We frequently saved models and evaluated their performance with a validation dataset in order to choose the best model for prediction on the testing dataset. The training process was executed using NVIDIA V100 GPU provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at the Goddard Space Flight Center (https://www.nccs.nasa.gov/systems/ADAPT/Prism).

Formatted: Superscript

At the time when we conduct this study, MERRA-2 and Stage IV hourly *P* data have a 20-year overlapping period from 2002 to 2021. We used the first 14 years (2002 to 2015) as the training dataset, the middle 3 years (2016 to 2018) as the validation dataset, and the more recent 3 years (2019 to 2021) as the testing dataset. Figure 2 shows the hourly mean or climatology for MERRA-2 and Stage IV for training and testing datasets, as well as the mean differences between the testing and the training periods. We can tell that there are large climatology differences (or biases) between MERRA-2 and Stage IV both for training and testing datasets, especially around the coastal area. Wetter conditions are observed in most of the study area in the testing period (average 0.03 mm/h) than in the training period, which is due to a higher percentage of rains (with values greater than 0.5mm/h) during the testing period than during the training period based on analyzing the Stage IV data (Table S1 in Supplement).~~Wetter conditions are observed in most of the study area in the testing period (average 0.03 mm/h) caused by a higher percentage of rains greater than 0.5mm/h (see Table S1 in Supplement) than the training period based on the Stage IV data~~~~Wetter conditions are observed in most of the study area in the test period (average 0.03 mm/h) comparing with the training period.~~. This allows us to assess the extrapolation capabilities of the different methods, which is particularly relevant in a changing climate.

[Insert Figure 2]

~~2.3~~3.2 Statistical approach

We used a widely accepted quantile delta mapping (QDM) as a benchmark approach for *P* bias correction. The QDM method corrects systematic biases at each grid cell in quantiles of a modelled series with respect to observed values. Compared to the regular quantile mapping method (Panofsky and Brier, 1968; Thrasher et al., 2012; Wood et al., 2002), QDM also ~~accounts~~ applies a relative difference ~~for the difference~~ between historical and future climate data (here, training and testing periods). Thus,~~, and thus~~ it is capable of preserving the trend of the future climate (Cannon et al., 2015), which is critical for this study since there are substantial differences between the precipitation during the training (2002 to 2015) and testing (2019 to 2021) periods (see Figure 2). This approach has been widely used to bias-correct~~bias correct~~ climate variables, including *P*, which indicated better performance compared to the other bias correction approaches (Cannon et al., 2015; Eden et al., 2012; Kim et al., 2021; Tegegne and Melesse, 2021; Tong et al., 2021). To be specific for QDM, the bias-corrected~~bias corrected~~ value $\hat{x}_{m,p}(t)$ for modeled data in the future projection at time t is given by applying the relative change $\Delta_m(t)$ multiplicatively to the historical bias corrected value $\hat{x}_{o:m,h:p}(t)$,

$$\hat{x}_{m,p}(t) = \hat{x}_{o:m,h:p}(t) \cdot \Delta_m(t) \tag{4}$$

where $\hat{x}_{o:m,h:p}(t) = F_{o,h}^{-1}[\tau_{m,p}(t)]$ and $\Delta_m(t) = \frac{x_{m,p}(t)}{F_{m,h}^{-1}[\tau_{m,p}(t)]}$. $x_{m,p}(t)$ represents uncorrected modeled data in the projection period and $\tau_{m,p}(t)$ is the percentile of $x_{m,p}(t)$ in the empirical cumulative density function (*F*) formulated by the modeled data in the projection period over a time window around t. $F_{o,h}^{-1}[\tau_{m,p}(t)]$ means applying inverse empirical cumulative density

Formatted: Font: Bold

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25" function formulated by the observed data in the historical period for $\tau_{m,p}(t)$ to obtain bias-corrected~~bias corrected~~ value [i.e., $x_{o:m,h:p}(t)$]. Similarly, $F_{m,h}^{-1}[\tau_{m,p}(t)]$ denotes applying inverse empirical cumulative density function formulated by the modeled data in the historical period for $\tau_{m,p}(t)$. The time window to construct the empirical cumulative density function around time t was set to be 45 days to preserve the seasonal cycle. In this study, the historical and projection periods correspond to the training and testing data periods, respectively. The modeled and observed data correspond to MERRA2 and coarsened Stage IV data, respectively. Details about this method are referred to Cannon et al. (2015).

The QDM bias correction was performed at the spatial resolution of MERRA2. The QDM-biased~~QDM biased~~ corrected P data at the coarse resolution was then bilinear interpolated into the high resolution, the same as the spatial resolution of Stage IV. This process of QDM and bilinear interpolation (He et al., 2016b) is named ~~as~~ QDM_BI in the following sections.

~~2.4~~3.3 Evaluation approaches

We evaluated model performance in different temporal scales, including hourly and aggregated (daily and monthly) time scales. The agreements between the observed and estimated (i.e., bias-corrected~~bias corrected~~ and downscaled) P for the different scales and extremes were quantified using the Kling-Gupta efficiency (KGE). The KGE is an objective performance metric combining correlation, bias, and variability, which was introduced by~~in~~ Gupta et al. (2009) and modified by~~in~~ Kling et al. (2012). KGE has been widely used for evaluating different datasets with observations (e.g., Beck et al., 2019b; Beck et al., 2019a; Wang et al., 2021) and as the standard evaluation metric in hydrology (Beck et al., 2017; Harrigan et al., 2018; Harrigan et al., 2020; Lin et al., 2019). The KGE is defined as follows:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad (5)$$

where the correlation component r is represented by correlation coefficient, the bias component β represented by the ratio of estimated and observed means, and the variability component γ represented by the estimated and observed coefficients of variation:

$$\beta = \frac{\mu_s}{\mu_o} \quad \text{and} \quad \gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \quad (6)$$

where μ_s and μ_o denote the distribution mean for the estimates and observations, and σ_s and σ_o denote the standard deviation for the estimates and observations, respectively. Note here that the variability component γ is not the ratio of σ_s and σ_o to ensure that the bias and variability ratios are not cross-correlated (Kling et al., 2012). KGE, r, β and γ represent perfect agreement when they equal one. In addition to KGE, the root mean square error (RMSE) and mean absolute error (MAE) metrics are also reported since they were often used to evaluate model performance on bias correction and downscaling (e.g., Maraun et al., 2015; Rodrigues et al., 2018).

Formatted: Font: Bold

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: 10 pt, Not Italic

Formatted: Font: 10 pt, Not Italic

To understand the performance on capturing *P* extremes, we assessed hourly *P* at 99th percentile and annual maximum wet spell in hours, as well as an extreme hurricane event that occurred during the testing period. These extreme indices and events are highly relevant to flooding (Pierce et al., 2014) and have a great environmental impact as well as impacts on property and human life.

Moreover, we evaluated *P* classification results from Scenario3 and Scenario6, the scenarios with multitask~~multitask~~ learning for bias correcting *P* categories, by comparing them with the four categories from the coarsened Stage IV observations. The four categories from the coarsened Stage IV were generated manually based on the ranges of the four classes. We also classified the results from QDM and raw MERRA2 into ~~the~~ four categories and compared the results with the categories from the coarsened Stage IV. A widely used metric, namely, Intersection ~~over~~ Over Union (~~IoU~~IOU) (Li et al., 2021), is applied to evaluate classification performance, which is defined by:

$$\mathrm{I\cancel{o}OU} = \frac{TP}{TP+FP+FN} \cdot 100 \qquad (7)$$

where *TP* represents true positives (prediction=1, truth=1), *FP* represents false positives (prediction=1, truth=0) and *FN* represents false negatives (prediction=0, truth=1). Taking the heavy rain category as an example, *TP* is an outcome where the model correctly predicts the heavy rain class; *FP* is an outcome where the model predicts it is a heavy rain class, but the true label is not a heavy rain class; *FN* is an outcome where the model predicts it is not a heavy rain category, but the true label is a heavy rain class. ~~where *TP* represents true positives, *FP* represents false positives and *FN* represents false negatives.~~ I~~o~~OU ranges from 0 to 100 and specifies the percentage of the amount of overlap between the predicted and ground truth bounding box.

~~3~~4 Results

In this section, we present the performance of the six DL model scenarios and the benchmark approach QDM_BI on bias correcting and downscaling hourly *P*, evaluated against Stage IV precipitation data during the testing period from 2019 to 2021.

~~3.2~~ 4.1 Overall agreement

The overall agreement between the observed and estimated *P* was quantified with KGE [Eq. (5)] as well as each component of KGE, which were calculated on an hourly basis for the entire testing period (2019 to 2021) and for all the grid cells over the study region. Table 3 shows that Scenario2 to Scenario6 have much higher KGE than Scenario1, indicating that the weighted loss function improved model performance through rebalancing hourly *P* data. Scenario1, however, highly overestimated the variability (i.e., γ is much greater than 1) and underestimated the mean (i.e., β is much smaller than 1), resulting in a negative KGE value. This indicates that using a regular loss function (i.e., MAE) tends to underestimate hourly

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Bold

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 3 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Normal, No bullets or numbering

P (relatively larger training loss than other scenarios during training, see Figure S1 in the Supplement). The KGE values are comparable for all the scenarios using the weighted loss function. The best KGE is obtained by Scenario5, with Scenario4 and Scenario6 performing consistently well in terms of KGE, which indicates that including atmospheric covariates as predictors further improved the model performance. However, the DL and benchmark approaches performed considerably worse in terms of the correlation component r of KGE than the other components (i.e., β and γ). The reason is that~~because~~ the correlation component r~~,~~ was estimated based on all the hour-to-hour~~hour to hour~~ *P* data, while the other two components (i.e., β and γ) ~~was~~were calculated based on ~~suggesting that~~ long-term climatological *P* statistics and were relatively easier to ~~be~~ estimate~~d~~ (Beck et al., 2019b)~~ are relatively easier to capture than hour to hour *P* dynamics (i.e., r)~~. The benchmark, QDM_BI, also highly overestimated the variability~~,~~ and has a lower KGE score than Scenario4, Scenario5, and Scenario6 of the DL approaches.

Formatted: Font: Italic

[Insert Table 3]

Table 3 also reports the results of RMSE and MAE, which are widely used to evaluate model performance on bias correction and downscaling. However, these two metrics are inadequate for pixel-wise comparison, particularly when comparing two datasets with spatial biases, due to the well-known "~~“~~double penalty problem"~~”~~ (Harris et al., 2022; Rossa et al., 2008). Specifically, for using RMSE or MAE metrics, the model estimates that correctly capture the right amounts of rain in slightly incorrect locations often score worse than estimates of no rain at all. For example, Scenario1 has the lowest RMSE and MAE, but it highly underestimated the ~~average~~ observed mean (i.e., β is much lower than 1), while it is the worst one in all the scenarios, including QDM_BI in terms of KGE scores. This illustrates the limitations of ~~the~~ grid point-based errors~~error~~ like RMSE and MAE as evaluation metrics.

~~3.3~~4.2 Hourly Climatology

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Due to climate variability and change, the climatology of hourly *P* over the testing period (2019 to 2021) is much higher than the training period (2002 to 2015) (Figure 2). We evaluated the long-term mean (i.e., climatology) during the testing period (Figure 3 and Figure 4a), which allows us to examine how well the methods could capture the *P* climatology but also the nonstationary changes of long-term *P*. Again, Scenario1 notably underestimated the climatology of observations (by 71% on average) (Figure 3 and Figure 4a)~~,~~ due to the use of MAE as a loss function. In general, all other DL scenarios and QDM_BI provide satisfactory results in~~on~~ capturing hourly *P* climatology. Scenario4 also slightly underestimated the climatology of Stage IV (12% on average, Figure 4a). This scenario only includes atmospheric covariates as model inputs without using the corrected *P* of MERRA-2, indicating the information from covariates only is~~are~~ not sufficient to estimate hourly *P*. The climatology of Scenario3, Scenario5, and Scenario6 appears~~appear~~ well matching with Stage IV in space, better than QDM_BI. Relative differences of climatology averaged over the study area between estimated and Stage IV are 1.5%, 1.8% and 0.38% for Scenario3, Scenario5, and Scenario6, respectively, while it is 2.5% for QDM_BI. Compared to Scenario3 and Scenario5,

~~the~~ Scenario2 underestimated the climatology, particularly around the coastal area (Figure 3)~~, which indicates the added value from multitask learning (Scenario3) and atmospheric covariates (Scenario5)~~. Figure 4a shows that QDM_BI has a relative larger variance and its KGE value is lower than the ones for Scenario 2, Scenario3, Scenario5, and Scenario6. Note that all the DL estimates appear~~appears~~ to be blurrier than Stage IV, similar to~~as~~ what has been found in previous studies (e.g., Ravuri et al., 2021), while the QDM_BI estimates are even blurrier than the DL estimates.

[Insert Figure 3]

[Insert Figure 4]

~~3.4~~4.3 Daily and Monthly *P* estimates

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

We aggregated the hourly *P* estimates into daily and monthly time scales to evaluate the performance of daily total *P* and monthly mean of hourly *P*. Overall, the KGE values for the daily total *P* are considerably greater than those for the hourly *P* (Table 3), which suggests temporal aggregation denoised the hourly precipitation data, leading to considerably higher correlation coefficient (r in Table 3), mainly contributing to higher KGE. Similarly, The KGE value for Scenario1 is the lowest since it highly underestimated the mean of daily total *P* (lower β), overestimated the variability (higher γ), and the correlation r is also lower compared to the other scenarios. The Scenario5 and Scenario6 have relatively~~relative~~ higher KGE scores than other DL scenarios and QDM_BI for daily total *P*. Daily total *P* from QDM_BI has a comparable KGE score with the DL models~~,~~ while overestimating~~overestimated~~ the variability (higher γ) compared to most of the DL scenarios.

Figure 5 shows the daily total *P* time series for each year during the testing period for the Stage IV, six DL scenarios, and QDM_BI~~,~~ averaged over the study area. The results show that the daily total *P* time series from the DL models closely matched with the daily total *P* time series from Stage IV except Scenario1. Again, Scenario1 highly underestimated the daily total *P* with the lowest KGE value, suggesting the difficulties of MAE in handling the highly unbalance feature of *P*. The daily total *P* from all the other five DL scenarios is~~are~~ much close to Stage IV with large~~r~~ KGE values (close to or larger than 0.9) ~~than QDM_BI~~. ~~For these five DL scenarios (Scenario2 to Scenario6),~~ Scenario5 and Scenario6 perform better than the other scenarios~~s~~ including QDM_BI, indicating incorporating covariates and corrected coarse resolution *P* ~~and/or multitask learning~~ further improved daily total *P* estimates. The bias-corrected~~bias corrected~~ and downscaled daily total *P* from QDM_BI, however, highly overestimated the daily total *P* of Stage IV for almost all the large precipitation events~~,~~ because the bias correction process for QDM_BI was executed individually at each grid cell and did not consider spatial dependencies and nonlinear relationships between covariates and observations, resulting in nonstable estimations (Wang and Tian, 2022).

[Insert Figure 5]

Table 3 also summarizes~~summarized~~ the statistics of the monthly mean of hourly *P*. The KGE values for the monthly mean of hourly *P* are greatly increased, higher than the daily total *P*. Except Scenario1, the KGE values for the monthly mean are very close to each other, with Scenario4 slightly lower than others including QDM_BI. The monthly~~Monthly~~ mean from QDM_BI had ~~a~~ relatively higher γ~~□~~, indicating overestimations of variability. Figure 6 presents the monthly mean time series of hourly precipitation for each month during the testing period for Stage IV, the six DL models,~~model~~ and QDM_BI, averaged over the study area. Similar to the daily total *P* time series, the monthly mean *P* from all the DL models closely matched with the monthly mean time series from Stage IV (KGE value greater than 0.9) except Scenario1, which highly underestimated the observations~~.~~. Scenario4 has the lowest KGE value and slightly underestimated the monthly mean, but all the scenarios (Scenario2 to Scenario6) ~~The Scenario6 has the highest KGE score, followed by Scenario3, Scenario5, Scenario2 and Scenario4, which~~ are consistently better than the KGE score from QDM_BI. These results indicate that incorporating the weighted loss function (Scenario2 to Scenario6 compared to Scenario1)~~), multitask learning (Scenario3 and Scenario6) and atmospheric covariates (Scenario4 to Scenario6)~~ improved monthly mean estimation, and the effects of the other customized components are not obvious at the monthly time scale. Similarly, the monthly mean from QDM_BI estimates has~~have~~ a relatively larger variability than Stage IV, resulting in a lower KGE value.

[Insert Figure 6]

~~3.5~~4.4 Extremes

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Table 4 summarizes~~summarized~~ the statistics of hourly *P* at 99th percentile and the annual maximum wet spell. The results show that Scenario1 highly underestimated hourly *P* at 99th percentile (lower β than 1) and overestimated variability (higher γ than 1), resulting in a negative KGE score, suggesting the inadequacy of using regular MAE loss function. Scenario2 has the highest KGE score with a higher correlation coefficient (higher *r*) than the other scenarios. This is probably because the number of trainable parameters for Scenario2 is the lowest, leading to a better regularization ability with limited data for extremes. The KGE values are similar for Scenario3, Scenario5, and Scenario6, and relatively ~~higher~~ lower ~~for~~ for Scenario4, suggesting the importance of incorporating observation-corrected~~observation corrected~~ *P* from coarse resolution as an input. The benchmark approach QDM_BI highly overestimated the variability of hourly *P* at 99th percentile compared to Stage IV, resulting in a lower KGE value~~a lower KGE values~~ than most of the DL scenarios except Scenario1.

Figure 4b shows the boxplots of the relative difference between hourly *P* estimates and Stage IV observations at the 99th percentile. On average, Scenario1 underestimated the 99th percentile hourly *P* by over 60%, while other DL scenarios underestimated by about 20%, with Scenario5 and Scenerio6 much closer to Stage IV. The 99th percentile estimated by QDM_BI has a much higher variance (as indicated by the distance between high 90% and low 10% bars in the boxplot, as well as high γ in Table 4) compared to DL models, while has a lower mean difference (underestimated by about 10%) due to bias correction through an explicit adjustment at each percentile. Figure 7 shows the spatial distribution of the hourly *P* at the 99th percentile for MERRA2, Stage IV, QDM_BI, and six DL models. We can see that the 99th percentile of MERRA-2 hourly *P* greatly underestimated Stage IV by 40% (spatial average 2.9mm for MERRA2 versus 4.8mm for Stage IV). While the hourly

P at the 99^{th} percentile from QDM_BI (area average 4.3mm) appears to be close to Stage IV, its spatial variability looks very different from Stage IV, probably due to QDM_BI correcting biases on a grid point basis. ~~The spatial average *P* at 99^{th} percentile for the six~~Scenario4 highly underestimated *P* values at the 99^{th} percentile compared with other scenarios except Scenario1, indicating excluding coarse resolution *P* as an input is not reasonable~~ deep learning models is 1.7mm, 3.9mm, 4.0mm, 3.7mm, 4.2mm and 4.1mm for Scenario1 to Scenario6, respectively, indicating that increasing model complexity decreased hourly *P* mean biases (i.e., β in Table 4) at 99^{th} percentile~~.

Formatted: Font: Italic

Formatted: Superscript

Formatted: Font: Italic

[Insert Figure 7]

The DL models treated each hourly *P* spatial data as a 2D image and did not explicitly account for temporal dependence between images. We assumed that the DL models could potentially preserve the temporal dependence of observations if the DL models were well bias-corrected~~bias corrected~~ and downscaled each 2D image~~The DL models treated hourly spatial *P* data independently and did not explicitly account for temporal dependence. However, the DL models could potentially well reduce temporal biases if spatial *P* data for each hour can be well corrected and downscaled~~. The annual maximum wet spell is a widely used extreme index for~~on~~ evaluating temporal dependence (e.g., Maraun et al., 2015). The wetness threshold for calculating the annual maximum wet spell index was set to 0.1mm/h, which is commonly used for hourly radar~~radar hourly~~ data (e.g., Tao et al., 2016). Table 4 shows that Scenario2 and Scenario3 have relatively~~relative~~ higher KGE scores for the annual maximum wet spell extreme index than the other DL scenarios, suggesting the usefulness of more parsimonious models with weighted loss function but without including atmospheric covariates as additional inputs. Further incorporating multitask~~multitask~~ learning (Scenario3 and Scenario6), however, slightly decreased the model performance compared to no multitask~~multitask~~ learning scenarios (Scenario2 and Scenario5), probably due to the increased parameters and decreased regularization ability. While scenario1 has the lowest KGE score than the other DL scenarios, it is still much higher than QDM_BI, which highly overestimated the mean of annual maximum wet spell for Stage IV observations (much higher β than 1). Boxplots in Figure 4c show the difference between model estimates and Stage IV observations for the annual maximum wet spell in hours during the testing period. Scenario1 highly underestimated the annual maximum wet spell by about 10 hours. Scenario2 and Scenario3 have the lowest differences with Stage IV in terms of the mean and variance of the annual maximum wet spells. On average, Scenario4, Scenario 5, and Scenario6 overestimated the annual maximum wet spell by about 10 hours, with Scenario4 and Scenario6 showing a relatively~~relative~~ larger variance. The benchmark approach QDM_BI has the largest difference (on average over 22 hours) and much larger variance compared to Stage IV, resulting in a negative KGE score. This is probably because QDM_BI corrected biases on a grid basis, which failed to account for the spatial and temporal dependence.

Formatted: Font: Italic

Figure 8 shows an extreme event occurred from 19:00 to 20:00 on 29 August 2021 in Universal Time Coordinated (UTC) time zone when Hurricane Ida landed at the Louisiana State in the United States from MERRA2, Stage IV, QDM_BI and the six DL scenarios. We can see that MERRA2 highly underestimated this extreme event and did not capture detailed features of

Stage IV. While QDM_BI estimates slightly enhanced the hourly *P* values, they~~it~~ still failed to capture detailed features. ~~The~~ Scenario1 to Scenario3 gradually enhanced hourly *P*, but these three models had difficulties capturing~~to capture~~ the center of the hurricane. By including atmospheric covariates, Scenario4 to Scenario6 roughly captured the center of the hurricane, and Scenario6 also reproduced the cyclones surrounding the center. These results suggest ~~the importance of incorporating weighted loss function, multitask learning, and atmospheric covariates~~the customized components improve the model performance on ~~for~~ bias correcting and downscaling specific extreme events.

[Insert Figure 8]

~~3.6~~4.5 *P* categories

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Figure 9 shows that Scenario3 and Scenario6, the scenarios with multitask~~multitask~~ learning for bias correcting *P* categories, have larger ~~IoU~~ IOU values (e.g., 19.63% for Scenario3 and 19.91% for Scenario6 for moderate rain 2.5-10mm) than QDM method (but 15.30% for moderate rain) particularly for the three categories with rain, indicating that the two DL models results ~~well~~ better matched with the wet categories of the coarsened Stage IV observations~~, better~~ than the QDM method. Furthermore, Scenario6 has relatively~~relative~~ larger ~~IoU~~ IOU scores than Scenario3, indicating incorporating atmospheric covariates improved classification accuracy. For example, there is 8.15% of the heavy rain category matched the coarsened Stage IV observations for Scenario3, while, for Scenario6, 11.07% of the heavy rain category matched the coarsened Stage IV observations. These results suggest that, ~~with an~~the auxiliary classification task, incorporated in the Scenario3 and Scenario6 of the DL model can ~~well~~ better estimate the four categories of hourly *P* during the testing period than the traditional bias correction method QDM.

[Insert Figure 9]

~~4~~5 Discussion

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 3 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

This study explored customized DL for bias correcting and downscaling hourly *P* through a set of experiments with or without customized loss functions, multitask~~multitask~~ learning, and inputs from atmospheric covariates of precipitation. Scenario1, which used regular MAE as a loss function, highly underestimated *P* for all the temporal scales as well as extremes, showing the lowest performance. Since most ~~of~~ hourly *P* are no rain, the regular loss function very likely leads the model to learn no rain events while neglecting rainy events. Regular MAE has been used for downscaling daily precipitation data with limited biases in previous studies (e.g., Sha et al., 2020a), but to our knowledge, there are no successful cases using regular MAE for downscaling hourly precipitation data with large biases. However, the scenarios with customized loss functions~~function~~ with weighted MAE (Scenario2 to Scenario6) consistently showed much better performance than Scenario1. This result suggests that penalizing more towards heavy *P* on a grid basis makes the optimization algorithm focus more on the grids where rainfalls occurred and, therefore, inherently rebalance the hourly *P* for model training. While this study explored bias correcting and downscaling hourly precipitation from climate reanalysis data, this algorithm with customized loss function can be potentially integrated with precipitation data from the Global Precipitation Measurement (GPM) mission to generate more accurate operational precipitation data at a finer resolution.

(e.g., Sha et al., 2020a)

The scenarios with multitask~~multitask~~ learning indicate~~indicates~~ limited added values ~~perform generally better than the other scenarios in terms of hourly climatology (Figure 4a), and daily and monthly assessments (Figure 5 and Figure 6). Multitask learning model with covariates can enhance extreme events and is the best model for application of bias correcting and downscaling *P* extreme events. Their added values, however, are limited~~ and performed worse than other scenarios without multitask~~multitask~~ learning ~~(Scenario2 and Scenario5)~~ in terms of extreme indices ~~(see Figure 4b, 4c and Table 4)~~. The reason for that is probably because adding multitask~~multitask~~ learning increased 30% trainable parameters with limited extreme data decreased the model regularization ability. Baño-Medina et al. (2020) designed a series of DL models with plain CNN architecture and different model complexity (i.e., increasing the number of trainable model~~model trainable~~ parameters) to downscale daily ERA5 reanalysis dataset and found that increasing model complexity makes~~make~~ model performance worse, particularly for extreme indices (98th percentile and annual maximum wet spell), which is consistent with our study.

Traditional methods (e.g., QDM_BI) mainly use coarse resolution *P* data as the only predictor for downscaling and bias correction, which cannot fully utilize nonlinear relationships between covariates and observations (Rasp and Lerch, 2018) during the bias correction and downscaling process. DL models with covariates as auxiliary variables, however, have indicated success in~~on~~ improving model performance for postprocessing~~postprocessing~~ temperature and precipitation forecasts due to the capability of learning nonlinear relationships between covariates and response variable automatically (Li et al., 2022; Rasp and Lerch, 2018). Scenario4 to Scenario6 incorporated physically relevant covariates of precipitation, with only Scenario4 excluding the coarse resolution *P* as Baño-Medina et al. (2020) did for downscaling daily precipitation. The results indicate that incorporating auxiliary predictors of atmosphere circulations and moisture conditions can help improve *P* bias correcting and downscaling skills~~skill~~ (see Figure 3 to Figure 8). However, only using covariates without coarse resolution *P* (Scenario4) is not sufficient to well estimate hourly *P*, while using coarse resolution *P* as additional input (Scenario5 and Scenario6) shows improved performance. This result is consistent with a recent study focusing on CNN-based postprocessing~~postprocessing~~ of *P* forecasts from numerical weather prediction models, showing total precipitation itself is the most important predictor (Li et al., 2022). Note that we did not explore the importance of rank among these covariates in improving the model performance in this study, which could be a potential avenue for future work. Furthermore, static variables, such as elevations, long-term~~long term~~ climatology (Sha et al., 2020a), soil texture, and land cover, could be helpful for resolving local details. However, our study region has little topographic variations, and therefore including elevation data cannot add any additional information to the model.

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Moreover, we compared the customized DL scenarios with a traditional method QDM_BI and found that most of the ~~all~~ DL experiments remarkably outperform QDM_BI in all the temporal scales as well as extremes. QDM_BI executed bias correction at each grid point without considering spatial dependencies and only used coarse resolution *P* as a predictor, and thus does not have the capability of capturing spatial features (e.g., detailed spatial features for the hurricane Ida in Figure 9) and accounting for the atmosphere and moisture covariates of precipitation. Furthermore, the proposed customized DL models are fully convolutional, and the trained models can potentially~~potentially can~~ be easily used to estimate hourly *P* in other places through transfer learning where high-resolution~~high resolution~~ data are not available [e.g., Stage IV radar coverage is limited in the western United States as a result of the scarcity of the radar network and blockage from the mountains (Nelson et al., 2016)]. There are many questions that need to be explored under this topic about transferability under various climate zones and the impact of spatial distance, which ~~The performance of transfer learning under various climate zones with different types of *P* events~~ deserves a separate study. The trained models also have the potential to generate high-resolution~~high resolution~~ hourly *P* estimates beyond the time range covered by Stage IV radars (e.g., before 2002). Furthermore, the SRDRN architecture can be further customized to downscale different gridded precipitation, including downscaling precipitation from GCM projections, which can be a future study.

Due to the stochastic nature of DL models, we ran each DL scenario for additional three times (four times in total) to evaluate the effects of stochasticity compared~~comparing~~ with the added value of each customized component of DL models (see Table S2 and Table S3 in the Supplement). The results show that KGE values for each scenario are significantly different at the p-value of 0.05 at the hourly time scale, which indicates that the added value of each customized component is not caused by model stochasticity. Scenario1 is significantly worse than other scenarios, including QDM_BI at hourly and aggregated time scales as well as extreme indices, emphasizing the added value of the weighted loss function. Scenario5 and Scenario 6 are significantly better than other scenarios, including QDM_BI, in terms of KGE values at hourly and aggregated time scales, and Scenario4 is significantly worse at the monthly time scale. For the 99th percentile extreme index, Scenario4 is significantly worse than Scenario3, Sceanrio5, and Scenario6. For the annual maximum wet spell index, Scenario2 and Scenario3 are significantly better than other scenarios. All these stochastic significance evaluation results are consistent with the findings in Section 4. Due to computational demand (20 to 22 hours for running each scenario once) and resource limits, we ran limited times for each scenario to consider the stochasticity of DL models, and incorporating DL models with Bayesian inference is a potential way to quantify systematic uncertainty caused by model itself as indicated by Vandal et al. (2018a).

~~5~~6 Conclusions

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 3 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Various gridded precipitation (*P*) data at different spatiotemporal scales have been developed to address the limitations of ground-based *P* observations. These gridded *P* data products, however, suffer from systematic biases and spatial resolutions are mostly too coarse to be used in local scale studies. Many studies based on DL approaches have been conducted to bias correct and downscale coarse resolution *P* data. However, it is still challenging for traditional approaches, as well as current DL approaches to capture small-scale~~small scale~~ features, especially for *P* extremes, due to the complexity of *P* data (e.g., highly unbalanced and skewed), particularly at fine temporal scale (e.g., hourly). To address these challenges, this study developed customized DL models by incorporating customized loss functions, multitask~~multitask~~ learning, and physically~~physical~~ relevant atmospheric covariates. We designed a set of model scenarios to evaluate the added values of each component of the customized DL models. Our results show that customized loss functions greatly improved model performance compared to the model scenario with regular loss function in all the temporal scales as well as extremes (on average, improved by over 70% for climatology and over 50% at the 99^{th} percentile). ~~The scenarios with multitask learning performed generally better than other scenarios on hourly climatology and aggregated time scales (daily and monthly), while the improvement is not as large as incorporating weighted loss function.~~ While multitask~~multitask~~ learning ~~greatly~~ improved model performance on capturing detailed features of extreme events (e.g., hurricane Ida), the scenarios with multitask~~multitask~~ learning performed worse than other scenarios in terms of extreme indices potentially due to the increased number of trainable parameters. The added value of incorporating atmospheric covariates is remarkable, likely because these scenarios took full advantage~~advantages~~ of nonlinear relationships between large-scale covariates, precipitation, and fine-scale observations. The results also indicated that the role of coarse resolution *P* as a predictor is very important for improving model performance despite the added values from the covariates. The DL scenarios with customized loss function and coarse resolution *P* as the only predictor are the best models at places where no covariate data are available. Moreover, ~~all the~~most of the DL scenarios with customized loss functions~~function~~ performed much better in all the temporal scales as well as extremes than the benchmark approach QDM_BI, which is not able to account for spatial dependence and nonlinear relationships. These results highlight the advantages of the customized DL model compared with regular DL models as well as traditional approaches, which provide~~provides~~ a promising tool to fundamentally improve precipitation bias correction and downscaling and better estimate *P* at high resolutions.

Code Availability Statement

The code of regular and customized SRDRN models is available at: https://osf.io/whefu/ (DOI: 10.17605/OSF.IO/WHEFU).

Data Availability Statement

The MERRA2 product is accessible through the Goddard Earth Sciences Data Information Services Center (GES DISC; http://disc.sci.gsfc.nasa.gov/mdisc/overview). The Stage IV radar precipitation data can be acquired via the National Center for Atmospheric Research (NCAR) data portal (https://data.eol.ucar.edu/dataset/21.093).

Author contributions

FW: Methodology, Conceptualization, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original draft preparation, Visualization. DT: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing- Original draft preparation, Supervision, Funding acquisition, Project administration. MC: Resources, Writing – Review & Editing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgments

This work is supported ~~by~~ in part by the NASA EPSCoR R3 program (No. NASA-AL-80NSSC21M0138), by the NSF CAREER program (No. NSF-EAR-2144293), and by the NOAA RESTORE program (No. NOAA-NA19NOS4510194).

References

~~Aadhar, S., & Mishra, V. (2017). High-resolution near real-time drought monitoring in South Asia. Scientific Data, 4(1), 1-14.~~

~~AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. Journal of Geophysical Research: Atmospheres, 116(D2).~~

~~AghaKouchak, A., Mehran, A., Norouzi, H., & Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets. Geophysical research letters, 39(9).~~ Aadhar, S. and Mishra, V.: High-resolution near real-time drought monitoring in South Asia, Scientific Data, 4, 1-14, 2017.

AghaKouchak, A., Mehran, A., Norouzi, H., and Behrangi, A.: Systematic and random error components in satellite precipitation data sets, Geophysical Research Letters, 39, 2012.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E.: Evaluation of satellite-retrieved extreme precipitation rates across the central United States, Journal of Geophysical Research: Atmospheres, 116, 2011.

Ashouri, H., Sorooshian, S., Hsu, K.-L., Bosilovich, M. G., Lee, J., Wehner, M. F., and Collow, A.: Evaluation of NASA's~~NASA's~~ MERRA precipitation product in reproducing the observed trend and distribution of extreme precipitation events in the United States, Journal of Hydrometeorology, 17, 693-711, 2016.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, Geoscientific Model Development, 13, 2109-2124, 2020.

Formatted: Justified, Indent: Left: 0", Hanging: 0.25", Line spacing: 1.5 lines

Formatted: Font: 10 pt, English (UK), Check spelling and grammar

Beck, H. E., Van Dijk, A. I., De Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrology and Earth System Sciences, 21, 2881-2903, 2017.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473-500, 2019a.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207-224, 2019b.

Bhattacharyya, S., Sreekesh, S., and King, A.: Characteristics of extreme rainfall in different gridded datasets over India during 1983–2015, Atmospheric Research, 267, 105930, 2022.

Bitew, M. M. and Gebremichael, M.: Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, Water Resources Research, 47, 2011.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?, Journal of Climate, 28, 6938-6959, 2015.

Cavalcante, R. B. L., da Silva Ferreira, D. B., Pontes, P. R. M., Tedeschi, R. G., da Costa, C. P. W., and de Souza, E. B.: Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia, Atmospheric Research, 238, 104879, 2020.

Chen, D., Mak, B., Leung, C.-C., and Sivadas, S.: Joint acoustic modeling of triphones and trigraphemes by multitask~~multi-task~~ learning deep neural networks for low-resource speech recognition, 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5592-5596,

Chen, Y.: Increasingly uneven intra-seasonal distribution of daily and hourly precipitation over Eastern China, Environmental Research Letters, 15, 104068, 2020.

Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., and Khadka, N.: Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya, Atmospheric Research, 250, 105365, 2021.

Daw, A., Karpatne, A., Watkins, W., Read, J., and Kumar, V.: Physics-guided neural networks (pgnn): An application in lake temperature modeling, arXiv preprint arXiv:1710.11431, 2017.

DeGaetano, A. T., Mooers, G., and Favata, T.: Temporal Changes in the Areal Coverage of Daily Extreme Precipitation in the Northeastern United States Using High-Resolution Gridded Data, Journal of Applied Meteorology and Climatology, 59, 551-565, 2020.

Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., and Vorogushyn, S.: Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, Hydrology and Earth System Sciences, 17, 2415-2434, 2013.

Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, correction, and downscaling of GCM-simulated precipitation, Journal of Climate, 25, 3970-3984, 2012.

Emmanouil, S., Langousis, A., Nikolopoulos, E. I., and Anagnostou, E. N.: An ERA-5 Derived CONUS-Wide High-Resolution Precipitation Dataset Based on a Refined Parametric Statistical Downscaling Framework, Water Resources Research, 57, e2020WR029548, 2021.

Fernando, K. R. M. and Tsokos, C. P.: Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks, IEEE Transactions on Neural Networks and Learning Systems, 2021.

Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, Nature Climate Change, 6, 986-991, 2016.

François, B., Thao, S., and Vrac, M.: Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks, Climate Dynamics, 57, 3323-3353, 2021.

Girshick, R.: Fast r-cnn, Proceedings of the IEEE international conference on computer vision, 1440-1448,

Goodfellow, I., Bengio, Y., and Courville, A.: Deep learning, MIT press2016.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80-91, 2009.

Habib, E., Henschke, A., and Adler, R. F.: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA, Atmospheric Research, 94, 373-388, 2009.

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

Hamal, K., Sharma, S., Khadka, N., Baniya, B., Ali, M., Shrestha, M. S., Xu, T., Shrestha, D., and Dawadi, B.: Evaluation of MERRA-2 precipitation products using gauge observation in Nepal, Hydrology, 7, 40, 2020.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrology and Earth System Sciences, 22, 2023-2039, 2018.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth System Science Data, 12, 2043-2060, 2020.

Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., and Palmer, T. N.: A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts, arXiv preprint arXiv:2204.02028, 2022.

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

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778,

He, X., Chaney, N. W., Schleiss, M., and Sheffield, J.: Spatial downscaling of precipitation using adaptable random forests, Water Resources Research, 52, 8217-8237, 2016b.

Hong, Y., Hsu, K. l., Moradkhani, H., and Sorooshian, S.: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response, Water resources research, 42, 2006.

Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, International conference on machine learning, 448-456,

Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., and Wang, J.: Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland, Journal of hydrology, 595, 125660, 2021.

Jury, M. R.: An intercomparison of observational, reanalysis, satellite, and coupled model data on mean rainfall in the Caribbean, Journal of Hydrometeorology, 10, 413-430, 2009.

Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., and Singh, A.: Physics-informed machine learning: case studies for weather and climate modelling, Philosophical Transactions of the Royal Society A, 379, 20200093, 2021.

Kim, I.-W., Oh, J., Woo, S., and Kripalani, R.: Evaluation of precipitation extremes over the Asian domain: observation and modelling studies, Climate Dynamics, 52, 1317-1342, 2019.

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

King, A. D., Alexander, L. V., and Donat, M. G.: The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia, International Journal of Climatology, 33, 2376-2387, 2013.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424, 264-277, 2012.

Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K., and Barve, A.: Deep learning–based downscaling of summer monsoon rainfall data over Indian region, Theoretical and Applied Climatology, 143, 1145-1156, 2021.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, nature, 521, 436-444, 2015.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., and Wang, Z.: Photo-realistic single image super-resolution using a generative adversarial network, Proceedings of the IEEE conference on computer vision and pattern recognition, 4681-4690,

Legasa, M., Manzanas, R., Calviño, A., and Gutiérrez, J.: A Posteriori Random Forests for Stochastic Downscaling of Precipitation by Predicting Probability Distributions, Water Resources Research, 58, e2021WR030272, 2022.

Li, W., Pan, B., Xia, J., and Duan, Q.: Convolutional neural network-based statistical postprocessing~~post-processing~~ of ensemble precipitation forecasts, Journal of Hydrology, 605, 127301, 2022.

Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y., and Lambrigtsen, B.: Advancing satellite precipitation retrievals with data driven approaches: Is black box model explainable?, Earth and Space Science, 8, e2020EA001423, 2021.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., and Allen, G. H.: Global reconstruction of naturalized river flows at 2.94 million reaches, Water resources research, 55, 6499-6516, 2019.

Lin, Y. and Mitchell, K. E.: 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications, Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA, Liu, Y., Ganguly, A. R., and Dy, J.: Climate downscaling using YNet: A deep convolutional network with skip connections and fusion, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3145-3153,

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., and Shi, C.: Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution, Remote Sensing of Environment, 233, 111364, 2019.

Mamalakis, A., Langousis, A., Deidda, R., and Marrocu, M.: A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall, Water Resources Research, 53, 2149-2170, 2017.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's~~Earth's~~ Future, 3, 1-14, 2015.

Mei, Y., Maggioni, V., Houser, P., Xue, Y., and Rouf, T.: A nonparametric statistical technique for spatial downscaling of precipitation over High Mountain Asia, Water Resources Research, 56, e2020WR027472, 2020.

Nelson, B. R., Prat, O. P., Seo, D.-J., and Habib, E.: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons, Weather and Forecasting, 31, 371-394, 2016.

Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J., Lee, J., Tian, Y., and Ma, H. Y.: Learning to correct climate projection biases, Journal of Advances in Modeling Earth Systems, 13, e2021MS002509, 2021.

Panda, K. C., Singh, R., Thakural, L., and Sahoo, D. P.: Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall, Journal of Hydrology, 605, 127381, 2022.

Panofsky, H. and Brier, G.: Some applications of statistics to meteorology, Pa. State Univ., University Park, Pa, 1968.

Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Miralles, D. G., Vicente-Serrano, S. M., and Funk, C.: A pan-African high-resolution drought index dataset, Earth System Science Data, 12, 753-769, 2020.

Pierce, D. W., Cayan, D. R., and Thrasher, B. L.: Statistical downscaling using localized constructed analogs (LOCA), Journal of Hydrometeorology, 15, 2558-2585, 2014.

Pour, S. H., Shahid, S., and Chung, E.-S.: A hybrid model for statistical downscaling of daily rainfall, Procedia Engineering, 154, 1424-1430, 2016.

Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C., and Le Moigne, P.: Evaluation of gridded meteorological datasets for hydrological modeling, Journal of Hydrometeorology, 18, 3027-3041, 2017.

Rasp, S. and Lerch, S.: Neural networks for postprocessing~~postprocessing~~ ensemble weather forecasts, Monthly Weather Review, 146, 3885-3900, 2018.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., and Madge, S.: Skilful precipitation nowcasting using deep generative models of radar, Nature, 597, 672-677, 2021.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., and Partyka, G. S.: Land surface precipitation in MERRA-2, Journal of Climate, 30, 1643-1664, 2017.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 2019.

Rivoire, P., Martius, O., and Naveau, P.: A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets, Earth and Space Science, 8, e2020EA001633, 2021.

Rodrigues, E. R., Oliveira, I., Cunha, R., and Netto, M.: DeepDownscale: a deep learning strategy for high-resolution weather forecast, 2018 IEEE 14th International Conference on e-Science (e-Science), 415-422,

Rossa, A., Nurmi, P., and Ebert, E.: Overview of methods for the verification of quantitative precipitation forecasts, in: Precipitation: Advances in measurement, estimation and prediction, Springer, 419-452, 2008.

Ruder, S.: An overview of multitask~~multi-task~~ learning in deep neural networks, arXiv preprint arXiv:1706.05098, 2017.

Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J., and Kumar, V.: Multi-Task Deep Learning of Daily Streamflow and Water Temperature, Water Resources Research, 58, e2021WR030138, 2022.

Schoof, J. T. and Pryor, S. C.: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, International Journal of Climatology: A Journal of the Royal Meteorological Society, 21, 773-790, 2001.

Seltzer, M. L. and Droppo, J.: Multitask~~Multi-task~~ learning in deep neural networks for improved phoneme recognition, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 6965-6969,

Seyyedi, H., Anagnostou, E., Beighley, E., and McCollum, J.: Satellite-driven downscaling of global reanalysis precipitation products for hydrological applications, Hydrology and Earth System Sciences, 18, 5077-5091, 2014.

Sha, Y., Gagne II, D. J., West, G., and Stull, R.: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation, Journal of Applied Meteorology and Climatology, 59, 2075-2092, 2020a.

Sha, Y., Gagne II, D. J., West, G., and Stull, R.: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature, Journal of Applied Meteorology and Climatology, 59, 2057-2073, 2020b.

Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water Resources Research, 54, 8558-8593, 2018.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c.: Deep learning for precipitation nowcasting: A benchmark and a new model, Advances in neural information processing systems, 30, 2017.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., and Bolton, A.: Mastering the game of go without human knowledge, nature, 550, 354-359, 2017.

Suliman, A. H. A., Awchi, T. A., Al-Mola, M., and Shahid, S.: Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq, Atmospheric Research, 242, 105007, 2020.

Sun, A. Y. and Tang, G.: Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets, Frontiers in Water, 2, 536743, 2020.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Reviews of Geophysics, 56, 79-107, 2018.

Tao, Y., Gao, X., Ihler, A., Hsu, K., and Sorooshian, S.: Deep neural networks for precipitation estimation from remotely sensed information, 2016 IEEE Congress on Evolutionary Computation (CEC), 1349-1355,

Tegegne, G. and Melesse, A. M.: Comparison of Trend Preserving Statistical Downscaling Algorithms Toward an Improved Precipitation Extremes Projection in the Headwaters of Blue Nile River in Ethiopia, Environmental Processes, 8, 59-75, 2021.

Thrasher, B., Maurer, E. P., McKellar, C., and Duffy, P. B.: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrology and Earth System Sciences, 16, 3309-3314, 2012.

Tong, K., Su, F., Yang, D., and Hao, Z.: Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau, Journal of hydrology, 519, 423-437, 2014.

Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., and Giorgi, F.: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods, Climate Dynamics, 57, 1425-1443, 2021.

Trinh, T., Do, N., Nguyen, V., and Carr, K.: Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam, Climate Dynamics, 57, 2713-2735, 2021.

Tripathi, S., Srinivas, V., and Nanjundiah, R. S.: Downscaling of precipitation for climate change scenarios: a support vector machine approach, Journal of hydrology, 330, 621-640, 2006.

Vandal, T., Kodra, E., and Ganguly, A. R.: Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation, Theoretical and Applied Climatology, 137, 557-570, 2019.

Vandal, T., Kodra, E., Dy, J., Ganguly, S., Nemani, R., and Ganguly, A. R.: Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2377-2386. 2018.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: Generating high resolution climate change projections through single image super-resolution: An abridged version, International Joint Conferences on Artificial Intelligence Organization,

Wang, F. and Tian, D.: On deep learning-based bias correction and downscaling of multiple climate models simulations, Climate Dynamics, 1-18, 2022.

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J.: Deep learning for daily precipitation and temperature downscaling, Water Resources Research, 57, e2020WR029308, 2021.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, Journal of Geophysical Research: Atmospheres, 107, ACL 6-1-ACL 6-15, 2002.

Xu, H., Xu, C.-Y., Chen, S., and Chen, H.: Similarity and difference of global reanalysis datasets (WFD and APHRODITE) in driving lumped and distributed hydrological models in a humid region of China, Journal of Hydrology, 542, 343-356, 2016.

Xu, M., Liu, Q., Sha, D., Yu, M., Duffy, D. Q., Putman, W. M., Carroll, M., Lee, T., and Yang, C.: PreciPatch: A dictionary-based precipitation downscaling method, Remote Sensing, 12, 1030, 2020.

Xu, X., Frey, S. K., and Ma, D.: Hydrological performance of ERA5 and MERRA-2 precipitation products over the Great Lakes Basin, Journal of Hydrology: Regional Studies, 39, 100982, 2022.

Xu, X., Frey, S. K., Boluwade, A., Erler, A. R., Khader, O., Lapen, D. R., and Sudicky, E.: Evaluation of variability among different precipitation products in the Northern Great Plains, Journal of Hydrology: Regional Studies, 24, 100608, 2019.

Yilmaz, K. K., Hogue, T. S., Hsu, K.-l., Sorooshian, S., Gupta, H. V., and Wagener, T.: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, Journal of Hydrometeorology, 6, 497-517, 2005.

Zhang, X., Anagnostou, E. N., and Schwartz, C. S.: NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS, Remote Sensing, 10, 642, 2018.

Zhong, R., Chen, X., Lai, C., Wang, Z., Lian, Y., Yu, H., and Wu, X.: Drought monitoring utility of satellite-based precipitation products across mainland China, Journal of hydrology, 568, 343-359, 2019.

~~Ashouri, H., Sorooshian, S., Hsu, K. L., Bosilovich, M. G., Lee, J., Wehner, M. F., & Collow, A. (2016). Evaluation of NASA's MERRA precipitation product in reproducing the observed trend and distribution of extreme precipitation events in the United States. Journal of Hydrometeorology, 17(2), 693-711.~~

~~Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. Geoscientific Model Development, 13(4), 2109-2124.~~

~~Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrology and Earth System Sciences, 23(1), 207-224.~~

~~Beck, H. E., Van Dijk, A. I., De Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrology and Earth System Sciences, 21(6), 2881-2903.~~

~~Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., et al. (2019). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. Bulletin of the American Meteorological Society, 100(3), 473-500.~~

~~Bhattacharyya, S., Sreekesh, S., & King, A. (2022). Characteristics of extreme rainfall in different gridded datasets over India during 1983–2015. Atmospheric Research, 267, 105930.~~

~~Bitew, M. M., & Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. Water Resources Research, 47(6).~~

~~Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? Journal of Climate, 28(17), 6938-6959.~~

~~Cavalcante, R. B. L., da Silva Ferreira, D. B., Pontes, P. R. M., Tedeschi, R. G., da Costa, C. P. W., & de Souza, E. B. (2020). Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia. Atmospheric Research, 238, 104879.~~

~~Chen, D., Mak, B., Leung, C. C., & Sivadas, S. (2014). Joint acoustic modeling of triphones and trigraphemes by multi-task learning deep neural networks for low-resource speech recognition. Paper presented at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).~~

~~Chen, Y. (2020). Increasingly uneven intra-seasonal distribution of daily and hourly precipitation over Eastern China. Environmental Research Letters, 15(10), 104068.~~

~~Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., et al. (2021). Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. Atmospheric Research, 250, 105365.~~

~~Daw, A., Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). Physics-guided neural networks (pgnn): An application in lake temperature modeling. arXiv preprint arXiv:1710.11431.~~

~~DeGaetano, A. T., Mooers, G., & Favata, T. (2020). Temporal Changes in the Areal Coverage of Daily Extreme Precipitation in the Northeastern United States Using High-Resolution Gridded Data. Journal of Applied Meteorology and Climatology, 59(3), 551-565.~~

~~Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., & Vorogushyn, S. (2013). Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling. Hydrology and Earth System Sciences, 17(7), 2415-2434.~~

~~Eden, J. M., Widmann, M., Grawe, D., & Rast, S. (2012). Skill, correction, and downscaling of GCM-simulated precipitation. Journal of Climate, 25(11), 3970-3984.~~

~~Emmanouil, S., Langousis, A., Nikolopoulos, E. I., & Anagnostou, E. N. (2021). An ERA-5 Derived CONUS-Wide High-Resolution Precipitation Dataset Based on a Refined Parametric Statistical Downscaling Framework. Water Resources Research, 57(6), e2020WR029548.~~

~~Fernando, K. R. M., & Tsokos, C. P. (2021). Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks. IEEE Transactions on Neural Networks and Learning Systems.~~

~~Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models. Nature Climate Change, 6(11), 986-991.~~

~~François, B., Thao, S., & Vrac, M. (2021). Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. Climate dynamics, 57(11), 3323-3353.~~

~~Girshick, R. (2015). Fast r-cnn. Paper presented at the Proceedings of the IEEE international conference on computer vision.~~

~~Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning: MIT press.~~

~~Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.~~

~~Habib, E., Henschke, A., & Adler, R. F. (2009). Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. Atmospheric Research, 94(3), 373-388.~~

~~Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. Nature, 573(7775), 568-572.~~

~~Hamal, K., Sharma, S., Khadka, N., Baniya, B., Ali, M., Shrestha, M. S., et al. (2020). Evaluation of MERRA-2 precipitation products using gauge observation in Nepal. Hydrology, 7(3), 40.~~

~~Harrigan, S., Prudhomme, C., Parry, S., Smith, K., & Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. Hydrology and Earth System Sciences, 22(3), 2023-2039.~~

~~Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., et al. (2020). GloFAS-ERA5 operational global river discharge reanalysis 1979–present. Earth System Science Data, 12(3), 2043-2060.~~

~~Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., & Palmer, T. N. (2022). A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts. arXiv preprint arXiv:2204.02028.~~

~~He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Paper presented at the Proceedings of the IEEE international conference on computer vision.~~

~~He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.~~

~~He, X., Chaney, N. W., Schleiss, M., & Sheffield, J. (2016). Spatial downscaling of precipitation using adaptable random forests. Water Resources Research, 52(10), 8217-8237.~~

~~Hong, Y., Hsu, K. l., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resources Research, 42(8).~~

~~Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Paper presented at the International conference on machine learning.~~

~~Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., et al. (2021). Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. Journal of hydrology, 595, 125660.~~

~~Jury, M. R. (2009). An intercomparison of observational, reanalysis, satellite, and coupled model data on mean rainfall in the Caribbean. Journal of Hydrometeorology, 10(2), 413-430.~~

~~Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., et al. (2021). Physics-informed machine learning: case studies for weather and climate modelling. Philosophical Transactions of the Royal Society A, 379(2194), 20200093.~~

~~Kim, I.-W., Oh, J., Woo, S., & Kripalani, R. (2019). Evaluation of precipitation extremes over the Asian domain: observation and modelling studies. Climate dynamics, 52(3), 1317-1342.~~

~~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. Journal of hydrology, 596, 125685.~~

~~King, A. D., Alexander, L. V., & Donat, M. G. (2013). The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. International Journal of Climatology, 33(10), 2376-2387.~~

~~Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of hydrology, 424, 264-277.~~

~~Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K., & Barve, A. (2021). Deep learning–based downscaling of summer monsoon rainfall data over Indian region. Theoretical and Applied Climatology, 143(3), 1145-1156.~~

~~LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.~~

~~Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.~~

~~Legasa, M., Manzanas, R., Calviño, A., & Gutiérrez, J. (2022). A Posteriori Random Forests for Stochastic Downscaling of Precipitation by Predicting Probability Distributions. Water Resources Research, 58(4), e2021WR030272.~~

~~Li, W., Pan, B., Xia, J., & Duan, Q. (2022). Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts. Journal of hydrology, 605, 127301.~~

~~Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y., & Lambrigtsen, B. (2021). Advancing satellite precipitation retrievals with data driven approaches: Is black box model explainable? Earth and Space Science, 8(2), e2020EA001423.~~

~~Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resources Research, 55(8), 6499-6516.~~

~~Lin, Y., & Mitchell, K. E. (2005). 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications. Paper presented at the Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA.~~

~~Liu, Y., Ganguly, A. R., & Dy, J. (2020). Climate downscaling using YNet: A deep convolutional network with skip connections and fusion. Paper presented at the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.~~

~~Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., et al. (2019). Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sensing of Environment, 233, 111364.~~

~~Mamalakis, A., Langousis, A., Deidda, R., & Marrocu, M. (2017). A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall. Water Resources Research, 53(3), 2149-2170.~~

~~Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., et al. (2015). VALUE: A framework to validate downscaling approaches for climate change studies. Earth's Future, 3(1), 1-14.~~

~~Mei, Y., Maggioni, V., Houser, P., Xue, Y., & Rouf, T. (2020). A nonparametric statistical technique for spatial downscaling of precipitation over High Mountain Asia. Water Resources Research, 56(11), e2020WR027472.~~

~~Nelson, B. R., Prat, O. P., Seo, D.-J., & Habib, E. (2016). Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. Weather and Forecasting, 31(2), 371-394.~~

~~Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J., Lee, J., et al. (2021). Learning to correct climate projection biases. Journal of Advances in Modeling Earth Systems, 13(10), e2021MS002509.~~

~~Panda, K. C., Singh, R., Thakural, L., & Sahoo, D. P. (2022). Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall. Journal of hydrology, 605, 127381.~~

~~Panofsky, H., & Brier, G. (1968). Some applications of statistics to meteorology. Pa. State Univ., University Park, Pa.~~

~~Panofsky, H. A., & Brier, G. W. (1968). Some applications of statistics to meteorology: Earth and Mineral Sciences Continuing Education, College of Earth and ….~~

~~Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Miralles, D. G., et al. (2020). A pan-African high-resolution drought index dataset. Earth System Science Data, 12(1), 753-769.~~

~~Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical downscaling using localized constructed analogs (LOCA). Journal of Hydrometeorology, 15(6), 2558-2585.~~

~~Pour, S. H., Shahid, S., & Chung, E.-S. (2016). A hybrid model for statistical downscaling of daily rainfall. Procedia Engineering, 154, 1424-1430.~~

~~Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C., & Le Moigne, P. (2017). Evaluation of gridded meteorological datasets for hydrological modeling. Journal of Hydrometeorology, 18(11), 3027-3041.~~

~~Rasp, S., & Lerch, S. (2018). Neural networks for postprocessing ensemble weather forecasts. Monthly Weather Review, 146(11), 3885-3900.~~

~~Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. Nature, 597(7878), 672-677.~~

~~Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., & Partyka, G. S. (2017). Land surface precipitation in MERRA-2. Journal of Climate, 30(5), 1643-1664.~~

~~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(7743), 195-204.~~

~~Rivoire, P., Martius, O., & Naveau, P. (2021). A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. Earth and Space Science, 8(4), e2020EA001633.~~

~~Rodrigues, E. R., Oliveira, I., Cunha, R., & Netto, M. (2018). DeepDownscale: a deep learning strategy for high-resolution weather forecast. Paper presented at the 2018 IEEE 14th International Conference on e-Science (e-Science).~~

~~Rossa, A., Nurmi, P., & Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In Precipitation: Advances in measurement, estimation and prediction (pp. 419-452): Springer.~~

~~Ruder, S. (2017). An overview of multi-task learning in deep neural networks. ArXiv Preprint ArXiv:1706.05098.~~

~~Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J., & Kumar, V. (2022). Multi-Task Deep Learning of Daily Streamflow and Water Temperature. Water Resources Research, 58(4), e2021WR030138.~~

~~Schoof, J. T., & Pryor, S. C. (2001). Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. International Journal of Climatology: A Journal of the Royal Meteorological Society, 21(7), 773-790.~~

~~Seltzer, M. L., & Droppo, J. (2013). Multi-task learning in deep neural networks for improved phoneme recognition. Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.~~

~~Seyyedi, H., Anagnostou, E., Beighley, E., & McCollum, J. (2014). Satellite-driven downscaling of global reanalysis precipitation products for hydrological applications. Hydrology and Earth System Sciences, 18(12), 5077-5091.~~

~~Sha, Y., Gagne II, D. J., 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. Journal of Applied Meteorology and Climatology, 59(12), 2057-2073.~~

~~Sha, Y., Gagne II, D. J., West, G., & Stull, R. (2020b). Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation. Journal of Applied Meteorology and Climatology, 59(12), 2075-2092.~~

~~Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54(11), 8558-8593.~~

~~Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. Advances in neural information processing systems, 30.~~

~~Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of go without human knowledge. Nature, 550(7676), 354-359.~~

~~Suliman, A. H. A., Awchi, T. A., Al-Mola, M., & Shahid, S. (2020). Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq. Atmospheric Research, 242, 105007.~~

~~Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Reviews of geophysics, 56(1), 79-107.~~

~~Tao, Y., Gao, X., Ihler, A., Hsu, K., & Sorooshian, S. (2016). Deep neural networks for precipitation estimation from remotely sensed information. Paper presented at the 2016 IEEE Congress on Evolutionary Computation (CEC).~~

~~Tegegne, G., & Melesse, A. M. (2021). Comparison of Trend Preserving Statistical Downscaling Algorithms Toward an Improved Precipitation Extremes Projection in the Headwaters of Blue Nile River in Ethiopia. Environmental Processes, 8(1), 59-75.~~

~~Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Bias correcting climate model simulated daily temperature extremes with quantile mapping. Hydrology and Earth System Sciences, 16(9), 3309-3314.~~

~~Tong, K., Su, F., Yang, D., & Hao, Z. (2014). Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau. Journal of hydrology, 519, 423-437.~~

~~Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., & Giorgi, F. (2021). Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. Climate dynamics, 57(5), 1425-1443.~~

~~Trinh, T., Do, N., Nguyen, V., & Carr, K. (2021). Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam. Climate dynamics, 57(9), 2713-2735.~~

~~Tripathi, S., Srinivas, V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. Journal of hydrology, 330(3-4), 621-640.~~

~~Vandal, T., Kodra, E., & Ganguly, A. R. (2019). Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. Theoretical and Applied Climatology, 137(1), 557-570.~~

~~Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2018). Generating high resolution climate change projections through single image super-resolution: An abridged version. Paper presented at the International Joint Conferences on Artificial Intelligence Organization.~~

~~Wang, F., & Tian, D. (2022). On deep learning-based bias correction and downscaling of multiple climate models simulations. Climate dynamics, 1-18.~~

~~Wang, F., Tian, D., Lowe, L., Kalin, L., & Lehrter, J. (2021). Deep learning for daily precipitation and temperature downscaling. Water Resources Research, 57(4), e2020WR029308.~~

~~Wood, A. W., Maurer, E. P., Kumar, A., & Lettenmaier, D. P. (2002). Long-range experimental hydrologic forecasting for the eastern United States. Journal of Geophysical Research: Atmospheres, 107(D20), ACL 6-1-ACL 6-15.~~

~~Xu, H., Xu, C. Y., Chen, S., & Chen, H. (2016). Similarity and difference of global reanalysis datasets (WFD and APHRODITE) in driving lumped and distributed hydrological models in a humid region of China. Journal of hydrology, 542, 343-356.~~

~~Xu, M., Liu, Q., Sha, D., Yu, M., Duffy, D. Q., Putman, W. M., et al. (2020). PreciPatch: A dictionary-based precipitation downscaling method. Remote Sensing, 12(6), 1030.~~

~~Xu, X., Frey, S. K., Boluwade, A., Erler, A. R., Khader, O., Lapen, D. R., & Sudicky, E. (2019). Evaluation of variability among different precipitation products in the Northern Great Plains. Journal of Hydrology: Regional Studies, 24, 100608.~~

~~Xu, X., Frey, S. K., & Ma, D. (2022). Hydrological performance of ERA5 and MERRA-2 precipitation products over the Great Lakes Basin. Journal of Hydrology: Regional Studies, 39, 100982.~~

~~Yilmaz, K. K., Hogue, T. S., Hsu, K. l., Sorooshian, S., Gupta, H. V., & Wagener, T. (2005). Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. Journal of Hydrometeorology, 6(4), 497-517.~~

~~Zhang, X., Anagnostou, E. N., & Schwartz, C. S. (2018). NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS. Remote Sensing, 10(4), 642.~~

~~Zhong, R., Chen, X., Lai, C., Wang, Z., Lian, Y., Yu, H., & Wu, X. (2019). Drought monitoring utility of satellite-based precipitation products across mainland China. Journal of hydrology, 568, 343-359.~~

Table 1. Deep Learning (DL) Experimental Design

Experimental Runs (Scenarios)	Input	Output	Loss
Scenario1	hourly precipitation (*P*)	*P*	MAE
Scenario2	*P*	*P*	Weighted MAE
Scenario3	*P*	*P* + categorical *P*	Weighted MAE +λ∗Weighted cross-entropy
Scenario4	Covariates w/o *P*	*P*	Weighted MAE
Scenario5	Covariates w/ *P*	*P*	Weighted MAE
Scenario6	Covariates w/ *P*	*P* + categorical *P*	Weighted MAE + λ∗Weighted cross-entropy

Table 2. Selected atmospheric covariates for DL downscaling and bias correction

~~NO~~	~~Name~~	~~Description~~
~~1~~	~~H250~~	~~Geopotential height at 250 hPa~~
~~2~~	~~H500~~	~~Geopotential height at 500 hPa~~
~~3~~	~~H850~~	~~Geopotential height at 850 hPa~~
~~4~~	~~Q250~~	~~Specific humidity at 250 hPa~~
~~5~~	~~Q500~~	~~Specific humidity at 500 hPa~~
~~6~~	~~Q850~~	~~Specific humidity at 850 hPa~~
~~7~~	~~T250~~	~~Air temperature at 250 hPa~~
~~8~~	~~T500~~	~~Air temperature at 500 hPa~~
~~9~~	~~T850~~	~~Air temperature at 850 hPa~~
~~10~~	~~U250~~	~~Eastward wind at 250 hPa~~
~~11~~	~~U500~~	~~Eastward wind at 500 hPa~~
~~12~~	~~U850~~	~~Eastward wind at 850 hPa~~
~~13~~	~~V250~~	~~Northward wind at 250 hPa~~
~~14~~	~~V500~~	~~Northward wind at 250 hPa~~
~~15~~	~~V850~~	~~Northward wind at 250 hPa~~
~~16~~	~~OMEGA500~~	~~Omega (vertical wind) at 500 hPa~~
~~17~~	~~SLP~~	~~Sea level pressure~~
~~18~~	~~T2M~~	~~2-meter air temperature~~

NO	Other variables	Variable description	Units
1	H250	Geopotential height at 250 hPa	m
2	H500	Geopotential height at 500 hPa	m
3	H850	Geopotential height at 850 hPa	m
4	Q250	Specific humidity at 250 hPa	kg/kg
5	Q500	Specific humidity at 500 hPa	kg/kg
6	Q850	Specific humidity at 850 hPa	kg/kg
7	T250	Air temperature at 250 hPa	K
8	T500	Air temperature at 500 hPa	K
9	T850	Air temperature at 850 hPa	K
10	U250	Eastward wind at 250 hPa	m/s
11	U500	Eastward wind at 500 hPa	m/s
12	U850	Eastward wind at 850 hPa	m/s
13	V250	Northward wind at 250 hPa	m/s
14	V500	Northward wind at 250 hPa	m/s
15	V850	Northward wind at 250 hPa	m/s
16	OMEGA500	Omega (vertical wind) at 500 hPa	Pa/s
17	SLP	Sea level pressure	Pa
18	T2M	2-meter air temperature	K

Formatted Table ... [1]
Formatted ... [2]
Formatted ... [3]
Formatted ... [4]
Formatted ... [5]
Formatted ... [6]
Formatted ... [7]
Formatted ... [8]
Formatted ... [9]
Formatted ... [10]
Formatted ... [11]
Formatted ... [12]
Formatted ... [13]
Formatted ... [14]
Formatted ... [15]
Formatted ... [16]
Formatted ... [17]
Formatted ... [18]
Formatted ... [19]
Formatted ... [20]
Formatted ... [21]
Formatted ... [22]
Formatted ... [23]
Formatted ... [24]
Formatted ... [25]
Formatted ... [26]
Formatted ... [27]
Formatted ... [28]
Formatted ... [29]
Formatted ... [30]
Formatted ... [31]
Formatted ... [32]
Formatted ... [33]
Formatted ... [34]
Formatted ... [35]
Formatted ... [36]
Formatted ... [37]
Formatted ... [38]
Formatted ... [39]
Formatted ... [40]
Formatted ... [41]
Formatted ... [42]
Formatted ... [43]
Formatted ... [44]
Formatted ... [45]
Formatted ... [46]
Formatted ... [47]
Formatted ... [48]
Formatted ... [49]
Formatted ... [50]
Formatted ... [51]
Formatted ... [52]
Formatted ... [53]
Formatted ... [54]

Table 3. Overall assessment for hourly, daily total, and monthly mean of hourly precipitation. KGE represents the modified Kling-Gupta efficiency (KGE) and it includes three components (correlation component r, bias component β and variability component γ). The correlation component *r* is represented by correlation coefficient, the bias component β is represented by the ratio of estimated and observed means, and the variability component γ is represented by the estimated and observed coefficients of variation.

Temporal scales	Scenarios*	KGE	r	β	γ	RMSE (mm)	MAE (mm)
Hourly precipitation	Scenario1	-0.0584	0.267	0.288	1.28	1.20	0.189
	Scenario2	0.218	0.297	0.958	0.660	1.25	0.258
	Scenario3	0.203	0.278	1.02	0.664	1.28	0.269
	Scenario4	0.250	0.331	0.883	0.682	1.21	0.240
	Scenario5	0.283	0.358	1.02	0.682	1.22	0.248
	Scenario6	0.262	0.356	1.00	0.639	1.20	0.247
	QDM_BI	0.248	0.332	1.02	1.35	1.36	0.256
Daily precipitation	Scenario1	0.0935	0.615	0.288	1.409	10.19	3.54
	Scenario2	0.644	0.685	0.958	0.840	8.76	3.42
	Scenario3	0.626	0.675	1.02	0.815	8.94	3.54
	Scenario4	0.618	0.642	0.883	0.935	9.37	3.55
	Scenario5	0.688	0.701	1.02	0.914	8.89	3.40
	Scenario6	0.668	0.701	1.00	0.855	8.65	3.34
	QDM_BI	0.644	0.689	1.02	1.17	10.50	3.42
Monthly mean of hourly precipitation	Scenario1	0.0206	0.567	0.289	1.52	0.162	0.133
	Scenario2	0.766	0.778	0.958	0.941	0.0721	0.0512
	Scenario3	0.784	0.791	1.02	0.951	0.0713	0.0505
	Scenario4	0.690	0.712	0.883	0.991	0.0835	0.0592
	Scenario5	0.778	0.782	1.02	0.964	0.0734	0.0519
	Scenario6	0.776	0.783	1.00	0.945	0.0719	0.0511
	QDM_BI	0.717	0.777	1.02	1.17	0.0850	0.0553

*Scenarios have different settings: Scenario1 is with a regular MAE loss function and coarse precipitation as a predictor; Scenario2 is with a weighted MAE loss and coarse precipitation as a predictor; Scenario3 is the same as Scenario2 except with a classification as an ~~additional~~ auxiliary task; Scenario4 is with a weighted loss function ~~but~~ and covariates as predictors; Scenario5 is the same as Scenario4 except also including coarse precipitation as predictors; Scenario 6 is the same as Scenario5 but including a classification as an ~~additional~~ auxiliary task.

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Line spacing: single

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Table 4. Performance of extreme indices including hourly *P* at 99% percentile and annual maximum wet spell in hours. KGE represents the modified Kling-Gupta efficiency (KGE) and it includes three components (correlation component r, bias component β and variability component γ). The correlation component *r* is represented by correlation coefficient, the bias component β is represented by the ratio of estimated and observed means, and the variability component γ is represented by the estimated and observed coefficients of variation.

Formatted: Font: Italic

Extreme indices	Scenarios*	KGE	r	β	γ	RMSE	MAE
99th percentile (mm)	Scenario1	-1.306	0.352	0.358	3.12	3.150	3.101
	Scenario2	0.367	0.415	0.806	1.14	1.049	0.946
	Scenario3	0.243	0.264	0.828	1.04	0.978	0.876
	Scenario4	0.204	0.242	0.763	1.06	1.255	1.153
	Scenario5	0.255	0.284	0.863	1.15	0.858	0.744
	Scenario6	0.245	0.271	0.845	1.12	0.922	0.800
	QDM_BI	0.158	0.244	0.900	1.36	0.793	0.655
Annual maximum wet spell (hours)	Scenario1	0.153	0.275	0.621	1.22	12.2	10.3
	Scenario2	0.293	0.302	1.11	0.988	9.17	7.14
	Scenario3	0.291	0.302	1.07	1.10	9.33	7.03
	Scenario4	0.121	0.282	1.46	1.21	17.0	12.7
	Scenario5	0.193	0.335	1.44	1.11	15.8	12.2
	Scenario6	0.152	0.306	1.47	1.14	16.6	12.6
	QDM_BI	-0.209	0.173	1.88	1.09	26.6	22.2

*Scenarios have different settings: Scenario1 is with a regular MAE loss function and coarse precipitation as a predictor; Scenario2 is with a weighted MAE loss and coarse precipitation as a predictor; Scenario3 is the same as Scenario2 except with a classification as an auxiliary task; Scenario4 is with a weighted loss function and covariates as predictors; Scenario5 is the same as Scenario4 except also including coarse precipitation as predictors; Scenario 6 is the same as Scenario5 but including a classification as an auxiliary task.

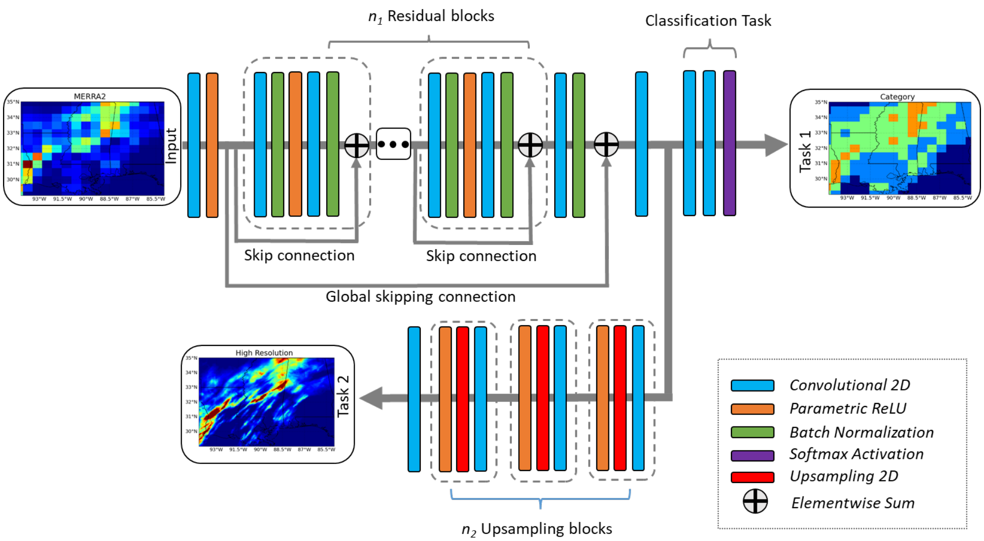


Figure 1. The ~~modified~~ customized SRDRN architecture with multitask~~multitask~~ learning, which~~,~~ ~~which~~ include~~s~~d the classification of *P* categories as an auxiliary task (Task 1) in addition to downscaling and bias correcting actual *P* values (Task 2). Note that this figure is ~~a modified version of~~modified from the SRDRN architecture shown in Wang et al. (2021).

Formatted: Font: Italic

Formatted: Font: Italic

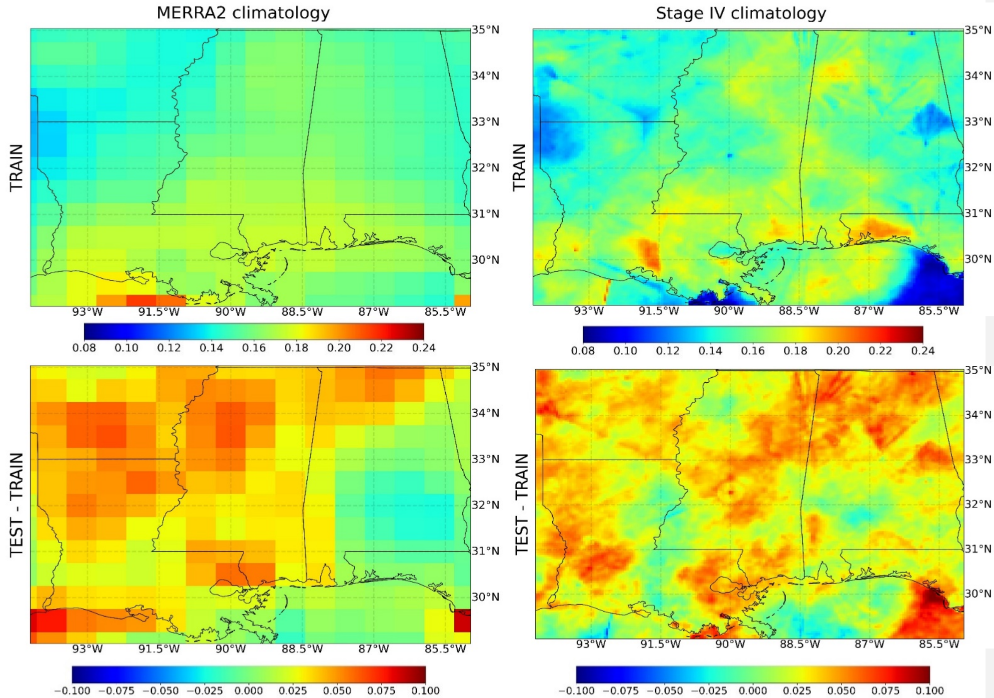


Figure 2. Climatology of hourly precipitation (in a unit of mm/h) from MERRA2 and Stage IV during the training period (2002 to 2015; first row) and their differences (second row) between the testing (2019 to 2021) and training periods.

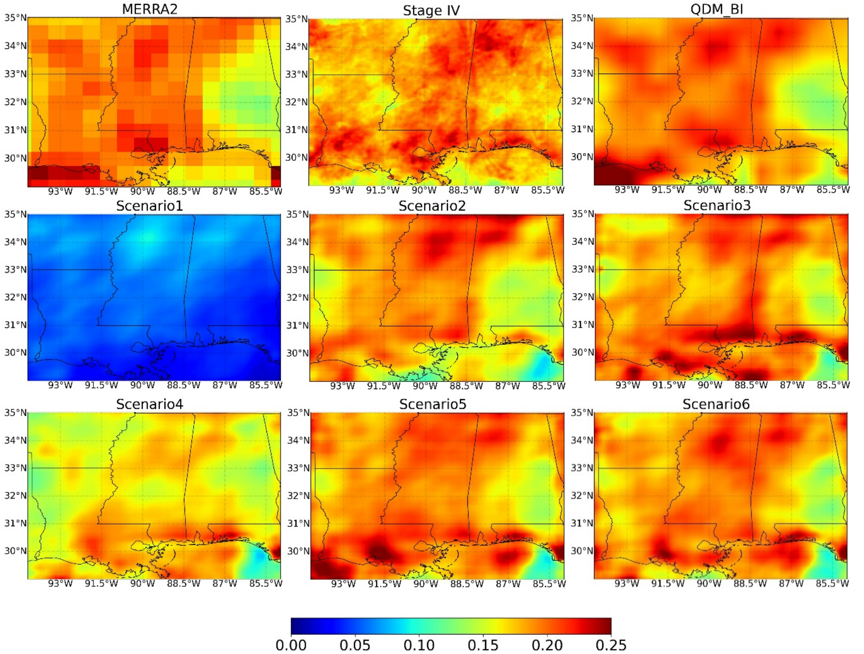
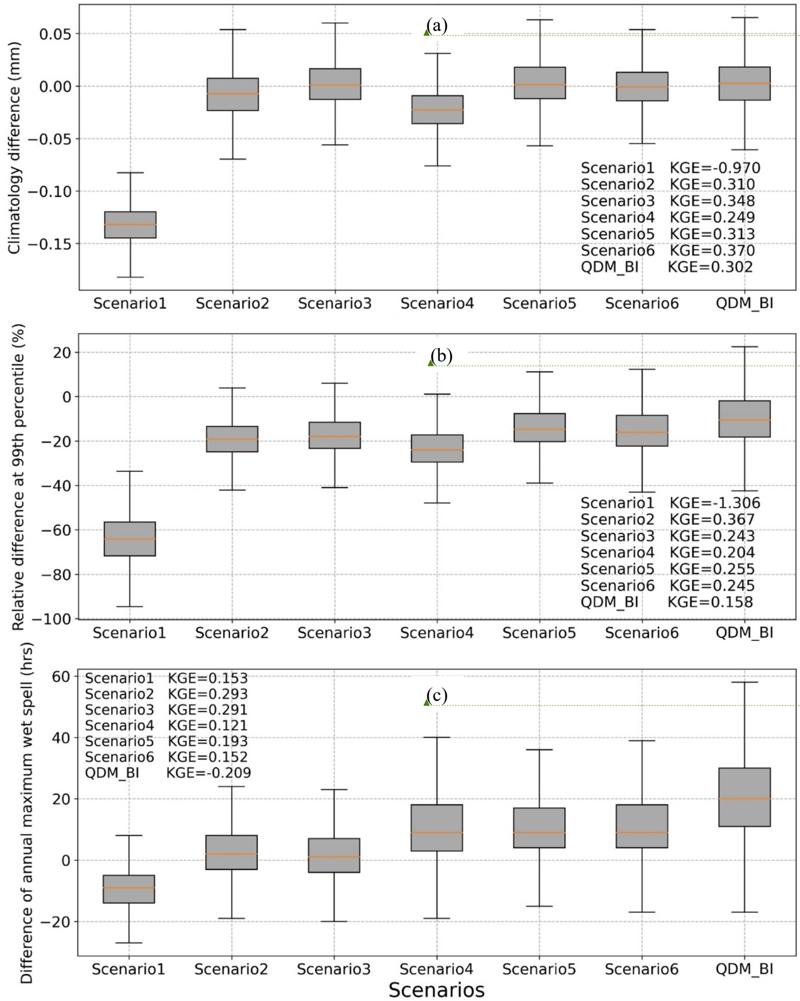


Figure 3. Hourly precipitation climatology (in a unit of mm/h) during the testing period (2019 to 2021), which includes MERRA2, Stage IV, QDM_BI, and six DL experimental runs (Scenario1 to Scenario6).



Formatted: Font: 10 pt

Formatted: Font: 10 pt

Formatted: Font: 10 pt

1130

Figure 4. Boxplots showing hourly precipitation estimates minus Stage IV observations based on~~,~~ (a) climatology, (b) extreme at 99% percentile, and (c) annual maximum wet spell in hours during the testing period (2019 to 2021). Precipitation estimates are produced from the QDM_BI approach and 6 DL experimental runs (Scenario1 to Scenario6).

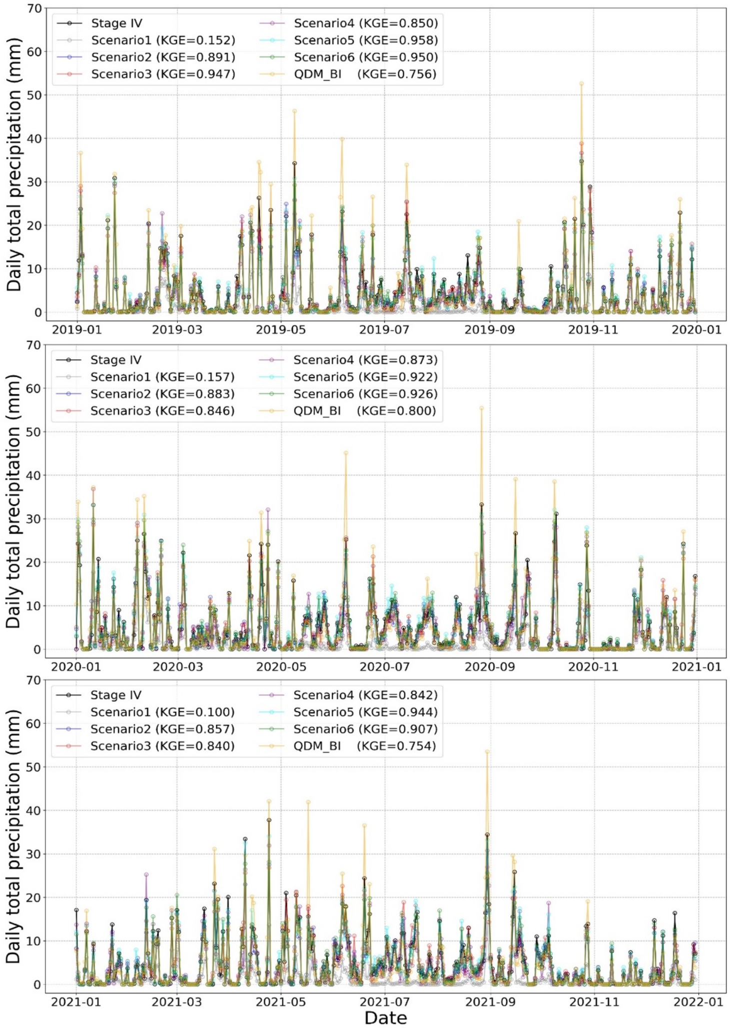


Figure 5. Daily total precipitation during the testing period (2019 to 2021) from Stage IV, QDM_BI, and 6 DL experimental runs (Scenario1 to Scenario6).

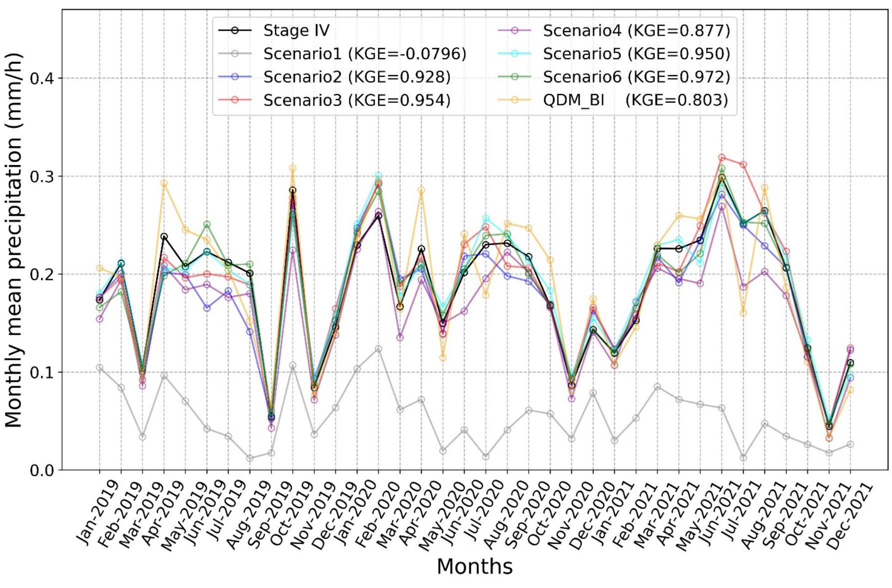


Figure 6. Monthly mean of hourly precipitation time series during the testing period (2019 to 2021) from Stage IV, QDM_BI, and 6 DL experimental runs (Scenario1 to Scenario6).

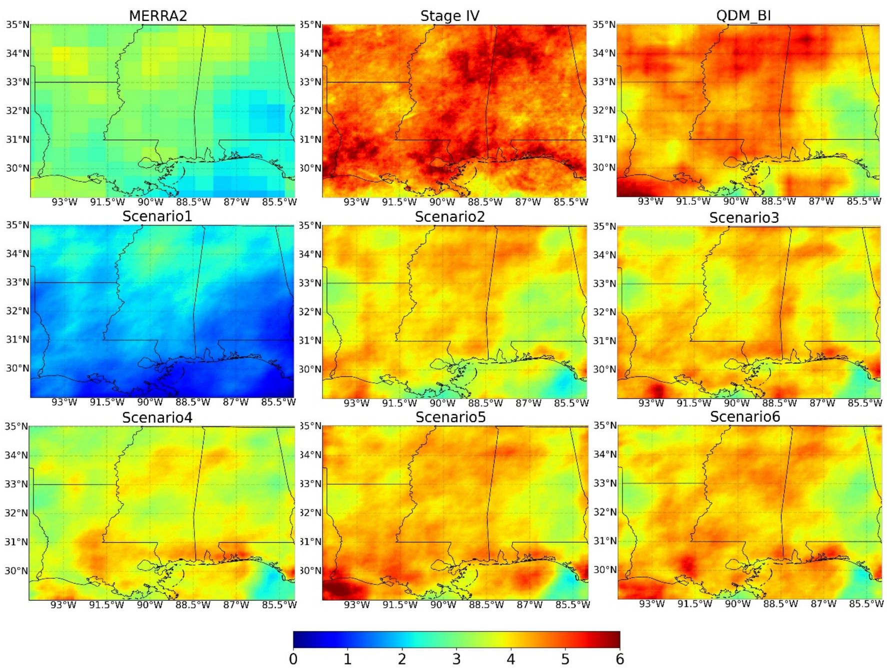


Figure 7. Spatial map of hourly precipitation extremes~~extreme~~ at 99th percentile (in a unit of mm/h) from raw MERRA2, Stage IV, QDM_BI, and 6 DL experimental runs (Scenario1 to Scenario6).

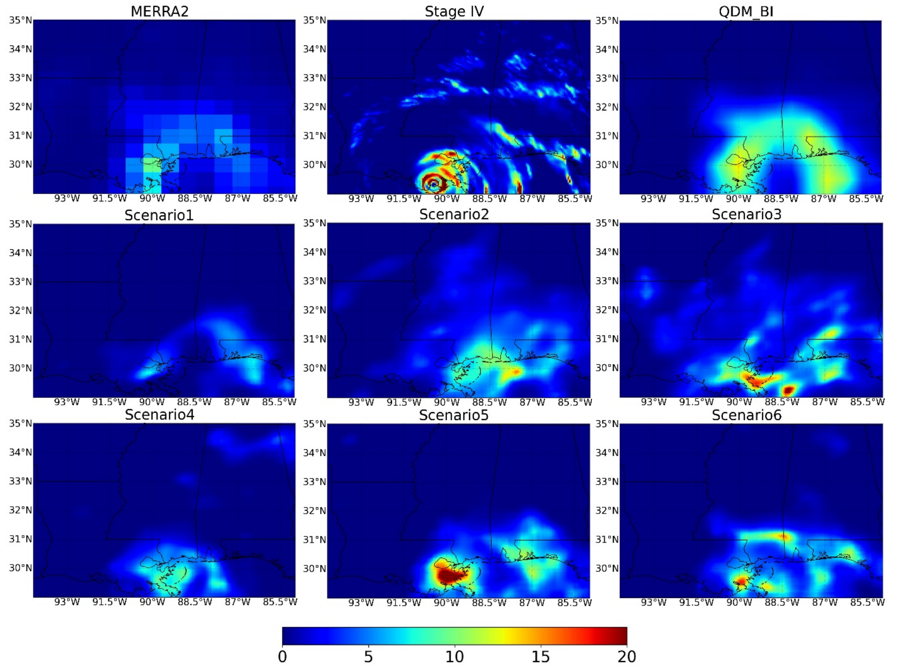


Figure 8. Hourly precipitation (in a unit of mm/h) from 19:00 to 20:00 on 29 August 2021 in UTC time zone when Hurricane Ida landed in Louisiana, including raw MERRA2, Stage IV, QDM_BI and six DL experimental runs (Scenario1 to Scenario6).

Categories	MERRA2	QDM	Scenario3	Scenario6
0-0.1mm	80.54	88.10	81.00	86.44
0.1-2.5mm	27.10	23.60	25.93	27.91
2.5-10mm	14.94	15.30	19.63	19.91
>10mm	4.32	7.12	8.15	11.07

Figure 9. Heat map showing the Intersection ~~over~~ Over Union (~~IoU~~IOU) comparing coarsened Stage IV with raw MERRA2, QDM, two deep learning experiment runs with classification task (Scenario3 and Scenario6)

~~Aadhar, S. and Mishra, V.: High-resolution near real-time drought monitoring in South Asia, Scientific Data, 4, 1-14, 2017.~~

~~AghaKouchak, A., Mehran, A., Norouzi, H., and Behrangi, A.: Systematic and random error components in satellite precipitation data sets, Geophysical Research Letters, 39, 2012.~~

~~AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E.: Evaluation of satellite-retrieved extreme precipitation rates across the central United States, Journal of Geophysical Research: Atmospheres, 116, 2011.~~

~~Ashouri, H., Sorooshian, S., Hsu, K. L., Bosilovich, M. G., Lee, J., Wehner, M. F., and Collow, A.: Evaluation of NASA's MERRA precipitation product in reproducing the observed trend and distribution of extreme precipitation events in the United States, Journal of Hydrometeorology, 17, 693-711, 2016.~~

~~Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, Geoscientific Model Development, 13, 2109-2124, 2020.~~

~~Beck, H. E., Van Dijk, A. I., De Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrology and Earth System Sciences, 21, 2881-2903, 2017.~~

~~Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473-500, 2019a.~~

~~Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207-224, 2019b.~~

~~Bhattacharyya, S., Sreekesh, S., and King, A.: Characteristics of extreme rainfall in different gridded datasets over India during 1983-2015, Atmospheric Research, 267, 105930, 2022.~~

~~Bitew, M. M. and Gebremichael, M.: Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, Water Resources Research, 47, 2011.~~

~~Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?, Journal of Climate, 28, 6938-6959, 2015.~~

~~Cavalcante, R. B. L., da Silva Ferreira, D. B., Pontes, P. R. M., Tedeschi, R. G., da Costa, C. P. W., and de Souza, E. B.: Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia, Atmospheric Research, 238, 104879, 2020.~~

~~Chen, D., Mak, B., Leung, C.-C., and Sivadas, S.: Joint acoustic modeling of triphones and trigraphemes by multi-task learning deep neural networks for low-resource speech recognition, 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5592-5596,~~

~~Chen, Y.: Increasingly uneven intra-seasonal distribution of daily and hourly precipitation over Eastern China, Environmental Research Letters, 15, 104068, 2020.~~

~~Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., and Khadka, N.: Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya, Atmospheric Research, 250, 105365, 2021.~~

~~Daw, A., Karpatne, A., Watkins, W., Read, J., and Kumar, V.: Physics-guided neural networks (pgnn): An application in lake temperature modeling, arXiv preprint arXiv:1710.11431, 2017.~~

~~DeGaetano, A. T., Mooers, G., and Favata, T.: Temporal Changes in the Areal Coverage of Daily Extreme Precipitation in the Northeastern United States Using High-Resolution Gridded Data, Journal of Applied Meteorology and Climatology, 59, 551-565, 2020.~~

~~Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., and Vorogushyn, S.: Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, Hydrology and Earth System Sciences, 17, 2415-2434, 2013.~~

~~Eden, J. M., Widmann, M., Grawe, D., and Rast, S.: Skill, correction, and downscaling of GCM-simulated precipitation, Journal of Climate, 25, 3970-3984, 2012.~~

~~Emmanouil, S., Langousis, A., Nikolopoulos, E. I., and Anagnostou, E. N.: An ERA-5 Derived CONUS-Wide High-Resolution Precipitation Dataset Based on a Refined Parametric Statistical Downscaling Framework, Water Resources Research, 57, e2020WR029548, 2021.~~

~~Fernando, K. R. M. and Tsokos, C. P.: Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks, IEEE Transactions on Neural Networks and Learning Systems, 2021.~~

~~Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, Nature Climate Change, 6, 986-991, 2016.~~

~~François, B., Thao, S., and Vrac, M.: Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks, Climate Dynamics, 57, 3323-3353, 2021.~~

~~Girshick, R.: Fast r-cnn, Proceedings of the IEEE international conference on computer vision, 1440-1448,~~

~~Goodfellow, I., Bengio, Y., and Courville, A.: Deep learning, MIT press2016.~~

~~Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80-91, 2009.~~

~~Habib, E., Henschke, A., and Adler, R. F.: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA, Atmospheric Research, 94, 373-388, 2009.~~

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

~~Hamal, K., Sharma, S., Khadka, N., Baniya, B., Ali, M., Shrestha, M. S., Xu, T., Shrestha, D., and Dawadi, B.: Evaluation of MERRA-2 precipitation products using gauge observation in Nepal, Hydrology, 7, 40, 2020.~~

~~Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrology and Earth System Sciences, 22, 2023-2039, 2018.~~

~~Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth System Science Data, 12, 2043-2060, 2020.~~

~~Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., and Palmer, T. N.: A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts, arXiv preprint arXiv:2204.02028, 2022.~~

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

~~He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778,~~

~~He, X., Chaney, N. W., Schleiss, M., and Sheffield, J.: Spatial downscaling of precipitation using adaptable random forests, Water Resources Research, 52, 8217-8237, 2016b.~~

~~Hong, Y., Hsu, K. l., Moradkhani, H., and Sorooshian, S.: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response, Water resources research, 42, 2006.~~

~~Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, International conference on machine learning, 448-456,~~

~~Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., and Wang, J.: Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland, Journal of hydrology, 595, 125660, 2021.~~

~~Jury, M. R.: An intercomparison of observational, reanalysis, satellite, and coupled model data on mean rainfall in the Caribbean, Journal of Hydrometeorology, 10, 413-430, 2009.~~

~~Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., and Singh, A.: Physics-informed machine learning: case studies for weather and climate modelling, Philosophical Transactions of the Royal Society A, 379, 20200093, 2021.~~

~~Kim, I.-W., Oh, J., Woo, S., and Kripalani, R.: Evaluation of precipitation extremes over the Asian domain: observation and modelling studies, Climate Dynamics, 52, 1317-1342, 2019.~~

~~Kim, S., Joo, K., Kim, H., Shin, J.-Y., and Heo, J.-H.: Regional quantile delta mapping method using regional frequency analysis for regional climate model precipitation, Journal of Hydrology, 596, 125685, 2021.~~

~~King, A. D., Alexander, L. V., and Donat, M. G.: The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia, International Journal of Climatology, 33, 2376-2387, 2013.~~

~~Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424, 264-277, 2012.~~

~~Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K., and Barve, A.: Deep learning–based downscaling of summer monsoon rainfall data over Indian region, Theoretical and Applied Climatology, 143, 1145-1156, 2021.~~

~~LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, nature, 521, 436-444, 2015.~~

~~Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., and Wang, Z.: Photo-realistic single image super-resolution using a generative adversarial network, Proceedings of the IEEE conference on computer vision and pattern recognition, 4681-4690,~~

~~Legasa, M., Manzanas, R., Calviño, A., and Gutiérrez, J.: A Posteriori Random Forests for Stochastic Downscaling of Precipitation by Predicting Probability Distributions, Water Resources Research, 58, e2021WR030272, 2022.~~

~~Li, W., Pan, B., Xia, J., and Duan, Q.: Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts, Journal of Hydrology, 605, 127301, 2022.~~

~~Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y., and Lambrigtsen, B.: Advancing satellite precipitation retrievals with data driven approaches: Is black box model explainable?, Earth and Space Science, 8, e2020EA001423, 2021.~~

~~Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., and Allen, G. H.: Global reconstruction of naturalized river flows at 2.94 million reaches, Water resources research, 55, 6499-6516, 2019.~~

~~Lin, Y. and Mitchell, K. E.: 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications, Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA,~~

~~Liu, Y., Ganguly, A. R., and Dy, J.: Climate downscaling using YNet: A deep convolutional network with skip connections and fusion, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3145-3153,~~

~~Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., and Shi, C.: Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution, Remote Sensing of Environment, 233, 111364, 2019.~~

~~Mamalakis, A., Langousis, A., Deidda, R., and Marrocu, M.: A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall, Water Resources Research, 53, 2149-2170, 2017.~~

~~Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 1-14, 2015.~~

~~Mei, Y., Maggioni, V., Houser, P., Xue, Y., and Rouf, T.: A nonparametric statistical technique for spatial downscaling of precipitation over High Mountain Asia, Water Resources Research, 56, e2020WR027472, 2020.~~

~~Nelson, B. R., Prat, O. P., Seo, D. J., and Habib, E.: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons, Weather and Forecasting, 31, 371-394, 2016.~~

~~Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J., Lee, J., Tian, Y., and Ma, H. Y.: Learning to correct climate projection biases, Journal of Advances in Modeling Earth Systems, 13, e2021MS002509, 2021.~~

~~Panda, K. C., Singh, R., Thakural, L., and Sahoo, D. P.: Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall, Journal of Hydrology, 605, 127381, 2022.~~

~~Panofsky, H. and Brier, G.: Some applications of statistics to meteorology, Pa. State Univ., University Park, Pa, 1968.~~

~~Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Miralles, D. G., Vicente-Serrano, S. M., and Funk, C.: A pan-African high-resolution drought index dataset, Earth System Science Data, 12, 753-769, 2020.~~

~~Pierce, D. W., Cayan, D. R., and Thrasher, B. L.: Statistical downscaling using localized constructed analogs (LOCA), Journal of Hydrometeorology, 15, 2558-2585, 2014.~~

~~Pour, S. H., Shahid, S., and Chung, E.-S.: A hybrid model for statistical downscaling of daily rainfall, Procedia Engineering, 154, 1424-1430, 2016.~~

~~Raimonet, M., Oudin, L., Thieu, V., Silvestre, M., Vautard, R., Rabouille, C., and Le Moigne, P.: Evaluation of gridded meteorological datasets for hydrological modeling, Journal of Hydrometeorology, 18, 3027-3041, 2017.~~

~~Rasp, S. and Lerch, S.: Neural networks for postprocessing ensemble weather forecasts, Monthly Weather Review, 146, 3885-3900, 2018.~~

~~Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., and Madge, S.: Skilful precipitation nowcasting using deep generative models of radar, Nature, 597, 672-677, 2021.~~

~~Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., and Partyka, G. S.: Land surface precipitation in MERRA-2, Journal of Climate, 30, 1643-1664, 2017.~~

~~Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 2019.~~

~~Rivoire, P., Martius, O., and Naveau, P.: A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets, Earth and Space Science, 8, e2020EA001633, 2021.~~

~~Rodrigues, E. R., Oliveira, I., Cunha, R., and Netto, M.: DeepDownscale: a deep learning strategy for high-resolution weather forecast, 2018 IEEE 14th International Conference on e-Science (e-Science), 415-422,~~

~~Rossa, A., Nurmi, P., and Ebert, E.: Overview of methods for the verification of quantitative precipitation forecasts, in: Precipitation: Advances in measurement, estimation and prediction, Springer, 419-452, 2008.~~

~~Ruder, S.: An overview of multi-task learning in deep neural networks, arXiv preprint arXiv:1706.05098, 2017.~~

~~Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J., and Kumar, V.: Multi-Task Deep Learning of Daily Streamflow and Water Temperature, Water Resources Research, 58, e2021WR030138, 2022.~~

~~Schoof, J. T. and Pryor, S. C.: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, International Journal of Climatology: A Journal of the Royal Meteorological Society, 21, 773-790, 2001.~~

~~Seltzer, M. L. and Droppo, J.: Multi-task learning in deep neural networks for improved phoneme recognition, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 6965-6969,~~

~~Seyyedi, H., Anagnostou, E., Beighley, E., and McCollum, J.: Satellite-driven downscaling of global reanalysis precipitation products for hydrological applications, Hydrology and Earth System Sciences, 18, 5077-5091, 2014.~~

~~Sha, Y., Gagne II, D. J., West, G., and Stull, R.: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation, Journal of Applied Meteorology and Climatology, 59, 2075-2092, 2020a.~~

~~Sha, Y., Gagne II, D. J., West, G., and Stull, R.: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature, Journal of Applied Meteorology and Climatology, 59, 2057-2073, 2020b.~~

~~Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water Resources Research, 54, 8558-8593, 2018.~~

~~Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c.: Deep learning for precipitation nowcasting: A benchmark and a new model, Advances in neural information processing systems, 30, 2017.~~

~~Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., and Bolton, A.: Mastering the game of go without human knowledge, nature, 550, 354-359, 2017.~~

~~Suliman, A. H. A., Awchi, T. A., Al-Mola, M., and Shahid, S.: Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq, Atmospheric Research, 242, 105007, 2020.~~

~~Sun, A. Y. and Tang, G.: Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets, Frontiers in Water, 2, 536743, 2020.~~

~~Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Reviews of Geophysics, 56, 79-107, 2018.~~

~~Tao, Y., Gao, X., Ihler, A., Hsu, K., and Sorooshian, S.: Deep neural networks for precipitation estimation from remotely sensed information, 2016 IEEE Congress on Evolutionary Computation (CEC), 1349-1355,~~

~~Tegegne, G. and Melesse, A. M.: Comparison of Trend Preserving Statistical Downscaling Algorithms Toward an Improved Precipitation Extremes Projection in the Headwaters of Blue Nile River in Ethiopia, Environmental Processes, 8, 59-75, 2021.~~

~~Thrasher, B., Maurer, E. P., McKellar, C., and Duffy, P. B.: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrology and Earth System Sciences, 16, 3309-3314, 2012.~~

~~Tong, K., Su, F., Yang, D., and Hao, Z.: Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau, Journal of hydrology, 519, 423-437, 2014.~~

~~Tong, Y., Gao, X., Han, Z., Xu, Y., Xu, Y., and Giorgi, F.: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods, Climate Dynamics, 57, 1425-1443, 2021.~~

~~Trinh, T., Do, N., Nguyen, V., and Carr, K.: Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam, Climate Dynamics, 57, 2713-2735, 2021.~~

~~Tripathi, S., Srinivas, V., and Nanjundiah, R. S.: Downscaling of precipitation for climate change scenarios: a support vector machine approach, Journal of hydrology, 330, 621-640, 2006.~~

~~Vandal, T., Kodra, E., and Ganguly, A. R.: Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation, Theoretical and Applied Climatology, 137, 557-570, 2019.~~

~~Vandal, T., Kodra, E., Dy, J., Ganguly, S., Nemani, R., and Ganguly, A. R.: Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2377-2386,~~

~~Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: Generating high resolution climate change projections through single image super-resolution: An abridged version, International Joint Conferences on Artificial Intelligence Organization,~~
~~Wang, F. and Tian, D.: On deep learning-based bias correction and downscaling of multiple climate models simulations, Climate Dynamics, 1-18, 2022.~~
~~Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J.: Deep learning for daily precipitation and temperature downscaling, Water Resources Research, 57, e2020WR029308, 2021.~~
~~Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, Journal of Geophysical Research: Atmospheres, 107, ACL 6-1-ACL 6-15, 2002.~~
~~Xu, H., Xu, C.-Y., Chen, S., and Chen, H.: Similarity and difference of global reanalysis datasets (WFD and APHRODITE) in driving lumped and distributed hydrological models in a humid region of China, Journal of Hydrology, 542, 343-356, 2016.~~
~~Xu, M., Liu, Q., Sha, D., Yu, M., Duffy, D. Q., Putman, W. M., Carroll, M., Lee, T., and Yang, C.: PreciPatch: A dictionary-based precipitation downscaling method, Remote Sensing, 12, 1030, 2020.~~
~~Xu, X., Frey, S. K., and Ma, D.: Hydrological performance of ERA5 and MERRA-2 precipitation products over the Great Lakes Basin, Journal of Hydrology: Regional Studies, 39, 100982, 2022.~~
~~Xu, X., Frey, S. K., Boluwade, A., Erler, A. R., Khader, O., Lapen, D. R., and Sudicky, E.: Evaluation of variability among different precipitation products in the Northern Great Plains, Journal of Hydrology: Regional Studies, 24, 100608, 2019.~~
~~Yilmaz, K. K., Hogue, T. S., Hsu, K.-l., Sorooshian, S., Gupta, H. V., and Wagener, T.: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, Journal of Hydrometeorology, 6, 497-517, 2005.~~
~~Zhang, X., Anagnostou, E. N., and Schwartz, C. S.: NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS, Remote Sensing, 10, 642, 2018.~~
~~Zhong, R., Chen, X., Lai, C., Wang, Z., Lian, Y., Yu, H., and Wu, X.: Drought monitoring utility of satellite-based precipitation products across mainland China, Journal of hydrology, 568, 343-359, 2019.~~

Page 37: [1] Formatted Table **Fang Wang** **11/1/22 3:12:00 PM**

Formatted Table

Page 37: [2] Formatted **Fang Wang** **11/1/22 3:10:00 PM**

Font: 10 pt

Page 37: [3] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [4] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [5] Formatted **Fang Wang** **11/1/22 3:10:00 PM**

Font: 10 pt

Page 37: [6] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [7] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [8] Formatted **Fang Wang** **11/1/22 3:10:00 PM**

Font: 10 pt

Page 37: [9] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [10] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [11] Formatted **Fang Wang** **11/1/22 3:10:00 PM**

Font: 10 pt

Page 37: [12] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [13] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [14] Formatted **Fang Wang** **11/1/22 3:10:00 PM**

Font: 10 pt

Page 37: [15] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [16] Formatted **Fang Wang** **11/1/22 3:09:00 PM**

Centered

Page 37: [18] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [19] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [20] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [21] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [22] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [23] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [24] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [25] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [26] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [27] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [28] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [29] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [30] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [31] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [32] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [33] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Centered

Page 37: [35] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [36] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [37] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [38] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [39] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [40] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [41] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [42] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [43] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [44] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [45] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [46] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [47] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [48] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [49] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [51] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [52] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [53] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [54] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [55] Formatted Table Fang Wang 11/1/22 3:12:00 PM

Formatted Table

Page 37: [56] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [57] Formatted Fang Wang 11/1/22 3:10:00 PM

Font: 10 pt

Page 37: [58] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered

Page 37: [59] Formatted Fang Wang 11/1/22 3:09:00 PM

Centered