Responses to Reviewer #1

We are grateful to reviewer #1 for his/her constructive comments and suggestions which are helpful to improve the quality of our manuscript. And we will make great efforts to address all the comments, with the details explained as follows.

General comment: Presentation of the performance of BMA historical projections: I have the feeling that BMA is slightly oversold. I agree that it performs better than AEM, but cases where it does not perform better are most of the time not mentioned. I spotted some of them in my miscellaneous remarks. I think this is a pity, as understanding and explaining why this is the case could be very interesting. I generally feel that the manuscript lacks a lot of analysis and discussion of the results.

Response: According to the reviewer's constructive comments, we will revise the manuscript with additional explanations and discussions. Our detailed responses to the reviewer's comments are provided as follows.

Comment #1: Validity of the method under climate change: the authors analyse the performance of the method under historic data. However, the method is then applied to future climate projections. As the change of climate is rather important in the future, the validity of the method in terms of extrapolation must be questioned. Usually, especially in hydrology under change (see the many "split sample" papers and applications of this method), the historic data is split into two parts, one part being used for optimising the parameters of the model/relation and the second part being used as an independent evaluation. While this does not guarantee that the method might be performing well in the future (since the historic data might be less contrasted that future data), this is a condition that is necessary to satisfy. If the method does not perform well on historic data that is independent, then it is very likely that its behaviour will not be satisfying in the future.

Response: To address the reviewer's comment, a split-sample test was performed to evaluate the robustness of BMA weights. The BMA climate simulations were calibrated during the 20-year period from 1975 to 1994, and then validated during the 10-year period from 1995 to 2004 by comparing against the CRU observations. Figures R1 and R2 (as shown below) depict the spatial patterns of absolute model biases of annual and seasonal precipitation and PET, respectively, generated from the BMA and ensemble mean (AEM) simulations. Results show that there is no significant difference between model biases in the calibration (1975−1994) and validation (1995−2004) periods for the BMA ensemble simulation. The BMA approach significantly reduces the model biases in the calibration and validation periods except for the winter mean precipitation over Southeast China, which has a significant wet bias (Figures R1e and R1f). Although the BMA-simulated summer mean precipitation has a dry bias, it has been largely reduced compared against the AEM simulation. We also observe that the BMA-simulated PET is highly consistent with the CRU PET observations for the calibration and validation periods. This indicates that the robustness of the BMA climate simulations is acceptable for the study area and can lead to reasonable climate projections.

To better address the reviewer's comment, we will add the abovementioned discussions in the revised version of the manuscript and will also investigate the cause of the wet bias in the BMA-simulated winter precipitation.

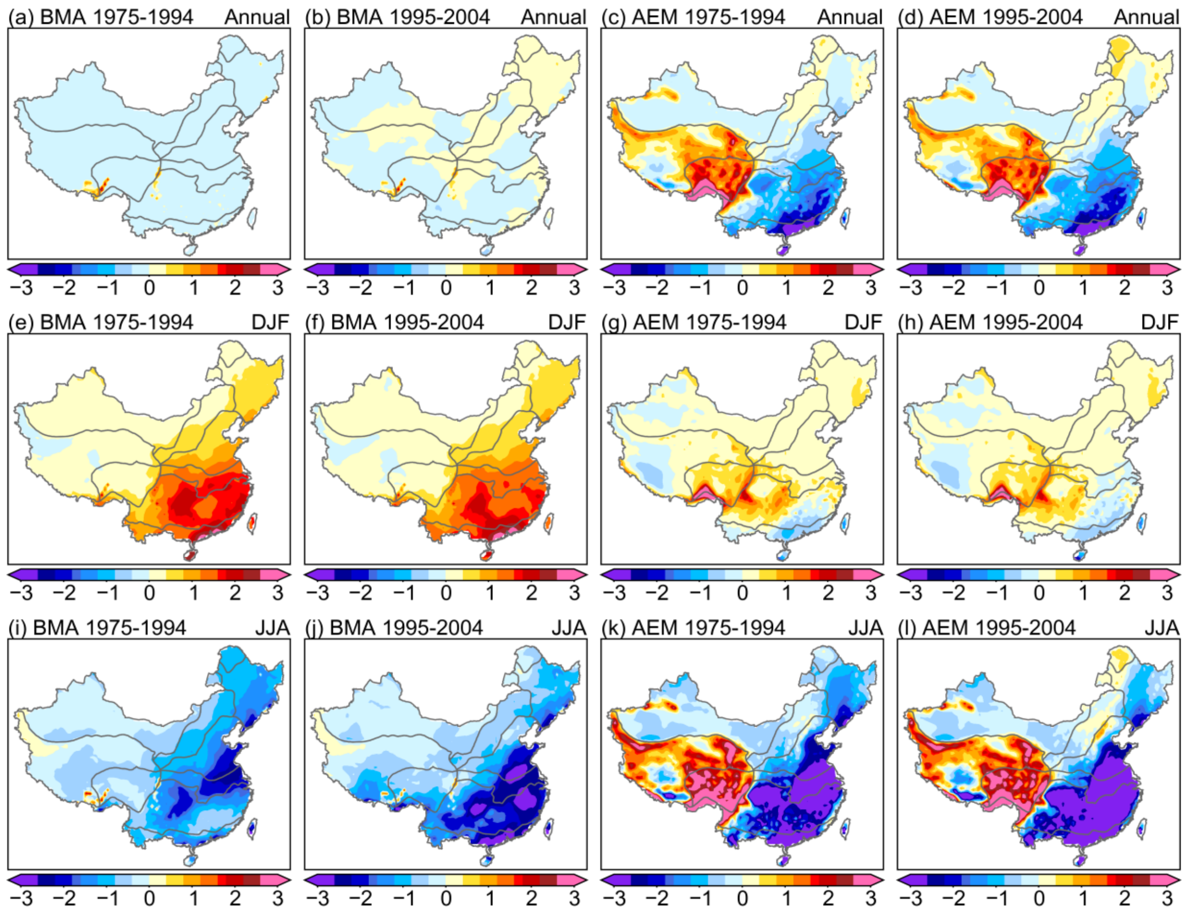

Figure R1. Spatial patterns of absolute model biases of annual and seasonal mean precipitation generated from the BMA and ensemble mean (AEM) simulations over a period of 30 years (1975−2004). The BMA weights calibrated during 1975−1994 (**a**, **e**, **i**) were used to validate the precipitation simulation during 1995−2004 (**b**, **f**, **j**).

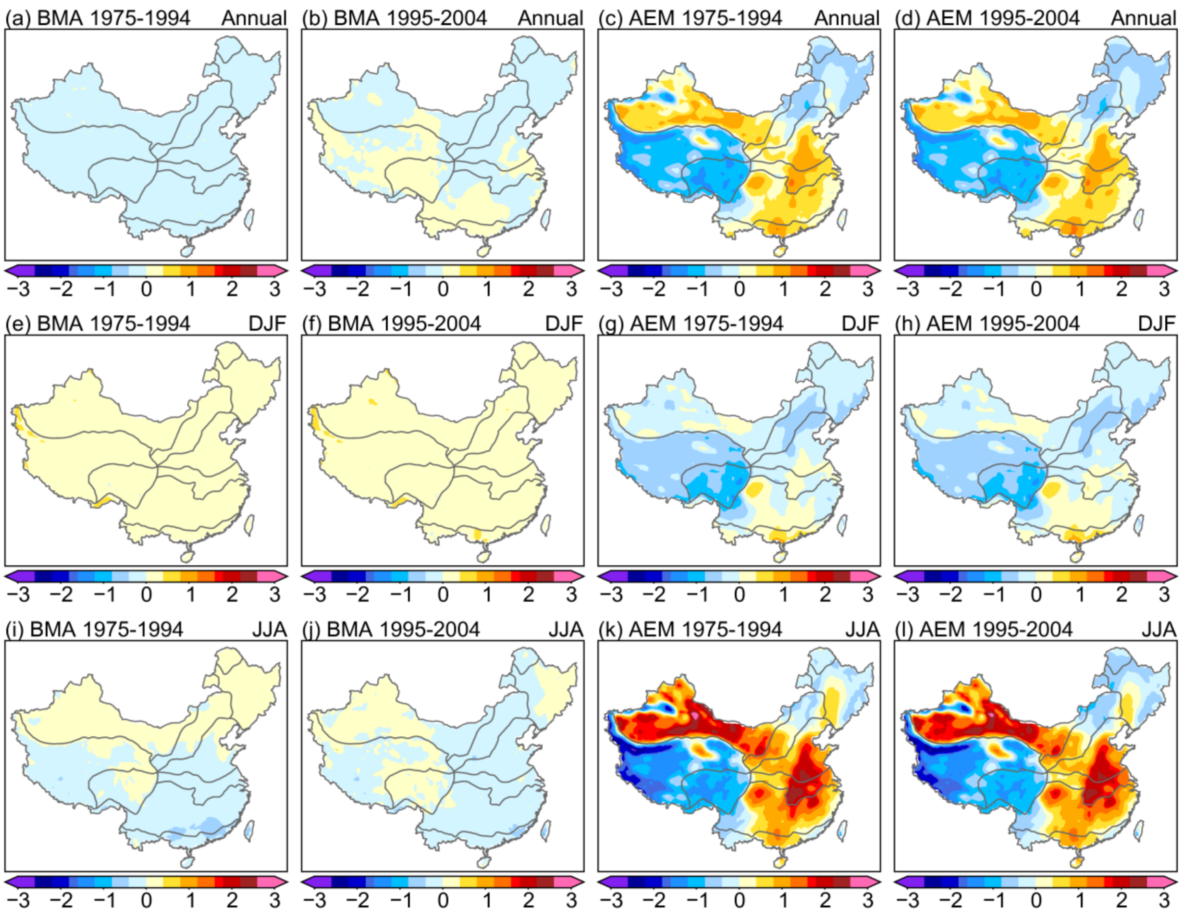

Figure R2. Spatial patterns of absolute model biases of annual and seasonal mean PET generated from the BMA and ensemble mean (AEM) simulations over a period of 30 years (1975−2004). The BMA weights calibrated during 1975−1994 (**a**, **e**, **i**) were used to validate the precipitation simulation during 1995−2004 (**b**, **f**, **j**).

Comment #2: One point that became evident to me when seeing the figures were even not mentioned by the authors. Namely, the two RCPs show very close results with BMA! Check for example Figure 9 (precipitation): if you compare RCP4.5 and RCP8.5 for a specific indicator, the spatial pattern is very similar and magnitudes are also very close. For PET, this is a bit less the case, but still, strongly similar spatial patterns can be observed. Later on, in Figure 14, several boxplots of the drought duration are identical for both RCPs. This is a bit worrying I would say, as I doubt that raw projections also show this behaviour. Could it be that BMA constraints too much projections? To understand that, similar plots without BMA could help.

Response: To address the reviewer's comment, we assessed future changes in precipitation and PET generated from the ensemble mean climate simulations. Figures R3 and R4 (as shown below) depict the absolute and relative changes for annual, winter (DJF) and summer (JJA) mean precipitation and PET, respectively, generated from the ensemble mean climate simulations between past and future climates under RCP4.5 and RCP8.5. Results show that ensemble mean simulations lead to a large difference between RCP4.5 and RCP8.5 in projecting future changes in the precipitation and PET. For example, the winter precipitation is projected to increase by more than 25% in North China under RCP4.5 (see Figure R3h), but such an increase is further intensified under RCP8.5 (see Figure R3k). The summer PET projected under RCP8.5 is also significantly larger than under RCP4.5. Therefore, the ensemble mean climate projections under RCP4.5 and RCP8.5 can show different spatial patterns.

To better address the reviewer's comment, an ensemble mean drought projection was performed to compare with the BMA-based drought projection. Figure R5 depicts the box-and-whisker plots of the drought duration, severity, and frequency (i.e., the number of drought episodes) generated from the CRU observations and the ensemble mean simulations for the past and future climates over the 10 climate divisions. Results show that the drought duration, severity, and the number of drought episodes are projected to increase under RCP8.5 compared against the historical simulations for most climate divisions in China. In comparison, the change signal under RCP4.5 is not as strong as RCP8.5. The drought severity and duration are projected to remain unchanged or decline under RCP4.5 compared against the historical simulations for several climate divisions (e.g., Divisions 4, 5, 8, and 9). Figure R5 also indicates that the drought evolution in China for the past climate is not well reproduced by the ensemble mean simulations, especially for the drought frequency that is generally underestimated. This motivates us to use the BMA technique to improve the reliability of drought projections. Our findings reveal that the BMA technique can improve the regional climate simulations, but it did not lead to an all-round enhancement upon the AEM approach. There is a substantial difference between the AEM and BMA techniques in projecting future changes in drought characteristics. Therefore, we will further discuss the difference between the AEM- and BMA-based drought projections in the revised version of the manuscript.

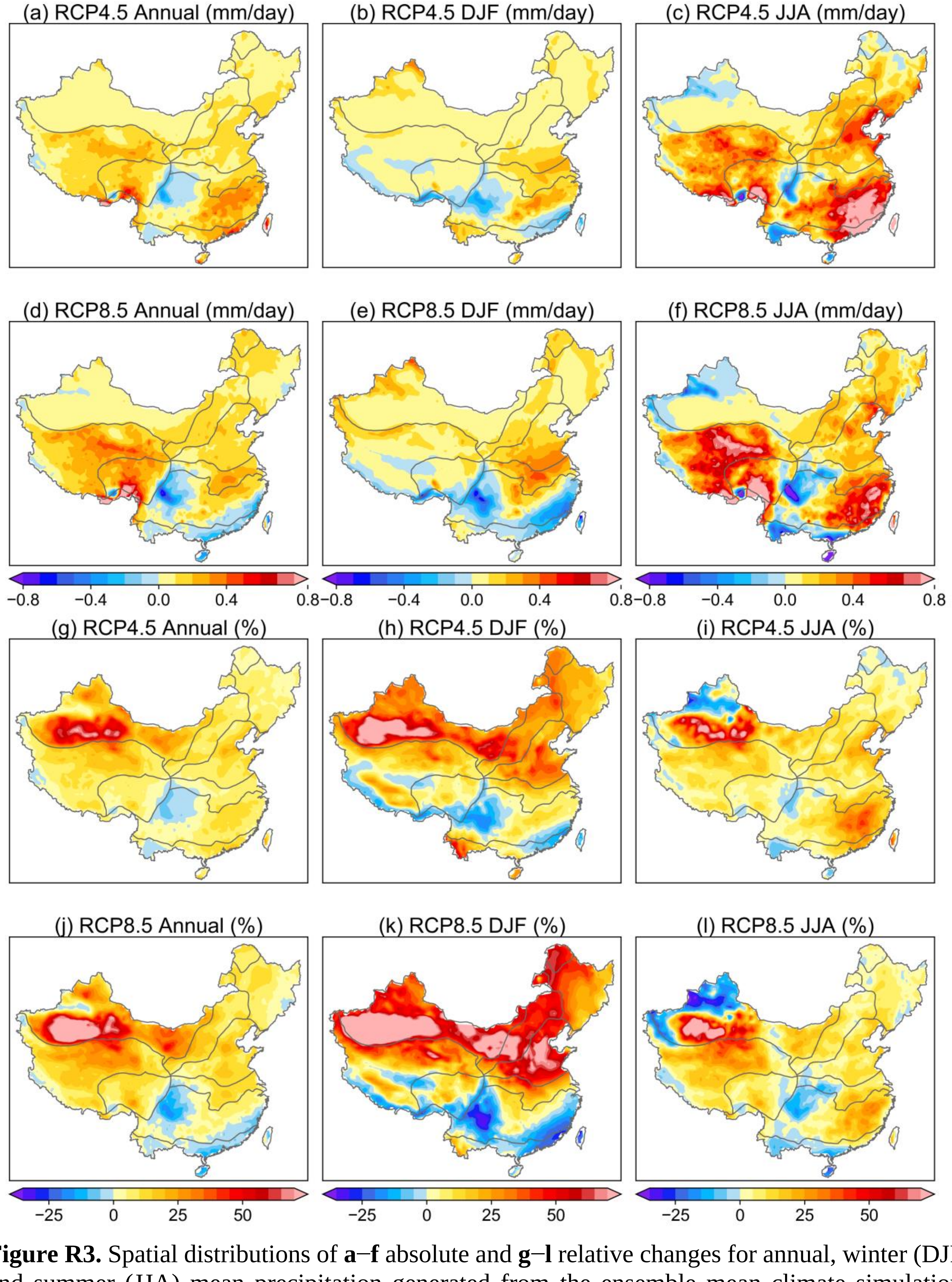

Figure R3. Spatial distributions of **a**−**f** absolute and **g**−**l** relative changes for annual, winter (DJF) and summer (JJA) mean precipitation generated from the ensemble mean climate simulations between past and future climates under RCP4.5 and RCP8.5.

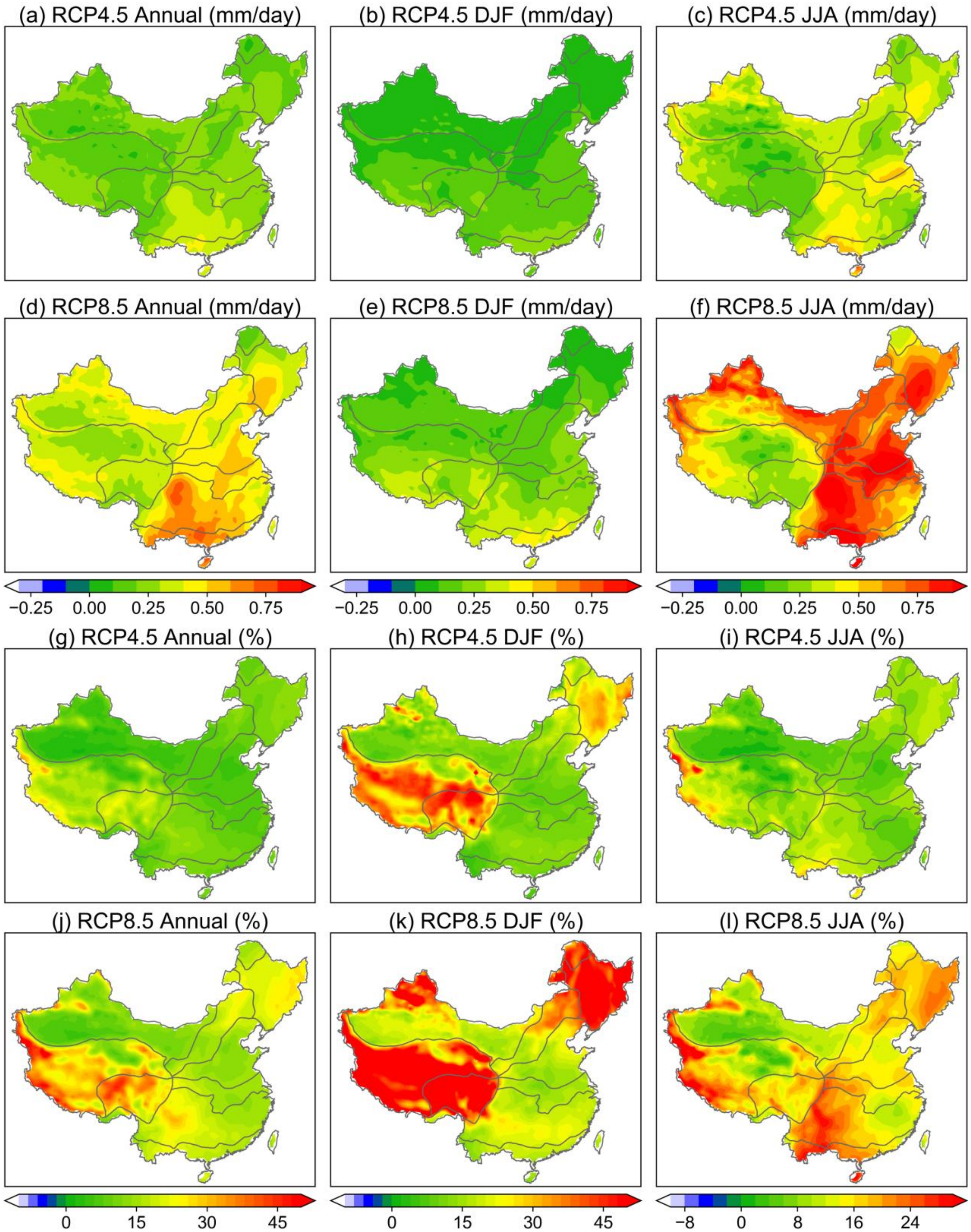

Figure R4. Spatial distributions of **a**–**f** absolute and **g**–**l** relative changes for annual, winter (DJF) and summer (JJA) mean PET generated from the ensemble mean climate simulations between past and future climates under RCP4.5 and RCP8.5.

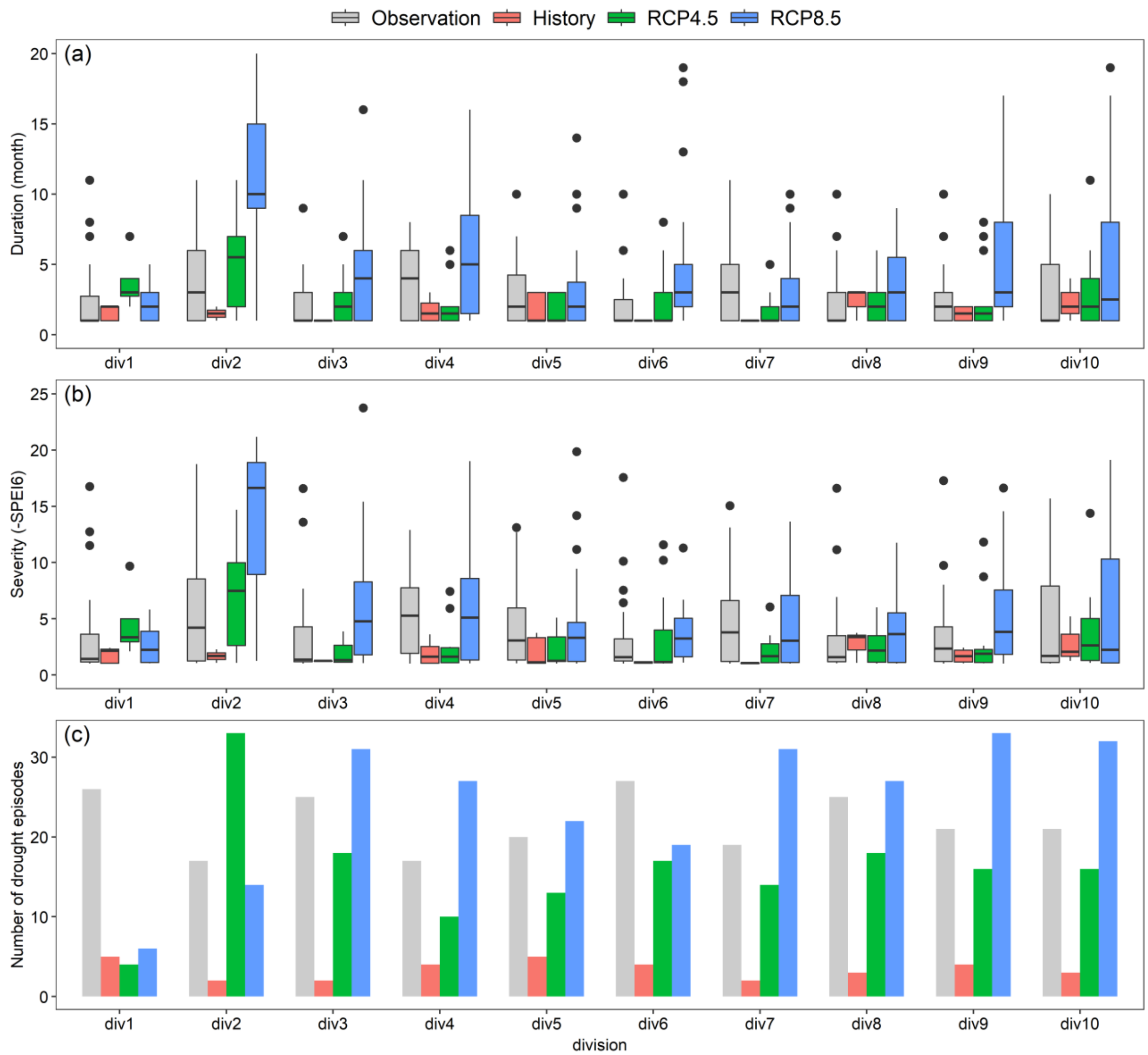

Figure R5. Box-and-whisker plots of the drought duration and severity as well as the number of drought episodes generated from the CRU observations and the ensemble mean simulations for the past and future climates over the 10 climate divisions. The thick black horizontal bars represent the median value, and the lower and upper edges of the box represent the 25th (Q1) and 75th (Q3) percentile values, respectively. The upper and lower whiskers represent the values of Q3 + 1.5 × IQR and Q1 − 1.5 × IQR, respectively, where IQR denotes the interquartile range that is equal to Q3 – Q1. The values beyond the end of the whiskers are indicated by outlier points.

Comment #3: A second point is seen in Figure 15: some RCPs boxplots (well many in fact) are flat. This means that all projections give the same number of drought episodes. While this could be for an indicator showing small number (e.g. <5), here we are often around 30! This is a bit strange too.

Response: To address the reviewer's comment, we will further analyze the difference of drought projections generated from the AEM and BMA techniques.

Comment #4: Abstract, line 11: that is not entirely true, as some adaptation strategies are dedicate to tackle flood issues. Line 22-23: which aspect of the risk is expected to double? Frequency? Duration? Intensity? Please be more specific here.

Response: According to the reviewer's comment, the statement has been revised as follows.

> "Understanding future drought risk plays a crucial role in developing climate change adaptation strategies and in enhancing disaster resilience. In this study,…The estimated joint risks from the drought duration and drought severity in China are expected to become more than double under both emission scenarios…."

Comment #5: Line 34 and others: there are some surnames in the citations, please correct.

Response: These have been corrected in the revised version of the manuscript.

Comment #6: Line 36-38: using ensembles and not restraining the analysis to the mean is already standard in hydrology, including for droughts. Please check for example some examples in HESS: Vidal et al. (2016) and Parajka et al. (2016).

Response: We appreciate the reviewer's comment. The statement has been revised as follows.

> "Each individual RCM has strengths and weaknesses in characterizing the hydroclimatic regimes. Thus, a multi-model ensemble simulation is commonly used to improve the reliability of drought projections. The arithmetic ensemble mean (AEM) of drought variables (e.g., precipitation) and the inter-model spread derived from multiple RCMs are widely used to assess climate change impacts on regional droughts, which has been proven to enable more accurate simulation than a single climate model (Parajka et al., 2016; Rajsekhar and Gorelick, 2017; Vidal et al., 2016)."

Comment #7: Line 56: while the statement is true, Ramos et al. (2013) is about forecasts, not projections. Extrapolating the conclusion from Ramos to projections is far from trivial I believe. I suggest removing this citation.

Response: According to the reviewer's suggestion, the citation has been removed.

Comment #8: Line 131: please define the pdf acronym.

Response: The pdf acronym represents the probability density function, which has been defined in the revised version of the manuscript.

Comment #9: Line 232: I quite disagree regarding precipitation. On Figure 4l, we see rather high differences on Eastern China.

Response: We agree with the reviewer that the BMA-simulated summer precipitation has a dry bias in Southeast China, but the bias has been substantially reduced compared with the AEM simulation. To address the reviewer's comment, the statement has been revised as follows.

> "In general, there are considerable discrepancies between the AEM simulations and the CRU observations in reproducing the spatial pattern of the mean precipitation and PET, especially for the summer season. Compared to the AEM simulations, the BMA ensemble simulations have significantly lower absolute model biases except for the winter mean precipitation over Southeast China."

Comment #10: Line 250: please define COR.

Response: COR indicates the correlation coefficient. This definition has been added in the revised version of the manuscript as follows.

> "The simulated pattern agrees better with observations if the model has a higher correlation coefficient (COR) and a more consistent standard deviation (SD) with the observation, as well as it lies nearer the "OBS"

Comment #11: Line 252-254: this is not so clear for the other ones. See for example the winter PET.

Response: According to the reviewer's constructive comment, the statement has been deleted and the entire paragraph has been revised as follows.

> "To further evaluate the performance of AEM and BMA in simulating the annual, summer and winter mean precipitation and PET, the Taylor Diagram is used to quantify the consistency between the patterns from two simulations and the CRU observation (Taylor, 2001). The simulated pattern agrees better with observations if the model has a higher correlation coefficient (COR) and a more consistent standard deviation (SD) with the observation, as well as it lies nearer the "OBS". Figure 6 presents the relative performance of AEM and BMA in simulating the annual, summer, and winter mean precipitation and PET for 10 climate divisions in China. The solid grey circles around the "OBS" indicate the normalized root-mean-square error (NRMSE) between the climate simulations and the CRU observations. In general, the BMA simulations have higher CORs than the AEM simulations for both annual and seasonal results. For example, the CORs for the AEM-simulated annual precipitation are below 0.9 for the 10 climate divisions, and nearly all of them are larger than 0.9 for the corresponding BMA simulation (Fig. 6a). The BMA simulation also generally lowers the NRMSE compared to the AEM simulation. For example, the AEM-simulated summer PET for Divisions 8 and 9 have the NRMSE of ~1.5 and ~1.25 mm/day, while they drop to ~0.75 mm/day (Fig. 6f). In addition, the AEM simulation has a poor performance of reproducing the winter and summer mean precipitation, as well as the summer mean PET since nearly all of their CORs are smaller than 0.6 (see Figs. 6c, 6e, and 6f). But the corresponding CORs for the BMA simulation are increased for all the climate divisions, especially for the summer mean precipitation (Fig. 6e). It should be also noted that the BMA simulation does not necessarily outperform the AEM simulation in terms of the variation amplitude, but the CORs are generally improved, leading to lower NRMSE values (e.g., Fig. 6d). For example, the BMA simulation is not as reliable as the AEM simulation in reproducing the variation amplitude of the winter PET for Division 9, but the CORs are improved from ~0.65 to ~0.85, leading to the decline of NRMSE from ~0.75 to ~0.6 mm/day. Therefore, it is evident that the BMA simulation outperforms the AEM simulation in simulating the annual, winter, and summer precipitation and PET for 10 climate divisions in China."

Comment #12: Line 252-254 and lines 260-262 are contradictory

Response: We regret for the contradictory statement. The statement in Line 252-254 has been deleted and the entire paragraph has been revised (see our response to *Comment #11*).

Comment #13: Line 264-265: I rather disagree; there is quite often a factor 2 between OBS and BMA. In addition, some divisions show a peak timing that is not adequately represented.

Response: We agree with the reviewer that the BMA simulation did not completely correct the error in the climate simulation. Thus, it is desired to include more climate simulation datasets and improve the climate model averaging technique. To address the reviewer's comment, the statement has been deleted in the revised version of the manuscript.

Comment #14: Line 267-268: nice to finally see some attempt of discussion of the results. However, I feel we need more: why is this the case? Explain! Please also provide a reference.

Response: We appreciate the reviewer's comment, and agree that the explanation is not enough in the manuscript. We will add discussions and relevant references in the revised version of the manuscript.

Comment #15: Line 269-270: again, I think that the presentation is unfair: the number of opposite behaviour is rather similar from what I see on the figure.

Response: We appreciate the reviewer's comment, and agree that the presentation is too strong and inappropriate. We will revise the statement accordingly.

Comment #16: Line 306-313: this is methods, not results. Please move that part in Methods.

Response: This part has been moved into section 2.3 in the revised version of the manuscript.

Comment #17: Line 312-313: there is no justification why these 3 copulas were chosen for these divisions.

Response: The optimal copula families were selected according to the AIC values. The statement has been revised as follows.

> "The MCMC simulations showed that the Marshall-Olkin copula was optimal for describing the dependence between drought severity and duration in Divisions 1−3 and 8 according to the AIC values, while the Clayton and Gumbel copulas were chosen for Divisions 4−7 and Divisions 9−10, respectively."

Comment #18: Line 352-354: please remove, this is in the caption already

Response: It has been removed.

Comment #19: Line 362: occurrenceS

Response: It has been corrected.

Comment #20: References: Chambers et al. Is it a book? Please specify the type of work.

Response: We regret for the unclear reference. Chambers et al. (2018) is a book. The type of work has been specified in the reference list as shown below.

> "Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A.: Graphical methods for data analysis, Chapman and Hall/CRC Press, New York, 2018."

Comment #21: Figure 1: we miss a legend in order to have the possibility to understand panels i and j.

Response: The legends in Figures 1i and 1j have been added according to the reviewer's suggestions. Figure R6 (as shown below) is the revised version of Figure 1.

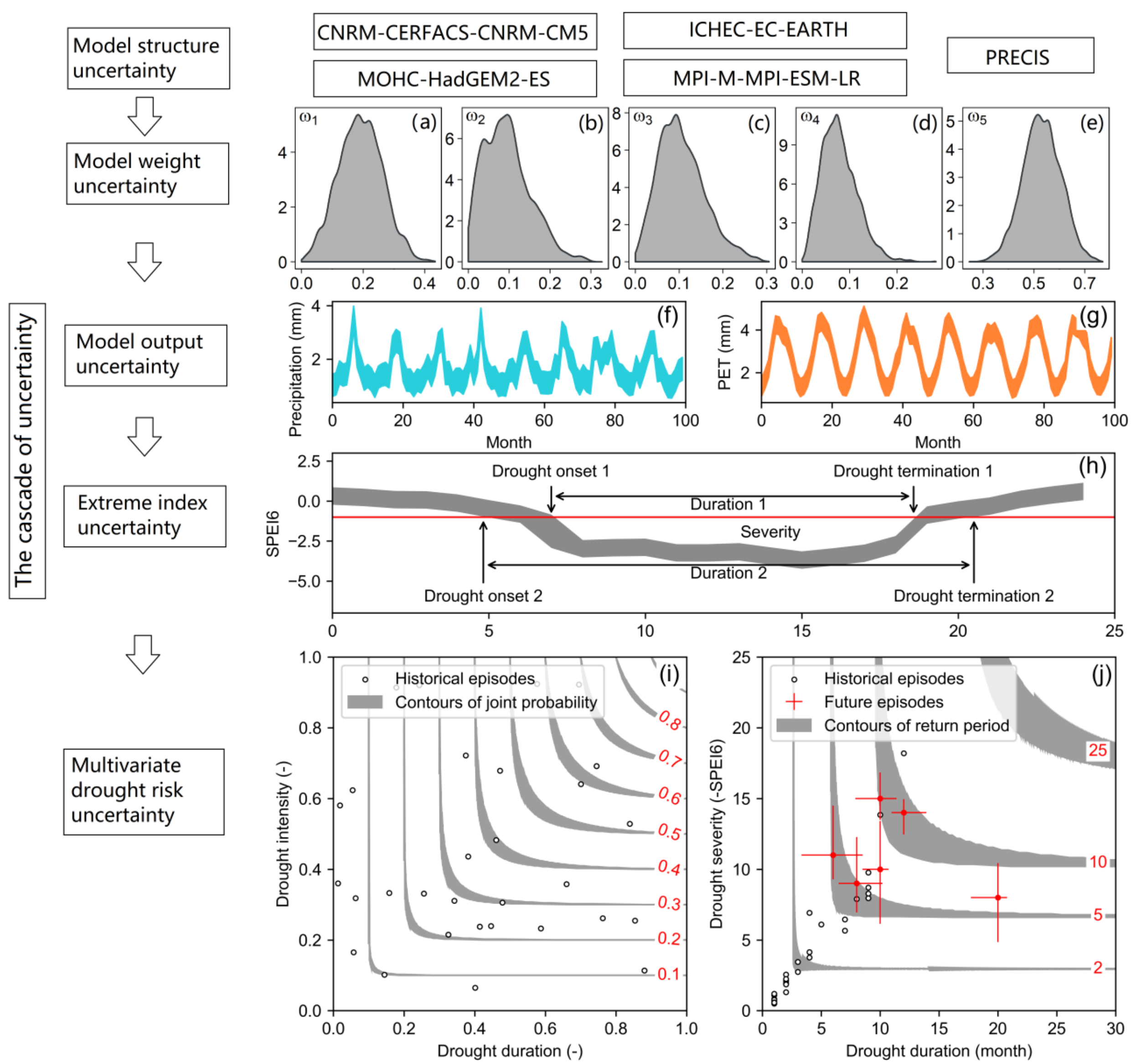

Figure R6. Schematic of the two-stage Bayesian multi-model framework for probabilistic multidimensional drought risk projections. The model structure uncertainty consists of five RCM datasets. **a**−**e** denote the model weight uncertainty, i.e., the posterior distributions of BMA weights derived from the MCMC simulation. **f** and **g** represent the BMA-derived precipitation and potential evapotranspiration (PET), respectively, as well as their uncertainty ranges. **h** denotes the uncertainty range of drought index (SPEI used in this study) owing to the uncertainty in the BMA-derived precipitation and PET, leading to a probabilistic delineation of drought episodes. **i** represents the contours of the copula-based joint probability with uncertainty intervals between drought characteristics (drought duration and severity) derived from the MCMC simulation, leading to the uncertainty in the return period of drought episodes as shown in **j**. The red "whiskers" in **j** represent the uncertainty of drought characteristics resulting from the climate projection.

Comment #22: Figure 2: while there is only one colour scale, it seems to me that the flat low lands in Eastern China are plotted in different greens. Can you please check?

Response: We have double checked the 1-km DEM dataset and the color scheme in Figure 2, which is consistent with China elevation maps in previous studies. For example, Figures R7a and R7b (as shown below) depict China elevation maps reprinted from Zhu et al. (2018) and Ma et al. (2009), respectively. It can be seen that lowland plains in Eastern China are plotted in similar colors.

To better address the reviewer's comment, we have revised Figure 2a with a finer color scale (as shown in Figure R7d) and compared it with the original version of Figure 2a (as shown in Figure R7c). Although different color schemes are used, Figures R7c and R7d show the same elevation pattern in Eastern China. Therefore, Figure 2 remains unchanged in the revised version of the manuscript.

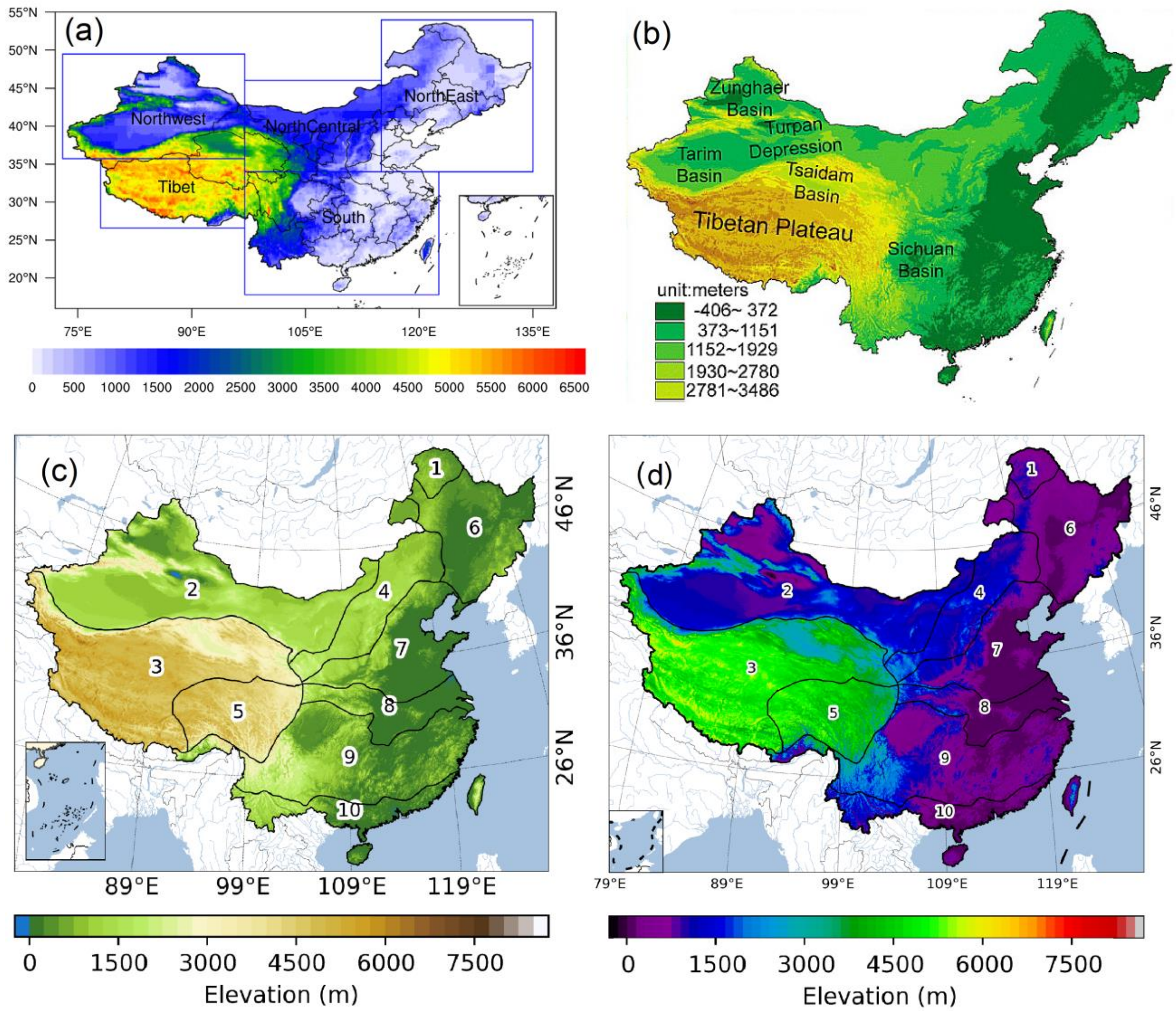

Figure R7. a and **b** China elevation maps reprinted from Zhu et al. (2018) and Ma et al. (2009), respectively. **c** Figure 2a in the manuscript. **d** China elevation map generated using a fine color scale

References:

Ma, L., Zhang, T., Frauenfeld, O. W., Ye, B., Yang, D. and Qin, D.: Evaluation of precipitation from the ERA-40, NCEP-1, and NCEP-2 reanalyses and CMAP-1, CMAP-2, and GPCP-2 with ground-based measurements in China, J. Geophys. Res. Atmos., 114(9), doi:10.1029/2008JD011178, 2009.

Zhu, J., Huang, G., Wang, X., Cheng, G. and Wu, Y.: High-resolution projections of mean and extreme precipitations over China through PRECIS under RCPs, Clim. Dyn., 50(11–12), 4037–4060, doi:10.1007/s00382-017-3860-1, 2018.

Comment #23: From Figure 4 onwards: please specify the period of study used for these figures

Response: We appreciate the reviewer's comment. The period of study has been specified in the captions of these figures.

Comment #24: Figure 4 and 5: we need the AEM plots in order to more easily compare the 3 datasets. That can help us understand the spatial differences, as here for example we have a very unclear idea of of far/close AEM and BMA are from each other.

Response: We appreciate the reviewer's comment. The AEM plots of precipitation and potential evapotranspiration (PET) have been added into Figures 4 and 5, respectively. The revised version of Figures 4 and 5 are also shown in Figures R8 and R9 below. The AEM simulations lead to a considerable difference in the spatial patterns of precipitation and PET compared with the CRU observations. For example, the AEM-simulated precipitation has a significant dry bias in Southeast China for the summer season (see Figures R8c and R8f), while the summer PET generated from the AEM simulation shows a different spatial pattern compared with the CRU observations (see Figures R9c and R9f).

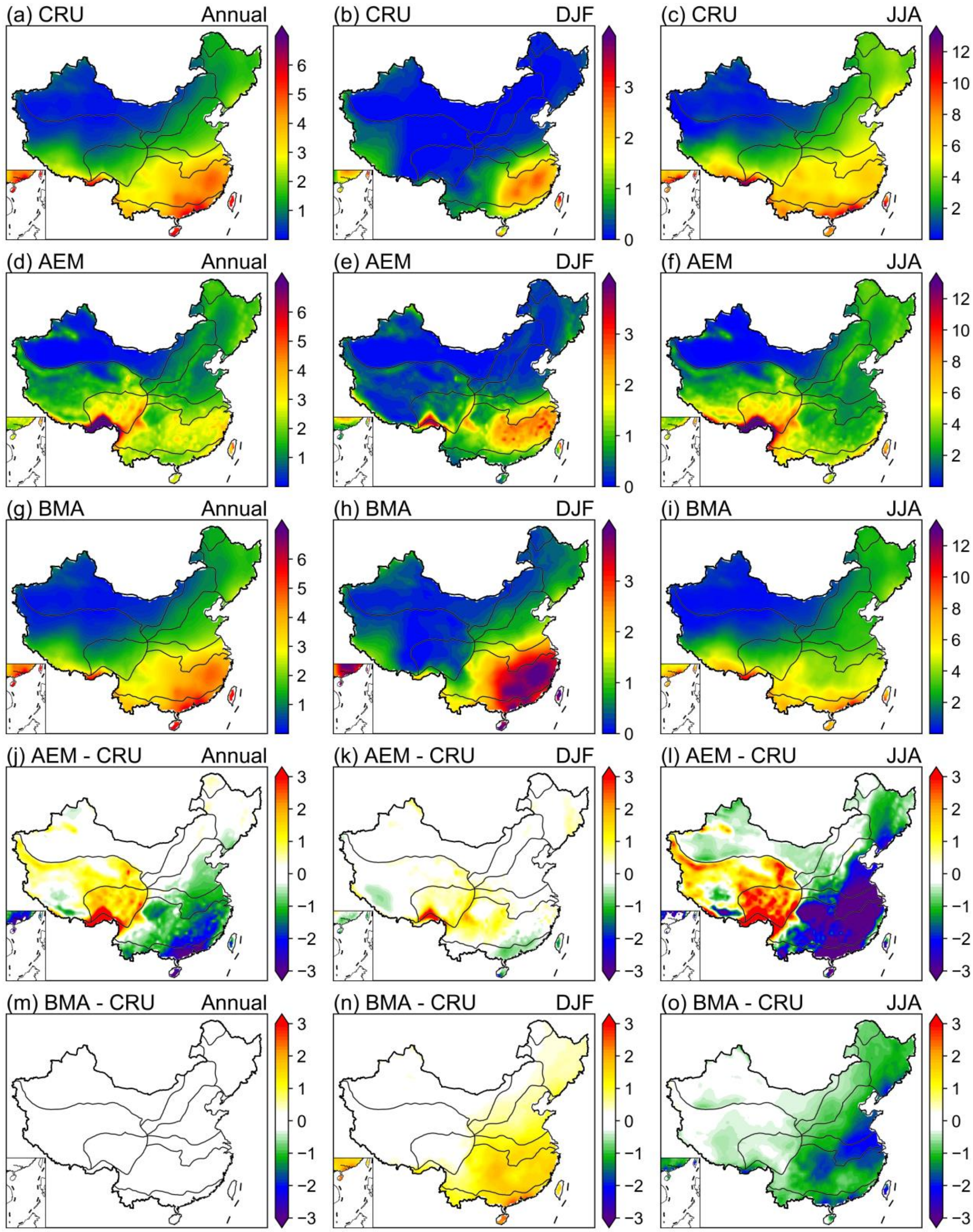

Figure R8. Spatial patterns of 30-year (1975−2004) annual and seasonal mean precipitation (unit: mm/day) generated from **a**−**c** the CRU observation, **d**−**f** the ensemble mean (AEM) simulation, **g**−**i** the MCMC-based BMA simulation. **j**−**l** and **m**−**o** correspond to the absolute model biases for the AEM and BMA simulations. DJF stands for December, January, and February; JJA stands for

June, July, and August. BMA is the Bayesian model averaging, CRU is the Climatic Research Unit, and AEM is the arithmetic ensemble mean. MCMC is the Markov chain Monte Carlo.

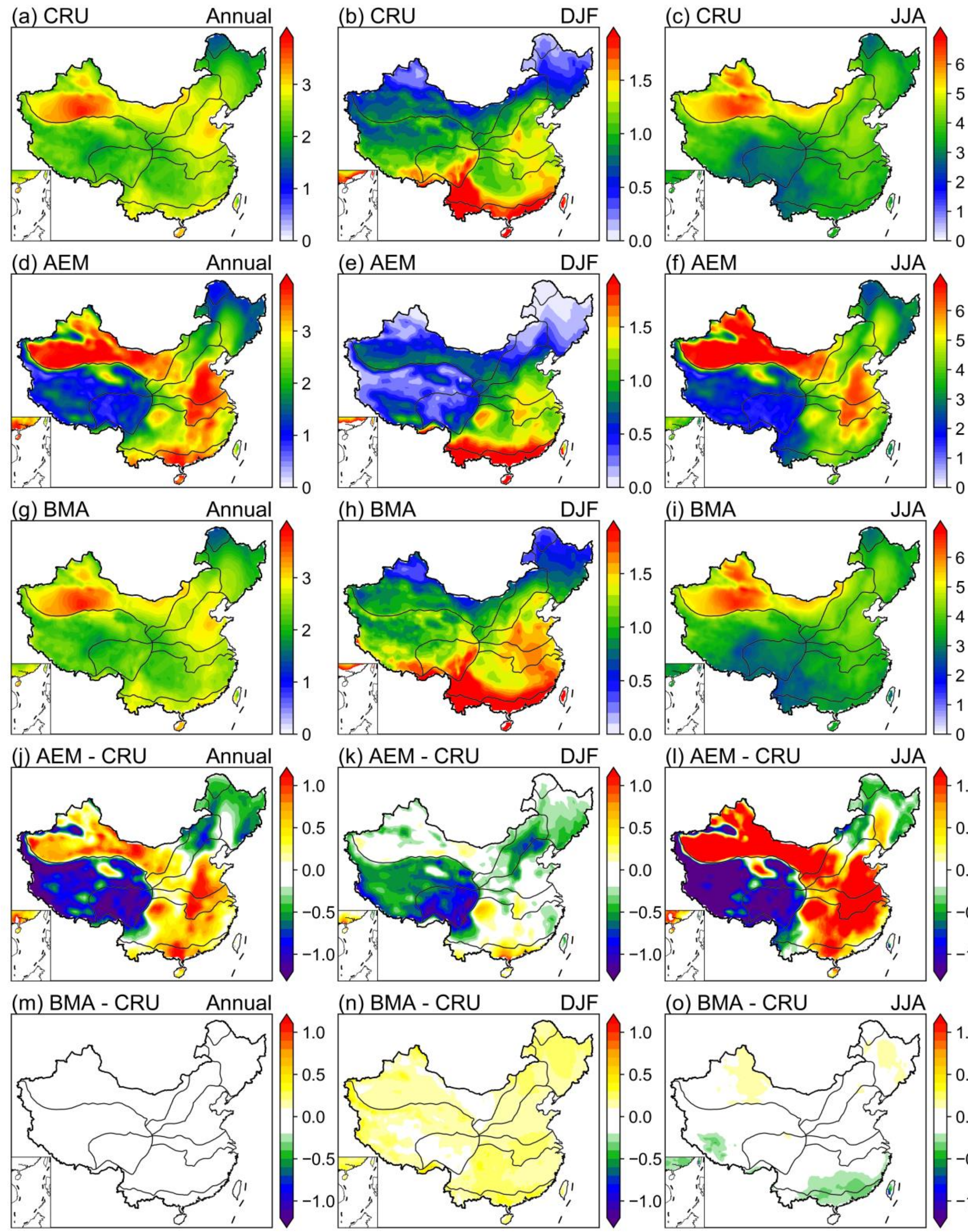

Figure R9. Spatial patterns of 30-year (1975−2004) annual and seasonal mean potential evapotranspiration (PET) (unit: mm/day) generated from **a**−**c** the CRU observation, **d**−**f** the ensemble mean (AEM) simulation, **g**−**i** the MCMC-based BMA simulation. **j**−**l** and **m**−**o** correspond to the absolute model biases for the AEM and BMA simulations.

Comment #25: Figure 6: why there are no line around OBS to help use assess the error, as in classical Taylor diagrams? That would help a lot.

Response: According to the reviewer's constructive comment, the contours of the normalized root-mean-square (RMS) differences between the climate simulations and the CRU observations have been added into Figure 6. The revised version of Figure 6 is also shown in Figure R10 below.

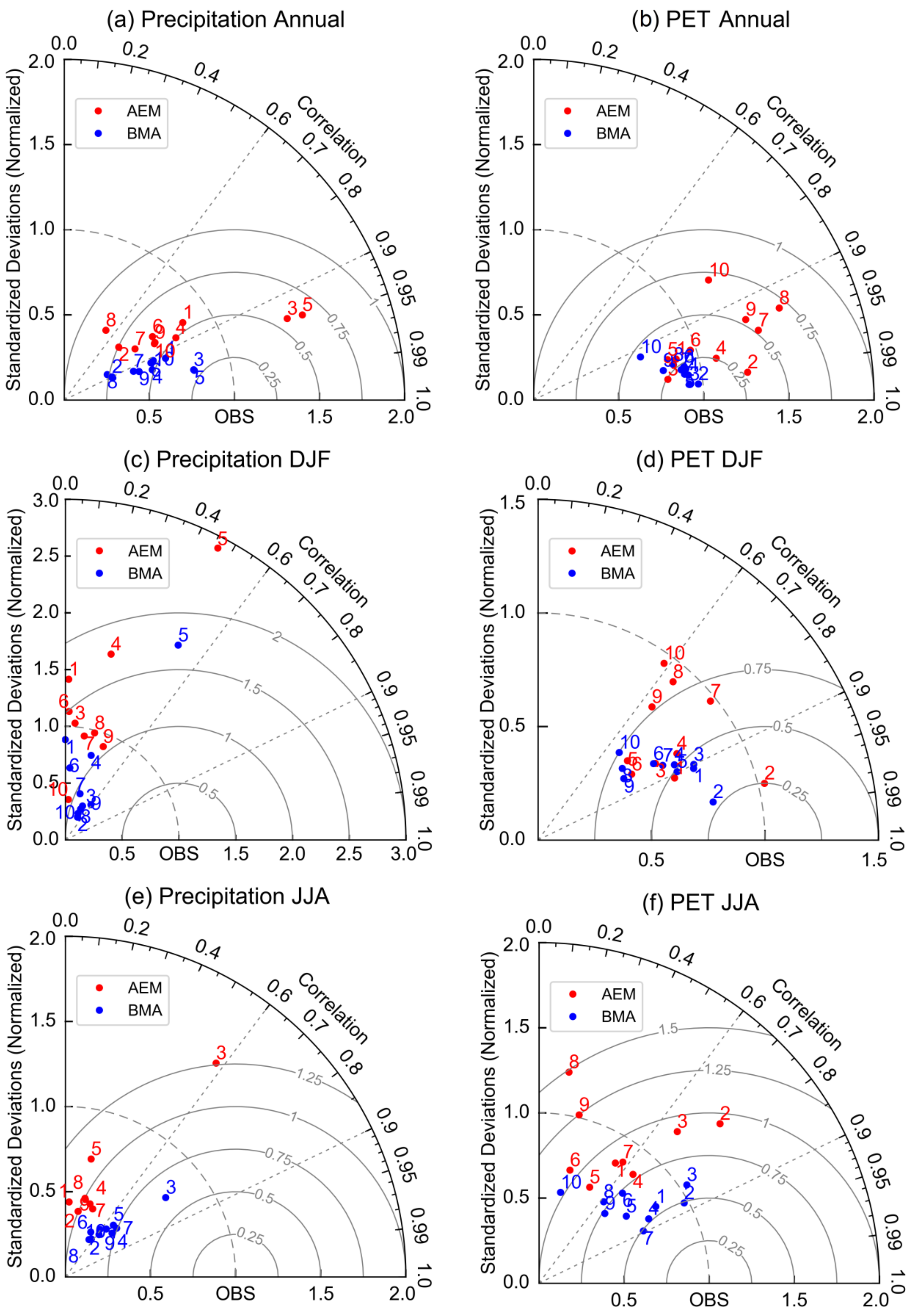
(a) Precipitation Annual
(b) PET Annual
(c) Precipitation DJF
(d) PET DJF
(e) Precipitation JJA
(f) PET JJA
Standardized Deviations (Normalized)
Correlation
OBS
AEM
BMA

Figure R10. Comparison of precipitation and PET derived from the BMA and AEM simulations for 10 climate divisions in China during the historical 30-year period (1975−2004). (a−b), (c−d), and (e−f) correspond to Annual, DJF, and JJA, respectively. Each point represents the simulation for a climate division, which agrees better with the observation when it has a higher correlation and a more consistent standard deviation with the observation, as well as it lies nearer the “OBS”.