Anonymous Referee #1

Comment

This article proposes an ensemble of GCM model for simulation of precipitation based on spatial assessment metrics. The article presents original research works and outputs. The work is relevant to the interests of the readership of HESS and is well-written. However, there are few issues that need to be addressed. Therefore, authors are encouraged to revise the manuscript accordingly.

Reply

Thanks for your highly constructive comments on our manuscript. The manuscript has now been revised according to the comments. The details of the revisions made are given under each comment. Revisions are marked in Red.

Comment 1

In Section 3, authors introduce different GCM performance assessment metrics. For almost all parameters except Kling-Gupta, the range of the metric and the meaning of the extreme values are elaborated. To be consistent, it is recommended to revise section 3.1.6 accordingly.

Reply

Thanks for your suggestion. We have now defined the range of KGE as below.

"In the present study, KGE was calculated between historical observed data and GCM simulated data using Eq. (12). KGE values can range between –infinity to 1, where values closer to 1 are preferred."

Comment 2

In section 3.3, it is highlighted that the RM values for annual, monsoon, and winter precipitations are averaged to derive overall rank for each GCM. Does this approach flatten the effect of extreme cases? Was it necessary to average them? How were the individual rankings? Authors need to explain the impact of this approach on their final conclusion.

Reply

Thank you very much for this interesting comment. In the original manuscript, in order to derive an overall rank for each GCM, RM values corresponding to annual, monsoon and winter precipitation were first averaged and then based on the average of RM values an overall rank was assigned to each GCM. This procedure helped in assigning one single rank to each GCM while taking into account precipitation for annual, monsoon, and winter seasons all together.

Following your above comment, in the revised manuscript, we ranked each GCM for each season (i.e. annual, monsoon, winter, pre-monsoon, and post-monsoon precipitation) to derive ranks for each variable (precipitation, maximum and minimum temperature) separately by applying comprehensive rating metric. Later, comprehensive rating metric was again applied on precipitation, maximum and minimum temperature ranks to derive an overall rank of GCMs for the whole study area. This procedure helps us to avoid averaging. The obtained results are discussed in section 4.3 as below.

4.3 Overall Ranks of GCMs for Precipitation, Maximum Temperature and Minimum Temperature

The application of various evaluation metrics has yielded different ranks for the same GCM (Ahmadalipour et al., 2017;Raju et al., 2017). The ranks attained by GCMs corresponding to different metrics and seasons (annual, monsoon, winter, pre-monsoon and post-monsoon) were used to calculate the RM values for each GCM. The ranks of GCMs for P, T_{max} and T_{min} are presented in Table 4 along with the RM values. As seen in Table 4, EC-EARTH, BCC-CSM1.1 (m) and CSIRO-Mk3-6-0 were the most skillful GCMs in reproducing the spatial characteristics of P, T_{max} and T_{min} respectively. On the other hand, IPSL-CM5B-LR, CMCC-CM, and INMCM4 displayed the least skill in reproducing the spatial characteristics of P, T_{max} and T_{min} respectively.

Table 4. Ranks of GCMs for P, T_{max} and T_{min} based on rating metric values

GCM	P	Rank	GCM	T_{max}	Rank	GCM	T_{min}	Rank
EC-EARTH	0.823	1	BCC-CSM1.1(m)	0.702	1	CSIRO-Mk3-6-0	0.750	1
NorESM1-M	0.794	2	NorESM1-M	0.663	2	GFDL-ESM2G	0.720	2
GFDL-CM3	0.714	3	HadGEM2-ES	0.656	3	CMCC-CMS	0.692	3
CCSM4	0.689	4	IPSL-CM5B-LR	0.630	4	BCC-CSM1.1(m)	0.684	4
MIROC5	0.685	5	HadGEM2-AO	0.626	5	GFDL-ESM2M	0.681	5
GFDL-ESM2G	0.673	6	CMCC-CMS	0.616	6	MIROC-ESM-CHEM	0.657	6
CESM1-CAM5	0.654	7	HadGEM2-CC	0.608	7	NorESM1-M	0.656	7
HadGEM2-AO	0.651	8	FGOALS-g2	0.600	8	ACCESS1-3	0.656	8
GFDL-ESM2M	0.643	9	CSIRO-Mk3-6-0	0.594	9	MIROC-ESM	0.654	9
FGOALS-g2	0.607	10	ACCESS1-0	0.577	10	MIROC5	0.646	10
MIROC-ESM	0.589	11	IPSL-CM5A-LR	0.566	11	CCSM4	0.631	11
ACCESS1-0	0.555	12	INMCM4	0.561	12	CESM1-BGC	0.628	12
ACCESS1-3	0.555	12	GISS-E2-H	0.556	13	CESM1-CAM5	0.595	13
MIROC-ESM-CHEM	0.532	14	MIROC5	0.551	14	MRI-CGCM3	0.584	14
HadGEM2-CC	0.531	15	BNU-ESM	0.538	15	CanESM2	0.577	15
HadGEM2-ES	0.514	16	BCC-CSM1-1	0.534	16	BNU-ESM	0.569	16
BCC-CSM1-1	0.506	17	GISS-E2-R	0.532	17	FGOALS-g2	0.569	16
CESM1-WACCM	0.482	18	MPI-ESM-LR	0.532	17	MPI-ESM-MR	0.569	16
CNRM-CM5	0.480	19	FIO-ESM	0.524	19	MPI-ESM-LR	0.566	19
CESM1-BGC	0.467	20	CESM1-WACCM	0.522	20	EC-EARTH	0.506	20
INMCM4	0.464	21	ACCESS1-3	0.520	21	IPSL-CM5A-MR	0.490	21
FIO-ESM	0.462	22	GFDL-ESM2M	0.514	22	HadGEM2-ES	0.487	22
MPI-ESM-MR	0.437	23	MPI-ESM-MR	0.513	23	ACCESS1-0	0.481	23
IPSL-CM5A-LR	0.426	24	CCSM4	0.466	24	FIO-ESM	0.446	24
CanESM2	0.406	25	CESM1-BGC	0.459	25	CMCC-CM	0.428	25
MPI-ESM-LR	0.395	26	CanESM2	0.442	26	GISS-E2-R	0.418	26
BCC-CSM1.1(m)	0.394	27	MIROC-ESM	0.442	26	GISS-E2-H	0.416	27
IPSL-CM5A-MR	0.382	28	CNRM-CM5	0.434	28	HadGEM2-AO	0.416	27
CMCC-CMS	0.381	29	EC-EARTH	0.427	29	IPSL-CM5A-LR	0.416	27
MRI-CGCM3	0.381	29	MIROC-ESM-CHEM	0.427	29	BCC-CSM1-1	0.413	30
CMCC-CM	0.353	31	GFDL-ESM2G	0.416	31	HadGEM2-CC	0.413	30
BNU-ESM	0.337	32	GFDL-CM3	0.398	32	CNRM-CM5	0.361	32
GISS-E2-H	0.319	33	CESM1-CAM5	0.371	33	CESM1-WACCM	0.356	33
CSIRO-Mk3-6-0	0.273	34	IPSL-CM5A-MR	0.326	34	IPSL-CM5B-LR	0.275	34
GISS-E2-R	0.253	35	MRI-CGCM3	0.319	35	GFDL-CM3	0.231	35
IPSL-CM5B-LR	0.144	36	CMCC-CM	0.249	36	INMCM4	0.226	36

The better performance of EC-EARTH, BCC-CSM1.1 (m) and CSIRO-Mk3-6-0 in simulating P, T_{max} and T_{min} over Indo-Pak sub-continent has also been reported in several past studies. Latif et al. (2018) reported the relatively better performance of EC-EARTH, and BCC-CSM1.1 (m) out of 36 CMIP5 GCMs in simulating precipitation over Indo-

Pakistan sub-continent based on spatial correlation. Rehman et al. (2018) conducted a study to assess the performance of CMIP5 GCMs in simulating the mean precipitation and temperature over south Asia. The study reported the better performance of EC-EARTH in simulating precipitation and CSIRO-Mk3-6-0 in simulating temperature. Khan et al. (2018) assessed the performance of 31 CMIP5 GCMs in simulating the mean precipitation and temperature over Pakistan using multiple daily gridded datasets and identified EC-EARTH as the best GCM for simulating precipitation and CSIRO-Mk3-6-0 for simulating temperature. Better performance of CSIRO-Mk3-6-0 in simulating maximum and minimum temperature is also reported in the study by (Ahmed et al., 2019b).

Regarding the issue of deriving rank based on average of RM values (given in the original manuscript) and overall ranks based on individual ranks; a comparison was made. The comparison of overall ranks obtained with the above two approaches is shown in Table below (not included in the manuscript). As seen in Table below, it was understood that the differences between the overall ranks derived based on; average of RM values and overall ranks is mostly quite small. Therefore, it can be stated that either the overall ranks can be derived based on average of RM values or individual ranks. However, derivation of overall ranks based on ranks of individual variable is relatively simple and hence recommended.

GCM	P	T_{max}	T_{min}	Avg RM value	Rank based on average RM value	Rank based on overall RM values	Difference
ACCESS1-0	0.555	0.577	0.481	0.537	17	13	4
ACCESS1-3	0.555	0.520	0.656	0.577	9	4	5
BCC-CSM1-1	0.506	0.534	0.413	0.484	6	3	3
BCC-CSM1.1(m)	0.394	0.702	0.684	0.594	22	23	-1
BNU-ESM	0.337	0.538	0.569	0.481	23	24	-1
CanESM2	0.406	0.442	0.577	0.475	25	26	-1
CCSM4	0.689	0.466	0.631	0.595	5	7	-2
CESM1-BGC	0.467	0.459	0.628	0.518	18	19	-1
CESM1-CAM5	0.654	0.371	0.595	0.540	14	18	-4
CESM1-WACCM	0.482	0.522	0.356	0.454	27	29	-2
CMCC-CM	0.353	0.249	0.428	0.343	36	36	0
CMCC-CMS	0.381	0.616	0.692	0.563	11	6	5
CNRM-CM5	0.480	0.434	0.361	0.425	31	34	-3
CSIRO-Mk3-6-0	0.273	0.594	0.750	0.539	15	12	3
EC-EARTH	0.823	0.427	0.506	0.585	8	16	-8
FGOALS-g2	0.607	0.600	0.569	0.592	7	10	-3
FIO-ESM	0.462	0.524	0.446	0.477	24	25	-1
GFDL-CM3	0.714	0.398	0.231	0.448	28	28	0
GFDL-ESM2G	0.673	0.416	0.720	0.603	4	8	-4
GFDL-ESM2M	0.643	0.514	0.681	0.613	3	5	-2
GISS-E2-H	0.319	0.556	0.416	0.430	29	30	-1
GISS-E2-R	0.253	0.532	0.418	0.401	33	32	1
HadGEM2-AO	0.651	0.626	0.416	0.564	10	9	1
HadGEM2-CC	0.531	0.608	0.413	0.517	19	17	2
HadGEM2-ES	0.514	0.656	0.487	0.552	13	11	2
inmcm4	0.464	0.561	0.226	0.417	32	27	5
IPSL-CM5A-LR	0.426	0.566	0.416	0.469	26	20	6
IPSL-CM5A-MR	0.382	0.326	0.490	0.399	34	35	-1
IPSL-CM5B-LR	0.144	0.630	0.275	0.350	35	31	4
MIROC-ESM-CHEM	0.532	0.427	0.657	0.539	2	14	-12
MIROC-ESM	0.589	0.442	0.654	0.561	12	15	-3
MIROC5	0.685	0.551	0.646	0.627	16	2	14
MPI-ESM-LR	0.395	0.532	0.566	0.498	21	21	0
MPI-ESM-MR	0.437	0.513	0.569	0.506	20	22	-2
MRI-CGCM3	0.381	0.319	0.584	0.428	30	33	-3
NorESM1-M	0.794	0.663	0.656	0.704	1	1	0

Comment 3

It is needed to give some background knowledge about Random Forest method. Why is it selected? It is needed to give some reasoning for this selection. Also in the results and discussion, more explanation is needed for this method.

Reply

Thank you very much for this comment. We have now added a new sub-section covering the information related to Random Forest algorithm as show below. .

3.5.2 Random Forest (RF)

Random Forest (RF) algorithm (Breiman, 2001) was used in the calculation of the mean time series of P, T_{max} and T_{min} corresponding to an MME of four top ranked GCMs. RF is a relatively new machine learning algorithm widely used in modelling non-linear relationships between predictors and predictands (Ahmed et al., 2019a). RF algorithm is found to perform well with spatial data sets and less prone to over-fitting (Folberth et al., 2019). Most importantly Folberth et al. (2019) reported that RF is less sensitive to multivariate correlation.

RF is an ensemble technique where regression is done using multiple decision trees. RF algorithm uses the following steps in regression.

1. A bootstrap resampling method is used to select sample sets from training data.
2. Classification And Regression Tree (CART) technique is used to develop unpruned trees using the bootstrap sample.
3. A large number of trees are developed with the samples selected repetitively from training data so that all training data have equal probability of selection.
4. A regression model is fitted for all the trees and the performance of each tree is assessed.
5. Ensemble prediction is estimated by averaging the predictions of all trees which is considered as the final prediction.

Wang et al. (2017) and He et al. (2016) reported that the performance of RF varies with the number of trees (*ntree*) and the number of variables randomly sampled at each split in developing the trees (*mtry*). It was observed that RF performance increases with the increase in *ntree*. However, in the present study the performance was not found to increase significantly in term of root mean square error when the *ntree* was greater than 500. Therefore, *ntree* was set to 500 while the *mtry* was set to $p/3$ where p is the number of variables (i.e. GCMs) used for developing RF-based MME.

The MME prediction can be improved by assigning larger weight to the GCMs which show better performance (Sa'adi et al., 2017). RF regression models developed using historical P, T_{max} and T_{min} simulations of GCMs as independent variable and historical observed P, T_{max} and T_{min} as dependent variable provide weights to the GCMs according to their ability to simulate historical observed P, T_{max} and T_{min}.

The “Random Forest” package written in R programming language was employed in this study for developing RF-based MMEs. RF-based MMEs were calibrated with the first 70% of the data and validated with the rest of the data.

Comment 4

Section 4.1, it is suggested to present NRMSE formula.

Reply

Thank you very much for your suggestion. We have now added a sub-section entitled “Accuracy Assessment of Gridded Precipitation Data” under the method section 3.1 and provided details on NRMSE and *md*.

3.1 Accuracy Assessment of Gridded Precipitation and Temperature Data

The accuracy of gridded GPCC precipitation data and CRU temperature data was assessed by comparing them with observed station data using NRMSE and *md*. NRMSE is a non-dimensional form of Root Mean Square Error (RMSE) which is derived by normalizing RMSE by variance of observations. 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 closer to 0 are preferred as they denote smaller errors (Chen and Liu, 2012). In this study, NRMSE was calculated Eq. 1.

$$NRMSE = \frac{\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_{sim,i} - x_{obs,i}\right)^2\right]^{1/2}}{x_{max} - x_{min}} \qquad (1)$$

Where $x_{sim,i}$ and $x_{obs,i}$ refer to the i^{th} value in the gridded and observed time series of the climate variable (i.e. precipitation or temperature) respectively, and N is the number of data points in each time series.

The ‘*md*’ shown in Eq. 2 is widely used to estimate the agreement between observed and gridded data of climate variables (Noor et al., 2019;Ahmed et al., 2019a). It varies between 0 (no agreement) and 1 (perfect agreement) (Willmott, 1981).

$$md = 1 - \frac{\sum_{i=1}^{n}\left(x_{obs,i} - x_{sim,i}\right)^{j}}{\sum_{i=1}^{n}\left(\left|x_{sim,i} - \overline{x_{obs}}\right| + \left|x_{obs,i} - \overline{x_{obs}}\right|\right)^{j}} \quad (2)$$

Where $x_{sim,i}$ and $x_{obs,i}$ are the i^{th} data point in the gridded data and observed data series of a climate variable.

Comment 5

To have a better understanding about the site, it would be good to add the location of stations on the map.

Reply

Thank you very much for your suggestion. We have revised Figure 1 and included the locations of the stations. Also, we have provided the names of stations in Table 2.

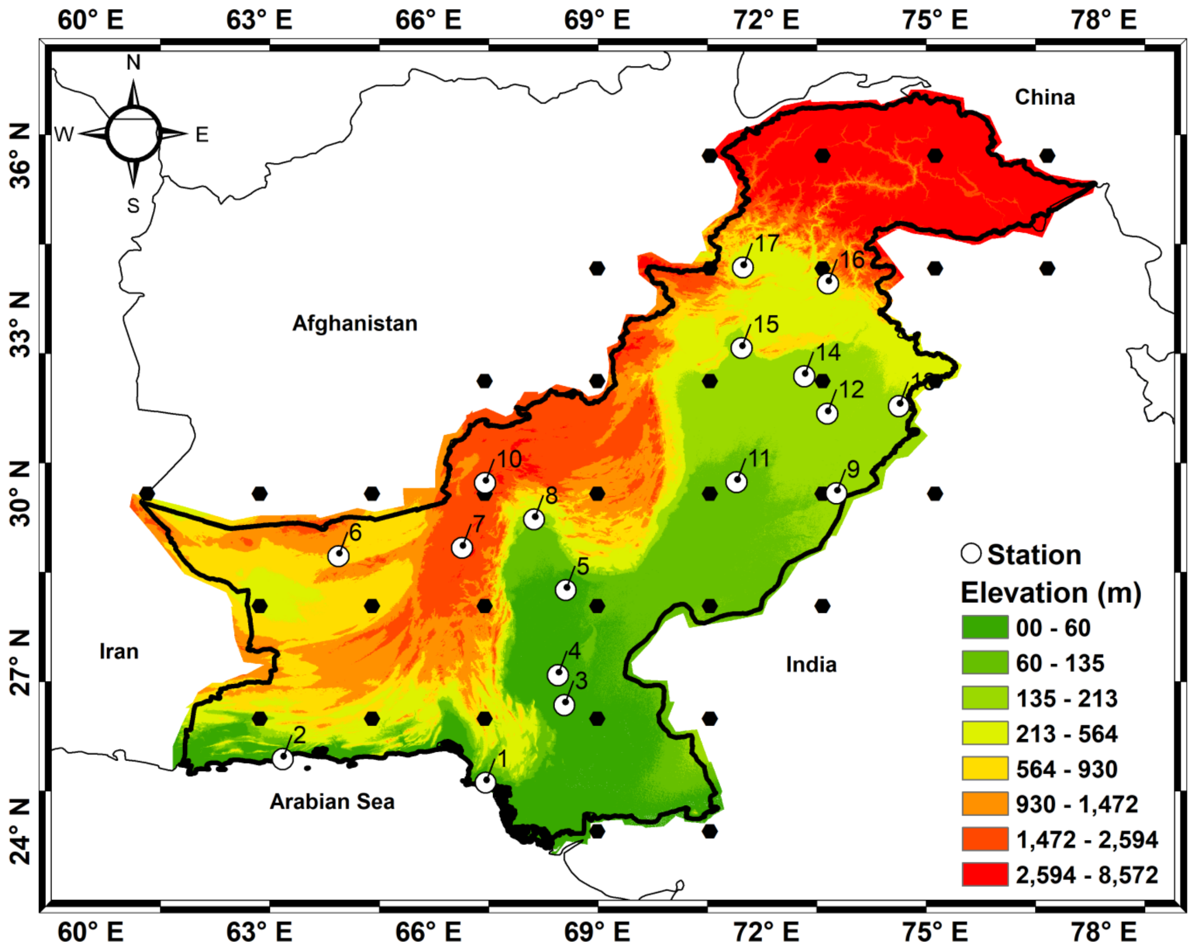

Figure 1. The location of Pakistan in central-south Asia and the GCM grid points over the country along with locations of precipitation and temperature observation stations. The names of the stations are given in Table 2.

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

		Precipitation (P)		Maximum Temperature (T_{max})		Minimum Temperature (T_{min})	
Station No	Station Name	NRMSE	*md*	NRMSE	*md*	NRMSE	*md*
1	Karachi	0.530	0.840	0.270	0.880	0.180	0.919
2	Pasni	0.470	0.890	0.310	0.840	0.260	0.879
3	Nawabshah	0.740	0.740	0.300	0.850	0.170	0.919
4	Padidan	0.590	0.780	0.190	0.920	0.150	0.939
5	Jacobabad	0.520	0.840	0.100	0.960	0.090	0.959
6	Dalbandin	0.090	0.960	0.140	0.940	0.230	0.889
7	Kalat	0.970	0.870	0.240	0.900	0.470	0.779
8	Sibbi	0.590	0.880	0.390	0.810	0.260	0.889
9	Bahawalnagar	0.530	0.810	0.310	0.899	0.270	0.881
10	Quetta	0.750	0.760	0.240	0.890	0.120	0.949
11	Multan	0.730	0.740	0.120	0.950	0.120	0.949
12	Faisalabad	0.700	0.740	0.210	0.900	0.170	0.919
13	Lahore	0.710	0.700	0.140	0.940	0.110	0.959
14	Sargodha	0.790	0.680	0.160	0.930	0.170	0.919
15	Mianwali	0.720	0.750	0.240	0.890	0.120	0.949
16	Islamabad	0.450	0.840	0.160	0.930	0.190	0.909
17	Peshawar	0.690	0.720	0.190	0.920	0.110	0.949

Comment 6

It is also recommended to highlight the limitations of the study in the discussion part

Reply

Thanks for your suggestion. We have now added the following paragraph to the manuscript to highlight the limitations of this study and recommendation for future work.

"In this study performance of GCMs was assessed based on their ability to simulate past observed P, T_{max} and T_{min} and hence the best performing GCMs were identified and used for the development of MMEs. However, it is found that past and future climate may have a weak association hence it is not necessary that if a GCM performs well in the past will give reliable results in future (Knutti et al., 2010). In other words, the best GCMs selected for the MMEs considering their ability to simulate past climate may not be the best in the future under changing climate (Ruane and McDermid, 2017;Ahmed et al., 2019b). This is due to the large uncertainties associated with GHG emission scenarios and GCMs. As a solution to this limitation, Salman et al. (2018) selected an ensemble of GCMs based on past performance as well as the degree of agreement between their future projections. The study detailed in the present manuscript can be repeated in future to select GCMs considering their past performance and the degree of agreement in their future projections.

In the present study, the MME of P, T_{max} and T_{min} were developed by considering top four ranked GCMs. In the past, MMEs were developed considering 3 to 10 top ranked GCMs. However, none of the study showed the performance of MME by varying the number of GCMs in MME. The performance of an MME can be sensitive to the choice of the number of GCMs. Hence, in future, a study should be conducted to investigate the impact of the number of GCMs used for the development of the MME.

Only RF algorithm was used in this study for the development of MMEs. Other machine learning algorithms (e.g. Artificial Neural Networks, Support Vector Machine, Relevance Vector Machine, K-nearest neigbour, Extreme Learning method) can also be used for the development of MMEs. A comparison of the performance of MMEs developed with different machine learning algorithms can assist in identification of pros and cons of different algorithms in relation to development of MMEs.

In the present study, GCM ranking and MME development was conducted only considering P, T_{max} and T_{min} pertaining to annual, monsoon, winter, pre-monsoon and post-monsoon seasons. However, several studies reported that the ranking of GCMs based on different climate variables may assist in the identification of a more dependable set of GCMs for ensemble generation (Johnson and Sharma, 2012;Xuan et al., 2017). In future, the ranking of GCMs can be conducted considering several climate variables (e.g. precipitation, mean temperature, maximum temperature, minimum temperature, wind speed, evapotranspiration and solar radiation)."

Comment 7

For figures 4, 5, and 7 a performance measure such as r-squared is needed for each scatter plot of observed vs simulated data points.

Reply

Thanks for your suggestion. We have indicated the performance of MMEs in terms of the modified index of agreement (*md*) in all plots in Figures 4, 5 and 7 as shown below.

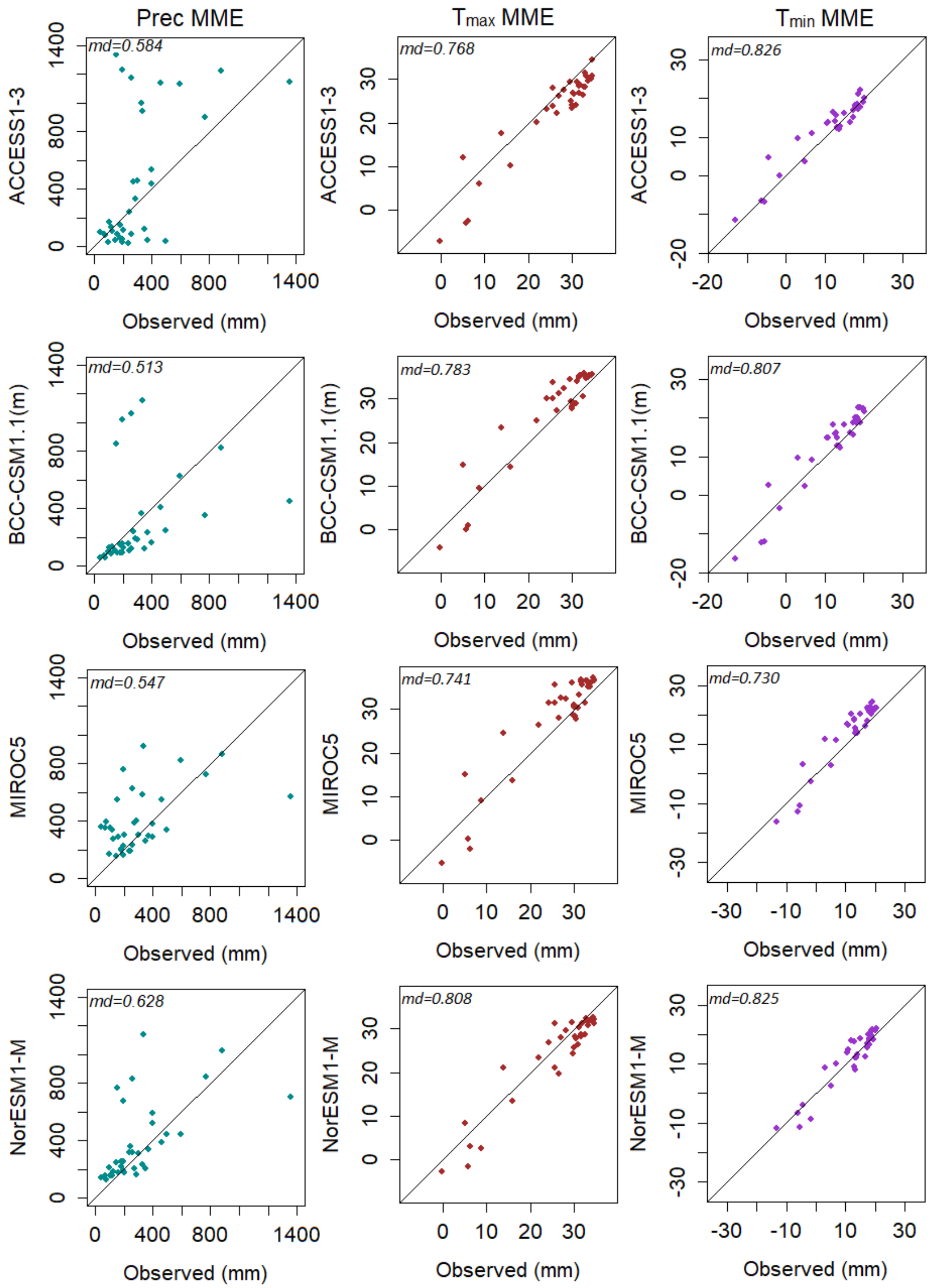

Figure 4. Scatter of spatially averaged annual P, T_{max} and T_{min} of four top ranked GCMs against GPCC P, CRU T_{max} and CRU T_{min} for the period 1961 to 2005.

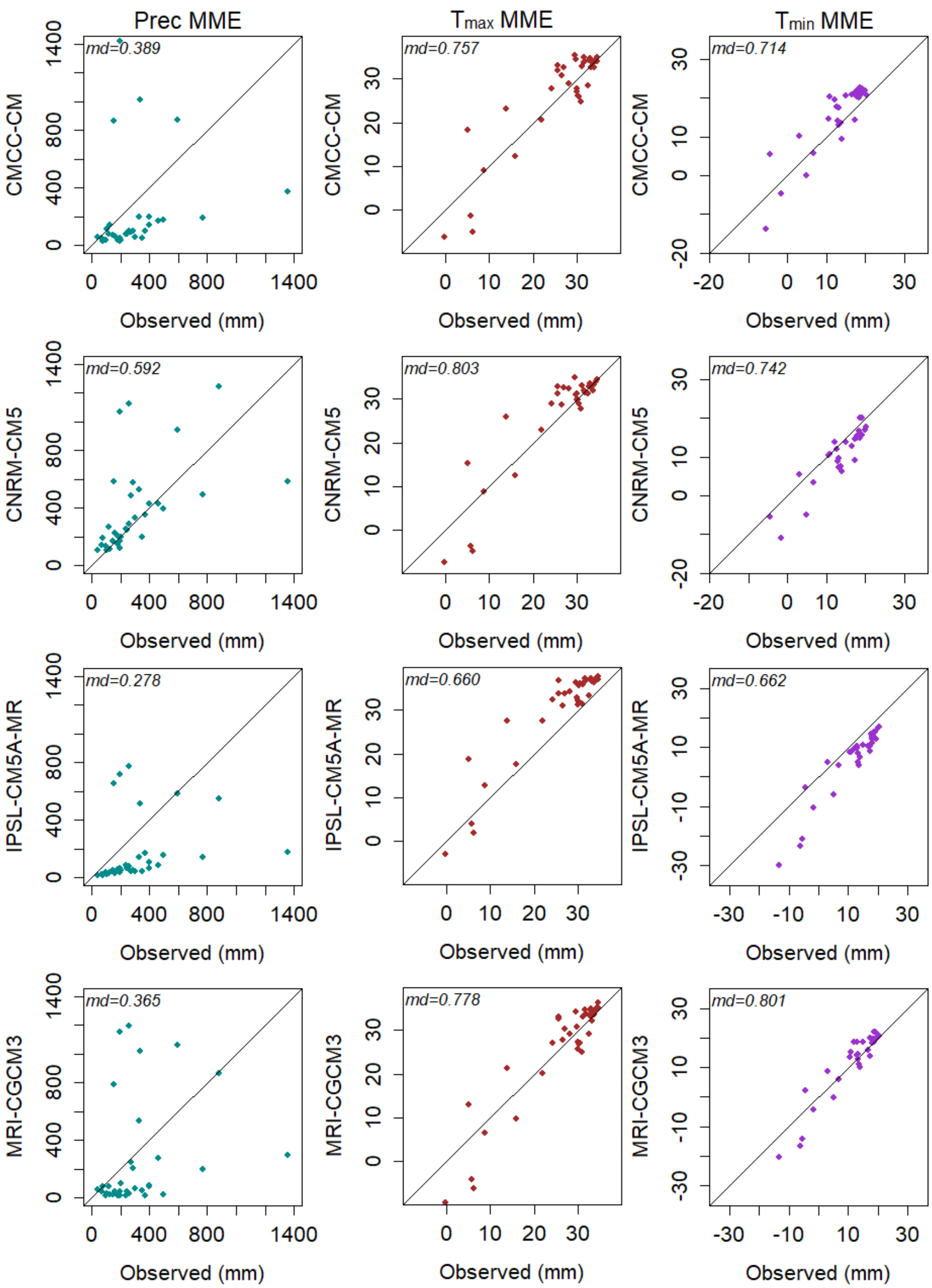

Figure 5. Scatter of spatially averaged annual P, T_{max} and T_{min} of four lowest ranked GCMs against GPCC P, CRU T_{max} and CRU T_{min} for the period 1961 to 2005.

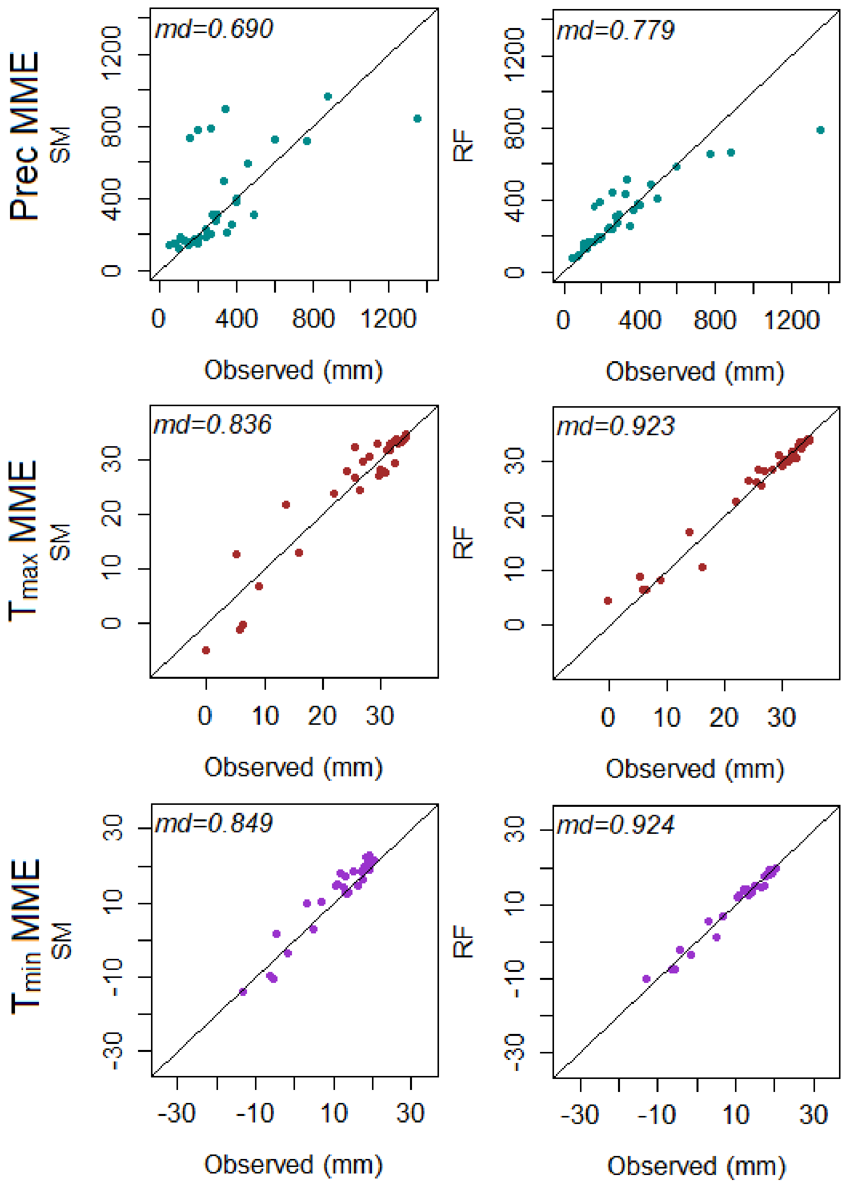

Figure 7. Scatter of spatially averaged mean annual GPCC *P*, CRU T_{max} and CRU T_{min} MME of four top ranked GCMs against *P*, CRU T_{max} and CRU T_{min} using Simple Mean (SM) and Random Forest (RF) for the period 1961 to 2005.

References

Ahmadalipour, A., Rana, A., Moradkhani, H., and Sharma, A.: Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis, Theor. Appl. Climatol., 128, 71-87, 10.1007/s00704-015-1695-4, 2017.

Ahmed, K., Shahid, S., Nawaz, N., and Khan, N.: Modeling climate change impacts on precipitation in arid regions of Pakistan: a non-local model output statistics downscaling approach, Theor. Appl. Climatol., 137, 1347-1364, 10.1007/s00704-018-2672-5, 2019a.

Ahmed, K., Shahid, S., Sachindra, D. A., Nawaz, N., and Chung, E.-S.: Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method, J. Hydrol., 573, 281-298, https://doi.org/10.1016/j.jhydrol.2019.03.092, 2019b.

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

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.

Folberth, C., Baklanov, A., Balkovič, J., Skalský, R., Khabarov, N., and Obersteiner, M.: Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning, Agr. Forest Meteorol., 264, 1-15, 2019.

He, X., Chaney, N. W., Schleiss, M., and Sheffield, J.: Spatial downscaling of precipitation using adaptable random forests, Water Resour. Res., 52, 8217-8237, 2016.

Johnson, F., and Sharma, A.: A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations, Water Resour. Res., 48, 2012.

Khan, N., Shahid, S., Ahmed, K., Ismail, T., Nawaz, N., and Son, M.: Performance Assessment of General Circulation Model in Simulating Daily Precipitation and Temperature Using Multiple Gridded Datasets, Water, 10, 1793, 2018.

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.

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, Int. J. Climatol., 38, e577-e595, 2018.

Noor, M., Ismail, T. b., Shahid, S., Ahmed, K., Chung, E.-S., and Nawaz, N.: Selection of CMIP5 multi-model ensemble for the projection of spatial and temporal variability of rainfall in peninsular Malaysia, Theor. Appl. Climatol., 10.1007/s00704-019-02874-0, 2019.

Raju, K. S., Sonali, P., and Kumar, D. N.: Ranking of CMIP5-based global climate models for India using compromise programming, Theor. Appl. Climatol., 128, 563-574, 2017.

Rehman, N., Adnan, M., and Ali, S.: Assessment of CMIP5 climate models over South Asia and climate change projections over Pakistan under representative concentration pathways, International Journal of Global Warming, 16, 381-415, 2018.

Ruane, A. C., and McDermid, S. P.: Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment, Earth Perspectives, 4, 1, 10.1186/s40322-017-0036-4, 2017.

Sa'adi, Z., Shahid, S., Chung, E.-S., and bin Ismail, T.: Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models, Atmos. Res., 197, 446-460, 2017.

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, Atmos. Res., 213, 509-522, https://doi.org/10.1016/j.atmosres.2018.07.008, 2018.

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, Int. J. Climatol., 0, doi:10.1002/joc.5705, 2017.

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, Bull. Am. Meteorol. Soc., 63, 1309-1313, 1982.

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, Theor. Appl. Climatol., 128, 207-222, 10.1007/s00704-015-1704-7, 2017.