Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics

Kamal Ahmed[1, 2], Dhanapala Arachchige Sachindra[3], Shamsuddin Shahid[1], Mehmet Cüneyd Demirel[4], Eun-Sung Chung[5]

[1]Faculty of Civil Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310, Malaysia
[2]Faculty of Water Resource Management, Lasbela University of Agriculture, Water and Marine Sciences, Balochistan, 90150, Pakistan
[3]Institute for Sustainability and Innovation, College of Engineering and Science, Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia
[4]Department of Civil Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey
[5]Department of Civil Engineering, Seoul National University of Science and Technology, Seoul, 139-743, Korea

Correspondence to: Eun S. Chung (eschung@seoultech.ac.kr)

Abstract. The climate modelling community has trialled a large number metrics to evaluate the temporal performance of the Global Circulation Models (GCMs) for the selection of GCMs, while very little attention has been given to spatial performance of GCMs which is equally important. This study evaluated the performance of 20 Coupled Model Intercomparison Project 5 (CMIP5) GCMs pertaining to their skills in simulating mean annual, monsoon and winter precipitation over Pakistan using state-of-the-art spatial metrics; SPAtial EFficiency, Goodman–Kruskal's lambda, Fractions Skill Score, Cramer's V, Mapcurves, and Kling-Gupta efficiency for the period 1961-2005. The multi-model ensemble (MME) precipitation was generated through intelligent merging of simulated precipitation of selected GCMs employing Random Forest (RF) regression and Simple Mean (SM). The results indicated some differences in the ranks of GCMs for different metrics. The overall ranks indicated NorESM1-M, CESM1-CAM5, GFDL-CM3 and GFDL-ESM2G as the best GCMs in simulating the spatial patterns of mean annual, monsoon and winter precipitation over Pakistan. MME precipitation generated based on the best performing GCMs showed more similarities with observed precipitation compared to precipitation simulated by individual GCMs. The MME developed using RF displayed better performance than the MME-based on SM. Multiple spatial metrics have been used for the first time for selecting GCMs based on their capability to mimic the spatial patterns of annual and seasonal precipitation. The approach suggested in the present study can be extended to any number of GCMs and climate variables and applicable to any region for the suitable selection of an ensemble of GCMs to reduce uncertainties in climate projections.

1 Introduction

Climate change is a complex, multidimensional phenomenon that is being critically studied over the last few decades (Byg and Salick, 2009;Cameron, 2011). The changes in climate are mostly observed by studying the variations in precipitation and

temperature regimes (Sheffield and Wood, 2008). Several studies reported increase in severity and frequency of droughts, floods, heatwaves and cold snaps in the recent years which are indicative of abrupt variations in the precipitation and temperature regimes (Duffy et al., 2015). According to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5), the average global land and ocean temperature has risen by around 0.72℃ (0.49–0.89℃) during 1951–2012. It is projected that it will further increase by 1.8 ℃ to 4 ℃ by the end of the 21st century (IPCC, 2014). The climate modelling community has widely agreed that the sharp temperature rise in the post-industrial revolution era is significantly affecting the global hydrologic cycle (Hegerl et al., 2018). The spatiotemporal variations in the global hydrologic cycle are influential on the humans and the environment. Therefore, it is important to study the variations in spatiotemporal patterns of climate variables such as precipitation and temperature (Akhter et al., 2016).

Global Circulation Models (GCMs) are principally utilised to simulate and project climate on global scale (Wright et al., 2015;Sachindra et al., 2014). Over the years, a large number of GCMs have been developed and used for the simulation and projection of global climate. The Coupled Model Intercomparison Project Phase 5 (CMIP5) is a set of GCMs available from the IPCC AR5. The CMIP5 GCMs showed significant improvements in climate simulations compared to its previous generation of CMIP3 models (Wang et al., 2016). Currently, over 40 GCMs are available in the CMIP5 suite with different spatial resolutions (Demirel and Moradkhani, 2015). Human and computational resources pose a restriction on the size of the sub-set of GCMs used in a climate change impact assessment (Ekström et al., 2016). Salman et al. (2018b) and Pour et al. (2018b) reported that a multi-model ensemble (a sub-set) of GCMs selected considering their skills in reproducing past observed characteristics of climate can reduce the GCM associated uncertainties in climate change impact assessment. The multi-model ensembles (MME) also enhance the reliability of prediction using information from several sources or GCMs (Pavan and Doblas-Reyes, 2000;Knutti et al., 2010).

The methods used for the generation of MME are broadly divided into two groups; (1) simple composite method (SCM) and (2) weighted ensemble method (WEM) (Wang et al., 2017b). In SCM all ensemble members are equally weighted while in the WEM, ensemble members are weighted according to their performance in simulating the past climate (Wang et al., 2017a;Oh and Suh, 2017;Giorgi and Mearns, 2002). The SCM is relatively simple to use and found to perform better than individual GCMs (Weigel et al., 2010;Fu et al., 2018;Dong et al., 2018). However, WEM is preferred as it has the capability to remove the systematic biases and improve the prediction capability since better GCMs are assigned higher weightages (Krishnamurti et al., 2000;Krishnamurti et al., 1999). Salman et al. (2018a) reported that prediction capability of a MME improves if it is based on WEM method. Thober and Samaniego (2014) also showed that sub-ensembles generated using WEM has the better capability to capture the historical characteristics of precipitation and temperature extremes. The performances of MMEs depend on the performance of ensemble members in simulating historical climate (Pour et al., 2018a). Therefore, selection of a sub-ensemble is a major challenge is climate change modelling.

Numerous endeavours have been made to examine the adequacy of climate models in simulating various climate variables (e.g. precipitation) (McMahon et al., 2015;Gu et al., 2015). Smith et al. (1998) stated that selection of an appropriate set of GCMs in a climate change impact assessment can be achieved considering 4 criteria; (1) Vintage, where only the latest

generation of GCMs are considered, (2) Spatial resolution, where fine resolution GCMs are preferred over coarser ones, (3) Validity, where performance of GCMs are considered, and (4) Representativeness, an ensemble of GCMs covering a wide range of projections of a climate variable (e.g. precipitation) is considered. In the above criteria, assessment and selection of GCMs based on their validity is the most widely adopted criterion where GCMs are ranked and selected according to their skill in simulating observed past climate (Mendlik and Gobiet, 2016).

A wide variety of methods has been used to assess climate models based on their ability to simulate the observed historical climate (past performance) such as such as reliability ensemble averaging approach (Giorgi and Mearns, 2002) relative entropy (Shukla et al., 2006), Bayesian approach (Min and Hense, 2006;Tebaldi et al., 2005;Chandler, 2013), probability density function (Perkins et al., 2007), hierarchical ANOVA models (Sansom et al., 2013), clustering (Knutti et al., 2013), correlation (Xuan et al., 2017;Jiang et al., 2015), and symmetrical uncertainty (Salman et al., 2018b). Johnson and Sharma (2009) assessed the performance of GCMs in replicating inter-annual variability. Thober and Samaniego (2014) evaluated the performance of GCMs in reproducing extreme indices of precipitation and temperature. Apart from that, some studies combined several performance measures such as root means square error, mean absolute error, correlation coefficient, and skill scores into one performance index to assess the accuracy of GCMs in reproducing past climate (Gu et al., 2015;Barfus and Bernhofer, 2015;Gleckler et al., 2008b;Wu et al., 2016;Ahmadalipour et al., 2015;Raju et al., 2016). Moreover, the past performance assessment of GCM is performed at different temporal scales; daily (Perkins et al., 2007), monthly (Raju et al., 2016), seasonal (Ahmadalipour et al., 2015) and annual (Murphy et al., 2004). Besides temporal scales, a number of studies ranked GCMs based on spatial areal average (Ahmadalipour et al., 2015;Abbasian et al., 2018) while some studies considered GCM performances at all the grid points covering the study area (Raju et al., 2016;Salman et al., 2018b).

It is also observed in the literature that there is no consensus on the choice of the GCM selection approach and temporal scale at which the performance assessment is done. Raäisaänen (2007), Smith and Chandler (2010) and McMahon et al. (2015) also argued that there is no universally accepted criterion for the assessment of GCMs. However, McMahon et al. (2015) reported that GCM simulations at annual time scale can better reproduce long-term annual mean statistics compared to that at daily time scale. Gleckler et al. (2008a) stated that assessment of GCMs with respect to a climate variable like precipitation over multiple time scale or seasons may provide vital information to water resources managers especially in the regions where climate variability is high. Moreover, Raju et al. (2016) and Salman et al. (2018b) demonstrated that GCM assessment provides more useful information when the evaluation is conducted at individual grid points covering the study area of interest. Selection of GCMs based on their performance in individual grid points over a region does not guarantee its capability to simulate spatial pattern of regional climate. It is expected that GCMs should able to capture the spatial pattern of major features of climate of a region such as monsoon and western disturbances. Koch et al. (2018) and Demirel et al. (2018) argued that climate modelling community is mostly focused on the temporal performance of GCMs and ignores explicit assessment of their spatial performance which is also equally important. They also emphasized on the importance of the use of multiple spatial metrics for GCM performance assessment. Furthermore, the metrics should be insensitive to the units of the variables compared.

Overall, review of literature revealed that several studies assessed the performance of GCMs considering several grid points over the whole study area; however they ignored the capability of GCMs to replicate the spatial patterns. Spatial patterns of GCMs provide better understanding on the occurrences of hydro-climatic phenomena such as precipitation distributions, floods and droughts. Therefore, it is imperative to assess the skills of GCMs to replicate the historical spatial patterns of climate variables. Within this framework, the current study hypothesized that the sub-ensemble members identified based on their ability to mimic the spatial pattern of observed precipitation of a region can be used for generation of a reliable MME for precipitation for that region. This study for the first time, employed five state of the art spatial performance metrics; SPAtial EFficiency metric (SPAEF) (Demirel et al., 2018), Goodman–Kruskal's lambda (Goodman and Kruskal, 1954), Fractions Skill Score (FSS) (Roberts and Lean, 2008), Cramer's V (Cramér, 1999), Mapcurves (Hargrove et al., 2006), Kling-Gupta efficiency (KGE) (Gupta et al., 2009) for the assessment of performance of 20 CMIP5 GCM in simulating observed annual, monsoon and winter precipitation over Pakistan. Then based on the above spatial performance metrics the most skilful GCMs were identified and hence multi-model ensemble (MME) means of precipitation using Simple Mean (SM) and Random Forest (RF) were generated.

2 Study Area and Datasets

2.1 Study area

As shown in Figure 1, Pakistan located in south Asia shares its border with India in the east, China in the north, Afghanistan and Iran in the west and Arabian Sea in the south. Pakistan has a rugged topography ranging from 0 m in the south to 8572 m in the north. Pakistan is overwhelmed by arid and semi-arid climate, and displays significant climatic variations. Pakistan receives summer monsoon precipitation during the period June-September and winter precipitation during the period December-March. Besides that, there are two intermediate rainy seasons called the pre-monsoon and the post-monsoon during the periods April-May and October-November, respectively (Sheikh 2001).

The bulk of the summer precipitation is caused by the monsoon winds that arise from the Bay of Bengal while westerly disturbances in the Mediterranean Sea are responsible for the winter precipitation. The average precipitation in Pakistan widely varies from southwest to northern parts in the range of < 100 to > 1000 mm/year. Since the country is mostly characterized by arid and semi-arid climate; the bulk of the country receives less than 500 mm/year of precipitation while only a very limited area in the north receives more than 1,000 mm/year of precipitation (Ahmed et al., 2017).

Figure 1. The location of Pakistan in central-south Asia and the GCM grid points over the country.

2.2 Datasets

2.2.1 Gridded Precipitation Data

The lack of long records of climate observations with an extensive spatial coverage is a major issue in hydro-climatological investigations in many regions. As a solution to this problem, gridded data sets based on observations and various interpolation and data assimilation techniques have been created (Kishore et al., 2015). In this investigation, gridded monthly precipitation data of the Global Precipitation Climatology Center (GPCC) (Schneider et al., 2013) were used as the surrogates of observed precipitation for the period 1961-2005. GPCC precipitation data are available at a spatial resolution of 0.5°.

As stated in the existing literature GPCC data are of high quality (Shiru et al., 2018;Salman et al., 2018c) and have an excellent seamless spatial and temporal coverage (Spinoni et al., 2014). Most importantly, GPCC precipitation data have shown high correlation with observed precipitation over Pakistan (Kazmi et al., 2016).

2.2.2 GCM precipitation data

Monthly precipitation data simulated by the 20 CMIP5 GCMs were extracted from the IPCC data distribution center for period 1961-2005. The monthly precipitation projections for all Representative Concentration Pathways (RCP) are only available for 20 of the CMIP5 GCMs. Hence, only those 20 GCMs were used for the current investigation. The modelling centres, names of GCMs and spatial resolution of each of the selected GCMs are provided in Table 1. For the sake of fair comparison, monthly precipitation simulations of all selected CMIP5 were remapped to a common grid with a resolution of $2^{o}\times2^{o}$ (approximately the average resolution of the GCMs considered in the present study), using bilinear interpolation technique.

Table 1. CMIP5 GCMs considered in this study.

3 Methodology

In this study, GCMs for annual, monsoon and winter season were first ranked separately (individual ranking) using five spatial performance measures; SPAEF, Lambda, FSS, Cramer-V, Mapcurves, and KGE. Then a comprehensive rating metric (RM) (Jiang et al., 2015) was used to rank the GCMs considering the individual ranks determined corresponding to all above spatial performance measures. The RM values of GCMs obtained for annual, monsoon and winter precipitations were finally averaged for the overall ranking of GCMs. Finally, a sub-set of GCMs (MME) based on the overall ranks was selected and a precipitation data set for the MME was derived. The procedure used for the ranking, identification of the ensemble of GCMs and derivation of precipitation data from the multi-model ensemble of GCMs are outlined as follows.

1. All GCM simulated past precipitation for the period 1961-2005 were remapped to a common grid with a 2°×2° resolution.
2. SPAEF, Lambda, FSS, Cramer-V, Mapcurves, and KGE were individually applied to annual, monsoon and winter precipitation for the period 1961-2005.
3. The goodness of fit (GOF) estimated by SPAEF, Lambda, FSS, Cramer-V, Mapcurves, and KGE for annual, monsoon and winter precipitation were used to rank the GCMs separately.
4. A comprehensive rating metrics (RM) was used to combine the ranks of GCMs determined by above spatial performance measures separately for annual, monsoon and winter precipitation.
5. The RM values calculated in step 4 for annual, monsoon and winter precipitation were averaged to obtain the overall ranks of the GCMs in simulating precipitation over Pakistan.
6. The four top ranked GCMs based on their overall performance in replicating annual, monsoon and winter precipitation were identified.
7. Simple Average (SM) and Random Forest (RF) were used to generate MME precipitation mean with the precipitation simulated by the four top ranked GCMs identified in step 6.
8. Finally, the spatial patterns of MME precipitation generated from SM and RF were validated by visually comparing with the spatial patterns of observed precipitation.

Details of the methods and the determination of the best performing ensemble of GCMs are provided in the following sections.

3.1 GCM Performance Assessment

3.1.1 SPAtial EFficiency metric

SPAtial EFficiency metric (SPAEF), proposed by Koch et al. (2018) is a robust spatial performance metric which considers three statistical measures (1) Pearson correlation, (2) coefficient of variation and (3) histogram overlap in the assessment of GOF of a model. The major advantage of SPAEF is that, it combine the information derived from the above three independent statistical measures into one metric. The SPAEF between past observed precipitation (i.e. GPCC) and GCM simulated precipitation was calculated using Eq. (1). In Eq. (1), α is the Pearson correlation coefficient between observed and GCM simulated precipitation, β is the spatial variability and γ is the overlap between the histograms of observed precipitation and GCM simulated precipitation.

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

Equation (2) and (3) show the procedure for β and γ calculations respectively (for Pearson correlation (α) refer to (Pearson, 1948). In Eq. (2) σ_G and σ_O refer to standard deviation of GCM simulated and observed precipitation respectively and μ_G and μ_O refer to mean of GCM simulated and observed precipitation respectively.

$$\beta = \frac{\left(\frac{\sigma_G}{\mu_G}\right)}{\left(\frac{\sigma_O}{\mu_O}\right)} \quad (2)$$

In Eq. (3), *K*, *L* and *n* refer to histograms value of observed precipitation, histograms value of GCM simulated precipitation and the number of bins in a histogram.

$$\gamma = \frac{\sum_{j=1}^{n} \min(K_j, L_j)}{\sum_{j=1}^{n} K_j} \quad (3)$$

The SPAEF can have a value between $-\infty$ and 1, where value closer to 1 indicates higher spatial similarity between the observations and model simulations (Koch et al., 2018).

3.1.2 Goodman–Kruskal's lamba

Goodman–Kruskal's lamba also known as Lambda coefficient (λ) is used to measure the nominal/categorical association between categorical maps (Goodman and Kruskal, 1954). Lambda coefficient (λ) varies between 0 and 1, where a value closer to 1 refers to a higher similarity between the map of model simulations and that of observations of precipitation. The Lambda (λ) coefficient was calculated using Eq. (4), where c_{ij} is a contingency matrix (describes the relationships between the data classes), *i* and *j* are the class or categories in observed and simulated maps, *m* and *n* represent the number of classes in observed and simulated maps respectively.

$$\lambda = \frac{\sum_{i=1}^{m} max_j c_{ij} - max_j \sum_{i=1}^{m} c_{ij}}{N - max_j \sum_{i=1}^{m} c_{ij}} \quad (4)$$

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

3.1.3 Fractions Skill Score

The Fractions Skill Score (*FSS*) proposed by (Roberts and Lean, 2008) is another measure used for the assessment of spatial agreement between model simulations and observations. In this study, *FSS* between observed and GCM simulated precipitation was computed using Eq. (5). *FSS* varies between 0 and 1 where a value closer to 1 refers to higher agreement between observed and simulated precipitation.

$$FSS = 1 - \frac{\frac{1}{N}\sum_{N}\left(P_s - P_o\right)^2}{\frac{1}{N}\left(\sum_{N} P_s^2 + \sum_{N} P_o^2\right)} \quad (5)$$

In Eq. (5) P_s and P_o are simulated and observed precipitation respectively whereas N refers to the total number of grid points.

3.1.4 Cramer's V

Cramer's V (Cramér, 1999) statistic is a Chi-square-test-based measure which is used in assessing spatial agreement between observations and model simulations (Zawadzka et al., 2015). Its value ranges between 0 and 1 and can be calculated using Eq. (6).

$$V = \sqrt{\frac{x^2}{N(\min(m,n)-1}} \quad (6)$$

where, x^2 is Chi-Square, N is the grand total of observations, m is the number of rows and n is the number of columns. In this exercise m = 42 (number of rows of data) and n = 2 (observed and modelled precipitation).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

3.1.5 Mapcurves

Mapcurves is another statistical measure, developed by Hargrove et al. (2006) for the measurement of similarity between categorical maps. Mapcurves provides the degree of spatial association between two maps. The value of Mapcurves can vary from 0 to 1 (perfect agreement). In the present study, the degree of spatial association between the historical observed precipitation map (i.e. GPCC precipitation) and each of the GCM simulated precipitation maps was determined using Eq. (7) where, *Y* refers the Mapcurves value, *C* is the degree of intersection between the two maps, *A* and *B* are the total area of historical and GCM simulated maps.

$$Y = \sum\left[\left(\frac{C}{B+C}\right)\left(\frac{C}{A+C}\right)\right] \quad (7)$$

3.1.6 Kling-Gupta efficiency

Kling-Gupta efficiency (KGE) is a GOF test developed by Gupta et al. (2009), for the model performance assessment. KGE considers three statistical measures (1) Pearson correlation, (2) variability ratio and (3) bias ratio in the assessment of model performance. In the present study, KGE was calculated between historical observed precipitation and GCM simulated precipitation using Eq. (8). In Eq. (8), α_P is the Pearson correlation (Pearson, 1948) between observed and GCM simulated precipitation, β_P is the is the bias ratio, and γ_{RP} is the variability ratio. Equation (9) and (10), show the calculation of β_P and γ_{RP} respectively. In Eq. (9), μ_G and μ_O refer to mean of GCM simulated and observed precipitation respectively, whereas in Eq. (10), CV_G and CV_O refer to coefficient of variation of GCM simulated and observed precipitation respectively.

$$KGE = 1 - \sqrt{(\alpha_P - 1)^2 + (\beta_P - 1)^2 + (\gamma_{RP} - 1)^2} \quad (8)$$

$$\beta_P = \frac{\mu_G}{\mu_O} \quad (9)$$

$$\gamma_{RP} = \frac{CV_G}{CV_O} = \frac{\left(\frac{\sigma_G}{\mu_G}\right)}{\left(\frac{\sigma_O}{\mu_O}\right)} \quad (10)$$

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

3.2 Comprehensive Rating Metrics

The ranking of GCMs with respect to a given climate variable using one single GOF measure is a relatively simple task. However, the ranking of GCMs becomes more challenging when multiple GOF measures are used with multiple climate variables, as different GCMs may display different degrees of accuracies for different GOF measures and climate variables. In such case, an information aggregation approach that combines information from several GOF measures can be used. In this study, a comprehensive rating metric (Chen et al., 2011) was used to obtain the overall ranks of GCMs. The overall ranks of GCMs based on different GOFs were obtained for each season separately using Eq. (11).

$$RM = 1 - \frac{1}{nm}\sum_{i=1}^{n} rank_i \quad (11)$$

In Eq. (11), *m* refers to the number of GCMs, *n* refers to the number of metrics or seasons and *i* refers to the rank of a GCM based on i^{th} GOF. A value of RM near to 1 refers to a better GCM in terms of its ability to mimic the spatial or temporal characteristics of observations.

3.3 Identification of Ensemble Members

The uncertainties in climate projections arise from GCM structure, assumptions and approximations, initial conditions, and parameterization can be reduced by identifying an ensemble of better performing GCMs (Kim et al., 2015). Lutz et al (2016) reported that one or a small ensemble of GCMs is suitable for climate change impact assessment. A number of studies (Weigel et al., 2010;Miao et al., 2012) have suggested that one GCM is not enough to assess the uncertainties associated with the future climate. Therefore, identification of an ensemble of GCMs is a necessity in climate change impact assessment. In the present study, the most appropriate ensemble of GCMs was identified by considering the four top ranked GCMs. The ensemble of GCMs was identified in two steps: (1) RM values of GCMs for annual, monsoon and winter precipitation were averaged to derive an overall rank for each GCM, and (2) four top ranked GCMs based on RM values for all seasons were considered for the ensemble. The selection of an appropriate set of GCMs considering their skills in different seasons enables the selection of an ensemble which can better simulate the observations in different seasons.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

3.4 Development of Multi-model Ensemble (MME) Mean

The uncertainties in projections of a climate variable can be reduce by using its mean time series calculated from a MME of better performing GCMs (You et al., 2018). Numerous approaches are documented in the literature for the calculation of mean time series from an ensemble of better performing GCMs starting from simple arithmetic mean to machine learning algorithms (Kim et al., 2015). In the present study, two approaches 1). Simple Mean (SM) and 2). Random Forest (RF) (Breiman, 2001) were used in the calculation of mean time series of precipitation corresponding to an ensemble of four top ranked GCMs.

4 Results and Discussion

4.1 Accuracy Assessment of Gridded Precipitation Data

As a preliminary analysis, the GPCC precipitation data were validated with the observed precipitation. The validation was carried out for the period 1961-2005. In the present study, two statistical metrics namely; Normalized Root Mean Square Error (NRMSE), and modified index of agreement (*md*) were used to assess the accuracy of GPCC precipitation in replicating the mean and the variability of observed precipitation.

NRMSE is a non-dimensional form of Root Mean Square Error (RMSE) which is derived by normalizing RMSE by variance. NRMSE is more reliable than RMSE in comparing model performance when the model outputs are in different units or the same unit but with different orders of magnitude (Willmott, 1982). NRMSE can have any positive value, however values near to zero are preferred (Chen and Liu, 2012). The '*md*' is widely used to estimate the errors between observed and simulated values and it varies between 0 (no agreement) and 1 (perfect agreement) (Willmott, 1981).

The NRMSE and *md* values between observed precipitation and GPCC precipitation (pertaining to the grid point closest to the observation station) obtained for 17 locations in Pakistan are given in Table 2. Overall, all the stations showed low and high NRMSE and *md* values respectively, indicating that the accuracy of the GPCC precipitation in replicating observed precipitation over Pakistan is high.

Table 2. Validation of accuracy of GPCC precipitation using NRMSE and *md*.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

4.2 Evaluation and Ranking of GCMs

The SPAEF, Lambda, FSS, Cramer-V, Map-curves, and KGE between GPCC and GCMs simulated mean annual, monsoon and winter precipitation of Pakistan were estimated for the period 1961 to 2005. As an example, Table 3 shows the GOF values that define the performance of each GCM in simulating GPCC mean annual precipitation (winter and monsoon not shown). The GOF values near to 1 refer to the better performance of the GCM of interest. For example, GFDL-ESM2G has a GOF value of 0.724 for SPAEF, and hence regarded as the best GCM in term of SPAEF, whereas CSIRO-Mk3.6.0 can be regarded as the poorest which has a GOF value of -0.412 in term of SPAEF. The GOF values for other metrics (i.e. Lambda, FSS, Cramer-V, Map-curves, and KGE) can be interpreted in the same manner.

Table 3. GOF values of GCMs obtained using different spatial metrics for mean annual precipitation.

The GCMs were then ranked based on GOF value of each metric shown in Table 3 and presented in Figure 2. Figure 2 shows the ranks attained by GCMs corresponding to different metrics. For example, BCC-CSM1.1 (m) attained ranks 12, 11, 12, 13, 14 and 19 for SPAEF, Lambda, FSS, Cramer-V, Mapcurves and KGE respectively. It was observed that none of the GCMs was able to secure the same rank for all metrics. However, NorESM1-M, CESM1-CAM5, GFDL-ESM2G secured rank 1, 2, and 3 respectively for four metrics (i.e. Lambda, FSS, Cramer-V and Mapcurves). Some of the GCMs attained the same rank for three metrics (e.g. CSIRO-Mk3.6.0, CESM1-CAM5, GFDL-CM3 and GFDL-ESM2G). Cramer-V and Mapcurve showed more or less similar ranks for GCMs. Similar results were also seen for monsoon and winter precipitation (not presented in the manuscript).

Figure 2. Ranks of GCMs according to their performance in replicating spatial patterns of mean annual precipitation.

4.3 Overall Ranks of GCMs based on Annual and Seasonal Precipitation

The application of various evaluation metrics have shown different ranks for the same GCM (Ahmadalipour et al., 2015;Raju et al., 2016). Thus, the procedure detailed in Section 3.2 was employed to combine the ranks of each GCM produced by each spatial performance metric into an overall rank. The ranks attained by GCMs (as an example see Figure 2 for annual precipitation) corresponding to different metrics were used to calculate the RM values for each GCM. The overall ranks of GCMs for mean annual, monsoon and winter precipitation are presented in Table 4 along with the RM values. As

seen in Table 4, NorESM1-M, CESM1-CAM5 and HadGEM2-AO were the most skilful GCMs in reproducing the spatial characteristics of mean annual, monsoon and winter precipitation respectively. On the other hand, MRI-CGCMs3 displayed the least skill in reproducing the spatial characteristics of annual precipitation, and CSIRO-Mk3.6.0 showed the least skill in reproducing the spatial characteristics of monsoon and winter precipitation.

Table 4. Overall ranks of GCMs for mean annual, monsoon and winter precipitation based on rating metric values.

The better performance of NorESM1-M, CESM1-CAM5 and HadGEM2-AO in simulating precipitation over Indo-Pak sub-continent has also been reported in several past studies. Babar et al. (2014) assessed 13 CMIP5 GCMs for simulating Indian summer monsoon precipitation and found that NorESM1-M can capture the seasonal cycle of precipitation. Anand et al. (2018) and Jena et al. (2015) concluded that CESM1-CAM5 is one of the GCMs capable of simulating Indian summer monsoon precipitation. Latif et al. (2018) reported the relatively better performance of HadGEM2-AO out of 36 CMIP5 GCMs in simulating precipitation over Indo-Pakistan sub-continent based on spatial correlation.

Table 4 also shows that ranks of many GCMs in simulating annual and monsoon precipitation are more or less the same, however; the ranks of GCMs corresponding to winter season are somewhat different. The difference in ranks of GCMs in winter compared to that of annual and monsoon seasons was probably due to differences in synoptic climatology. The winter precipitation occurs during the period December to March due to the Westerly winds that blow from Mediterranean Sea and enters Pakistan from the western side (Sheikh, 2001). The monsoon precipitation occurs during June to September caused by the monsoon winds that blow from the Bay of Bengal and enters Pakistan from the north eastern side (Sheikh, 2001;Sheikh et al., 2009). It can be inferred that the selection of an appropriate ensemble of GCMs also depends on the season and the mechanism which causes precipitation. The findings of the present study also support the results of Ahmadalipour et al. (2015) where they reported that the performance of GCMs differ from seasons to season.

The spatial patterns of precipitation simulated by the GCMs ranked 1 and ranked 20 were compared with the spatial patterns of GPCC precipitation, and presented in Figure 3. In Figure 3 it was seen that the GCMs that attained rank 1 showed spatial patterns more or less similar to that of GPCC precipitation. On the other hand, GCMs ranked 20 (i.e. rank 20) showed large differences compared to the spatial patterns of GPCC precipitation. The Figure 3 clearly shows that GCMs which attained rank 20 under-estimated the annual, monsoon and winter precipitation over a large region in the study area.

Figure 3. Spatial patterns of GPCC (a - c), GCM at rank 20 (d - f) and GCM at rank 1 (g - i) for mean annual, monsoon, and winter precipitation respectively.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

4.4 Identification of Ensemble Members

Based on the criteria mentioned in Section 3.3, average RM values for each GCM was estimated and then the GCMs were ranked based on the average RM values. Table 5 shows the average RM values of the 20 GCMs considered in this study. The four top ranked GCMs; NorESM1-M, CESM1-CAM5, GFDL-CM3 and GFDL-ESM2G indicated in bold in Table 5 were designated as the members of the ensemble for projecting precipitation over Pakistan.

Table 5. Averaged RM values of GCMs for the identification of ensemble members

The performances of the four top ranked GCMs (i.e. GCMs ranked 1, 2, 3 and 4) and four lowest ranked GCMs (i.e. GCMs ranked 17, 18, 19, and 20) were visually evaluated using scatter plots shown in Figures 4 and 5, pertaining to annual, monsoon and winter precipitation. In order to plot the scatter, the precipitation simulated by each GCM and GPCC precipitation pertaining to all grid point was averaged (spatially averaged precipitation). As expected, GCMs that attained ranks 1 to 4 showed closer agreements with the observed precipitation compared to that of GCMs which attained ranks 17, 18, 19, and 20. The scatter plots in Figure 5 indicated that the precipitation simulated by the least skilful GCMs heavily underestimated annual precipitation. Over and underestimation of precipitation can also be seen in the scatter plots of GCMs ranked 1, 2, 3 and 4. However, their scatter was found much aligned with the 45 degree line compared to that of GCMs ranked 17, 18, 19, and 20. Therefore, it is argued that the GCMs ranked 1, 2, 3 and 4 can be used as an ensemble for the simulation/projection of precipitation.

Figure 4. Scatter of precipitation of four top ranked GCMs against GPCC annual, monsoon and winter precipitation.

Figure 5. Scatter of precipitation of four lowest ranked GCMs against GPCC annual, monsoon and winter precipitation.

Some of the GCMs identified for the ensemble over Pakistan were found similar with GCMs that showed better performance in the neighboring countries such as India and Iran. Jena et al. (2015) used Z-value test, correlation coefficient, relative precipitation comparison test, probability function comparison, root mean square error, and Student's t-test to evaluate the performance of 20 CMIP5 GCMs in simulating Indian summer monsoon. They found that CCSM4, CESM1-CAM5, GFDL-CM3, and GFDL-ESM2G perform better compared to the other GCMs. Prasanna (2015) conducted a study to assess the

performance of 12 CMIP5 GCMs using mean and coefficient of variation over South Asia (5N–35N; 65E–95E) and identified ACCESS, CNRM, HadGEM2-ES, MIROC5, Can-ESM, GFDL-ESM2M, GISS, MPI-ESM and NOR-ESM as better performing GCMs. Sarthi et al. (2016) evaluated the performance of 34 CMIP5 GCMs using Taylor diagram, skill score, correlation and RMSE. They found that BCC-CSM1.1(m), CCSM4, CESM1(BGC), CESM1(CAM5), CESM1(WACCM), and MPI-ESM-MR were able to better capture the Indian summer monsoon precipitation. Afshar et al. (2016) applied Nash–Sutcliffe efficiency, percent of bias, coefficient of determination, and ratio of the RMSE to the standard deviation of observations for assessing performance of precipitation simulations of 14 CMIP5 GCMs over a mountainous catchment in north-eastern Iran which borders Pakistan. They recommend GFDL-ESM2G, IPSL-CM5A-MR, MIROC-ESM, and NorESM1-M as better GCMs. Mahmood et al. (2018) used correlation coefficient, error between observed and GCM means and standard deviation, root mean square error, to assess the performance of CMIP5 GCMs in simulating precipitation over Jhelum river basin, Pakistan and found the good performance of GFDL-ESM2G, HadGEM2-ES, NorESM1-ME, CanESM2, and MIROC5. Latif et al. (2018) reported better performance of HadGEM2-AO, INM-CM4, CNRM-CM5, NorESM1-M, CCSM4 and CESM1-WACCM out of 36 GCMs in simulating precipitation over Indo-Pakistan based on partial correlation. The above findings indicated that the GCMs identified in this study for the ensemble were also found to perform well in the other studies in nearby countries/regions.

4.5 Multi-model Ensemble (MME) Mean

The performances of GCM ensembles identified in Section 4.4 were validated considering two types of MME means. The MME mean of precipitation of the four top ranked GCMs was calculated with 1). Simple Mean (SM) and 2). Random Forest (RF). In SM, the time series of precipitation of the four top ranked GCMs were averaged to obtain the MME while in RF, the time series of precipitation of the four top ranked GCMs were considered as inputs to the RF based MME.

In Figure 6, the spatial patterns of precipitation corresponding to both MMEs derived with SM and RF were compared with that of GPCC precipitation. The spatial patterns of precipitation were created using Ordinary Kriging technique. Ordinary Kriging was selected as it was found to perform better than other Interpolation methods over the Pakistan (Ahmed et al., 2014). As seen in Figure 6, both MMEs captured the spatial patterns of observed precipitation to a good degree. However, the differences can be seen in both MMEs in replicating the spatial pattern of GPCC precipitation. The visual comparison provided in Figure 6 also indicated that RF-based MME performs better than the MME based on SM. The SM was found to underestimate annual precipitation in the south-western and the northern regions, and overestimate the monsoon and winter precipitation in the northern region, while the RF was found to produce spatial pattern almost identical to that of GPCC annual, monsoon and winter precipitation. The better performance of RF in generating MME has also been reported in several studies. Salman et al. (2018b) generated MME mean for maximum and minimum temperature over Iraq using four CMIP5 GCMs and reported RF performed better compared to individual GCMs. Likewise, Wang et al. (2017) conducted a

comprehensive study to evaluate the performance of different machine learning techniques including RF, support vector machine, Bayesian model averaging and the arithmetic ensemble mean in generating MME. They considered 33 CMIP5 GCMs for precipitation and temperature over 108 station located in Australia and concluded RF and SVM can generate better MME compared to other techniques.

Figure 6. Spatial patterns of observed (a - c), MME computed using Simple Mean (SM) (d - f) and MME computed using Random Forest (RF) (g - i) for mean annual, monsoon, and winter precipitation respectively during 1961 to 2005.

The performance of MME ensembles was further evaluated using scatter plots shown in Figure 7. Scatter plots were developed using spatially averaged GPCC and MME annual, monsoon and winter precipitation at all grid points for the period 1961-2005. According to scatter plots in Figure 7, RF-based MME performed significantly better compared to its counterpart SM-based MME in monsoon and winter seasons.

Figure 7. Scatter plots of GPCC and MME mean precipitation for annual, monsoon and winter seasons obtained using Simple Mean (SM) and Random Forest (RF) for the period 1961-2005.

5. Conclusion

This study quantitatively and qualitatively assessed the accuracy of 20 CMIP5 GCMs in simulating mean annual, monsoon and winter precipitation over Pakistan for the period 1961-2005. The quantitative evaluation was done based on five state-of-art spatial metrics; SPAtial EFficiency, Goodman–Kruskal's lambda, Fractions Skill Score, Cramer's V, Mapcurves, and Kling-Gupta efficiency and qualitative evaluation was done using scatter plots. A comprehensive rating metric was used to derive the overall ranks of GCMs based on their ranks pertaining to annual, monsoon, and winter seasons.
Following conclusions were drawn from this study:

1) The low Normalized Root Mean Square Error (NRMSE), and high modified index of agreement (md) confirmed the close agreement of monthly GPCC precipitation with the observed precipitation extracted from 17 stations located in different climate zones of Pakistan. The low NRMSE and high md values of GPCC precipitation can be associated with extensive data quality control measures and the use of a large number of stations for the development of GPCC precipitation dataset (Schneider et al., 2014).

2) Ranks of the 20 GCMs produced by all spatial metrics; SPAtial EFficiency, Goodman–Kruskal's lambda, Fractions Skill Score, Cramer's V, Mapcurves, and Kling-Gupta efficiency for the period 1961-2005 were found mostly similar to each other during a given season (i.e. mean annual, monsoon and winter). However, it was noticed that different GCMs performed better in simulating precipitation during different seasons (mean annual, monsoon and winter). NorESM1-M, CESM1-CAM5 and HadGEM2-AO were identified as the most skilful GCMs while MRI-CGCM3 and CSIRO-Mk3.6.0 were identified as the least skilful GCMs in simulating annual, monsoon and winter precipitation respectively. The overall ranks of GCMs based on comprehensive rating metric revealed that NorESM1-M, CESM1-CAM5, GFDL-CM3 and GFDL-ESM2G are the most suitable GCMs for projecting precipitation over Pakistan.

3) The spatial patterns of precipitation of four top ranked GCMs and their MME mean precipitation generated using Simple Mean (SM) and Random Forest (RF) showed more or less similar spatial patterns of Global Precipitation Climatology Center (GPCC) precipitation. Additionally, the comparison of MME mean precipitation generated with SM and RF clearly showed the superiority of RF in replicating the spatial patterns of the observed precipitation.

In this study, the GCMs were selected based on their capability to replicate the spatial patterns of annual and seasonal precipitation of an area. In the future, the present study can be extended to develop indices for the assessment of the performance of GCMs based on their capability to replicate both the temporal and spatial patterns of climate of a region.

Acknowledgement

This work is supported by the Post-Doctoral Fellowship Scheme of Universiti Teknologi Malaysia (PDRU) grant no. Q.J130000.21A2.04E10.

References

Abbasian, M., Moghim, S., and Abrishamchi, A.: Performance of the general circulation models in simulating temperature and precipitation over Iran, Theoretical and Applied Climatology, 10.1007/s00704-018-2456-y, 2018.

Afshar, A. A., Hasanzadeh, Y., Besalatpour, A. A., and Pourreza-Bilondi, M.: Climate change forecasting in a mountainous data scarce watershed using CMIP5 models under representative concentration pathways, Theoretical and Applied Climatology, 129, 683-699, 10.1007/s00704-016-1908-5, 2016.

Ahmadalipour, A., Rana, A., Moradkhani, H., and Sharma, A.: Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis, Theoretical and Applied Climatology, 1-17, 2015.

Ahmed, K., Shahid, S., and Harun, S. B.: Spatial interpolation of climatic variables in a predominantly arid region with complex topography, Environment Systems and Decisions, 34, 555-563, 2014.

Ahmed, K., Shahid, S., Chung, E.-S., Ismail, T., and Wang, X.-J.: Spatial distribution of secular trends in annual and seasonal precipitation over Pakistan, Clim. Res., 74, 95-107, 2017.

Akhter, J., Das, L., and Deb, A.: CMIP5 ensemble-based spatial rainfall projection over homogeneous zones of India, Climate Dynamics, 49, 1885-1916, 2016.

Anand, A., Mishra, S. K., Sahany, S., Bhowmick, M., Rawat, J. S., and Dash, S.: Indian Summer Monsoon Simulations: Usefulness of Increasing Horizontal Resolution, Manual Tuning, and Semi-Automatic Tuning in Reducing Present-Day Model Biases, Scientific reports, 8, 3522, 2018.

Babar, Z. A., Zhi, X.-f., and Fei, G.: Precipitation assessment of Indian summer monsoon based on CMIP5 climate simulations, Arabian Journal of Geosciences, 8, 4379-4392, 10.1007/s12517-014-1518-4, 2014.

Barfus, K., and Bernhofer, C.: Assessment of GCM capabilities to simulate tropospheric stability on the Arabian Peninsula, International Journal of Climatology, 35, 1682-1696, 2015.

Breiman, L.: Random Forests, Machine Learning, 45, 5-32, 10.1023/A:1010933404324, 2001.

Byg, A., and Salick, J.: Local perspectives on a global phenomenon—climate change in Eastern Tibetan villages, Global Environmental Change, 19, 156-166, 2009.

Cameron, F.: Climate change as a complex phenomenon and the problem of cultural governance, museum and society, 9, 84-89, 2011.

Chen, F.-W., and Liu, C.-W.: Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan, Paddy and Water Environment, 10, 209-222, 10.1007/s10333-012-0319-1, 2012.

Chen, W., Jiang, Z., and Li, L.: Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs, J. Clim., 24, 4741-4756, 2011.

Cramér, H.: Mathematical methods of statistics (PMS-9), Princeton university press, 1999.

Demirel, M. C., and Moradkhani, H.: Assessing the impact of CMIP5 climate multi-modeling on estimating the precipitation seasonality and timing, Clim. Change, 135, 357-372, 10.1007/s10584-015-1559-z, 2015.

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., and Stisen, S.: Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, Hydrology and Earth System Sciences, 22, 1299-1315, 2018.

Dong, T.-Y., Dong, W.-J., Guo, Y., Chou, J.-M., Yang, S.-L., Tian, D., and Yan, D.-D.: Future temperature changes over the critical Belt and Road region based on CMIP5 models, Advances in Climate Change Research, 9, 57-65, 2018.

Duffy, P. B., Brando, P., Asner, G. P., and Field, C. B.: Projections of future meteorological drought and wet periods in the Amazon, Proc Natl Acad Sci U S A, 112, 13172-13177, 10.1073/pnas.1421010112, 2015.

Ekström, M., Grose, M., Heady, C., Turner, S., and Teng, J.: The method of producing climate change datasets impacts the resulting policy guidance and chance of mal-adaptation, Climate Services, 4, 13-29, 2016.

Fu, Y.-H., Lu, R.-Y., and Guo, D.: Changes in surface air temperature over China under the 1.5 and 2.0° C global warming targets, Advances in Climate Change Research, 9, 112-119, 2018.

Giorgi, F., and Mearns, L. O.: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging"(REA) method, J. Clim., 15, 1141-1158, 2002.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres, 113, 2008a.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008b.

Goodman, L. A., and Kruskal, W. H.: Measures of association for cross classifications, Journal of the American statistical association, 49, 732-764, 1954.

Gu, H., Yu, Z., Wang, J., Wang, G., Yang, T., Ju, Q., Yang, C., Xu, F., and Fan, C.: Assessing CMIP5 general circulation model simulations of precipitation and temperature over China, International Journal of Climatology, 35, 2431-2440, 2015.

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

Hargrove, W. W., Hoffman, F. M., and Hessburg, P. F.: Mapcurves: a quantitative method for comparing categorical maps, Journal of Geographical Systems, 8, 187, 2006.

Hegerl, G. C., Black, E., Allan, R. P., Ingram, W. J., Polson, D., Trenberth, K. E., Chadwick, R. S., Arkin, P. A., Sarojini, B. B., and Becker, A.: Challenges in quantifying changes in the global water cycle, Bulletin of the American Meteorological Society, 99, 2018.

IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland,151 pp., pp. 979-1037, 2014.

Jena, P., Azad, S., and Rajeevan, M. N.: Statistical selection of the optimum models in the CMIP5 dataset for climate change projections of Indian monsoon rainfall, Climate, 3, 858-875, 2015.

Jiang, Z., Li, W., Xu, J., and Li, L.: Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation, J. Clim., 28, 8603-8619, 2015.

Johnson, F., and Sharma, A.: Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments, J. Clim., 22, 4373-4382, 2009.

Kazmi, D. H., Li, J., Ruan, C., Zhao, S., and Li, Y.: A statistical downscaling model for summer rainfall over Pakistan, Climate Dynamics, 1-14, 2016.

Kim, J., Ivanov, V. Y., and Fatichi, S.: Climate change and uncertainty assessment over a hydroclimatic transect of Michigan, Stochastic Environmental Research and Risk Assessment, 30, 923-944, 2015.

Kishore, P., Jyothi, S., Basha, G., Rao, S. V. B., Rajeevan, M., Velicogna, I., and Sutterley, T. C.: Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends, Climate Dynamics, 46, 541-556, 10.1007/s00382-015-2597-y, 2015.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in combining projections from multiple climate models, J. Clim., 23, 2739-2758, 2010.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194-1199, 2013.

Koch, J., Demirel, M. C., and Stisen, S.: The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models, Geoscientific Model Development, 11, 1873-1886, 2018.

Krishnamurti, T., Kishtawal, C., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Improved weather and seasonal climate forecasts from multimodel superensemble, Science, 285, 1548-1550, 1999.

Krishnamurti, T. N., Kishtawal, C., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., and Surendran, S.: Multimodel ensemble forecasts for weather and seasonal climate, J. Clim., 13, 4196-4216, 2000.

Latif, M., Hannachi, A., and Syed, F.: Analysis of rainfall trends over Indo-Pakistan summer monsoon and related dynamics based on CMIP5 climate model simulations, International Journal of Climatology, 2018.

Mahmood, R., Jia, S., Tripathi, N., and Shrestha, S.: Precipitation Extended Linear Scaling Method for Correcting GCM Precipitation and Its Evaluation and Implication in the Transboundary Jhelum River Basin, Atmosphere, 9, 160, 2018.

McMahon, T., Peel, M., and Karoly, D.: Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation, Hydrology and Earth System Sciences, 19, 361-377, 2015.

Mendlik, T., and Gobiet, A.: Selecting climate simulations for impact studies based on multivariate patterns of climate change, Clim. Change, 135, 381-393, 10.1007/s10584-015-1582-0, 2016.

Miao, C., Duan, Q., Yang, L., and Borthwick, A. G.: On the applicability of temperature and precipitation data from CMIP3 for China, PLoS One, 7, e44659, 10.1371/journal.pone.0044659, 2012.

Min, S. K., and Hense, A.: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models, Geophys. Res. Lett., 33, 2006.

Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768, 2004.

Pearson, K.: Early statistical papers, University Press, 1948.

Oh, S.-G., and Suh, M.-S.: Comparison of projection skills of deterministic ensemble methods using pseudo-simulation data generated from multivariate Gaussian distribution, Theoretical and Applied Climatology, 129, 243-262, 2017.

Pavan, V., and Doblas-Reyes, F.: Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features, Climate Dynamics, 16, 611-625, 2000.

Perkins, S., Pitman, A., Holbrook, N., and McAneney, J.: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, J. Clim., 20, 4356-4376, 2007.

Pour, S. H., Shahid, S., Chung, E.-S., and Wang, X.-J.: Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh, Atmospheric Research, 2018.

Prasanna, V.: Regional climate change scenarios over South Asia in the CMIP5 coupled climate model simulations, Meteorology and Atmospheric Physics, 127, 561-578, 2015.

Raäisaänen, J.: How reliable are climate models?, Tellus A: Dynamic Meteorology and Oceanography, 59, 2-29, 2007.

Raju, K. S., Sonali, P., and Kumar, D. N.: Ranking of CMIP5-based global climate models for India using compromise programming, Theoretical and Applied Climatology, 128, 563-574, 2016.

Roberts, N. M., and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Monthly Weather Review, 136, 78-97, 2008.

Sachindra, D., Huang, F., Barton, A., and Perera, B.: Statistical downscaling of general circulation model outputs to precipitation—part 1: calibration and validation, International Journal of Climatology, 34, 3264-3281, 2014.

Salman, S. A., Shahid, S., Ismail, T., Ahmed, K., and Wang, X.-J.: Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties, Atmospheric Research, 2018a.

Salman, S. A., Shahid, S., Ismail, T., Ahmed, K., and Wang, X.-J.: Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties, Atmospheric Research, 213, 509-522, 2018b.

Salman, S. A., Shahid, S., Ismail, T., Al-Abadi, A. M., Wang, X.-j., and Chung, E.-S.: Selection of gridded precipitation data for Iraq using compromise programming, Measurement, 132, 87-98, 2018c.

Sansom, P. G., Stephenson, D. B., Ferro, C. A., Zappa, G., and Shaffrey, L.: Simple uncertainty frameworks for selecting weighting schemes and interpreting multimodel ensemble climate change experiments, J. Clim., 26, 4017-4037, 2013.

Sarthi, P. P., Kumar, P., and Ghosh, S.: Possible future rainfall over Gangetic Plains (GP), India, in multi-model simulations of CMIP3 and CMIP5, Theoretical and Applied Climatology, 124, 691-701, 10.1007/s00704-015-1447-5, 2016.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15-40, 10.1007/s00704-013-0860-x, 2013.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15-40, 2014.

Sheffield, J., and Wood, E. F.: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, Climate dynamics, 31, 79-105, 2008.

Sheikh, M., Manzoor, N., and Khan, A.: Climate Profile and Past Climate Changes in Pakistan. Global Change Impact Studies Centre (GCISC), Islamabad Research Report: GCISC-RR-01, 2009.

Sheikh, M. M.: Drought management and prevention in Pakistan, COMSATS 1st meeting on water resources in the south: present scenario and future prospects, Islamabad, 2001, 1-2,

Shiru, M. S., Shahid, S., Alias, N., and Chung, E.-S.: Trend Analysis of Droughts during Crop Growing Seasons of Nigeria, Sustainability, 10, 871, 2018.

Shukla, J., DelSole, T., Fennessy, M., Kinter, J., and Paolino, D.: Climate model fidelity and projections of climate change, Geophys. Res. Lett., 33, 2006.

Smith, I., and Chandler, E.: Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance, Clim. Change, 102, 377-393, 2010.

Smith, J. B., Hulme, M., Jaagus, J., Keevallik, S., Mekonnen, A., and Hailemariam, K.: Climate change scenarios, UNEP Handbook on Methods for Climate Change Impact Assessment and Adaptation Studies, 2, 3-1, 1998.

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., and Vogt, J.: World drought frequency, duration, and severity for 1951-2010, International Journal of Climatology, 34, 2792-2804, 10.1002/joc.3875, 2014.

Thober, S., and Samaniego, L.: Robust ensemble selection by multivariate evaluation of extreme precipitation and temperature characteristics, Journal of Geophysical Research: Atmospheres, 119, 594-613, 2014.

Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A., and Yu, Q.: Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia, International Journal of Climatology, 0, doi:10.1002/joc.5705, 2017.

Wang, X., Yang, T., Li, X., Shi, P., and Zhou, X.: Spatio-temporal changes of precipitation and temperature over the Pearl River basin based on CMIP5 multi-model ensemble, Stochastic Environmental Research and Risk Assessment, 1-13, 2016.

Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C.: Risks of Model Weighting in Multimodel Climate Projections, J. Clim., 23, 4175-4191, 10.1175/2010jcli3594.1, 2010.

Willmott, C. J.: On the validation of models, Physical Geography, 2, 184-194, 10.1080/02723646.1981.10642213, 1981.

Willmott, C. J.: Some comments on the evaluation of model performance, Bulletin of the American Meteorological Society, 63, 1309-1313, 1982.

Wright, D. B., Knutson, T. R., and Smith, J. A.: Regional climate model projections of rainfall from US landfalling tropical cyclones, Climate Dynamics, 45, 3365-3379, 2015.

Wu, Z., Chen, X., Lu, G., Xiao, H., He, H., and Zhang, J.: Regional response of runoff in CMIP5 multi-model climate projections of Jiangsu Province, China, Stochastic Environmental Research and Risk Assessment, 31, 2627-2643, 2016.

Xuan, W., Ma, C., Kang, L., Gu, H., Pan, S., and Xu, Y.-P.: Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China, Theoretical and Applied Climatology, 128, 207-222, 10.1007/s00704-015-1704-7, 2017.

You, Q., Jiang, Z., Wang, D., Pepin, N., and Kang, S.: Simulation of temperature extremes in the Tibetan Plateau from CMIP5 models and comparison with gridded observations, Climate Dynamics, 51, 355-369, 2018.

Zawadzka, J., Mayr, T., Bellamy, P., and Corstanje, R.: Comparing physiographic maps with different categorisations, Geomorphology, 231, 94-100, 2015.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Discussions

Table 1. CMIP5 GCMs considered in this study.

Country	Modelling Centre	Model Name	Resolution in arc degrees (Lat × Lon)
Australia	Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence	CSIRO-Mk3.6.0	1.86 × 1.87
China	Beijing Climate Center	BCC-CSM1.1	2.812 x 2.812
		BCC-CSM1.1(m)	1.125 x 1.125
	The First Institute of Oceanography, SOA	FIO-ESM	2.81 x 2.78
France	Institut Pierre-Simon Laplace	IPSL-CM5A-LR	3.75 x 1.89
		IPSL-CM5A-MR	2.50 x 1.26
Japan	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC-ESM	2.81 x 2.78
		MIROC-ESM-CHEM	2.81 x 2.78
		MIROC5	1.40 x 1.40
	Meteorological Research Institute	MRI-CGCM3	1.12 x 1.12
Norway	Bjerknes Centre for Climate Research, Norwegian Meteorological Institute	NorESM1-M	2.5 x 1.9
South Korea	National Institute of Meteorological Research, Korea Meteorological Administration	HadGEM2-AO	1.87 × 1.25
UK	Met Office Hadley Centre	HadGEM2-ES	1.87 x 1.25
USA	NASA/GISS (Goddard Institute for Space Studies)	GISS-E2-H	2.5 x 2.0
		GISS-E2-R	2.5 x 2.0
	National Center for Atmospheric Research	CCSM4	1.25 x 0.94
		CESM1(CAM5)	1.25 x 0.94
	Geophysical Fluid Dynamics Laboratory	GFDL-CM3	2.5 x 2.01
		GFDL-ESM2G	2.5 x 2.01
		GFDL-ESM2M	2.5 x 2.01

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Table 2. Validation of accuracy of GPCC precipitation using NRMSE and *md*

Stations	**NRMSE**	***md***
Dalbandin	0.09	0.96
Jacobabad	0.52	0.84
Kalat	0.97	0.87
Karachi	0.53	0.84
Nawabshah	0.74	0.74
Padidan	0.59	0.78
Pasni	0.47	0.89
Quetta	0.75	0.76
Sibbi	0.59	0.88
Faisalabad	0.70	0.74
Islamabad	0.45	0.84
Lahore	0.71	0.70
Mianwali	0.72	0.75
Multan	0.73	0.74
Peshawar	0.69	0.72
Sargodha	0.79	0.68
Bahawalnagar	0.53	0.81

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Table 3. GOF values of GCMs obtained using different spatial metrics for mean annual precipitation

GCMs	SPAEF	Lambda	FSS	Cramer-V	Mapcurves	KGE
BCC-CSM1.1(m)	0.142	0.167	0.657	0.646	0.494	0.130
BCC-CSM1-1	0.149	0.194	0.256	0.571	0.405	-4.580
CCSM4	0.431	0.389	0.829	0.687	0.542	0.500
CESM1-CAM5	0.638	0.500	0.881	0.761	0.640	0.600
CSIRO-Mk3.6.0	-0.412	0.139	0.069	0.457	0.317	-5.020
FIO-ESM	0.097	0.139	0.745	0.601	0.436	-0.360
GFDL-CM3	0.433	0.389	0.837	0.731	0.609	0.400
GFDL-ESM2G	0.724	0.444	0.841	0.749	0.616	0.390
GFDL-ESM2M	0.667	0.389	0.820	0.733	0.599	0.350
GISS-E2-H	-0.179	0.111	0.316	0.438	0.300	-0.350
GISS-E2-R	-0.169	0.194	0.303	0.510	0.367	-0.560
HadGEM2-AO	0.321	0.194	0.658	0.573	0.435	0.390
HadGEM2-ES	0.192	0.139	0.606	0.548	0.407	0.270
IPSL-CM5A-LR	0.110	0.139	0.239	0.529	0.360	-3.350
IPSL-CM5A-MR	0.036	0.278	0.244	0.563	0.397	-2.720
MIROC5	0.692	0.222	0.790	0.696	0.547	0.230
MIROC-ESM	0.338	0.361	0.740	0.723	0.598	0.380
MIROC-ESM-CHEM	0.340	0.361	0.728	0.723	0.598	0.370
MRI-CGCM3	-0.319	0.139	0.193	0.395	0.264	-0.600
NorESM1-M	0.514	0.556	0.884	0.780	0.670	0.590

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Discussions

Table 4. Overall ranks of GCMs for mean annual, monsoon and winter precipitation based on rating metric values

Rank	GCM	Annual RM value	GCM	Monsoon RM value	GCM	Winter RM value
1	NorESM1-M	0.91	CESM1-CAM5	0.94	HadGEM2-AO	0.91
2	CESM1-CAM5	0.89	CCSM4	0.83	HadGEM2-ES	0.85
3	GFDL-ESM2G	0.85	GFDL-ESM2G	0.75	GFDL-CM3	0.71
4	GFDL-CM3	0.77	NorESM1-M	0.74	NorESM1-M	0.68
5	GFDL-ESM2M	0.73	GFDL-CM3	0.68	MIROC5	0.64
6	CCSM4	0.69	MIROC-ESM	0.63	MRI-CGCM3	0.61
7	MIROC-ESM	0.63	GFDL-ESM2M	0.63	CESM1-CAM5	0.58
8	MIROC5	0.62	MIROC-ESM-CHEM	0.61	BCC-CSM1-1	0.57
9	MIROC-ESM-CHEM	0.60	HadGEM2-ES	0.57	IPSL-CM5A-LR	0.53
10	HadGEM2-AO	0.47	BCC-CSM1-1-(m)	0.54	GISS-E2-H	0.51
11	BCC-CSM1-1-(m)	0.41	MIROC5	0.53	GFDL-ESM2G	0.47
12	FIO-ESM	0.38	HadGEM2-AO	0.52	MIROC-ESM	0.45
13	HadGEM2-ES	0.34	FIO-ESM	0.44	CCSM4	0.36
14	BCC-CSM1-1	0.29	IPSL-CM5A-LR	0.33	MIROC-ESM-CHEM	0.36
15	IPSL-CM5A-MR	0.27	BCC-CSM1-1	0.23	BCC-CSM1-1-(m)	0.28
16	GISS-E2-R	0.23	IPSL-CM5A-MR	0.18	GISS-E2-R	0.26
17	IPSL-CM5A-LR	0.16	GISS-E2-H	0.15	IPSL-CM5A-MR	0.25
18	GISS-E2-H	0.14	GISS-E2-R	0.11	FIO-ESM	0.23
19	CSIRO-Mk3.6.0	0.08	MRI-CGCM3	0.08	GFDL-ESM2M	0.16
20	MRI-CGCM3	0.06	CSIRO-Mk3.6.0	0.01	CSIRO-Mk3.6.0	0.12

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

Hydrology and Earth System Sciences
Discussions

Table 5. Averaged RM values of GCMs for the identification of ensemble members

GCM	RM value for each season			Average RM value	Overall Rank
	Annual	Monsoon	Winter		
CESM1-CAM5	0.892	0.942	0.575	0.803	**1**
NorESM1-M	0.908	0.742	0.683	0.778	**2**
GFDL-CM3	0.767	0.675	0.708	0.717	**3**
GFDL-ESM2G	0.850	0.750	0.467	0.689	**4**
HadGEM2-AO	0.467	0.517	0.908	0.631	5
CCSM4	0.692	0.825	0.358	0.625	6
MIROC5	0.617	0.533	0.642	0.597	7
HadGEM2-ES	0.342	0.567	0.850	0.586	8
MIROC-ESM	0.633	0.633	0.450	0.572	9
MIROC-ESM-CHEM	0.600	0.608	0.358	0.522	10
GFDL-ESM2M	0.725	0.625	0.158	0.503	11
BCC-CSM1-1-m-	0.408	0.542	0.275	0.408	12
BCC-CSM1-1	0.292	0.233	0.574	0.366	13
FIO-ESM	0.375	0.442	0.225	0.347	14
IPSL-CM5A-LR	0.158	0.333	0.525	0.339	15
GISS-E2-H	0.142	0.150	0.508	0.267	16
MRI-CGCM3	0.058	0.083	0.608	0.250	17
IPSL-CM5A-MR	0.267	0.183	0.250	0.233	18
GISS-E2-R	0.233	0.108	0.258	0.200	19
CSIRO-Mk3.6.0	0.075	0.008	0.117	0.067	20

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

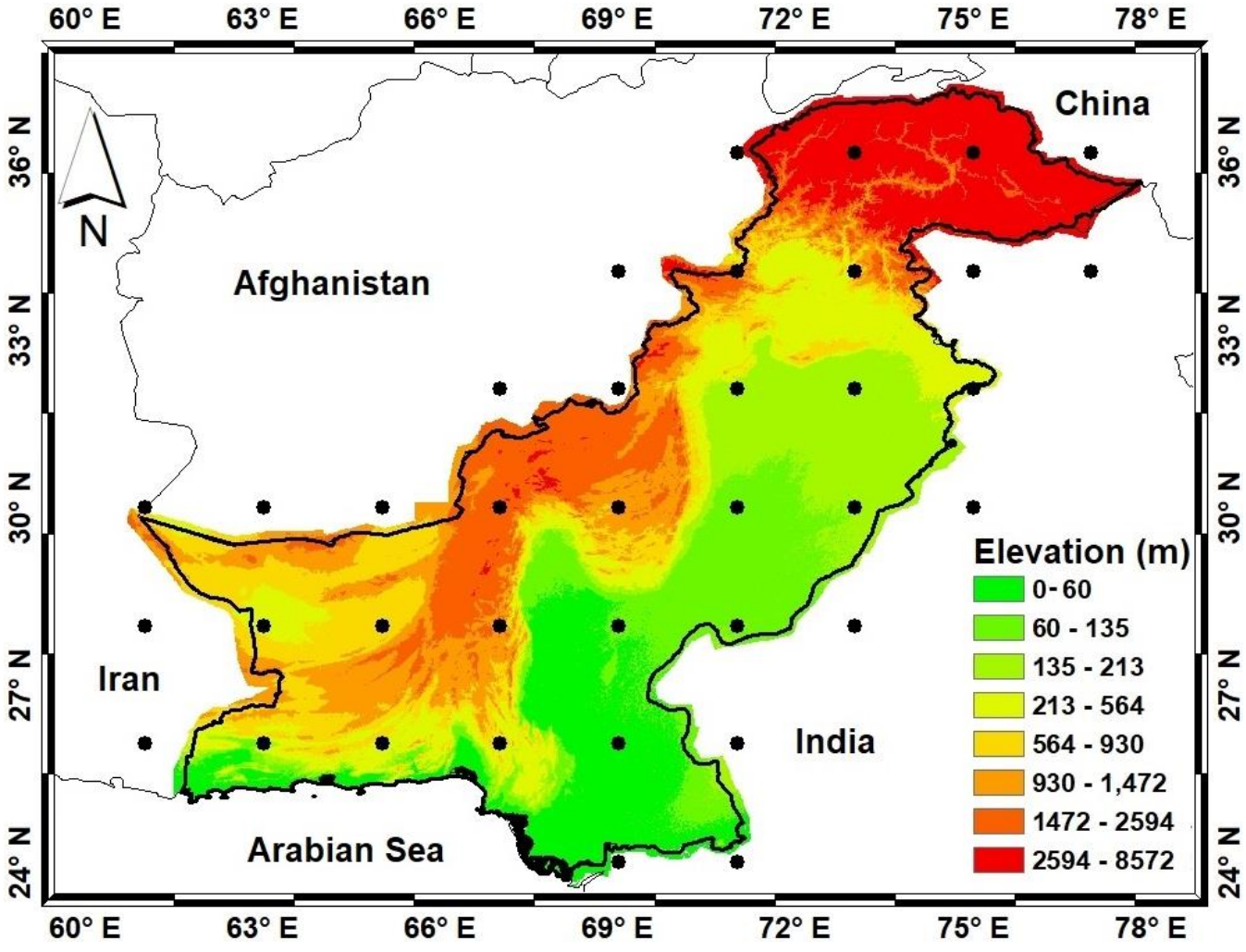

Figure 1. The location of Pakistan in central-south Asia and the GCM grid points over the country.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

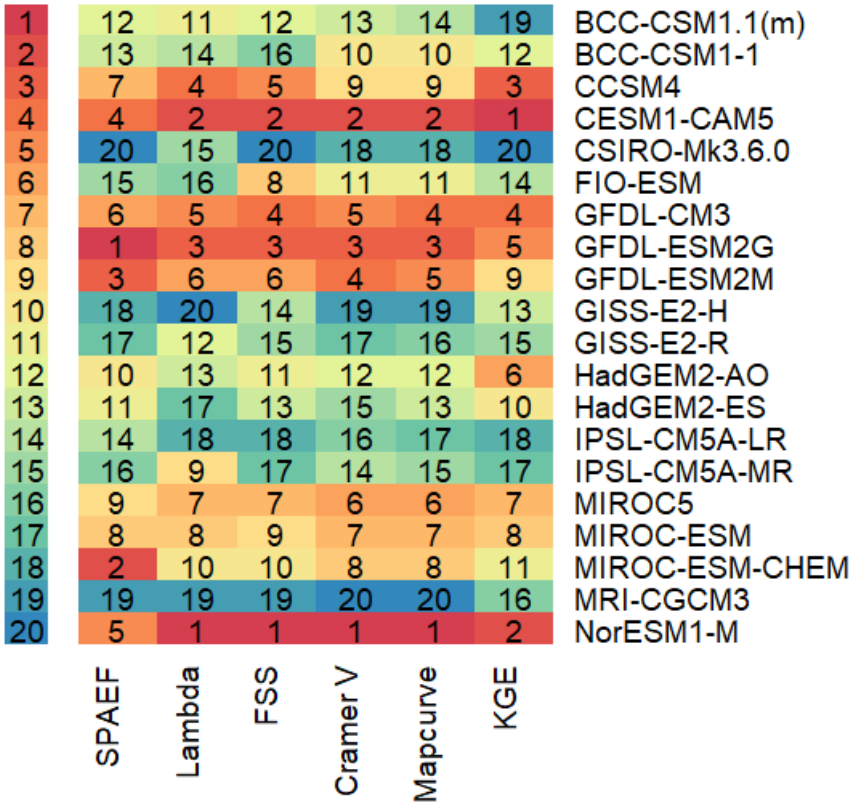

Figure 2. Ranks of GCMs according to their performance in replicating spatial patterns of mean annual precipitation.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

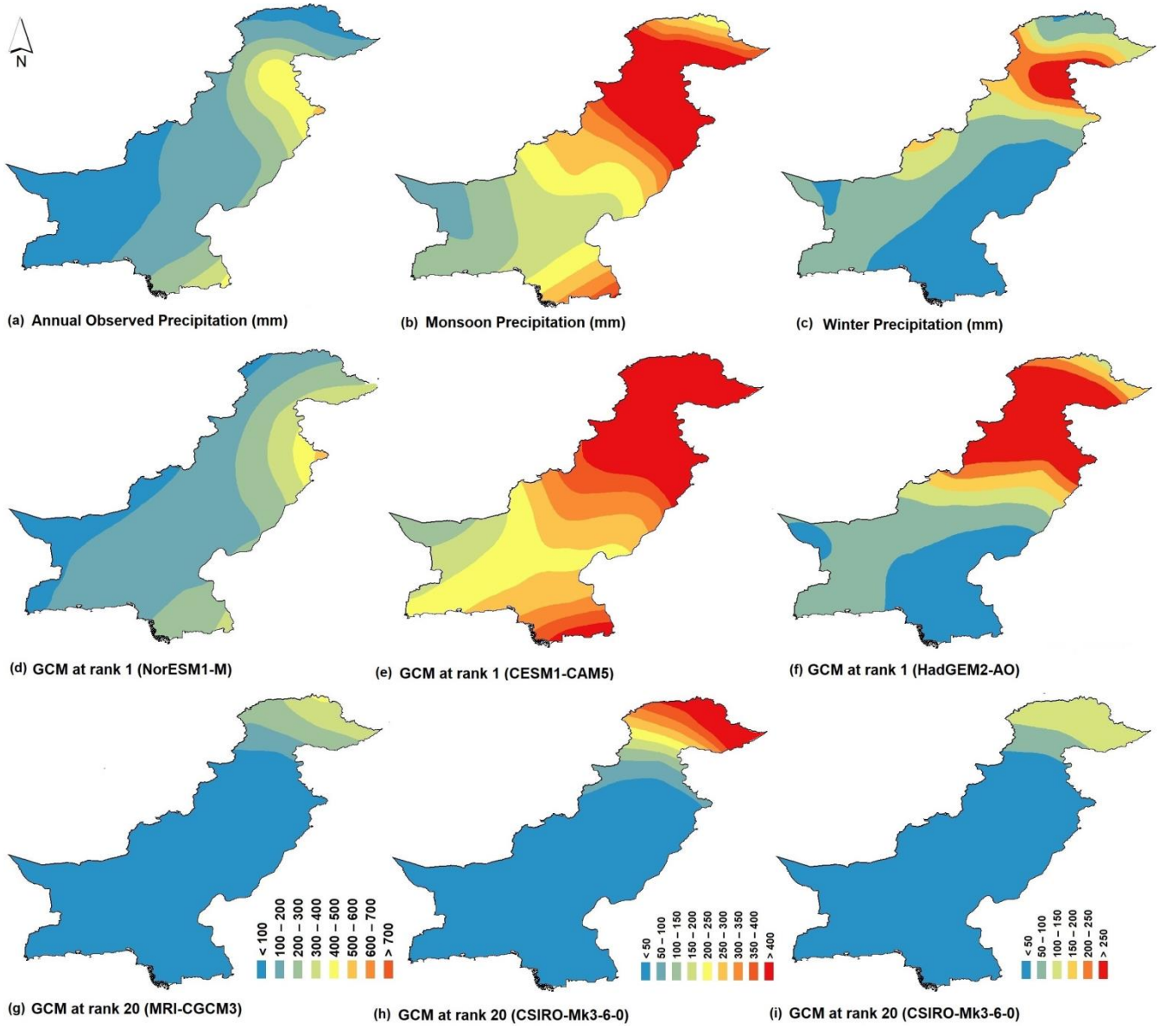

Figure 3. Spatial patterns of GPCC (a - c), GCM at rank 1 (d - f) and GCM at rank 20 (g - i) for mean annual, monsoon, and winter precipitation respectively.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

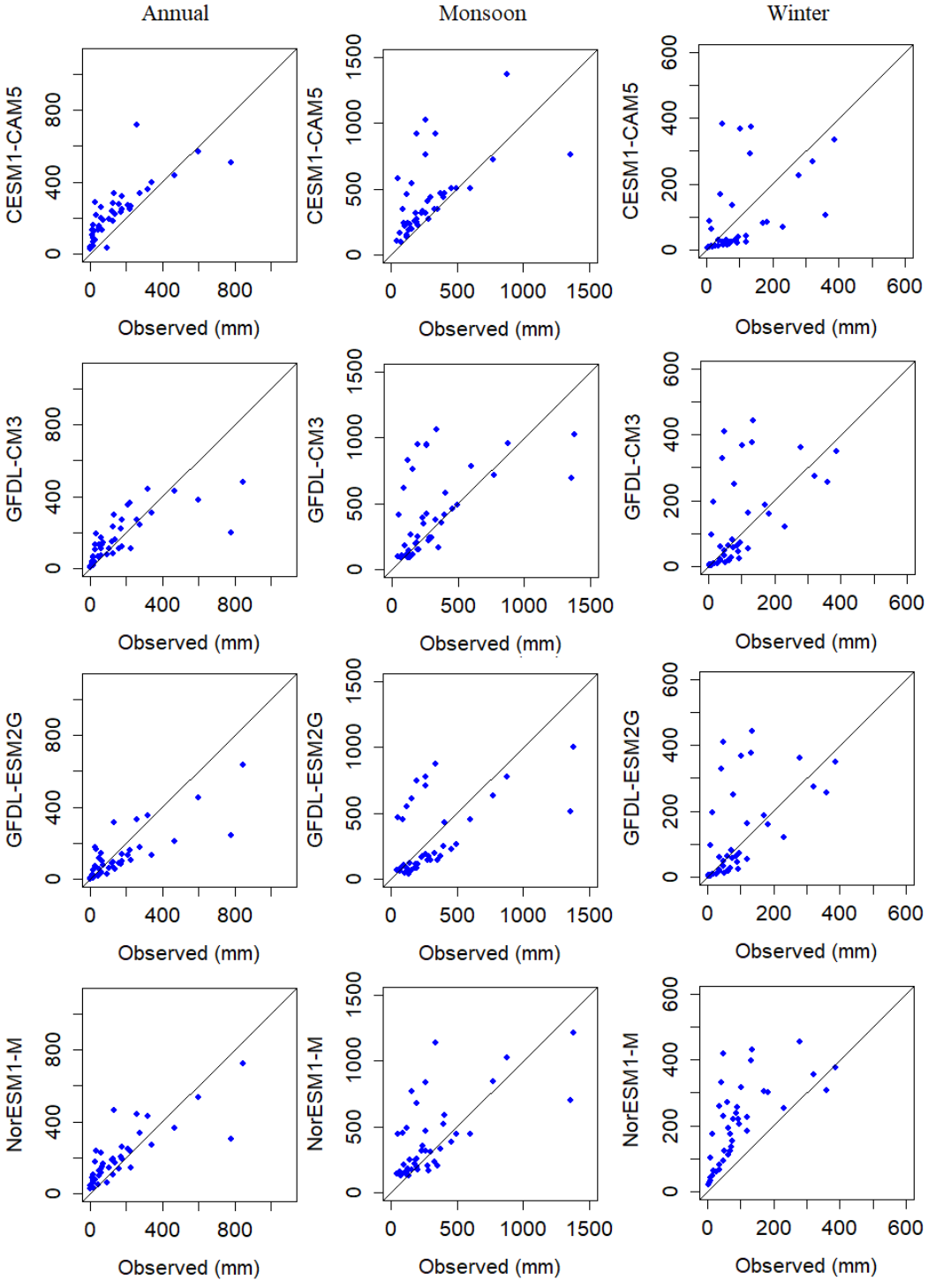

Figure 4. Scatter of precipitation of four top ranked GCMs against GPCC annual, monsoon and winter precipitation.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

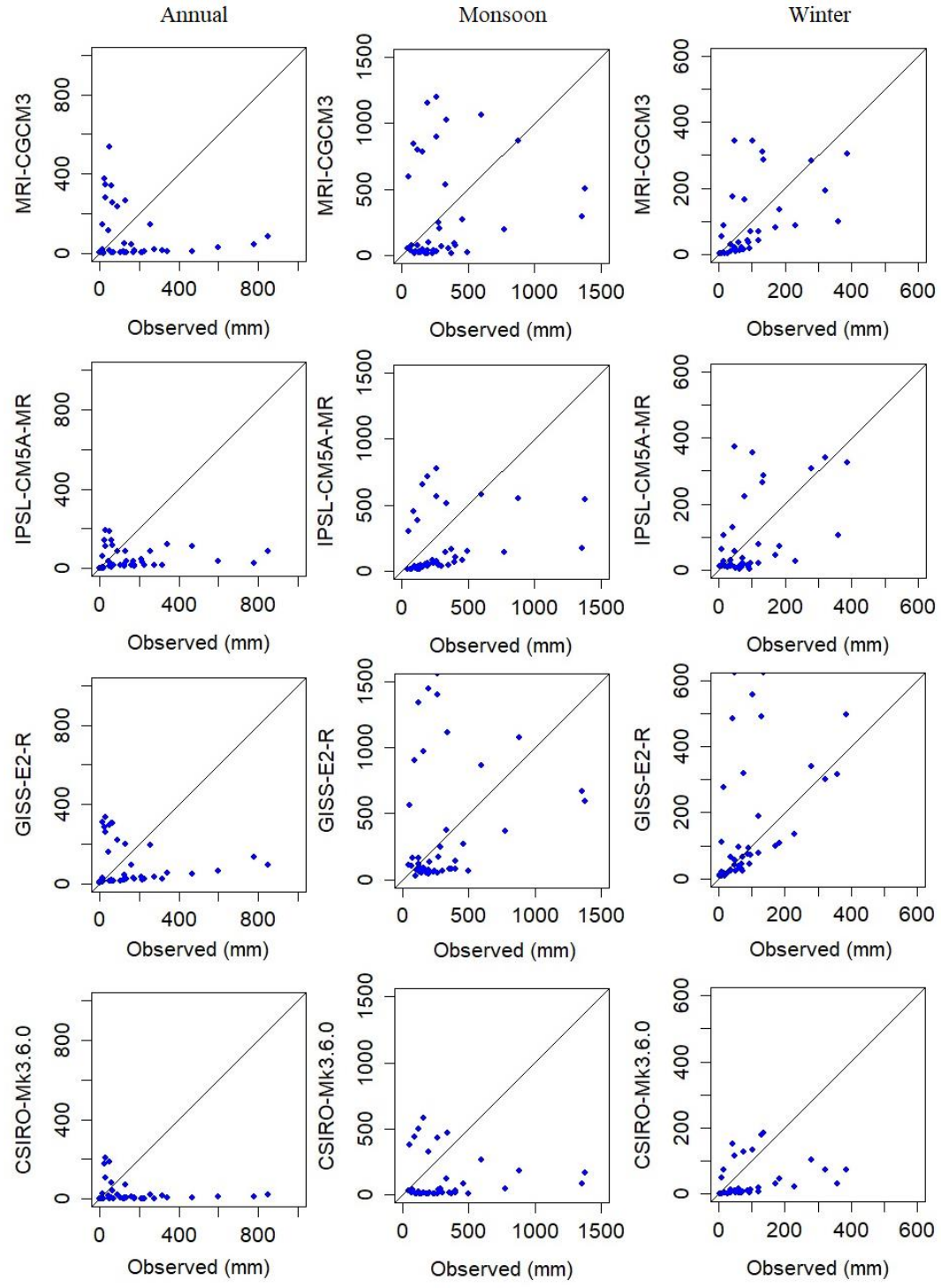

Figure 5. Scatter of precipitation of four lowest ranked GCMs against GPCC annual, monsoon and winter precipitation.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

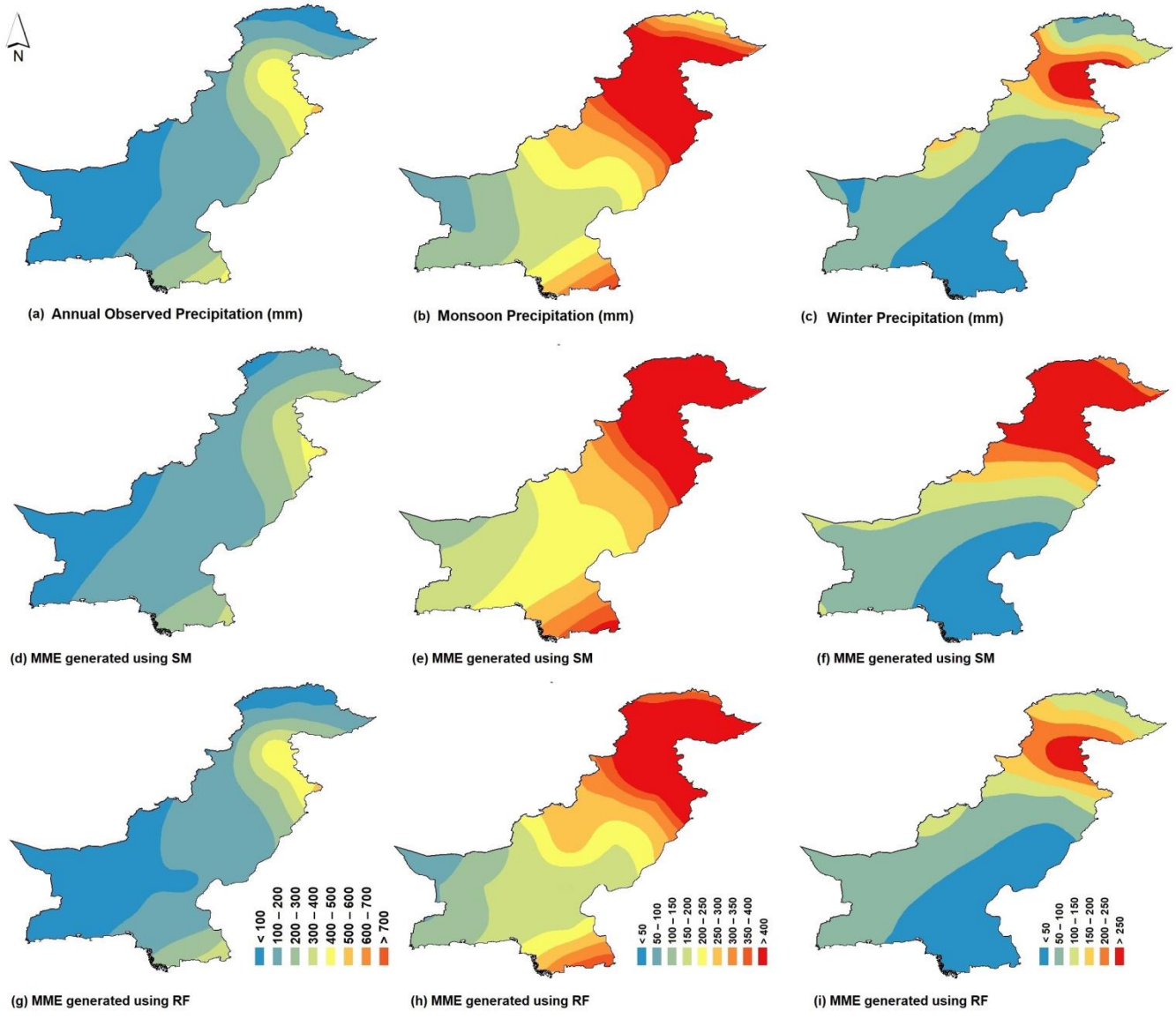

Figure 6. Spatial patterns of observed (a - c), MME computed using Simple Mean (SM) (d - f) and MME computed using Random Forest (RF) (g - i) for mean annual, monsoon, and winter precipitation respectively during 1961 to 2005.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2019

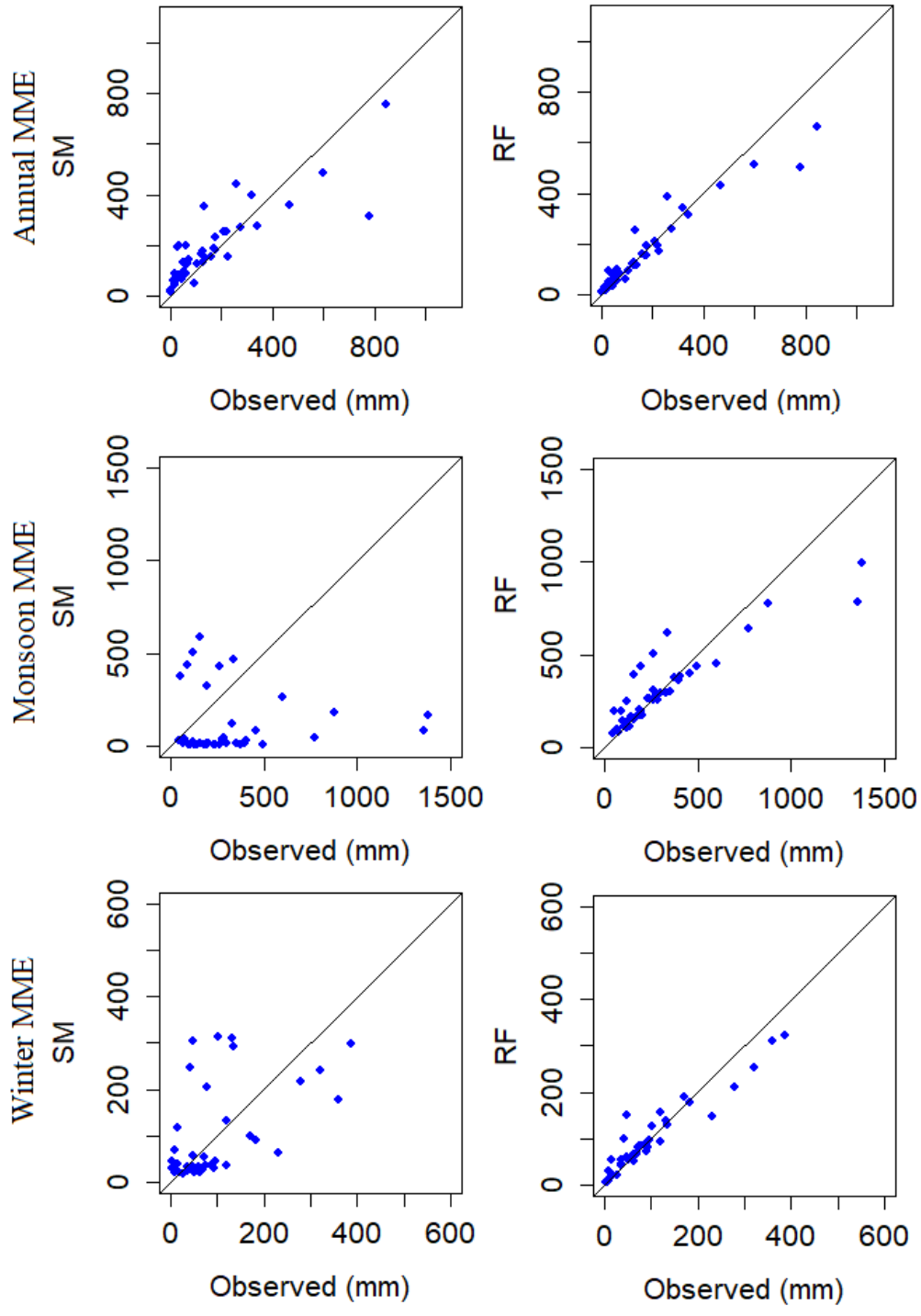

Figure 7. Scatter plots of GPCC and MME mean precipitation for annual, monsoon and winter seasons obtained using Simple Mean (SM) and Random Forest (RF) for the period 1961-2005.