UFNet: Joint U-Net and fully connected neural network to bias correct precipitation predictions from climate models

Shuang Yu, Indrasis Chakraborty, Gemma J. Anderson, Donald D. Lucas, Yannic Lops, Daniel Galea

a Lawrence Livermore National Laboratory, Livermore, CA, USA

Corresponding author: Gemma J. Anderson, anderson276@llnl.gov, Shuang Yu, yu42@llnl.gov

File generated with AMS Word template 2.0

Early Online Release: This preliminary version has been accepted for publication in Artificial Intelligence for the Earth Systems, may be fully cited, and has been assigned DOI 10.1175/AIES-D-23-0076.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2024 American Meteorological Society. This is an Author Accepted Manuscript distributed under the terms of the default AMS reuse license. For information regarding reuse and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
ABSTRACT

Precipitation values produced by climate models are biased due to the parameterization of physical processes and limited spatial resolution. Current bias correction approaches usually focus on correcting lower-order statistics (mean, standard deviation), which make it difficult to capture precipitation extremes. However, accurate modeling of extremes is critical for policymaking to mitigate and adapt to the effects of climate change. We develop a deep learning framework, leveraging information from key dynamical variables impacting precipitation to also match higher-order statistics (skewness and kurtosis) for the entire precipitation distribution, including extremes. The deep learning framework consists of a two-part architecture: a U-Net convolutional network to capture the spatiotemporal distribution of precipitation and a fully connected network to capture the distribution of higher-order statistics. The joint network, termed UFNet, can simultaneously improve the spatial structure of the modeled precipitation and capture the distribution of extreme precipitation values. Using climate model simulation data and observations that are climatologically similar but not strictly paired, the UFNet identifies and corrects the climate model biases, significantly improving the estimation of daily precipitation as measured by a broad range of spatiotemporal statistics. In particular, UFNet significantly improves the underestimation of extreme precipitation values seen with current bias-correction methods. Our approach constitutes a generalized framework for correcting other climate model variables which improves the accuracy of the climate model predictions, while utilizing a simpler and more stable training process.

1. Introduction

Accurate climate modeling is crucial for understanding climate change and implementing effective mitigation and adaptation strategies. General circulation models (GCMs) are one of the most important tools scientists use to predict weather and future climate change (Maraun et al., 2016). However, the accuracy of GCMs is often compromised by biases due to the parameterization and misrepresentation of physical processes, and coarse grid resolution, which does not explicitly resolve convective precipitation with more intense precipitation rates (Zelinka et al., 2020; Maraun, 2016; Flato et al., 2014; Christensen et al., 2008). Therefore, many users of GCM predictions leverage bias correction techniques for their applications.

The bias correction of GCM outputs is typically performed using one of two methods. In the first technique, the climate models are run in “weather forecast mode” with initial data from
multiple numerical weather prediction (NWP) center analyses or reanalysis, which helps identify and assess the climate errors development, and facilitates the comparison to detailed observation processes. (Ma et al., 2014; Xie et al., 2012; Phillips et al., 2004). This approach is mainly used for shorter time scales, such as short-range forecasts of precipitation, clouds, and convection, and is usually not utilized on longer time scales (Smith, 2001). Another technique is to reduce and correct the biases of GCM outputs through post-processing and statistical analysis. This technique finds statistical relationships between historical GCM simulations and climate observations, and then applies these relationships to reduce biases in the simulation data. Statistical bias correction methods directly improve the accuracy of GCM predictions with low computational cost and have been extensively used in climate science, including studies in climate impact analysis (Li et al., 2010; Wood et al., 2004) and climate-related decision making (Li et al., 2015; Teutschbein & Seibert, 2012).

Precipitation is one of the most important meteorological variables in the climate system. In 2022, precipitation-related extreme climate events, such as floods and droughts, caused more than 24 billion U.S. dollars in economic losses and 178 human casualties in the U.S. (NOAA, 2023). Therefore, accurate precipitation predictions are critical for policymaking to mitigate the impact of extreme weather events. However, this remains a challenge, as precipitation involves complex physical processes at multiple scales, from microscopic interaction of droplets in clouds and atmospheric convection to the large-scale atmospheric circulation. In addition, there are difficulties in simulating the spatial distribution of precipitation, especially precipitation extremes, which are crucial when modeling the occurrence of meteorological disasters (Fulton et al., 2023; Gimeno et al., 2022).

A popular statistical post-processing method for the bias correction of GCM outputs is quantile mapping (QM) (Maraun et al., 2016; Ehret et al., 2012). This is a simple statistical method which maps the one-dimensional cumulative distribution function (CDF) of the simulated precipitation to that of the observed precipitation for a given historical period (Cannon et al., 2015). QM often shows good results in correcting the distribution bias for a given grid point. However, QM does not consider the spatial correlations between neighboring grid points, meaning that it is difficult to use QM to improve the spatial structure of the GCM output (Tong et al., 2021; Gudmundsson et al., 2012; Déqué et al., 2007). For large-scale precipitation events, such as tropical or midlatitude cyclones, QM cannot both bias correct and coherently retain the spatial patterns. In addition, QM assumes that the distribution of
precipitation is non-variant (i.e., stationary) over the period over which the CDF is calculated and applied. Thus, QM by itself cannot capture climate change, and it is difficult to calibrate the bias of future climate predictions (Hess et al., 2022).

Machine learning methods for image-to-image translation provide a new direction for improving the performance of GCM output. Example applications include downscaling (Miao et al., 2019; Pan et al., 2019), forecasting (Pan et al., 2020; Weyn et al., 2019), and climate signal identification (Barnes et al., 2019). A recent review of machine learning applications to climate variability and weather is provided in Molina et al. (2023). Most of the applications above adopt a supervised learning problem setting and minimize a pixel-wise distance measure as a regression task. The limitation of this approach is that it is difficult to train powerful data-driven models with limited GCM simulation and observation data that are paired in space and time, as satellite precipitation observations are only available from 1979 (Xie et al., 2010). Moreover, free running GCMs without nudging can capture important climate features and patterns but are not specifically designed to predict weather on individual days. By relaxing the requirement of paired data, GCM biases can also be learned from a larger amount of historical simulation and observation data that are climatologically similar using unpaired image-to-image translation networks. Francois et al. (2021) applied a cycle-consistent generative adversarial network (CycleGAN; Zhu et al., 2017) to bias correct winter temperature and precipitation over Paris. Based on the same CycleGAN architecture, Pan et al. (2021) bias corrected precipitation over the contiguous United States (CONUS). Compared with the previous bias correction methods which only use single fields, Pan et al. (2021) also incorporated dynamical variables affecting precipitation, such as 500 hPa geopotential height and specific humidity. The use of additional fields helps provide dynamical consistency to the bias correction results. Hess et al. (2022) also used a CycleGAN architecture to build a bias correction method for GCM global precipitation data. They found that a CycleGAN is superior in modeling the intermittency of precipitation compared to QM, and described the shortcomings of CycleGANs in capturing non-stationarity in the bias correction of future precipitation predictions. More recently, Fulton et al. (2023) used the Unsupervised Image-to-Image Translation (UNIT), a neural network that incorporates the architecture of GANs, to improve modeling of the South Asian monsoon using five variables. Although the UNIT neural network reduces the dispersion of the data, it was not sufficient by itself for bias correcting data. They also applied QM post-processing to correct the under dispersion of precipitation produced by the UNIT neural network. However, this makes the results reliant on the QM bias
correction, which does not predict corrections for climate in the future. Furthermore, there are some general problems with GAN-based methods when applied to bias correction. GANs are notoriously difficult to train because they require generators and discriminators to compete constantly, making training unstable and slow. They also have issues with robustness (Arjovsky et al., 2017), which can often make their predictions unreliable. In addition, GANs usually require a large amount of training data, which is not easily available in long-term climate simulations (Salimans et al., 2016).

To enable our bias correction method to capture multi-scale spatial patterns, we choose an convolution-based U-Net (Ronneberger et al., 2015) architecture. U-Net is a convolution-based neural network containing encoding and decoding components that has the advantage of being stable, more robust, and easier to train than a GAN. Three dynamical variables, geopotential height, surface specific humidity and surface level pressure, which affect precipitation were used as inputs to the U-Net. To improve the performance of higher order moments of precipitation, we trained a separate fully connected deep neural network (DNN) to correct the temporal distribution of the precipitation. Finally, this DNN is nested into the U-Net architecture to produce the UFNet, which is an end-to-end bias correction method for GCM precipitation predictions. In this study, we use the Energy Exascale Earth System Model as the GCM simulation and use an observational dataset as the ground truth. We compare the UFNet methodology with QM to evaluate the performance of the precipitation bias correction, especially for the extreme precipitation events. The data and methods are described in Section 2; the results are presented in Section 3, and main conclusions and discussion are given in Section 4.

2. Data and Method

2.1. Data

In this study, we develop a deep learning precipitation bias correction method for outputs of historical climate simulations from the first version of the Energy Exascale Earth System Model (E3SM, Golaz et al., 2019). The E3SM components include atmosphere and land (110-km grid spacing), ocean and sea ice (60 km in the midlatitudes and 30 km at the equator and poles), and river transport (55 km) models. We use simulated daily averaged surface level pressure (SLP), geopotential height at 500 hPa (Z500), and surface specific humidity (Q) as input features. These features are dynamical field variables that are correlated with
precipitation in the climate system and help to improve the accuracy of the bias correction of precipitation. The U.S. National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) unified gauge-based analysis of daily precipitation (0.25°, Xie et al., 2010) during 1950-2010 is used as ground truth. We limit the geographical extent for precipitation to the area over the contiguous United States (CONUS), which is bounded by 25°N-50°N and 65°W-125°W. Observed precipitation of grid points over the ocean are all set to zero. To better capture spatially separated connections between dynamics and precipitation, we also add another 5° of spatial padding around the dynamical variables, which are hence bounded by 20°N-55°N and 60°W-130°W. The spatial resolution of E3SM and observed datasets are both regridded to a common 1-degree resolution grid using conservative interpolation, because E3SM is at coarser resolution than CPC.

2.2. **UFNet Model Configuration**

The bias correction method, UFNet, consists of two parts: a U-Net architecture and a fully-connected network architecture. The U-Net is used to capture the multi-scale spatial structure of precipitation while the fully-connected network is trained to correct the statistical distribution of bias corrected precipitation (Fig. 1), like the QM process.

First, we train the fully-connected network (DNN in Fig. 1) to capture the cumulative distribution function (CDF) of the simulated precipitation. During 1950-1994 (our training dataset), at each grid point, we randomly sample 300 values (which we chose to correspond with the batch size of the U-Net training described below) of simulated daily average precipitation and calculate a CDF. We then partition each CDF into 500 bins, which are used as input features to the DNN. Due to the non-Gaussian nature of precipitation data, the CDF of precipitation changes sharply at low values between 0-2 mm/day. Using a CDF with 500 bins, which has more bins than samples, resolves small precipitation values and improves predictions because the cumulative probability changes are relatively smooth. To account for spatial variations, the latitude, longitude, and maximum observed precipitation of each grid cell are also included in the feature vector, resulting in a total of 503 input features during DNN training. The target values for the DNN are derived by computing the CDF from observed precipitation values that are seasonally and climatologically consistent with the simulations. Following Pan et al. (2021), for a given simulation precipitation input belonging to temporal time step $T$ (with year $T_y$ and day of year $T_d$), we assume that observations will be seasonally and climatologically consistent if they are drawn randomly from $T_d \pm 15$ and $T_y \pm 5$. To
estimate robust CDFs, this random selection procedure was repeated at each grid point 1000 times via bootstrap resampling with different random seeds, thus obtaining $724 \times 1000 = 724000$ samples (724 land grid points are present in the research region). 70% of these samples are used as the training set and 30% as the test set. The DNN has 4 fully-connected layers. We use the Adam optimizer (Kingma & Ba, 2014) with a batch size of 32, learning rate of $10^{-3}$, and train the DNN for 100 epochs, which is sufficient to achieve convergence. The loss function is calculated using two terms as:

$$L = \frac{1}{\alpha} \sum_{i=1}^{\alpha} (y_i - \hat{y}_i)^2 + \frac{1}{\beta} \sum_{i=1}^{\beta} (y_i - \hat{y}_i)^2$$  \hspace{1cm} \text{eq.1}$$

where $y$ and $\hat{y}$ are the precipitation CDFs from observations and predictions, respectively; the first term in the loss function is a summation over the full CDF ($\alpha = 500$); the second term is a summation over low percentile bins ($\beta = 120$) to better capture the probability dependence of the lowest precipitation values of the CDF. After 100 epochs of training, the DNN is saved to be used in UFNet.

Next, we build a U-Net architecture to capture the spatial distribution of precipitation. The U-Net uses convolutions and pooling layers to perform large-scale feature extraction and skip-connections to preserve small-scale, high-frequency information. For training the U-Net, we choose six channels (SLP, Z500, Q for the current and the previous days) with a spatial map size of $36 \times 72$ as training features. Similar to the DNN described above, the U-Net model was trained using resampled daily data as follows: input features belonging to temporal time stamp $T$ (with year $T_y$ and day of year $T_d$) are assumed to be consistent with an observation at day $T_d \pm 15$ and year $T_y \pm 5$. The year and day values are selected from the available dates within the training data. When nearing the edges of training years, such as 1994, we selected a random value from altered intervals, $T_y - 5$ on the lower side and $T_y$ on the upper side, to avoid selecting a sample outside of the training period. Thus, the input feature and the target observation precipitation belong to the same climate window, rather than the same day. We split the resampled data into training and test sets. The training set covers the period of 1950–1994 (16,435 days) while data from 1995 to 2010 (5,840 days) are held out as a test set.

To train the UFNet, we combine the pre-trained DNN with the U-Net. The weights of U-Net and pre-trained DNN are updated in this step. For the U-Net architecture, there are six input channels. The output of the U-Net is referred to as Output1 (in Fig. 1). Then we calculate the CDF of Output1 (referred to as CDF1). In addition, we calculate the CDF of the GCM
precipitation data and combine it with the latitude, longitude, and the maximum observed precipitation of the grid point across the training dataset to build the input for the trained DNN, which is used to infer the predicted precipitation CDF (referred to as CDF2). We calculate the final output (referred to as Output2) of the UFNet as follows:

$$Output2 = CDF2^{-1}(CDF1(Output1)) \quad \text{eq.2}$$

Here Output2 is the bias corrected precipitation output from UFNet, $CDF2^{-1}$ is the inverse of the empirical cumulative distribution function which was calculated based on the output of the DNN. The joint loss function, based on Output1 and Output2, is minimized:

$$L = MSE(y - Output1) + 5 \times MSE(y_{mean} - Output1_{mean}) + 5 \times MSE(y_{std} - Output1_{std}) + MSE(y_{mean} - Output2_{mean}) \quad \text{eq.3}$$

where $y$ and Output2 are the precipitation from the observation and prediction respectively, Output1 is the output from the U-Net architecture, $y_{mean}$ and $y_{std}$ are mean and standard deviation values of the observed precipitation, and $Output1_{mean}$ and $Output1_{std}$ are the mean and standard deviation values of the U-Net output. The coefficients of “5” that appear in eq.3 are used to adjust the weights of the components so that they are balanced in the loss. After 300 epochs of training, the UFNet converges until the loss value stabilizes in the training and test set.

Fig. 1. Architecture of the UFNet model.

2.3. Quantile Mapping

File generated with AMS Word template 2.0
For this work, we use quantile mapping as the baseline to compare the performance of UFNet. Quantile mapping is a univariate correction technique that maps the quantiles of historical climate simulations to the quantiles of corresponding observations (Mamalakis et al., 2017; Michelangeli et al., 2009). This is performed at each grid point using:

\[ x_c = CDF_Y^{-1}(CDF_X(x)) \]

where, \( x \) and \( x_c \) are the values at a grid point before and after correction by QM; \( CDF_Y^{-1} \) is the inverse of the empirical cumulative distribution function of the observations, \( CDF_X \) is the empirical cumulative distribution function of the original simulation. Both \( CDF_Y^{-1} \) and \( CDF_X \) are built on the same training dataset as the UFNet (i.e., over 1950-1994). Its performance is evaluated on the held-out test set.

2.4. Performance Evaluation

We employ several popular statistical indices (see Table 1) to evaluate the performance of our proposed UFNet. We use the spatial correlation (r) and spatially-averaged root mean square error (RMSE) between observations and simulation for these indices to evaluate the GCM simulation performance before and after bias correction.

\[
\rho = \frac{\sum_{i,j} (V_{i,j}^y - \bar{V}^y)(V_{i,j}^\hat{y} - \bar{V}^\hat{y})}{\sqrt{\sum_{i,j} (V_{i,j}^y - \bar{V}^y)^2} \sqrt{\sum_{i,j} (V_{i,j}^\hat{y} - \bar{V}^\hat{y})^2}}
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i,j} (V_{i,j}^y - V_{i,j}^\hat{y})}
\]

Here \( V_{i,j}^y \) and \( V_{i,j}^\hat{y} \) are the considered indices for simulation and observation at grid (i, j) and \( \bar{V}^y \) and \( \bar{V}^\hat{y} \) are the spatial average of the statistics, where N is the total number of grid points.

Table 1. Indices for Assessing the Bias Correction Performance. E shows the expected value.

<table>
<thead>
<tr>
<th>scope</th>
<th>statistic</th>
<th>equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment</td>
<td>Mean (µ)</td>
<td>( E_{Y\sim Y}(y) )</td>
</tr>
</tbody>
</table>

File generated with AMS Word template 2.0
Standard deviation ($\sigma$) 
\[ \sqrt{E_{y \sim pY} (y - m_Y)^2} \]

Skewness (Skew) 
\[ E_{y \sim pY} \left( \frac{y - m_Y}{s_Y} \right)^3 \]

Kurtosis (Kurt) 
\[ E_{y \sim pY} \left( \frac{y - m_Y}{s_Y} \right)^4 \]

Quantile 
1-3 tertile ($Q_{33\%,66\%,99\%}$) 
33%, 66%, 99% quantile

Frequency 
Probability of precipitation (PoP) 
\[ \frac{Total \ precipitation \ days}{Total \ days} \]

Intensity 
Average intensity 
\[ E_{Y > 0} (y) \]
1-day max 
max(y)
3-day max 
\[ \max \left( \sum_{i} y_i \right) \]
5-day max 
\[ \max \left( \sum_{i} y_i \right) \]

Seasonality 
Winter precipitation ratio 
\[ \frac{Dec - Feb \ precipitation}{Total \ precipitation} \]
Summer precipitation ratio 
\[ \frac{Jun - Aug \ precipitation}{Total \ precipitation} \]
3. Results

3.1. Correction of spatial pattern of precipitation.

We evaluate the UFNet by examining the extent that the distribution of the bias corrected precipitation matches the distribution of ground truth observations. We compare the UFNet results with QM-based post-processing, as well as the U-Net only bias correction methodology (Output1). When comparing the mean error of the spatial precipitation field from E3SM with observations, large mean errors are evident, especially in the western half of the region where overestimations can be seen (Fig. 2a, e). All the methods, namely U-Net, UFNet, and QM, corrected this overestimation. For a more detailed grid-scale evaluation, we considered the spatial distribution of the first to fourth order moments of the precipitation distribution (mean, standard deviation, skewness, and kurtosis, Fig. 3a-d). The results show that E3SM overestimates the mean precipitation (second column of Fig. 3a) as well as the variance in the western region (second column of Fig. 3c). One possible reason is that the orographic lifting and shadowing impacts from the narrow mountain ranges produce localized precipitation which is a known difficulty of coarse resolution GCM simulations (Gershunov et al., 2019; Kapnick et al., 2018). Also, for the skewness and kurtosis, E3SM produces a significant underestimation especially in the northwestern and eastern regions (second column of Fig. 3c and 3d).

The U-Net significantly corrects the mean and standard deviation of precipitation (Fig. 2b and 2e; third column of Fig. 3a and 3b) and improves the overestimation of the GCM simulated precipitation in the central CONUS. However, the U-Net shows significant underestimation of skewness and kurtosis (third column of Fig. 3c and 3d), implying that a U-Net alone cannot adequately capture the extreme tails of the ground truth distribution. For the mean and standard deviation, UFNet achieves considerable improvement of precipitation in the western mountains, especially in the Sierra Nevada region. In addition, UFNet corrects the precipitation distribution in inland areas, which improves the overestimation of precipitation that is seen in the E3SM simulations. Notably, UFNet provides a significant improvement in the correction of precipitation kurtosis and skewness, with a spatial correlation coefficient for skewness ($r = 0.76$), which far exceeds that of E3SM ($r = 0.65$), and the QM bias corrector ($r = 0.61$).
Fig. 2. Spatial precipitation statistics. Spatial distribution of mean errors with respect to the observational data for (a) E3SM simulation, and the bias corrected E3SM simulations using the (b) U-Net, (c) UFNet, and (d) QM methods. Precipitation rates were averaged over time. (e) Precipitation averaged over time and latitude.
Fig. 3. Bias correction results for the first to fourth order moments (rows a–d). Each column shows the spatial distributions of a considered moment from observation, E3SM simulation, U-Net bias corrector only, UFNet bias corrector and QM bias corrector. The spatial correlation coefficient and spatial average root mean square error between bias corrected precipitation and observation moments are calculated and denoted.

We also performed an analysis on the absolute values of precipitation obtained from the bias correction techniques. The results show that all the bias correction methods significantly underestimate low precipitation (< 1mm/day) (Fig. 4a). This is clearly seen for the U-Net and UFNet which assign no precipitation for observations of up to 1 mm/day. For the 66% quantile (Q66), the U-Net and UFNet simulate most of the precipitation better in the Rockies and Appalachians, but significantly underestimate precipitation in the central and southwestern CONUS (lack of blue shading in third and fourth columns of Fig. 4b). In the whole region, the UFNet improves the correlation coefficient, r, of the spatial distribution of Q66 by reducing the overestimation of the simulated precipitation at most grid points. For the 99% quantile (Q99), the UFNet shows excellent results (fourth column of Fig. 4c). It improves the underestimation of E3SM precipitation in the southeast region, showing better spatial correlation and lower bias. The U-Net captures the spatial variability well, but the values show a significant underestimation, especially in the central region, which reflects the shortcomings of the U-Net structure in capturing the tails of the precipitation distribution (third column of Fig. 4c). The DNN component within the UFNet improves this. The UFNet also improves the simulation precipitation frequency compared with E3SM, especially improving the
overestimation of precipitation in the Rocky Mountain region. QM slightly outperforms the UFNet in correcting Q66 and Q99, as can be noted by the higher correlation values and lower RMSE in the fourth and fifth columns of Fig. 4b and 4c. These are reasonable results as QM empirically aligns the simulated precipitation distribution to match the observed precipitation distribution for each grid cell.

Fig. 4. Similar to Fig. 3 but for correcting the 33%, 66%, 99% quantile (a–c), and the probability of precipitation (d).

To evaluate the regional differences for the precipitation distribution by the UFNet, we calculated the probability density function of precipitation for different states in the US (see Fig. 5). The results show that in all regions of the continental US, the UFNet significantly improves the results of the distribution of higher-order moments of precipitation, especially in the tails of the precipitation distribution, corresponding to extreme values. For the tail of the precipitation distribution, the results of UFNet show significant regional differences. For central regions such as Colorado, Nebraska, Arizona and Oklahoma, the UFNet shows improvement when comparing the bias corrected distribution of precipitation to that of the original E3SM simulation. However, for coastal regions, such as Oregon, Washington and Idaho, UFNet poorly characterizes the tail of the precipitation distribution. UFNet assigns no precipitation to low precipitation events (< 1mm/day), and underestimates the precipitation for events having an observed 1-2 mm/day of precipitation. Thus, the percentage of the number of days with less than 1mm/day of observed precipitation is overestimated.
3.2. Correction of temporal patterns of precipitation.

We also investigated how the bias-corrected precipitation predictions performed in a temporal sense, especially for the seasonality and annual cycle characteristics. As can be seen in Fig. 6, E3SM presents a significant overestimation of precipitation across the testing period. This is also consistent with the results shown earlier in Fig. 3a. As well as this overestimation, it could be noted that the timeseries obtained from E3SM data and observations do not track well and the seasonality changes characteristic is mismatched, with significant differences seen especially around 1996, 2000 and 2006. To quantify these differences, a correlation coefficient of 0.08 was calculated between these two timeseries, showing that they are not significantly correlated at the $\alpha = 0.05$ level. The UFNet model significantly improves the performance of E3SM for seasonal precipitation simulation, which can be quantified by the calculated correlation coefficient of 0.35. However, precipitation was still underestimated during the boreal winter, especially before 1998. As shown in the fourth column of Fig. 7a, the main area of underestimation is the central part of the continental United States, while the bias-corrected results show reasonable results for topographically complex regions, such as the western mountains. For summer precipitation, the UFNet correction results show a significant overestimation, especially in 2001, 2005 and 2006. The main areas of overestimation are the...
central region of the US, which usually experience hot dry spells during summer, and Eastern seaboard of the US, where landfall of tropical cyclones occurs during this time.

Fig. 6. Monthly changes of regional mean precipitation during 1995-2010 from observations (red), E3SM simulation (blue) and the UFNet bias corrector (green).

Fig. 7. Winter (a, December to February) and Summer (b, June to August) precipitation ratio comparison.

3.3. Bias correction of extreme precipitation.

We also carried out an evaluation of precipitation intensity and extreme precipitation. For this, we compared the mean intensity (Fig. 8a) and extreme intensity of precipitation in the forms of the 1-day maximum, 3-day maximum and 5-day maximum precipitation (Fig. 8b-d). The results show that the U-Net can capture the spatial variability of precipitation intensity and extreme precipitation well, however a significant underestimation in the simulation of extreme precipitation can be seen, especially in the southeastern part of the US. The UFNet achieves considerable improvement, particularly for the west coast and southeastern regions. A possible
The reason for this improvement is that although extreme precipitation events at a specific geographical point are rare events, they are more common when viewed from a broader, that is, continental, spatial scope. As such, the additional correction provided by the DNN, which captures the precipitation CDF over the whole area, is shown to be useful. Additionally, the U-Net itself works poorly for capturing the tails of the distribution for the target variable so we use the DNN to better capture the tails of the target variable so that the UFNet can more accurately generate the tails of distribution which corresponds to extreme precipitation. Although the QM algorithm also shows good bias correction for extreme precipitation, it still does not perform as well as UFNet, showing lower spatial correlation and higher RMSE for every extreme precipitation index.

![Image](image_url)

**Fig. 8.** Comparison between (a) average precipitation intensity, (b-d) 1-day, 3-day, and 5-day maximum precipitation.

In addition to the extreme intensity of precipitation analysis, we study the performance of UFNet bias correction techniques for extreme events of societal relevance in the research region. One possible use for modelling precipitation is for informing policymakers about the state of the future climate. Thus, accurately simulating the possibility of flooding or flash drought is important for the agriculture sector and food security in the region. For this, we use the standardized precipitation index, which is commonly used to monitor drought and flood hazards (Chandrasekara et al. 2021; Tirivarombo et al. 2018; Aadhar and Mishra 2017). In each grid, this index is defined by:

---

File generated with AMS Word template 2.0
where \( P_T \) is the mean precipitation over a time period length \( T \), \( P_* \) is the mean value of all \( P_T \) across the dataset, and \( \sigma_T \) is the standard deviation of the precipitation over a time period length \( T \). We use a period of thirty days as the time period \( T \), which is suitable to represent medium-term extremes in precipitation, relevant for both flooding and sub-seasonal flash drought (Christian et al. 2021; Mishra et al. 2021). We then produce a boxplot of all SPI values for all possible 30-day periods for different US states. Fig. 9 shows these plots for the states of Arizona and Georgia. For Arizona, the E3SM simulation underestimates the distribution of both dry periods (SPI < 0) and wet periods (SPI > 0) while the U-Net produces an even tighter distribution. The QM method shows minor improvement compared with the E3SM simulation, which also significantly underestimates the distribution of SPI for dry periods (SPI<0). The UFNet performed the best, correcting for the underestimation of simulated drought periods and the distribution of the SPI values was closer to the observed values. For regions with high precipitation such as Georgia, the E3SM simulation of the SPI distribution is excellent, and all bias-corrected methods present acceptable results close to observations. Thus, UFNet improves the SPI for both arid and humid regions.

![Fig. 9. Boxplots of the SPI at grid points contained in Arizona and Georgia during the wettest 30-day period each year for the five different datasets. We note that the data points in this plot originate from a 30-day rolling time window instead of selecting independent 30-day periods. The star symbol is the maximum value in each boxplot.](image)

4. Conclusion and Discussion
The accuracy of climate projections is hindered by biases present in GCMs due to their simplification or misrepresentation of unresolved climate processes. To improve the accuracy of climate simulations, we develop a deep learning approach combining a U-Net and DNN to correct climate simulation biases using samples that associate climate simulations with climatologically and seasonally consistent observations.

We find that the U-Net itself is insufficient to adequately correct biases in E3SM simulated precipitation data because it reduces the dispersion of the data, leading to significant underestimation at the extremes. However, the U-Net captures the spatial characteristics of the data well. In order to improve the effectiveness of the U-Net in capturing the distribution of precipitation data, we trained a DNN whose purpose is to capture the distribution of precipitation. In this study, we combine the U-Net and DNN to form a combined framework, termed UFNet. Results for UFNet demonstrate that it bias corrects the spatial distribution of precipitation and represents extreme precipitation well. Notably, Fulton et al. (2023) developed a bias correction method using the same premise, which also showed promising results. They used QM to post-process the data from a deep learning-based bias correcting method to improve the problem of insufficient simulation of precipitation dispersion encountered with their method. However, this post-processing method is not involved in training and cannot circumvent the inherent problems of the QM method, such as the inability to calibrate spatial correlation. Instead of using a post-processing approach, this study captures the distribution of precipitation by training an additional DNN and training an end-to-end deep learning bias-correction framework. Therefore, the resulting method produces an improved distribution of precipitation while avoiding the problems associated with QM post-processing.

The main contribution of this method is to combine two different deep learning architectures to effectively simulate highly realistic precipitation fields. Compared to current bias correction methods based on adversarial neural network architectures, UFNet is much easier to train and produces stable training results, which is important for the flexibility and repeatability of applications. In addition, unlike many other bias-correction methods that take GCM simulation precipitation as the only input, the inputs of UFNet are precipitation-related meteorological fields that make our results more dynamically consistent. Finally, UFNet allows for accurate correction of extreme precipitation data, which is essential for the study and risk assessment of extreme events. However, UFNet also shows limitations, especially for the underestimation of low precipitation grid points, and the underestimation of winter
precipitation when simulating seasonal changes. Future work will explore the introduction of physical constraints to make the bias correction results reliable for future climate.

Acknowledgments.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, was supported by the LLNL-LDRD Program under Project No. 22-SI-008, and it released as LLNL-JRNL-853831.

Data Availability Statement.

E3SM output data are accessible directly through the DOE Earth System Grid Federation (https://esgf-node.llnl.gov/projects/e3sm). The U.S. National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) is accessible through https://www.cpc.ncep.noaa.gov/.

REFERENCES


Flato G., Marotzke J., Abiodun B., et al. 2014: Evaluation of climate models//Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment


