Improving Seasonal Forecast of Summer Precipitation in Southeastern China using CycleGAN Deep Learning Bias Correction

Song Yang[1], Fenghua Ling[1], Lei Bai[2], Jing-Jia Luo[1*]

[1]Institute of Climate Application Research (ICAR)/CIC-FEMD/KLME/ILCEC, Nanjing University of Information Science and Technology, Nanjing 210044, China.
[2]Shanghai AI Laboratory, Shanghai, China.

[*]Corresponding author: Jing-Jia Luo (jingjia_luo@hotmail.com; jjluo@nuist.edu.cn)

Key Points:

- We use a CycleGAN-based deep learning model to correct NUIST-CFS1.0 climate model's forecasts of summer precipitation in southeastern China.
- The bias-corrected results of the CycleGAN model show a big reduction in the forecast error compared to traditional quantile mapping (QM).
- The CycleGAN model significantly enhances forecast accuracy for extreme precipitation events, including frequency, intensity, and duration.

Abstract

Accurate seasonal precipitation forecasts, especially for extreme events, are crucial to preventing meteorological hazards and its potential impacts on national development, social stability, and security. However, the intensity of summer precipitation is often significantly underestimated in many current dynamical models. This study uses a deep learning method called Cycle-Consistent Generative Adversarial Networks (CycleGAN) to enhance the seasonal forecast skill of the Nanjing University of Information Science & Technology Climate Forecast System (NUIST-CFS1.0) in predicting June-July-August precipitation in southeastern China. The results suggest that the CycleGAN-based model significantly improves the accuracy in predicting the spatial-temporal distribution of summer precipitation than traditional quantile mapping (QM) method. Due to the use of unpaired day-to-day correction models, we can pay more attention to the frequency, intensity, and duration of extreme precipitation events in the climate dynamical model forecast. This study expands the potential applications of deep learning models to improving seasonal precipitation forecasts.

Plain Language Summary

Despite significant efforts made in recent decades to enhance the seasonal precipitation forecast skills, accurately predicting summer precipitation over southeastern China and even in East Asia remains a great challenge. Currently, the statistical methods can be used as a cost-effective and efficient way to improve forecast accuracy by correcting model biases. In this study, we utilized a deep learning approach called CycleGAN to improve the seasonal prediction of the NUIST-CFS1.0 dynamical model system, specifically targeting summer precipitation over southeastern China. Of particular importance, the CycleGAN model can effectively improve the dynamical model's performances in predicting extreme precipitation events.

1 Introduction

Southeastern China, located in the East Asian monsoon region, experiences complex and diverse climate due to unique topography and land-sea distribution (Wang & Lin, 2002; Wu et al., 2003; Ding & Chan, 2005; Huang et al., 2007; Allan & Soden, 2008; Zhang & Zhou, 2015). The frequency of extreme weather events in the region has increased due to global warming (Allan &

Soden, 2008; Zhao, 2020). Typically, an extreme flooding event occurred during the June-July-August (JJA) of 2020 in the middle and lower reaches of the Yangtze River, resulting in significant human casualties and economic losses (Zhou et al., 2021; CMA, 2020). Therefore, accurate seasonal precipitation predictions are crucial for effective disaster prevention and mitigation. Currently, operational seasonal precipitation forecasts heavily rely on the performance of dynamical climate models. But the prediction skill of current-generation climate models for China and even East Asia is poor and requires much improvement to better meet societal requirements (Wang et al., 2009; Liu et al., 2015; Wang et al., 2015; Ding et al., 2020).

To improve the seasonal prediction skills of dynamical climate models, previous studies have explored methods such as quality control of observational data, ensemble techniques with perturbed initial conditions, and data assimilation to reduce the impact of initial state uncertainty (Gelaro et al., 2017; Zuo et al., 2017; Hersbach et al., 2020). In addition, improving the model's physical parameterization schemes and increasing its spatial resolution can also improve model predictions (Saha et al., 2010; 2014; Hersbach et al., 2020). However, implementing these methods requires the establishment of numerous observation sites, substantial data storage, computational resources, the development of new sub-grid process parameterization schemes or data processing algorithms, and so on (Rasp et al., 2018; Hersbach et al., 2020; François et al., 2021). To reduce computation resource demands, previous studies have proposed the cheap post-processing method (Klein et al., 1959; Glahn & Lowry, 1972). Using statistical methods to map model output with observation can significantly reduce model biases and improve prediction skills.

Most of the model post-processing approaches primarily rely on statistical methods, which include empirical orthogonal function (EOF) (Danforth et al., 2007; Feudale et al., 2011), singular value decomposition (SVD) (Ward et al., 1997; Hanssen-Bauer et al., 2005), and quantile mapping (QM) (Ines & Hansen, 2006; Themeßl et al., 2012). Among them, QM can improve the mean and variance of the cumulative distribution function (CDF). It is simple to understand and implement without complex processes, such as matrix decomposition and parameter setting (Ines & Hansen, 2006; Déqué 2007). However, its limitations include assumptions' stability (Ines & Hansen, 2006;

Teutschbein & Seibert, 2012), disregarding spatial correlation of precipitation (Thrasher et al., 2012), and neglecting physical mechanisms of precipitation (Déqué 2007; Teutschbein & Seibert, 2012). With the rapidly development of data science, the machine learning methods are widely used for bias-correction of dynamical models such as Random Forest (RF) (Li et al., 2019; Zarei et al., 2021), LightGBM (Zhong et al., 2021), Wavelet (Xu et al., 2019), and Support Vector Machines (SVM) (Pour et al., 2018). These methods often employ single-point or exponential corrections, which limits them to fully exploit spatial relationships. In contrast, deep learning methods can address it by identifying both temporal and spatial characteristics of the data. Previous studies have demonstrated that deep learning methods have strong potential for bias-correction tasks (Kim et al., 2021; Pan et al. 2019, Bretherton et al., 2022; Hess & Boers, 2022; Ling et al., 2022; François et al., 2021; Pan et al., 2021; Hess et al., 2022). In particular, the generative models can generate results that align better with the target distribution, capture local element characteristics, and thus offer greater potential for bias-correction in the model post-processing (Pan et al., 2021; Ling et al., 2022).

This study explored the goodness or potential of Cycle-Consistent Generative Adversarial Networks (CycleGAN) for correcting the dynamical model's bias in summer precipitation forecast produced by the NUIST-CFS1.0 real-time climate forecast system. It compares the results with the QM bias-correction method. Section 2 introduces the data, CycleGAN correcting method, and evaluation metrics. Section 3 analyzes the bias-correction results of CycleGAN. Section 4 presents a summary and discussion.

2 Data and Methods

2.1 Precipitation Data

Table S1 provides the details of the data and preprocessing. The precipitation (mm/day) prediction data to be corrected is produced by the NUIST-CFS1.0, which is a real-time dynamical climate model forecast system updated monthly. It is built based on the SINTEX-F ocean-atmosphere coupled global model that was originally developed at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and consists of 9-member ensemble forecasts ranging from 1 to 24

months (Luo et al., 2005a, 2005b, 2008; He et al. 2020). While the NUIST-CFS1.0 exhibits high skill in predicting climate anomalies at low latitudes, it significantly underestimates summer precipitation in East Asia, especially southeastern China (Luo et al., 2005a; Ying et al., 2022). We particularly focused on the forecasts initiated from 1st March for the summer months of JJA (i.e., lead time of 4-6 months). This start month is critical for China Meteorological Administration (CMA) to issue operational outlook for the summer precipitation in China. The precipitation data is obtained on daily temporal resolution and T106 (Gaussian grid) spatial resolution of 1.1º × 1.1º. The precipitation ground truth data is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) with a spatial resolution of 0.25º × 0.25º and processed into daily mean. Both the model forecasts and ERA5 precipitation data were adjusted to a common spatial resolution of 1º × 1º using bilinear interpolation with Climate Data Operator (CDO). The selected time period for the both datasets is JJA of 1982-2020 with the spatial domain of 72°E-135°E, 20°N-51°N.

2.2 CycleGAN Model and QM Method

Although dynamical seasonal forecast models can output daily or even six-hourly values, the daily conditions tend to be much random and hard to be predicted at lead times beyond weeks or even days. Therefore, the daily data produced by the seasonal prediction system cannot be paired with the daily observed data, and hence it is impossible to train the deep learning model with a supervised learning way to correct the model's daily outputs. To overcome it, we adopted the CycleGAN method to expand the training samples with daily data.

The CycleGAN, a variant of the original GAN (Goodfellow et al., 2014), is an unpaired model originally used for style migration and well-suited for addressing bias-correction in dynamical model's post-processing. The primary objective is to transform the input X from the climate model's output distribution to a distribution like the ERA5 observation domain Y, which is utilized for correcting climate prediction. CycleGAN incorporates "cycle-consistency loss", to ensure that data transformed to another domain can be accurately restored to the original domain. In addition, "identity loss" is employed to regulate the generator and prevent excessive migration. The main framework of the CycleGAN model, based on Zhu et al. (2017), is illustrated in Figure S1.

Note that the CycleGAN model may not be sensitive to small precipitation due to the wide range of magnitudes in the data. The logarithmic processing was applied to improve the model's ability to learn from small precipitation samples. Besides, the maximum-minimum normalization was used to transform the data to the [-1, 1] range, ensuring a more stable training process. To obtain prediction results for an extended period, a cross-validation approach was implemented (Table S1). By implementing five CycleGAN models, the results from each cross-validation period were seamlessly integrated to derive the bias-corrected precipitation forecasts with nine members spanning the years of 1991 to 2020.

The QM method is a probability distribution-based approach that corrects biases by matching the CDF between model output and observed data. Its advantages include capturing distribution characteristics well (Ines & Hansen, 2006; Déqué 2007; Themeßl et al., 2012; Cannon et al., 2015) and handling precipitation extremes (Themeßl et al., 2012; Cannon et al., 2015). Due to its simplicity and ease of operation, it is more effective than other traditional methods like EOF and can handle extreme events well, hence the QM method (Déqué, 2007) is selected as a baseline for skill comparison with the CycleGAN.

2.3 Evaluation Metrics

In this study, we selected several metrics to evaluate the bias-correcting ability of the CycleGAN and QM models (see Table 1). They consist of four basic descriptiveness metrics for "concentration, dispersion, skewness, and spike" level analysis, which are calculated using Formula 1-4 in Table 1. Skewness is a statistical measure that assesses the asymmetry of a data distribution, reflecting the departure from symmetry on both sides of the mean. Kurtosis quantifies the sharpness of a data distribution, characterizing the concentration of data around the mean and the thickness of the tails. We also adopted Root Mean Square Error (RMSE), Correlation Coefficient (CC), CDF, and three quantiles for the JJA precipitation prediction skill assessment. In addition, three indices were selected for assessing the prediction of extreme precipitation events, including 95th percentile and seasonal maximum daily precipitation as well as accumulated

amount of daily precipitation exceeding 20 mm/day. We focused on the southeastern China (Figure S2), which represents the principal zone of summer precipitation distribution, characterized by a north-south shift of the primary rainband in response to the seasonal migration of the East Asian monsoon. The intensity of precipitation there exhibits a significant variability in correlation with the strengthening and weakening of the monsoon. The strong precipitation variations lead to large prediction errors in many climate forecast systems (Huang et al., 2007; Sperber et al., 2013; Liu et al., 2015; Wang et al., 2015; Ding et al., 2020).

3 Results

3.1 Assessment of the Climatological Descriptiveness Indices

Climatological mean and standard deviation are widely used metrics for evaluating climate prediction performance. Figure 1a-1h and Figure S3a-S3f demonstrate the mean and variability of JJA precipitation in the southeastern China predicted by NUIST-CFS1.0 differ significantly from ERA5 during 1991-2020 and their maximum centers are not reproduced. Both the CycleGAN and QM methods successfully correct the bias in general. Notably, the $RMSE_P$s of the CycleGAN reach approximately a half of those achieved by the QM-correction, with the strong variability over many regions of the southeastern China being better reproduced. In addition, as depicted in Figure 1i-1p and Figure S3g-S3l, the CycleGAN method demonstrates a reduction in the pronounced positive skewness bias in the NUIST-CFS1.0 precipitation prediction over the northern part of the southeastern China, reaching to a level closer to that of ERA5 than the QM correction does. Similarly, the CycleGAN also exhibits improvements in the kurtosis values for the precipitation prediction in both northern and southeastern parts of the southeastern China, displaying larger improvements and smaller $RMSE_P$ values than QM.

3.2 Distribution of Temporal Variations

Similar to many other dynamical climate prediction systems, the NUIST-CFS1.0 forecasts initiated from 1st March also displayed large RMSEs in predicting JJA precipitation in the southeastern China (Figure 2c). In addition, the pronounced northward migration of the rainband from ~23ºN to ~33ºN during early June to early July is not reproduced (Figure 2g). After the

post-process correcting, both the CycleGAN and QM effectively reduce the RMSEs of the NUIST-CFS1.0 forecasts over a large area, particularly in the southern part of southeastern China (Figure 2a-2b). And the CycleGAN outperforms QM with smaller mean value of RMSEs. Both corrections also improve the forecast of the northward migration of the rainband, although the intensity is still underestimated (Figure 2e-2f). Again, the CycleGAN exhibits a superior ability compared to QM with smaller $RMSE_P$s overall. We also assessed the latitudinal distribution of the JJA mean precipitation along 99°E-128°E (Figure S4). The NUIST-CFS1.0 displays large errors across all the latitudes within 20°N-40°N with an averaged error of 1.93 (mm/day). The CycleGAN reduces the error down to 0.23 (mm/day) and produces a realistic distribution that is closer to that of ERA5 compared to the QM.

To further compare the skill in predicting the spatial-temporal distribution of precipitation, the kernel density function and CDF were analyzed (Figure 3). The former estimates the density of scatter points in different regions, allowing for a comparison of the Probability Density Function distributions between different datasets. The red shading of the filled area indicates a higher kernel density and a more concentrated distribution of JJA precipitation over the southeastern China (Figure 3a-3c). The results suggest that both the CycleGAN and QM corrections help reduce the biases in the original NUIST-CFS1.0 predictions. However, the CycleGAN exhibits better learning capabilities in capturing the disparities of the spatial-temporal distribution between the predicted and ERA5 precipitation, in comparison to QM. Consequently, the distribution of the predicted precipitation is better constrained, resulting in a smaller RMSE and a higher CC.

In addition, we also assessed the prediction of three precipitation quantiles (i.e., 33%, 66%, and 99%) based on the daily data during 1991-2020 (Figure S5-S6). Both the CycleGAN model and the QM method exhibit improvements over the original NUIST-CFS1.0 predictions in the southeastern China. Again, the CycleGAN is superior to the QM and achieves smaller $RMSE_P$ s for forecasts of the three categories. Notably, the correction by the CycleGAN displays much more improvement for predicting the 99% quantile. This suggests that the CycleGAN can successfully capture the distribution of extreme precipitation events.

3.3 Extreme Precipitation

A coarse-resolution climate model often has great difficulty in reproducing extreme precipitation events (IPCC, 2013; PaiMazumder, 2014; Bador et al., 2020), and thus requiring an effective postprocessing approach to reduce the bias. We assessed the performance of the CycleGAN and QM methods in correcting the predictions of the selected three indices related to extreme precipitation events (recall Table 1). The three indices are commonly used for assessing the intensity of extreme precipitation on each grid (Singh et al., 2013; Gupta et al., 2020; Luu et al., 2022). As expected, the original NUIST-CFS1.0 predictions with a spatial resolution of about 100km severely underestimates the extreme precipitation events especially in the Yangtze River basin and the southern part along the coast (see right column in Figure 4 and Figure S7). The RMSEs in predicting the spatial distribution of the three indices reach 11.33, 21.21, and 177.32 mm/day, respectively. The QM method can improve the models' bias but still with strong underestimation over the large part of the southeastern region (middle column in Figure 4 and Figure S7). The results show that the CycleGAN bias-correction method can realistically reproduce the intensive precipitation in the Yangtze River basin and the southern part along the coast, with spatial RMSEs of the three indices being largely reduced to 3.51, 10.21, and 50.09 mm/day, respectively (left column in Figure 4 and Figure S7). In addition, we also examined the performance in predicting the frequency of extreme events with daily precipitation exceeding 95th percentile. The results indicate a great challenge over many areas in the southeastern China (Figure S8). Nevertheless, the CycleGAN bias-correction increases the prediction skill over large parts of the southern and western regions (Figure S8d). In contrast, the improvement using the QM is negligible (Figure S8e). The above results clearly suggest that the CycleGAN outperforms QM and can help minimize the RMSEs in predicting the extreme precipitation.

4 Summary and Discussions

In this study, we used the CycleGAN deep learning algorithm to perform bias-correction based on the NUIST-CFS 1.0 JJA precipitation predictions. We compared its performance with the traditional statistical QM (i.e., a baseline). While the QM can help reduce the biases, our results illustrate that the CycleGAN model is superior to the QM in improving the predictions of climatology, seasonal northward migration of the rainband, interannual variability, skewness and

kurtosis of JJA mean precipitation in southeastern China, with the spatial RMSEs being further reduced by approximately 50% compared to the QM. Furthermore, the CycleGAN deep learning exhibits significant improvements in predicting daily extreme precipitation events, implying substantial potential for societal applications and disaster prevention.

CycleGAN is categorized as an unpaired domain adaptation model, utilizing daily precipitation data for training. It enables the deep learning model to grasp the differences between the high-dimensional distributions of daily NUIST-CFS 1.0 and ERA5 data through training both generator and discriminator components. The generator's role involves producing new data samples that mirror the pattern, while the discriminator discerns and identifies fresh data closely resembling the ERA5 distribution. Daily precipitation data offers some advantages. Firstly, it enhances the pool of available data samples for model training, thereby contributing to the stability and robustness of deep learning model training. In addition, the inclusion of daily precipitation data helps capture the traits of extreme events. Particularly valuable for forecasting extreme precipitation, which is a great challenge using current coarse-resolution dynamical model forecasts. In short summary, by harnessing CycleGAN and daily precipitation data, the model not only elevates overall forecast performance but also refines the prediction of daily extreme events. These improvements are vital for increasing socio-economic values of the seasonal forecasts.

It is worth noting that, although the two correction methods help reduce the errors in predicting the seasonal precipitation, improving the phase prediction of precipitation anomaly relative to the climatology appears to be rather difficult (Figure S9). The results show that the QM does not improve the ACC skill. While the CycleGAN significantly improves the skill in the southern and northwestern parts of the southeastern China, it degrades the skill in other areas, leading to a negligible increase of the spatially averaged ACC skill (Figure S9a and S9c). Similarly, the skill improvements in predicting the spatial pattern of the precipitation anomalies during 1991-2020 over the southeastern China are also negligible despite that the improvements in some years can be significant (Figure S9d). Further efforts are warranted to improve the dynamical understanding and deep learning bias-correction methods.

283

284 In our training, we initially set the learning rate of $2e^{-4}$, and we conducted a total of 600 training
285 epochs. It was gradually reduced starting from 200 epochs. Training GAN models involves a
286 dynamical process and reaching an optimal balance is key to obtaining best results. In theory,
287 longer training and more epochs increase the chance of finding this balance. However,
288 determining the model's best state is challenging. Currently, we used a qualitative assessment
289 based on the generated outcomes and analysis of the loss function trend to decide when to stop the
290 training. While insightful, this method can be intricate and resource intensive.

291

292 In future studies, creating a more efficient automatic stop-training method is essential. It would
293 move away from the fixed number of training rounds and simplify the model's training process. By
294 including dynamical criteria that assess the model's performance and convergence rate, this
295 stop-training method give us a clearer idea of when the optimal training state is reached. It enhance
296 the model's performance, reducing the need for manual adjustments during training. In addition,
297 we will try to improve the deep-learning model's accuracy in two ways. Firstly, incorporate
298 multi-level atmosphere circulation conditions that are responsible for precipitation into the
299 training data. Climate models, known for their ability to predict large-scale circulation patterns,
300 can help the deep-learning model better capture the physical processes. This refinement can
301 fine-tune the precipitation predictions. Secondly, adjust the model's loss function by introducing a
302 regularization term in order to enforce physical constraints. This strengthens the deep-learning
303 model's adherence to physical principles during the generative adversarial training. Typically, we
304 will achieve this by pre-training the deep-learning model to grasp the relationship between
305 atmosphere circulation and precipitation.

306 **Acknowledgments**

307 We appreciate Figshare for their open-source data storage support for the NUIST-CFS1.0. Data is
308 housed within the "CFS1.0 data" section of [Paper # 2023GL105943] project and can be accessed
309 through the citation DOI https://doi.org/10.6084/m9.figshare.24013767.v1 (Yang, 2023). This
310 work is supported by the National Key Research and Development Program of China (No.

2020YFA0608000) and the National Natural Science Foundation of China (Grant 42030605). Thanks to Baoxiang Pan, Jianxiang Xu, Jiaying He and Zeyu Lu for helpful discussion.

Open Research

The ERA5 data is available at https://doi.org/10.24381/cds.adbb2d47. The NUIST-CFS1.0 data is available on Figshare, and the private URL is https://figshare.com/s/e82408aa5e35b7e51765.

References

Allan, R. P., & Soden, B. J. (2008). Atmospheric warming and the amplification of precipitation extremes. *Science*, 321(5895), 1481-1484. https://doi.org/10.1126/science.1160787

Bador, M., Boé, J., Terray, L., *et al*. (2020). Impact of higher spatial atmospheric resolution on precipitation extremes over land in global climate models. *Journal of Geophysical Research: Atmospheres*, 125(13), e2019JD032184. https://doi.org/10.1029/2019JD032184

Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., *et al*. (2022). Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. *Journal of Advances in Modeling Earth Systems*, 14(2). https://doi.org/10.1029/2021MS002794

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. https://doi.org/10.1175/JCLI-D-14-00754.1

China Meteorological Administration (CMA) (2020). *Yearbook of Meteorological Disasters in China*. Beijing: Meteorological Press, 183–185. (In Chinese).

Danforth, C. M., Kalnay, E., & Miyoshi, T. (2007). Estimating and correcting global weather model error. *Monthly weather review*, 135(2), 281-299. https://doi.org/10.1175/MWR3289.1

Déqué, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, 57(1-2), 16-26. https://doi.org/10.1016/j.gloplacha.2006.11.030

Ding, Y. H., & Chan, J. C. L. (2005). The East Asian summer monsoon: an overview. *Meteorology and Atmospheric Physics*, 89(1-4), 117-142. https://doi.org/10.1007/s00703-005-0125-z

Ding, Y. H., Liang, P., Liu, Y. J., & Zhang, Y. C. (2020). Multiscale variability of Meiyu and its prediction: A new review. *Journal of Geophysical Research: Atmospheres*, 125(7). https://doi.org/10.1029/2019JD031496

Feudale, L., & Tompkins, A. M. (2011). A simple bias correction technique for modeled monsoon precipitation applied to West Africa. *Geophysical research letters*, 38(3). https://doi.org/10.1029/2010GL045909

François, B., Thao, S., & Vrac, M. (2021). Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks. *Climate Dynamics*, 57, 3323-3353. https://doi.org/10.1007/s00382-021-05869-8

Gelaro, R., McCarty, W., Suárez, M. J., *et al*. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419-5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, 11(8), 1203-1211. https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., *et al*. (2014). Generative Adversarial Networks. *arXiv preprint* arXiv:1406.2661. https://doi.org/10.48550/arXiv.1406.2661

Gupta, V., Singh, V., & Jain, M. K. (2020). Assessment of precipitation extremes in India during the 21st century under SSP1-1.9 mitigation scenarios of CMIP6 GCMs. *Journal of Hydrology*, 590, 125422. https://doi.org/10.1016/j.jhydrol.2020.125422

Hanssen-Bauer, I., Achberger, C., Benestad, R. E., Chen, D., & Førland, E. J. (2005). Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, 29(3), 255-268. https://doi.org/10.3354/cr029255

He J. Y., Wu J. Y., & Luo J. J. (2020). Introduction to climate forecast system version 1. 0 of Nanjing University of Information Science and Technology. *Transactions of Atmospheric Sciences*, 43(1), 128-143. https://doi.org/10.13878/j.cnki.dqkxxb.20191110007 (in Chinese)

Hersbach, H., Bell, B., Berrisford, P., *et al*. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999-2049. https://doi.org/10.1002/qj.3803

Hess, P., & Boers, N. (2022). Deep learning for improving numerical weather prediction of heavy rainfall. *Journal of Advances in Modeling Earth Systems*, 14(3). https://doi.org/10.1029/2021MS002765

Hess, P., Drüke, M., Petri, S., Strnad, F. M., & Boers, N. (2022). Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nature Machine Intelligence*, 4(10), 828-839. https://doi.org/10.1038/s42256-022-00540-1

Huang, R., Chen, J., & Huang, G. (2007). Characteristics and variations of the East Asian monsoon system and its impacts on climate disasters in China. *Advances in Atmospheric Sciences*, 24, 993-1023. https://doi.org/10.1007/s00376-007-0993-x

Ines, A. V., & Hansen, J. W. (2006). Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and forest meteorology*, 138(1-4), 44-53. https://doi.org/10.1016/j.agrformet.2006.03.009

IPCC (2013). In Stocker T. F., Qin D., Plattner G.-K., *et al*. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 1535). Cambridge, UK, and New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

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

Kim, H., Ham, Y. G., Joo, Y. S., & Son, S. W. (2021). Deep learning for bias correction of MJO prediction. *Nature Communications*, 12(1), 3087. https://doi.org/10.1038/s41467-021-23406-3

Klein, W. H., Lewis, B. M., & Enger, I. (1959). Objective prediction of five-day mean temperatures during winter. *Journal of Atmospheric Sciences*, 16(6), 672-682. https://doi.org/10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2

Li, H., Yu, C., Xia, J., Wang, Y., Zhu, J., & Zhang, P. (2019). A model output machine learning method for grid temperature forecasts in the Beijing area. *Advances in Atmospheric Sciences*, 36, 1156-1170. https://doi.org/10.1007/s00376-019-9023-z

Ling, F., Li, Y., Luo, J. J., Zhong, X., & Wang, Z. (2022). Two deep learning-based bias-correction pathways improve summer precipitation prediction over China. *Environmental Research Letters*, 17(12), 124025. https://doi.org/10.1088/1748-9326/aca68a

Liu, X., Wu, T., Yang, S., *et al*. (2015). Performance of the seasonal forecasting of the Asian summer monsoon by BCC_CSM1. 1 (m). *Advances in Atmospheric Sciences*, 32, 1156-1172. https://doi.org/10.1007/s00376-015-4194-8

Luo, J. J., Masson, S., Behera, S., Shingu, S., & Yamagata, T. (2005). Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts. *Journal of Climate*, 18(21), 4474-4497. https://doi.org/10.1175/JCLI3526.1

Luo, J. J., Masson, S., Roeckner, E., Madec, G., & Yamagata, T. (2005). Reducing climatology bias in an ocean–atmosphere CGCM with improved coupling physics. *Journal of Climate*, 18(13), 2344-2360. https://doi.org/10.1175/JCLI3404.1

Luo, J. J., Masson, S., Behera, S. K., & Yamagata, T. (2008). Extended ENSO predictions using a fully coupled ocean–atmosphere model. *Journal of Climate*, 21(1), 84-93. https://doi.org/10.1175/2007JCLI1412.1

Luu, L. N., Vautard, R., Yiou, P., & Soubeyroux, J. M. (2022). Evaluation of convection-permitting extreme precipitation simulations for the south of France. *Earth System Dynamics*, 13(1), 687-702. https://doi.org/10.5194/esd-13-687-2022

PaiMazumder, D., & Done, J. M. (2015). The roles of bias-correction and resolution in regional climate simulations of summer extremes. *Climate Dynamics*, 45, 1565-1581. https://doi.org/10.1007/s00382-014-2413-0

Pan, B., Anderson, G. J., Goncalves, A., *et al*. (2021). Learning to correct climate projection biases. *Journal of Advances in Modeling Earth Systems*, 13(10). https://doi.org/10.1029/2021MS002509

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation using convolutional neural network. *Water Resources Research*, 55(3), 2301-2321. https://doi.org/10.1029/2018WR024090

Pour, S. H., Shahid, S., Chung, E. S., & Wang, X. J. (2018). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric research*, 213, 149-162. https://doi.org/10.1016/j.atmosres.2018.06.006

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39), 9684-9689. https://doi.org/10.1073/pnas.1810286115

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference,* Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Saha, S., Moorthi, S., Pan, H. L., *et al*. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015-1058. https://doi.org/10.1175/2010BAMS3001.1

Saha, S., Moorthi, S., Wu, X., *et al*. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), 2185-2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Singh, D., Tsiang, M., Rajaratnam, B., & Diffenbaugh, N. S. (2013). Precipitation extremes over the continental United States in a transient, high-resolution, ensemble climate model experiment. *Journal of Geophysical Research: Atmospheres*, 118(13), 7063-7086. https://doi.org/10.1002/jgrd.50543

Sperber, K. R., Annamalai, H., Kang, I. S., *et al*. (2013). The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dynamics*, 41, 2711-2744. https://doi.org/10.1007/s00382-012-1607-6

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, 456, 12-29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Themeßl, M. J., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112, 449-468. https://doi.org/10.1007/s10584-011-0224-4

Thrasher, B., 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 https://doi.org/10.5194/hess-16-3309-2012

Wang, B., & Lin, H. (2002). Rainy season of the Asian–Pacific summer monsoon. *Journal of Climate*, 15(4), 386-398. https://doi.org/10.1175/1520-0442(2002)015<0386:RSOTAP>2.0.CO;2

Wang, B., Lee, J. Y., Kang, I. S., *et al*. (2009). Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dynamics*, 33, 93-117.https://doi.org/10.1007/s00382-008-0460-0

Wang, B., Lee, J. Y., & Xiang, B. (2015). Asian summer monsoon rainfall predictability: a predictable mode analysis. *Climate Dynamics*, 44, 61-74. https://doi.org/10.1007/s00382-014-2218-1

Ward, M. N., & Navarra, A. (1997). Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific. *Journal of Climate*, 10(9), 2210-2220. https://doi.org/10.1175/1520-0442(1997)010<2210:PAOSFV>2.0.CO;2

Wu, R., Hu, Z. Z., & Kirtman, B. P. (2003). Evolution of ENSO-related rainfall anomalies in East Asia. *Journal of Climate*, 16(22), 3742-3758. https://doi.org/10.1175/1520-0442(2003)016<3742:EOERAI>2.0.CO;2

Xu, L., Chen, N., Zhang, X., Chen, Z., Hu, C., & Wang, C. (2019). Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Climate Dynamics*, 53, 601-615. https://doi.org/10.1007/s00382-018-04605-z

Ying, W., Yan, H., & Luo, J. J. (2022). Seasonal predictions of summer precipitation in the middle-lower reaches of the Yangtze River with global and regional models based on NUIST-CFS1. 0. *Advances in Atmospheric Sciences*, 39(9), 1561-1578. https://doi.org/10.1007/s00376-022-1389-7

Zarei, M., Najarchi, M., & Mastouri, R. (2021). Bias correction of global ensemble precipitation forecasts by Random Forest method. *Earth Science Informatics*, 14, 677-689. https://doi.org/10.1007/s12145-021-00577-7

Zhang, L., & Zhou, T. (2015). Drought over East Asia: a review. *Journal of Climate*, 28(8), 3375-3399. https://doi.org/10.1175/JCLI-D-14-00259.1

Zhao, W. (2020). Extreme weather and climate events in China under changing climate. *National Science Review*, 7(5), 938-943. https://doi.org/10.1093/nsr/nwaa069

Zhou, Z. Q., Xie, S. P., & Zhang, R. (2021). Historic Yangtze flooding of 2020 tied to extreme Indian Ocean conditions. *Proceedings of the National Academy of Sciences*, 118(12). https://doi.org/10.1073/pnas.2022255118

Zhong, J., Zhang, X., Gui, K., *et al*. (2021). Robust prediction of hourly PM2. 5 from meteorological data using LightGBM. *National Science Review*, 8(10), nwaa307. https://doi.org/10.1093/nsr/nwaa307

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv preprint* arXiv:1703.10593. https://doi.org/10.48550/arXiv.1703.10593

Zuo, H., Balmaseda, M. A., & Mogensen, K. (2017). The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 49, 791-811. https://doi.org/10.1007/s00382-015-2675-1

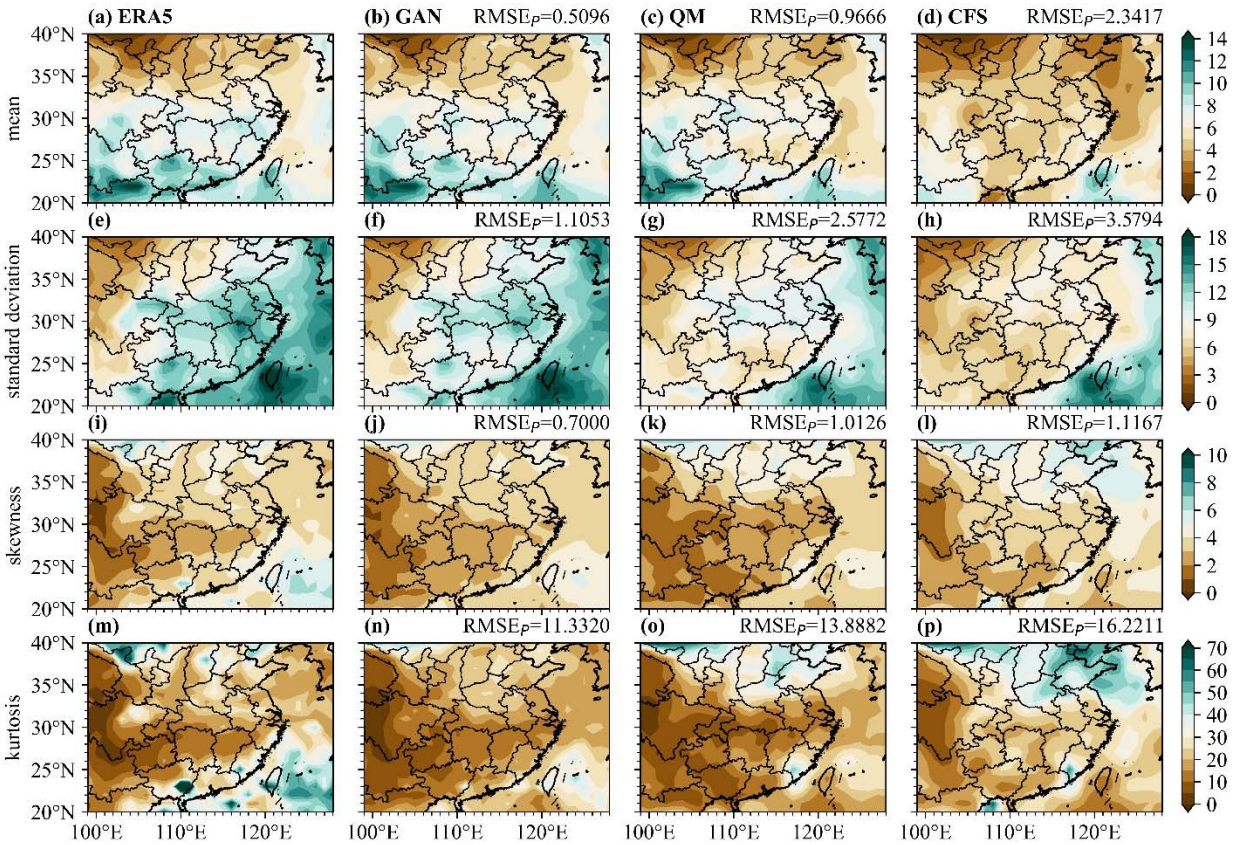

Figure 1. The spatial distribution of (a-d) climatological mean, (e-h) standard deviation, (i-l) skewness, and (m-p) kurtosis of JJA precipitation from 1991 to 2020. Each row corresponds to a different statistical metric, and a color bar at the rightest end represents the range of each metric. Each column represents a different data source: (a, e, i, m) ERA5, (b, f, j, n) the CycleGAN corrected predictions, (c, g, k, o) the QM corrected predictions, and (d, h, l, p) the original NUIST-CFS1.0 predictions. The RMSE$_P$ (RMSE of spatial patterns) in the upper right corner of each panel indicates the error of the spatial distribution relative to ERA5 observation for each metric.

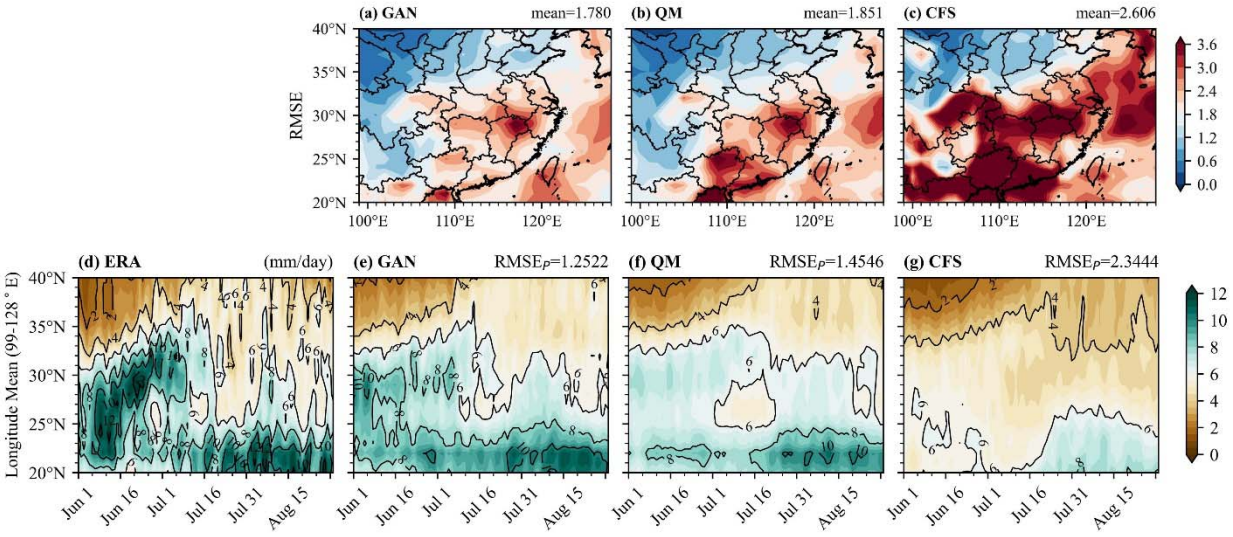

Figure 2. (a-c) The spatial distribution of RMSE (mm/day) of the three types of forecasts relative to ERA5 JJA mean precipitation during 1991-2020. The value in the upper right corner of each panel denotes the spatially averaged RMSE. (d-g) The latitudinal-migration of climatological mean precipitation of 1991-2020 averaged along 99°E-128°E based on the results of the ERA5, CycleGAN correction, QM correction and NUIST-CFS1.0 predictions. The value in the upper right corner of each panel indicates the error of the spatial patterns relative to ERA5 observation.

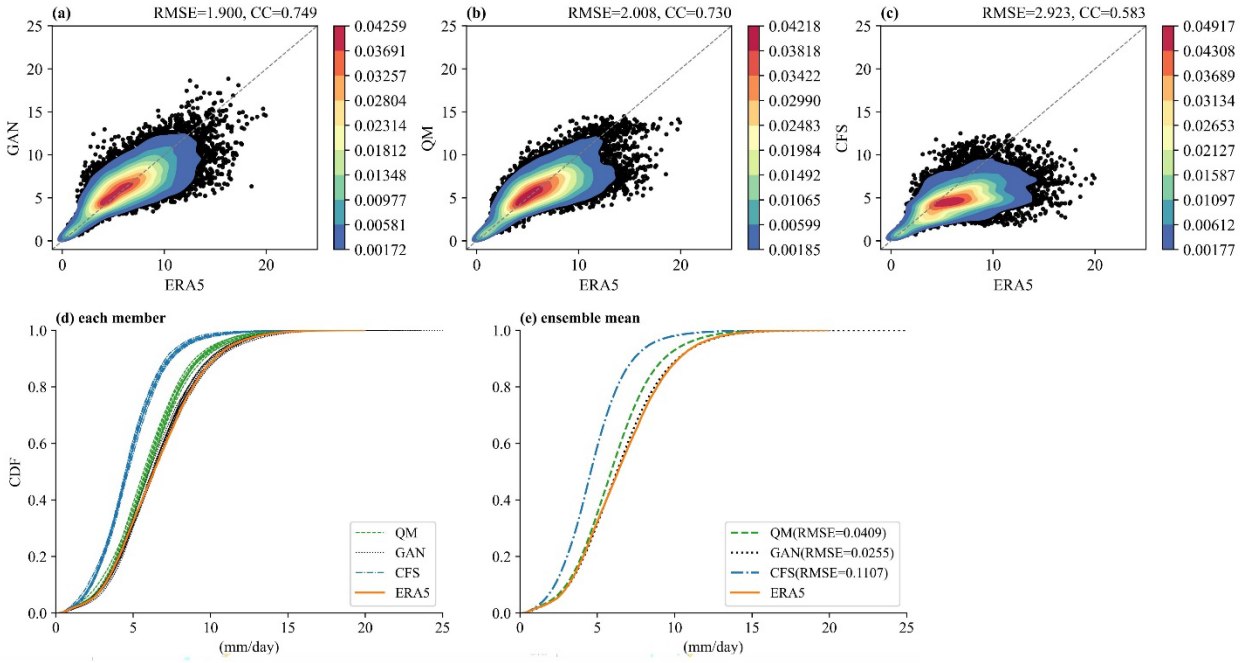

Figure 3. (a-c) Scattered kernel density maps between the CycleGAN-corrected, QM-corrected, original NUIST-CFS1.0 predictions and ERA5 data. The results are calculated based on all spatial-temporal points of JJA mean precipitation from 1991 to 2020. RMSE and CC values are also given in the upper right corner. The filled areas indicate the concentration of the scattered points. (d-e) The CDF curves of the JJA mean precipitation during 1991-2020 based on ERA5 data and individual member and 9-member ensemble mean predictions of the CycleGAN, QM, and NUIST-CFS1.0.

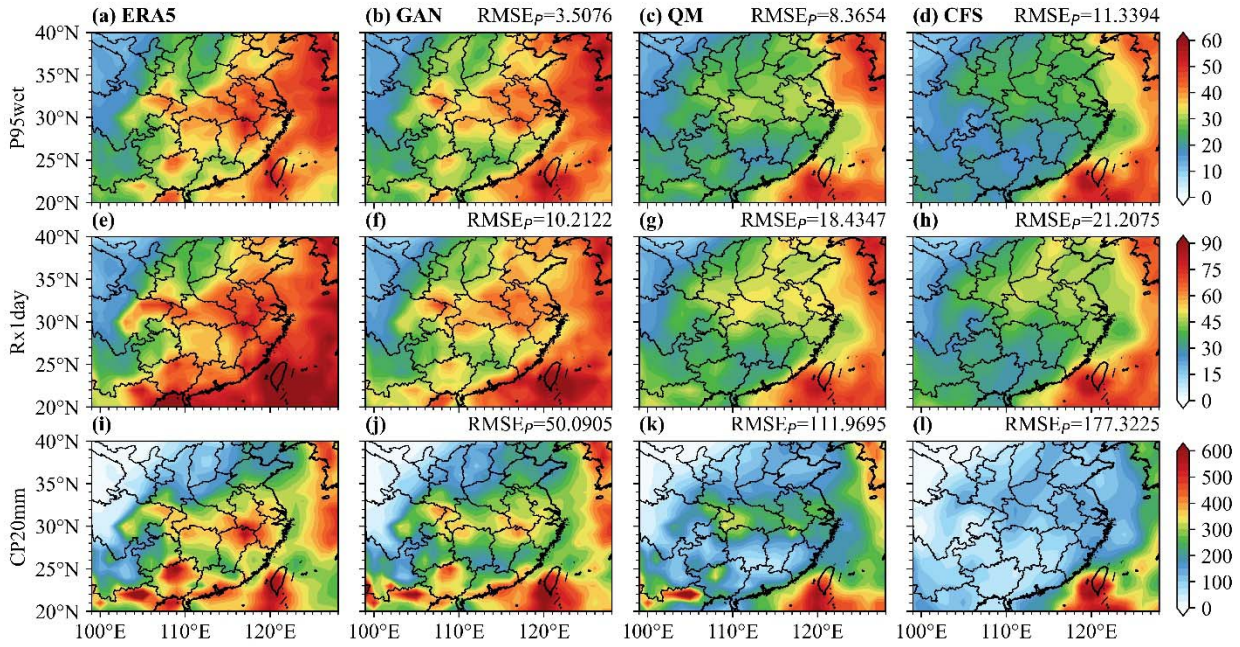

Figure 4. The spatial distributions of (a-d) P95wet, (e-h) climatological mean Rx1day, and (i-l) climatological mean CP20mm based on JJA daily precipitation during 1991-2020. Each row corresponds to a different statistical metric, and a color bar at the rightest end represents the range of each metric (units: mm/day). Each column represents a different data source: (a, e, i) ERA5, (b, f, j) the Cycle GAN corrected predictions, (c, g, k) the QM corrected predictions, and (d, h, l) the original NUIST-CFS1.0 predictions. The RMSE$_P$ in the upper right corner of each panel indicates the error of the spatial patterns relative to ERA5 observation for each metric.

Scope	Statistics	Calculation	
Climatological descriptiveness	Climatological mean (μ)	$\mathbb{E}_{x\sim P(x)}(x)$	**(1)**
	Standard deviation (σ)	$\sqrt{\mathbb{E}_{x\sim P(x)}(x-\mu)^2}$	**(2)**
	Skewness	$\mathbb{E}_{x\sim P(x)}\left(\frac{x-\mu}{\sigma}\right)^3$	**(3)**
	Kurtosis	$\mathbb{E}_{x\sim P(x)}\left(\frac{x-\mu}{\sigma}\right)^4$	**(4)**
Distribution	Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$	**(5)**
	Correlation Coefficient (CC)	$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$	**(6)**
	CDF	No fixed formula	
	$Q_{33\%,\ 66\%,\ 99\%}$	33%, 66%, 99% Quantile	
Extreme events	P95wet	95th percentile of daily precipitation, only wet (>= 1 mm/day) days are considered	
	Rx1day	Seasonal maximum 1-day precipitation (mm/day)	
	CP20mm	Seasonal cumulative precipitation of strong events (>= 20 mm/day)	

Table 1. The evaluation metrics related to "Climatological descriptiveness", "Distribution", and "Extreme events".