Original Manuscript ID: GMD-2023-158

Original Article Title: Deep-learning statistical downscaling of precipitation in the middle reaches of the Yellow River: A Residual in Residual Dense Block based network

Dear Referee #1,

We sincerely thank the Referee #1 for all your valuable comments and insightful suggestions to our manuscript to improve the manuscript. We have addressed all the specific comments in the revised manuscript, with the point-by-point responses detailed below.

Best regards,

Xiaoning Xie et al.

"The authors concluded that RRDBNet performed better than other methods in terms of difference, RMSE and correlartion coefficient (CC). However, the study lacks novelty and improvements are very trivial."

Response: Thanks for your comments and suggestions. Our purpose is to explore the performance of deep convolutional neural networks (CNNs) in precipitation downscaling and provide another idea and method for precipitation downscaling. The CNN proposed by Baño-Medina et al. (2020) has only 3 convolutional layers as shown in Figure R1. In the field of computer vision, due to the small number of convolutional layers, on the one hand, these shallow CNNs may not be able to perform in-depth feature extraction and abstraction, and may not adequately capture complex patterns and relationships in the data; on the other hand, shallow CNNs do not have the ability to extract features at more hierarchical levels, and may be difficult to fully learn higher-level, more abstract, and more discriminative feature information and are easier to produce overfitting (Simonyan and Zisserman, 2015; Szegedy et al., 2015; LeCun et al., 2015; He et al., 2016; Huang et al., 2017). Therefore, in this paper, we explore the performance of deep CNNs in precipitation downscaling to provide another idea and method for precipitation downscaling. Here, we develop a deep learning-based statistical downscaling method, i.e., Residual-in-Residual Dense Block based network (RRDBNet) model (Figure R2), to produce daily high-resolution precipitation over the MRYR region. Residual-in-Residual Dense Block (RRDB) is introduced in the proposed RRDBNet to optimize the structure of the network model. The multi-level residuals and dense connectivity strategies in RRDB help the model to adequately extract more abstract and advanced feature information (He et al., 2016; Huang et al., 2017; Zhang et al., 2018; Wang et al., 2018). Extensive experiments in this paper show that the proposed RRDBNet exhibits excellent performance in most metrics. This indicates that the proposed RRDBNet is effective and can be used as a reliable tool for precipitation downscaling.

Specifically, the proposed RRDBNet reduces RMSE by 58% (79%) and improves CC by 38% (145%) relative to GLM on climatology mean for R95P (R99P). It is also the best compared to other deep-learning models, with the smallest RMSE and the largest CC. In terms of temporal variation in annual extreme precipitation, RRDBNet decreases RMSE by 28% (49%) and enhances CC by 229% (355%) relative to GLM for R95P (R99P), and also has significant improvements over CNN and RDBNet (see Supplementary Table 4). Our results indicate the applicability of deep-learning algorithms in the precipitation downscaling and highlight the good performance of RRDBNet in extreme precipitation. The corresponding descriptions have been added in the revised version.

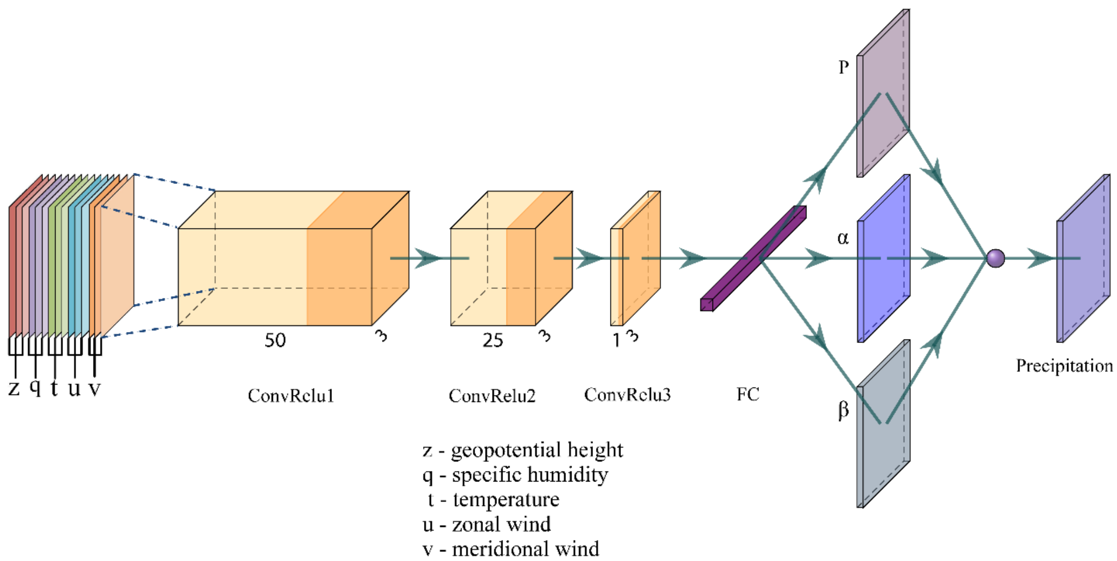

Figure R1. The CNN overall framework (adapted by Baño-Medina et al. (2020)).

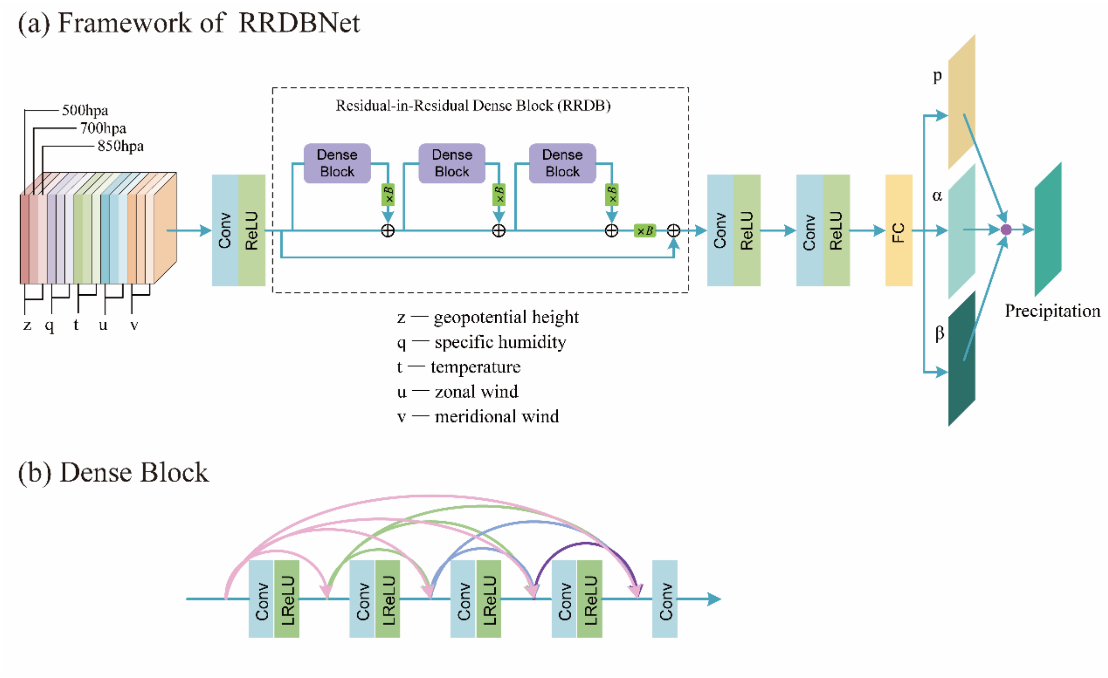

Figure R2. RRDBNet overall framework.

Comment 1: The models were trained with daily data, but all the evaluations were performed at aggregated time scale (annual, monthly and seasons). How does the model perform in daily time scale (both daily statistics and extremes)? The aggregated time scales may not be critical and may hide important information on model evaluation.

Author response: Thanks for your comments and suggestions. In the precipitation downscaling of this paper, we use probabilistic (stochastic) modeling (Cannon, 2008; Baño-Medina et al., 2020, 2022; Sun and Lan, 2021; Rampal et al., 2022) rather than deterministic modeling (Bürger, 1996, 2002). Deterministic modeling often leads to an underestimation of the variability and the extremes because the explained variance may be significantly smaller than the observed one (Williams, 1997; Cannon, 2008; Baño-Medina et al, 2020, 2022). This is particularly relevant for precipitation, whose local variability is often influenced by local phenomena that are not captured by the selected predictors (Schoof and Pryor, 2001; Maraun and Widmann, 2018). We also compare the effect of deterministic and stochastic modeling of precipitation on the probability density function (pdf) of daily precipitation frequency in Figure R3. It is obvious that deterministic precipitation modeling does not conform to the changing trend of observed precipitation frequency, while stochastic precipitation modeling is roughly consistent with the changing trend of observed precipitation frequency. Probabilistic modeling downscaling yields daily precipitation as random values derived from the Bernoulli-Gamma distribution. It may be more suitable for climatological statistical analysis. In the future we use the model also to do the assessment of future centennial precipitation changes in the region. In addition, based on your suggestions, we have done the validation of the daily precipitation in the area as shown in Table R1. We found that the RRDBNet model also performs well on daily precipitation compared to other models, having the smallest RMSE and largest correlation coefficient (CC). Since the five-year daily data we obtained after downscaling is 1827×100×159, when we use the five-year daily data to draw a scatter plot, the output is very slow. Therefore, we did a regional average before drawing the corresponding scatterplot (Figure R4). It can be seen that also RRDBNet performs well.

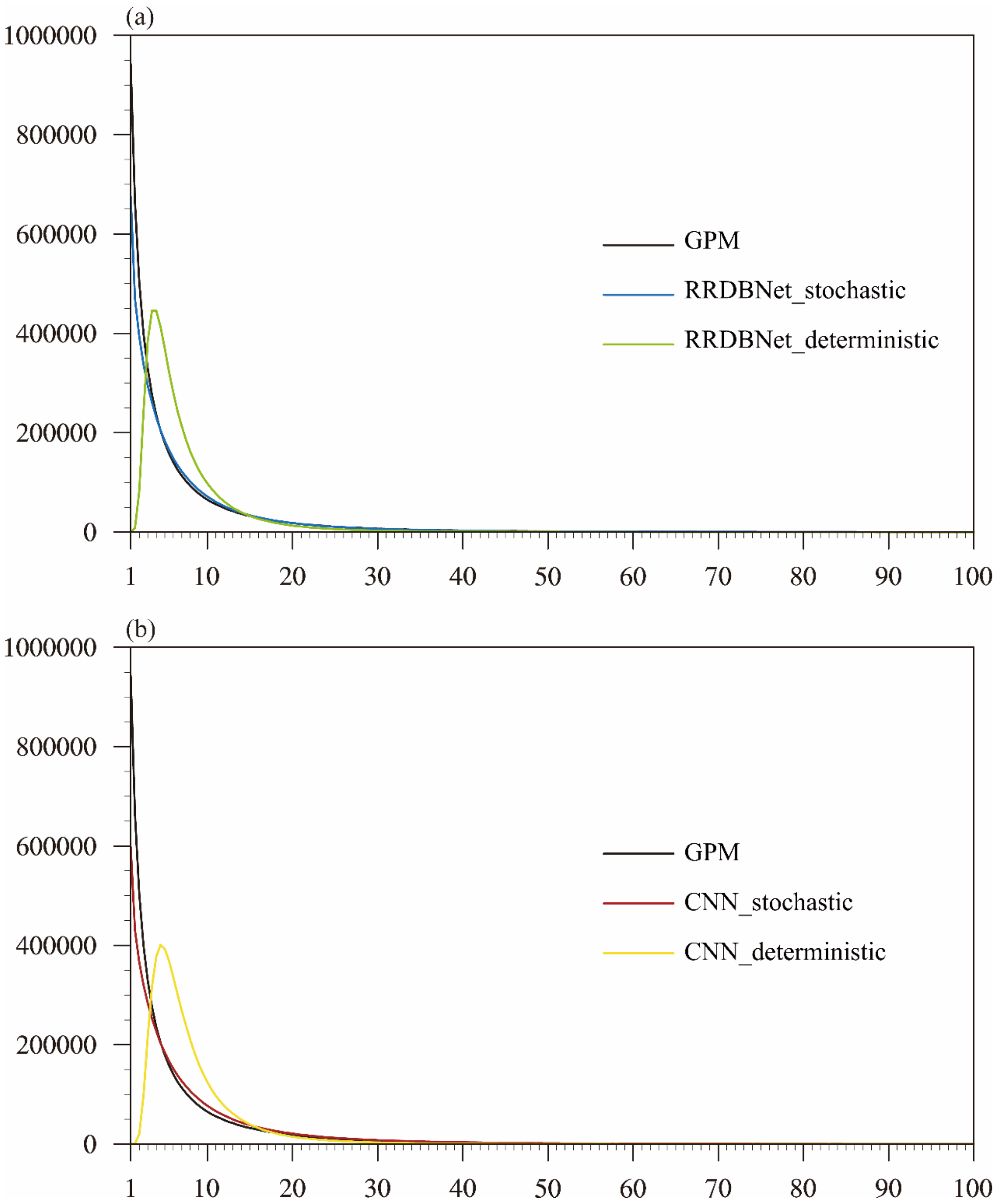

Figure R3. (a) Comparison of probability density functions of daily precipitation from 1 mm to 100 mm for all grids of GPM, and RRDBNet (stochastic and deterministic) in the MRYR from 2016 to 2020. (b) The same as (a) but GPM, and CNN (stochastic and deterministic).

Table R1. Evaluation metrics of daily precipitation in different models

Models	Difference	RMSE	CC
GLM	-0.12	11.65	0.17
CNN	0.07	6.30	0.39
RDBNet	0.03	6.55	0.37
RRDBNet	-0.10	5.93	0.40

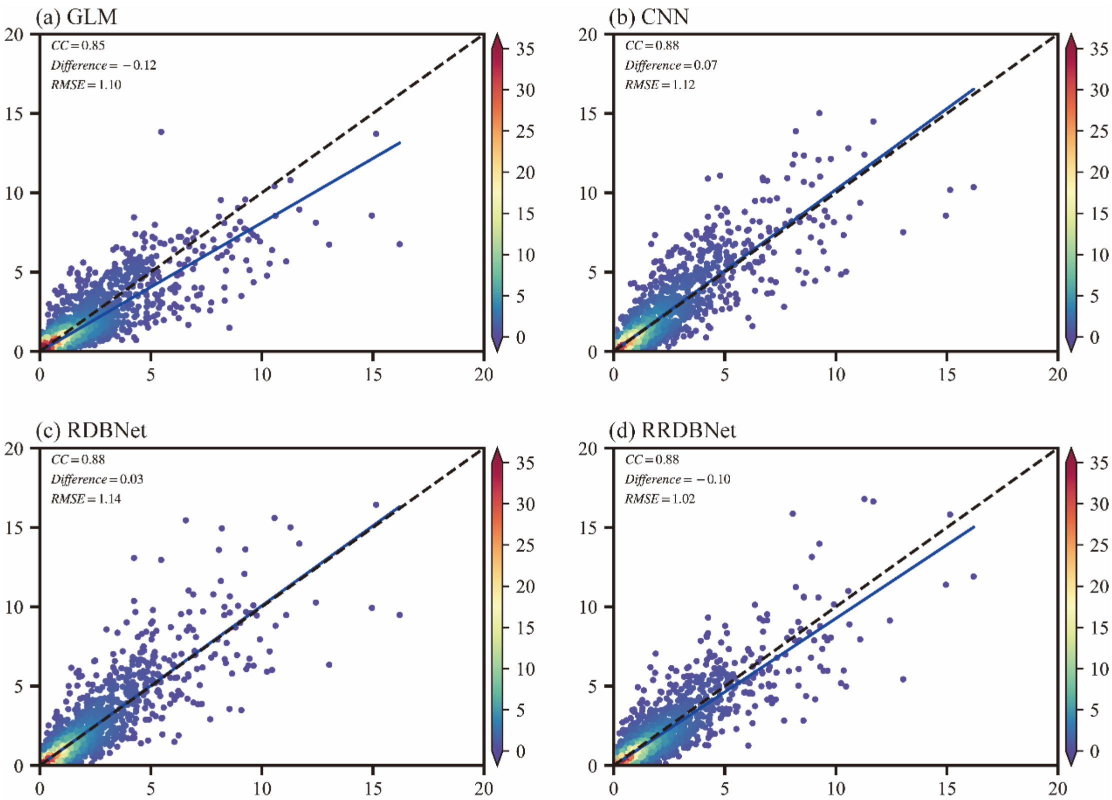

Figure R4. Scatter plot of daily precipitation after regional averaging in the MRYR from 2016 to 2020 for (a) GLM, (b) CNN, (c) RDBNet, and (d) RRDBNet.

Comment 2: The rationale of the proposed method is not clear. Compared to other deep learning methods, what are its advantages and why did the author propose this method? Without deep understanding the model itself, the manuscript gives readers limited insights.

Author response: Thanks for your comments and suggestions. To describe the rationale of the proposed RRDBNet more clearly, we have made corresponding changes to the methodological introduction in the revised manuscript (see Section 3). We have also explained the purpose and advantages of the proposed method in Instruction of the revised manuscript. Our purpose is to explore the performance of deep convolutional neural networks (CNNs) in precipitation downscaling and provide another idea and method for precipitation downscaling. The CNN proposed by Baño-Medina et al. (2020) has only 3 convolutional layers. In the field of computer vision, due to the small number of convolutional layers, on the one hand, these shallow CNNs may not be able to perform in-depth feature extraction and abstraction, and may not adequately capture complex patterns and relationships in the data; on the other hand, shallow CNNs do not have the ability to extract features at more hierarchical levels, and may be difficult to fully learn higher-level, more abstract, and more discriminative feature information and are easier to produce overfitting (Simonyan and Zisserman, 2015; Szegedy et al., 2015; LeCun et al., 2015; He et al., 2016; Huang et al., 2017). Therefore, in this paper, we explore the performance of deep CNNs in precipitation downscaling to provide another idea and method for precipitation downscaling. we develop a deep learning-based statistical downscaling method, i.e., Residual-in-Residual Dense Block based network (RRDBNet) model, to produce daily high-resolution precipitation over the MRYR region. Residual-in-Residual Dense Block (RRDB) is introduced in the proposed RRDBNet to optimize the structure of the network model. The multi-level residuals and dense connectivity strategies in RRDB help the model to adequately extract more abstract and advanced feature information (He et al., 2016; Huang et al., 2017; Zhang et al., 2018; Wang et al., 2018). Extensive experiments in this paper show that the proposed RRDBNet exhibits excellent performance in most metrics. This indicates that the proposed RRDBNet is effective and can be used as a reliable tool for precipitation downscaling.

Comment 3: The authors claimed the proposed method is much better than the other three methods, which may not be true. Given the stochastic nature of deep learning models and the slightly better statistics, the superiority may be purely due to stochasticity itself. Training the model multiple times may help discriminate the superiority and stochasticity. Furthermore, it is not fair to compare different deep learning models without giving model complexity (e.g., the number of trainable parameters).

Author response: We sincerely appreciate your valuable comments and suggestions. We aim to study the performance of deep CNNs in precipitation downscaling and provide another idea and method for precipitation downscaling. In this paper, we have conducted a large number of experiments. These experiments show that the proposed RRDBNet outperforms other models on most metrics, and not only on a single metric. We also repeated the experiment ten times for RRDBNet, and calculated the mean and variance of the ten experimental results in Tables R2 and R3. We found that the changes in the results of the ten experiments are very small and all outperformed other models. These may indicate that RRDBNet is an effective tool for downscaling precipitation. Furthermore, from the model structure analysis, the CNN proposed by Baño-Medina et al. (2020) has only three convolutional layers (Figure R1) and is a shallow CNN. Whereas, the multilevel residual and densely connected structure of RRDBNet (Figure R2) has been proven to capture more advanced, more abstract, and more discriminative features and more complex nonlinear relationships than these shallow CNNs in the field of computer vision (Simonyan and Zisserman, 2014; Szegedy et al., 2015; LeCun et al., 2015; He et al., 2016; Huang et al., 2017). We also compare the convergence of RRDBNet and CNN during training in subsection 4.4 of the original manuscript. The result is shown in Figure R5. We found that under the same experimental equipment and experimental parameter configuration, RRDBNet converges faster than CNN. Faster model convergence can save computational resources and time during training.

Table R2. Results of ten repeated experiments using RRDBNet.

	Difference	RMSE	CC
1	-0.09589823	0.29287103	0.93531325
2	-0.0973649	0.29267051	0.93549379
3	-0.09529275	0.29251509	0.93530962
4	-0.09570838	0.29507643	0.93433473
5	-0.09507929	0.29269518	0.93511726
6	-0.09596082	0.2931742	0.93505923
7	-0.09697484	0.29216068	0.93575122
8	-0.09673768	0.29315587	0.93529157
9	-0.09631896	0.29314359	0.93520193
10	-0.09771557	0.29449235	0.93481579
Ensemble	-0.0963±0.0008	0.2932±0.0009	0.9352±0.0004

Table R3. Evaluation metrics of annual mean precipitation in different models.

Models	Difference	RMSE	CC
GLM	-0.12	0.36	0.89
CNN	0.07	0.35	0.92
RDBNet	0.03	0.33	0.92
RRDBNet	-0.10	0.29	0.94
RRDBNet(10 times)	-0.0963±0.0008	0.2932±0.0009	0.9352±0.0004

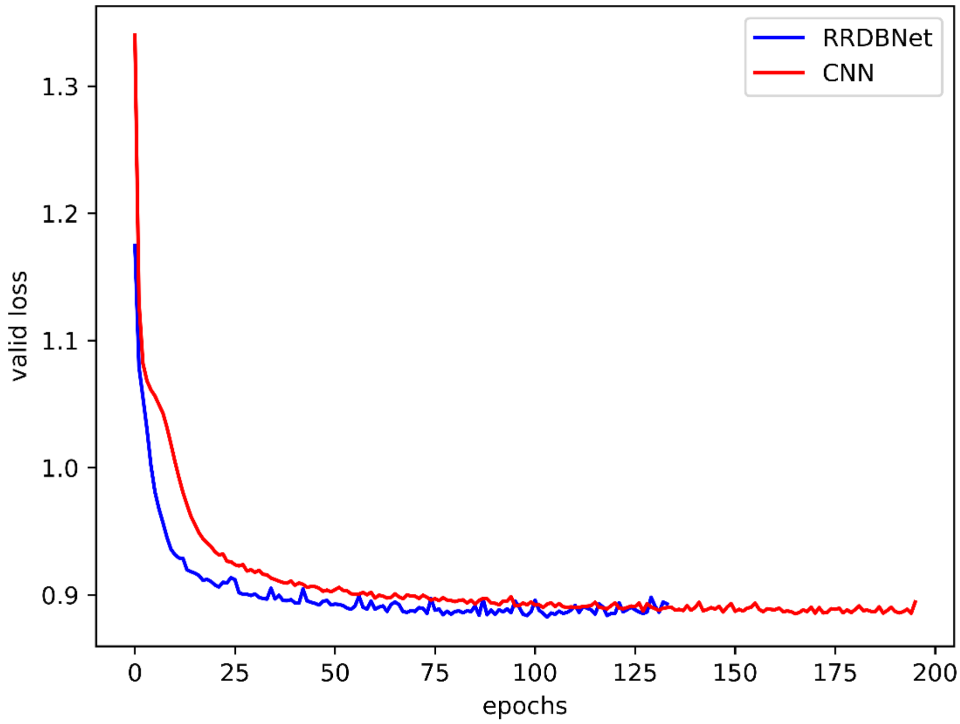

Figure R5. Validation loss of RRDBNet and CNN during training.

Comment 4: In the introduction section, the authors described many GCM downscaling. However, this study is not about GCM downscaling but reanalysis data, which is very different story and may mislead readers. Thus, the introduction needs to be rewritten.

Author response: Thanks for your nice suggestions. Different statistical downscaling (Maraun and Widmann, 2018) methods have been developed building on empirical relationships established between informative large-scale atmospheric variables (predictors) and local/regional variables of interest (predictands). Under the perfect-prognosis approach, these relationships are learned from (daily) data using simultaneous observations for both the predictors (from a reanalysis) and predictands (historical local or gridded observations), and are subsequently applied to GCM-simulated predictors (multidecadal climate change projections under different scenarios), to obtain locally downscaled values (see, e.g., Gutiérrez et al, 2013; Manzanas et al, 2018; Baño-Medina et al., 2020, 2022). Therefore, we first use RRDBNet to learn this statistical relationship from reanalysis data ($2^{\circ} \times 2^{\circ}$) and observation data ($0.1^{\circ} \times 0.1^{\circ}$), and then use the trained RRDBNet to act on the predictors ($2^{\circ} \times 2^{\circ}$) of GCM in the future. We have made instructions in Section 2 of the revised manuscript. Currently, we have also done experiments to use the trained RRDBNet for the GCM downscaling. The results of the experiment are shown in Figure R6. In variations of annual mean precipitation, the precipitation (MPI-ESM1-2-LR ($0.1° \times 0.1°$)) after the downscaling of RRDBNet is more consistent with the observation (CN05) than the precipitation of GCM (MPI-ESM1-2-LR ($2° \times 2°$)). This also proves that our RRDBNet is an effective and reasonable tool for downscaling precipitation.

The corresponding descriptions in the revised manuscript are as follows: **Under the "perfect-prognosis", the model learns statistical relationships from daily predictors (from reanalysis data) and predictands (from historical observations), and then works on the predictors of the GCM to obtain the corresponding regional or local downscaling results (Gutiérrez et al., 2013; Manzanas et al., 2018; Baño-Medina et al., 2020, 2022). Therefore, we selected predictors from ERA5 and used cumulative precipitation from Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) as predictand.**

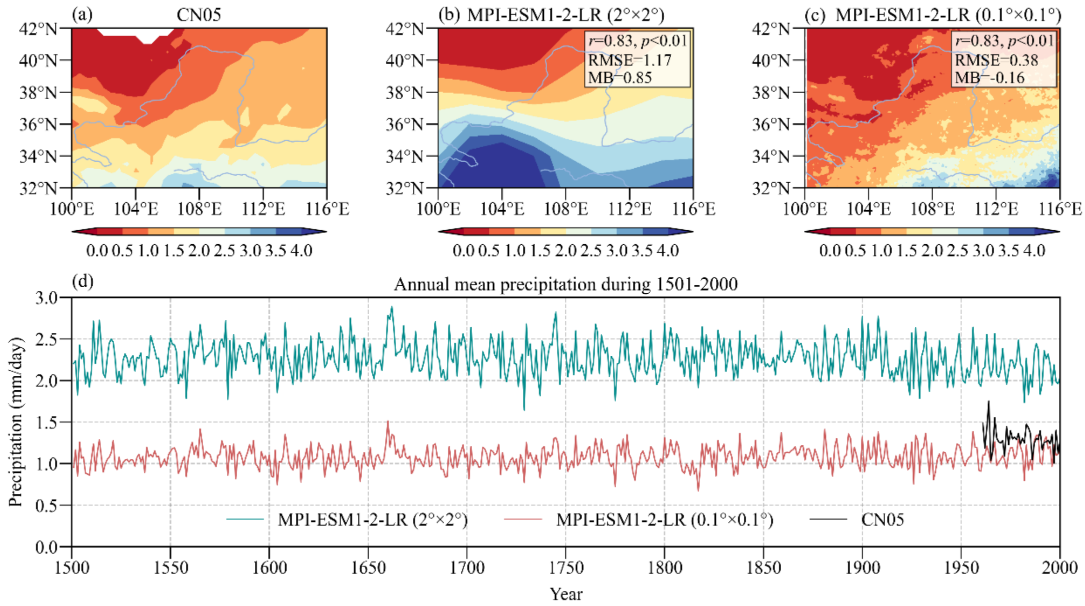

Figure R6. Spatial distribution of annual mean precipitation (mm/day) in the MRYR from 1501 to 2000 for (a) CN05, (b) MPI-ESM1-2-LR (2°×2°), and (c) MPI-ESM1-2-LR (0.1°×0.1°). (d) Variations of annual mean precipitation (mm/day) in the MRYR during 1501-2000 for CN05, MPI-ESM1-2-LR (2°×2°), and MPI-ESM1-2-LR (0.1°×0.1°).

Comment 5: Line 69: the reanalysis data ERA5 has spatial resolution of 0.25x0.25 degree not 2x2 degree. Furthermore, the authors selected 5 predictors without giving any reasons.

Author response: We sincerely appreciate your valuable comments. Note that most GCMs in CMIP5/6 have coarser resolution (almost 2 degree), we are going to downscale the GCM precipitation in the future. Therefore, we choose the ERA5 predictors ($2^{\circ} \times 2^{\circ}$) to train the model, which is upscaled from the default resolution 0.25×0.25 degree (as suggested by Baño-Medina et al. (2020)). The ERA5 data with $2^{\circ} \times 2^{\circ}$ can also be directly downloaded from https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5. These selected five predictors (Q, T, Z, U, and V) generally contain all the most of information related to the surface precipitation, which is consistent with the previous studies (Baño-Medina et al., 2020, 2022; Sun and Lan, 2022; Rampal et al., 2022). In addition, to facilitate the comparison between the proposed RRDBNet and the CNN proposed by Baño-Medina et al. (2020), so we choose the predictors selected by. i.e. the 5 predictors described in this paper. We have also clarified this point in Section 2 of the revised manuscript. The clarifications are as follows: **Specifically, the**

input dataset (predictor set) of statistic downscaling is derived from data with $2^{\circ} \times 2^{\circ}$ resolution in ERA5, which is upscaled from the default resolution 0.25×0.25 degree (consistent with the predictors selected by Baño-Medina et al. (2020)).

Comment 6: Line 89: the three parameters came out first time without any explanations. Y ∈ Rt°×100°×159 came out without further information.

Author response: Thanks for your great suggestions on improving our manuscript. Following your suggestion, we have made the explanation when these three parameters (p, α, and β) were first mentioned in Subsection 3.1 of the revised manuscript. $Y \in \mathbb{R}^{t\times100\times159}$ should be written as $Y \in \mathbb{R}^{100\times159}$. We have fixed this error in Subsection 3.1 of the revised manuscript. Y represents the studied precipitation area. 100×159 is the grid matrix after precipitation downscaling. The modification is as follows: **Overall, the predictors are processed by the statistical downscaling model (RRDBNet) to obtain the modeled parameters p, α, and β. Then precipitation is estimated using a mixed binomial-lognormal distribution with modeled p, α, and β. This is consistent with Baño-Medina et al., (2020, 2022), Sun et al., (2021), and Rampal et al., (2022). p denotes the probability of precipitation, and α and β represent the shape and scale of the gamma distribution, respectively. Finally, the precipitation $Y \in \mathbb{R}^{100\times159}$ is obtained by calculating the parameters p, α, and β (Baño-Medina et al., 2020, 2022; Sun and Lan, 2021; Rampal et al., 2022).**

Comment 7: Lines 114 to 119: Does the statement about batch normalization come from model testing? If that is true, this statement needs to be included in the results section. If it is not true, where this come from?

Author response: Thank you very much for your comments. We originally meant that there is no Batch Normalization (BN) layer (Ioffe and Szegedy, 2015) in the employed Residual-in-Residual Dense Block (RRDB) structure (Wang et al., 2018). The reason is that the BN layer may introduce artifacts that limit the generalization ability of the model when the statistics of the model's training and test sets are different. Therefore, there is no BN layer in our proposed RRDBNet. To avoid ambiguity and better express our meaning, we have deleted this content in the third paragraph of subsection 3.1.1 of the revised manuscript. In this paper, both training and test data are standardized before being input into the model.

Comment 8: Line 121: how did you get B=0.2?

Author response: Thank you very much for your comment. We set B to 0.2 based on the paper by Wang et al. (2018). We have added the corresponding reference for clarification in Subsection 3.1.1 of the revised manuscript. The modification is as follows: **in this paper we set B=0.2 (Wang et al., 2018).**

Comment 9: Line 139: The statement "The final precipitation is obtained by multiplying p with the random values of the distribution with shape α scale β." Why?

Author response: Thank you for your comments. Due to the mixed discrete-continuous nature of precipitation, the model predicts the precipitation by modelling the parameters of Bernoulli-Gamma distribution (Williams, 1997; Cannon, 2008; Baño-Medina et al., (2020, 2022), Sun et al., (2021), and Rampal et al., (2022)). The precipitation estimated through a mixed binomial-lognormal distribution of which the corresponding parameters (p, α, β) are modelled (Baño-Medina et al., (2020, 2022), Sun et al., (2021), and Rampal et al., (2022)). We have made an explanation of this in Subsection 3.1 of the revised manuscript. The explanation is as follows: **Overall, the predictors are processed by the statistical downscaling model (RRDBNet) to obtain the modeled parameters p, α, and β. Then precipitation is estimated using a mixed binomial-lognormal distribution with modeled p, α, and β. This is consistent with Baño-Medina et al., (2020, 2022), Sun et al., (2021), and Rampal et al., (2022). p denotes the probability of precipitation, and α and β represent the shape and scale of the gamma distribution, respectively. Finally, the precipitation $Y \in \mathbb{R}^{100\times159}$ is obtained by calculating the parameters p, α, and β (Baño-Medina et al., 2020, 2022; Sun and Lan, 2021; Rampal et al., 2022).**

Comment 10: Line 142: "In the training phase" was repeated.

Author response: Thanks for pointing out the mistake. We have modified it in Subsection 3.1.3 of the revised manuscript. The modification is as follows: **In the training phase, when the model outputs predicted values by forward propagation, the loss function calculates the**

difference between the predicted and actual values, i.e., the loss value.

References

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

Baño-Medina, J., Manzanas, R., Cimadevilla, E., Fernández, J., González-Abad, J., Cofiño, A. S., and Gutiérrez, J. M.: Downscaling multi-model climate projection ensembles with deep learning (DeepESD): contribution to CORDEX EUR-44, Geosci. Model Dev., 15, 6747–6758, 2022.

Bürger, G.: Expanded downscaling for generating local weather scenarios, Clim. Res., 7, 111–128, 1996.

Bürger, G.: Selected precipitation scenarios across Europe, J. Hydrol., 262, 99–110, 2002.

Cannon, A. J.: Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli-Gamma Density Network, J. Hydrometeorol., 9, 1284–1300, 2008.

Gutiérrez, J. M., San-Martín, D., Brands, S., Manzanas, R., and Herrera, S.: Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions, J. Clim., 26, 171–188, 2013.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit, pp. 770–778, 2016.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q.: Densely connected convolutional networks, in: Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit, pp. 4700–4708, 2017.

Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Int. Conf. Mach. Learn., ICML, pp. 448–456, pmlr, 2015.

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

Manzanas, R., Lucero, A., Weisheimer, A., & Gutiérrez, J. M. (2018). Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts?. Climate dynamics, 50, 1161-1176.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate

Research, Cambridge University Press, 2018.

Rampal, N., Gibson, P. B., Sood, A., Stuart, S., Fauchereau, N. C., Brandolino, C., Noll, B., and Meyers, T.: High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand, Weather and Climate Extremes, 38, 100 525, 2022.

Schoof, J. T. and Pryor, S.: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, Int. J. Climatol., 21, 773–790, 2001.

Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, 2015.

Sun, L. and Lan, Y.: Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods, International Journal of Climatology, 41, 1128–1147, 2021.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going deeper with convolutions, vol. 07-12-June-2015, pp. 1 – 9, 2015.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C.: ESRGAN: Enhanced super-resolution generative adversarial networks, in: Lect. Notes Comput. Sci., pp. 0–0, 2018.

Williams, P.: Modelling Seasonality and Trends in Daily Rainfall Data, NIPS, 10, 1997.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y.: Residual Dense Network for Image Super-Resolution, in: Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit, pp. 2472–2481, 2018.