Evaluation of CMIP5 GCMs Multimodel Ensemble Precipitation over Pakistan

Farah Ikram
Pakistan Meteorological Department, Islamabad, Pakistan

Abstract. This study provides an overview of the performance of the state-of-the-art global climate models of Coupled Model Intercomparison Project Phase 5 (CMIP5) in representing the present climate precipitation over Pakistan against the reanalysis dataset AgMERRA. The analysis have been done over the time period 1980-2010. The overall performance of the models are summarized by a heat map diagram based on RMSE's w.r.t observed dataset. The results show that ENSMEAN and ENSMEDIAN models outperform individual models in all months. EC-Earth show better performance for winter months and GFDL-ESM-2M, GFDL-ESM-2G, GFDL-CM3 for summer along with MRI-CGCM3 for all months. A seasonality index has also been calculated for comparing the representation of seasonal cycle with the reanalysis dataset. MIROC5 and EC-Earth has relatively better performance over Pakistan. Box and whisker analysis show that for winter (JFM) months CMIP5 has relatively better performance as compared to the reanalysis dataset while there is difference of up to 1.5mm/day for the rest of the months. There is large variability between the models for summer months (JJAS).

Key words: CMIP5, AgMERRA, RMSE, Seasonality Index, ENSMEAN, ENSMEDIAN.

Introduction

The Coupled Models Inter-comparison Project Phase 5 (CMIP5) provides an ultimate tool especially for researchers in developing countries to access the state of the art new generation Global Circulation Models (GCM) with much finer resolution than the previous generation of models (CMIP3) (Meehl, 2007; Taylor et al., 2012: 485). The evaluation of these models is of ultimate importance for global as well as regional studies based on climate change impacts since downscaling of the whole set of GCMs for such studies requires intensive computational and human resources. The climate model simulations are subject to uncertainties introduced through the model parameterizations, input parameters as well as the structure of the model (Knutti et al., 2010: 2739; IPCC AR5, 2013; Rupp et al., 2013: 884; Mc Sweeny et al., 2014). Precipitation and temperature are the key controls of the hydrological cycle.

Therefore, quantification of uncertainties and biases in GCM simulations of these two meteorological parameters are necessary for understanding the application of these simulations in hydrology as well as in climate change impact studies (Giorgi and Mearns, 2002: 1141; Gleckler et al., 2008; Gelaro et al., 2017: 5419).

The main contribution to the uncertainty of the model comes from our limited understanding of the processes taking place. These parametric errors, also caused by small-scale processes or missing processes in the models, are transformed into large-scale effects. As climate change has become more evident in recent decades, which is mainly caused by human activities, it has become a challenge for policymakers regarding adaptation and mitigation of future climate change and its impacts (Knutti et al., 2010: 2739). The model that fails to simulate the past climate correctly will also lose its credibility for the future projections as well (Rupp et al., 2013: 884; Ahmed et al., 2019: 4803; Flato et al., 2014: 741).

The main goal of this paper is to evaluate model performance for those who use their outputs for impact based studies involving downscaling of GCM data using several techniques and use as input for hydrological or crop modelling etc. Three sources of uncertainties have been discussed by (Hawkins and Sutton 2009: 1095; 2012) which are addressed in CMIP5 GCMs.

The first is uncertainty in global greenhouse gases forcing which is addressed by use of Representative Concentration Pathways (RCPs), response to this forcing depending on the model parameterizations by making available larger set of GCMs and internal variability coped by providing multiple ensemble members having different initial conditions (Li et al., 2016: 4253). As the different GCMs can simulate different changes at regional level with the same anthropogenic forcing, it is difficult to identify the most reliable GCM. Therefore, it is important to assess ensemble of GCMs based on their collective information of the regional spatial structure of climate change signal in sign and magnitude (Giorgi and Francisco, 2000: 1295; Giorgi et al., 2001).

The two quantitative reliability criteria (e.g Giorgi et al., 2001; Khan and Koch, 2018; McSweeney et al., 2015: 3237) of the regional climate change simulations asses the ability of GCMs to reproduce the present day climate referred as model performance (the closer the model's climate to observations, the more reliable it's climate simulations will be) and the convergence of multimodel simulations for a given anthropogenic forcing, referred as model convergence (the higher convergence, the more robust is the signals) (Giorgi and Linda, 2002; Meher et al., 2017: 7778; Ta et al., 2018: 1516). Cheng and Frauenfeld (2014a: 5767; 2014b: 3935) had suggested that the spatial pattern of annual temperatures in CMIP5 models show good agreement with the observed data, however, disagreement exists in the magnitude of maxima.

In this study we will be focusing on the model performance metrics in simulating the present climate over the domain of Pakistan. Several Studies based on GCM evaluation and their application in hydrological modelling have been done over the HKH (Hindukush-Karakoram-Himalaya) region as well on the Indus basin. Lutz et al. (2016: 3988) studied the basins of Indus, Ganges and Brahmaputra using skill based GCM selection criteria suggest disadvantage of the two contrasting techniques in envelop based approach of GCM evaluation i.e considering only annual means changes or high skill in simulating the past or present climate could lead to different set of multimodel ensembles for climate change projections.

Meher et al. (2017: 7778) analyzed CMIP3 and CMIP5 multimodel ensemble over the Western Himalayan Region. The sensitivity analysis showed that MIROC3.2 and MIROC5 outperformed in both CMIP3 and CMIP5 multimodel ensembles. Their study also showed that most of the models in both multimodel ensembles failed to reproduce the realistic pattern and magnitude of rainfall over the domain which is mainly dependent on inaccurate representation of topography in this region (Boos and Hurley, 2013: 2279). However, CMIP5 models as compared to the CMIP3 models were better in representing the rainfall patterns (Sperber et al., 2013).

Precipitation analysis similar relative root mean square error of CMIP5 multimode ensemble over the Central Asian domain revealed large variability among these models however, the top six models identified where HadCM3, MIROC5, MPI-ESM-LR, MPI-ESM-P, CMCC-CM and CMCC-CMS. The study also recommended the multimodel ensemble over the individual models. Wu et al., (2017: 176) also analyzed CMIP5 multimode ensemble over the HKH region for baseline (1976-2005) as well as for the future scenarios. The multimodel ensemble mean performed better than individual models with negative $RMSE_r$s over the study domain.

Data and Methodology

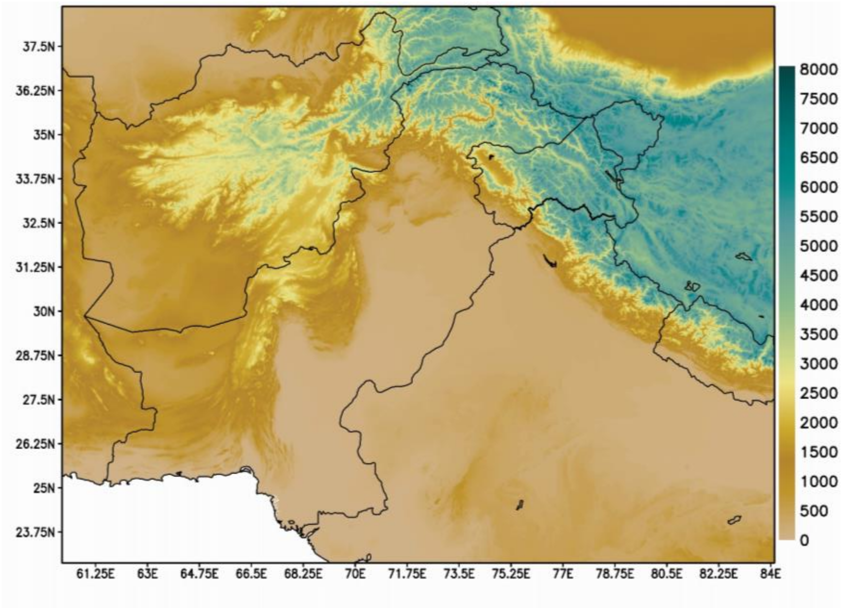

Fig. 1. Domain of Study, shaded is elevation in meters

The study region is 60.125E to 84.125E, 22.875N to 38.625N including the whole Indus Basin. The region is therefore, selected in order to use it for the impact based studies. Monthly precipitation data was obtained for 36 CMIP5 models from Earth System Grid (ESG) data portal. The data was downloaded for only first ensemble members of the historical runs at monthly frequency (Taylor et al., 2012). The reanalysis dataset used was daily product created for Agricultural Model Intercomparison and Improvement Project (AgMIP) in order to provide daily time series of climate variables required for agriculture models at global domain for the time period of 1980-2010. This dataset is produced by combining state-of-the-art reanalyses (NASA's Modern-Era Retrospective analysis for Research and Applications, MERRA, (Rienecker et al., 2011) from in situ observations and satellite data. The analysis were done both on monthly and seasonal basis. GCM data was interpolated at the same resolution as of the reanalysis dataset i.e 0.25º x 0.25º using the bilinear interpolation. Spatial as well as temporal averaging is applied in order get overall performance of the models. Seasonality index is defined in order to check the model performance in reproducing the seasonal cycle of precipitation. Two additional models, mean model and median model are also evaluated.

The performance metrics include Root Mean Square Error (RMSE) using climatologies of both observation and model over the domain of study (60.125E to 84.125E, 22.875N to 38.625N) as:

$$RMSE = \sqrt{[(x-y)^2]}$$

Where x is the model climatology and y is the observed/reanalysis climatology. The square brackets represent averaging over domain of study. $RMSE_r$ is used for assessing individual model performance with respect to other models given by:

$$RMSE_r = \frac{RMSE - RMSE_m}{RMSE_m}$$

Where $RMSE_r$ is the RMSE representing the relative model error. $RMSE_m$ is the median of all RMSE for all models (Glecker et al., 2008). The analysis have been done for both seasonal and annual time scales as well as for each month of the year. The values of RMSEr below zero shows that particular model performs better than the rest of the models whereas the models having $RMSE_r$s above zero shows that particular model performs worse than the rest of 50% of the models. The values of $RMSE_r$ are unitless. The models have been averaged both spatially and temporaly to see the overall behavior of models.

The seasonality index have been defined based on the methodology of William B Bull (2009: 126). The precipitation seasonality index (SI_p) has been defined as ratio of the average precipitation for the three wettest consective months (Pr_w) divided by the average total precipitation for three consecutive driest months (Pr_d).

$$\mathrm{SI_p} = \frac{\mathrm{Pr_w}}{\mathrm{Pr_d}}$$

Results and Discussion

Starting with relative root mean square error analysis on monthly climatology of multimodel ensemble precipitation. Most of the models show large variation in reproducing the precipitation (figure 1). Especially in the months of April, May, June and the peak monsoon season July, August, September. The models performing worse than rest of 50% models in the transition months of April, May, June, are ACCESS1-0, ACCESS1-3, HADGEM2-AO, HADGEM2-ES, INMCM4, MIROC-ESM-CHEM, HADGEM2-CC (AMJ) whereas EC-EARTH has worse performance in the month of June. For the months of July, August, September (JAS), the worst performing models are ($RMSE_r \geq 0.5$), IPSL-CM5B-LR, CSIRO-MK3-6-0, MRI-CGCM3, GISS-E2-H-p1, GISS-E2-R-p1. The winter transition period of October, November, December (OND), the models ACCESS1-0, BCC-CSM1-1, BNU-ESM, IPSL-CM5B-LR, MIROC-ESM, MIROC-ESM-CHEM, FIO-ESM, GISS-E2-R-p1 perform worse than the rest of the models whereas in the peak winter months of January and February, BNU-ESM and FIO-ESM has worse performance. Most of these models have RMSEr ≥ 0.5 in the months of transition period (OND, AMJ) and winter i.e performing worse than all other models presented in this study, while few have RMSEs lying within the range of 0.2 to 0.4.

The models performing better than the rest of the 50% in the winter transition period (OND) include CANESM2, IPSL-CM5A-MR, CESM1-CAM5, MIROC5, CSIRO-MK3-6-0, EC-EARTH, ENSMEAN and ENSMEDIAN. In the winter months (JFM) EC-EARTH, ENSMEAN and ENSMEDIAN perform better than the rest of 70% of the models. Whereas CANESM2, CESM-BCC, MPI-ESM-LR show smaller $RMSE_r$ in the summer transition period (AMJ) with IPSL-CM5A-LR, IPSL-CM5A-MR, EC-EARTH and CESM1-CAM5 having smallest $RMSE_r$ in the April and May months. Among these models, most perform well in OND months with CSIRO-Mk-3-6-0 and EC-Earth representing the lowest RMSEs (smaller than -0.5) hence better than all other models. The rest of the models have average behavior in winter months with $RMSE_r$ ranging between (-0.3 to -0.2) and EC-Earth showing lowest RMSE (≤ -0.5) along with ENSMEDIAN ($RMSE_r$ of -0.3 to -0.4). This implies most of the models perform well than the rest of 60-70% of the models in these seasons. Considering the summer monsoon months, JAS, most of the models have below 50% performance. However, the good models lie above the range of 50-60% ($RMSE_r$ ranging between 0 to -0.2) of the models with only GFDL-ESM2G having the lowest $RMSE_r$ i.e between -0.3 to -0.4.

From figure 2, at annual time scale, the models having highest $RMSE_r$ than the rest of the 50% models are MIROC-ESM, GISS-E2-H-p1 ($RMSE_r \geq 0.4$ to 0.5). At seasonal timescale, discussing the winter season first (DJFM). Most of the models have $RMSE_r$ ranging greater than 0.3 to 0.4 i.e performing worse than the rest of 70-80% models. These models include, BNU-ESM, FIO-ESM, and IPSL-CM5B-LR for DJFM and JF and IPSL-CM5B-LR, MRI-CGCM3, GISS-E2-H-p1, GISS-E2-R-p1 and CSIRO-Mk-3-6-0 for JJAS and JA.

The lowest $RMSE_r$ for annual timescale range from -0.4 to -0.3 for the models GFDL-CM3, GFDL-ESM2G. Whereas for DJFM, MRI-CGCM3, GFDL-CM3, GFDL-ESM2G, and GFDL-ESM-2M has $RMSE_r$ between -0.4 to -0.3 with model having lowest RMSE is EC-Earth. For the peak winter season JF, the ENSMEDIAN has RMSE less than -0.4 to -0.3 with EC-Earth again having the lowest $RMSE_r$. For summer monsoon months JJAS and peak monsoon months JA, only three models show lowest $RMSE_r$ (-0.4 to -0.5 for JJAS and -0.3 to -0.4 for JA) which are GFDL-CM3, GFDL-ESM2G, and GFDL-ESM-2M.

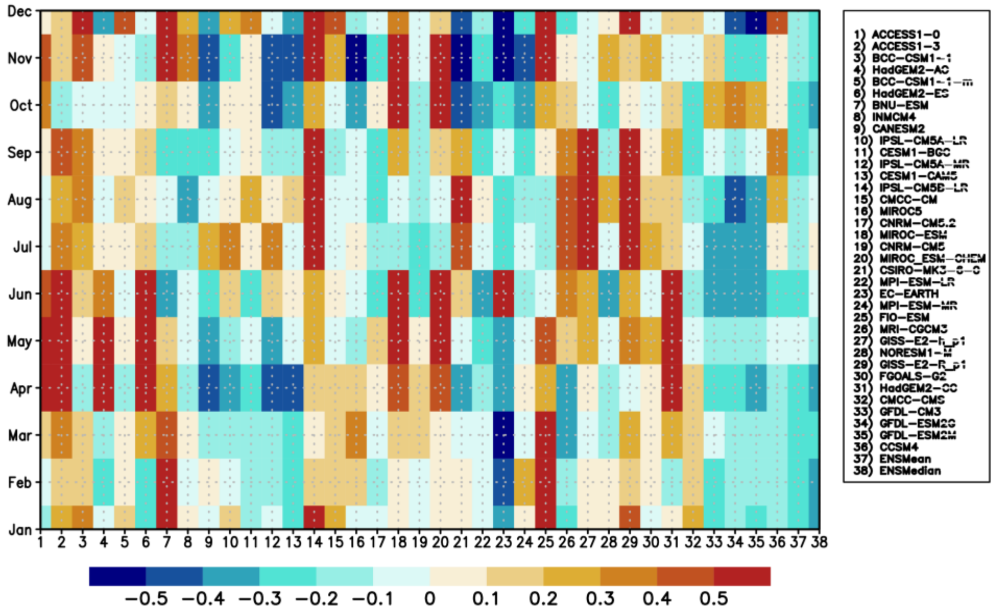

Fig. 2. Monthly precipitation climatology Relative RMSE's of CMIP5 multimodel ensemble (1980-2005). X-axis represent CMIP5 models and shading represent Relative RMSE's

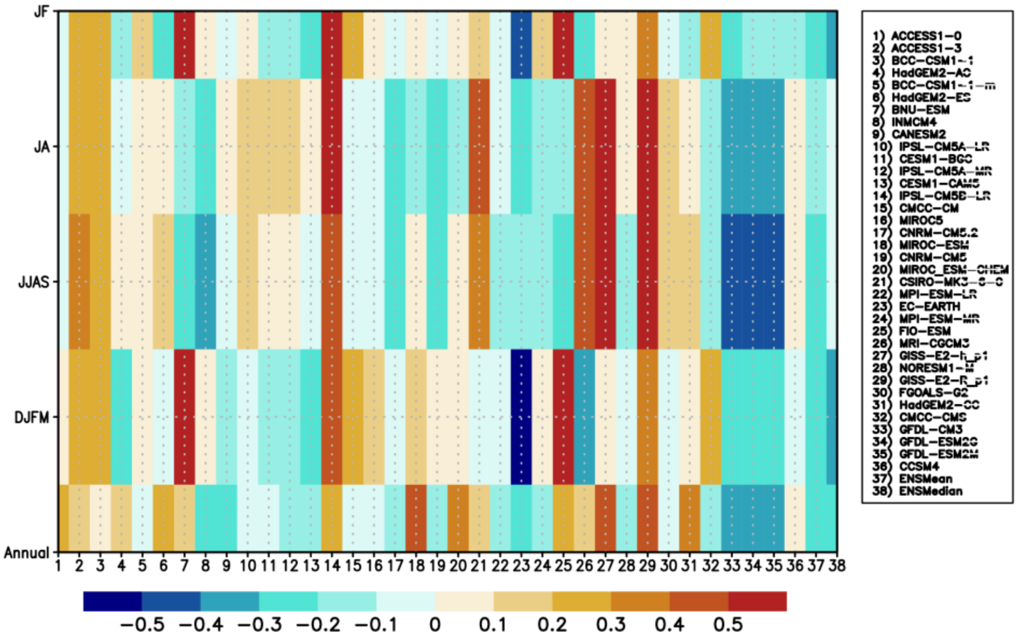

Fig. 3. The same as Fig. 2, but for Annual and seasonal Precipitation

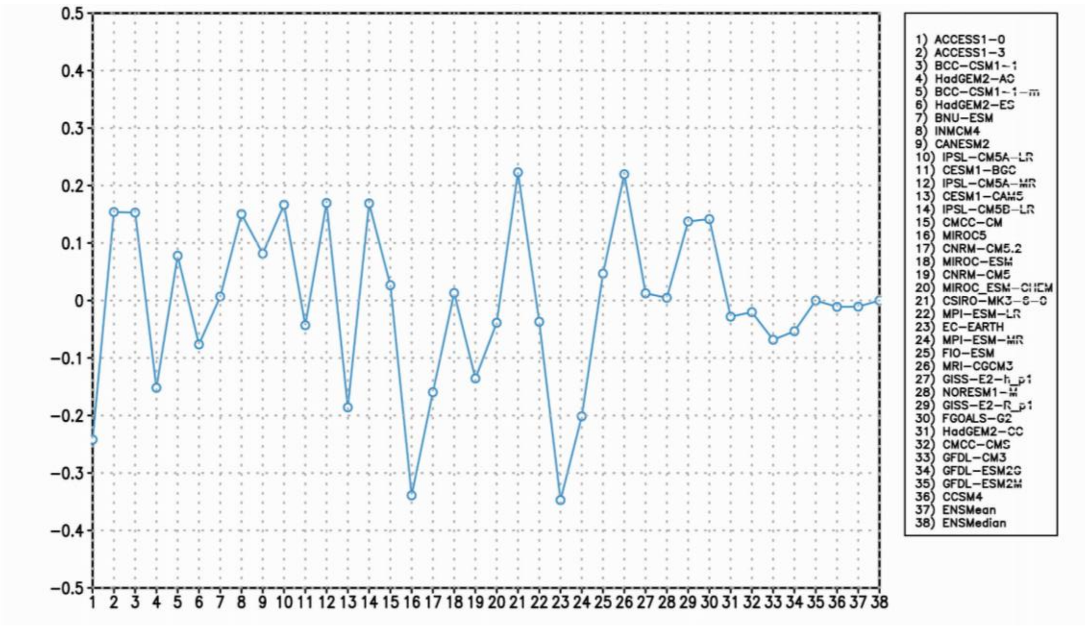

Fig. 4. RMSEr of precipitation seasonality index (1980-2005) calculated from CMIP5 multimodel ensemble

Fig. 4 show $RMSE_r$ of the seasonality index. This index is helpful in assessing that how well the seasonal cycle is represented in the models as compared to the observation for the area under study. The $RMSE_r$ of precipitation seasonality index (blue line) show 18 out of 38 models fall in the region of below zero $RMSE_r$ with MIROC5 and EC-Earth models showing lowest values (≤0.3). 16 out of 38 show above zero RMSE values. The models having larger $RMSE_r$ are CSIRO-Mk-3-6-0 and MRI-CGCM3 (>0.2). The ENSMEAN and ENSMEDIAN lie about zero $RMSE_r$ line indicating neither good nor bad in representing the seasonal cycle. As depicted in the figure, there is large variability between the models in representing the seasonality of precipitation. This particularly depends on the parameterization schemes in the models to represent the precipitation process. Since this region is influenced both by convective systems (Monsoonal systems) and large scales systems (western disturbances) as well as interaction of both, it becomes the source of uncertainty in the models.

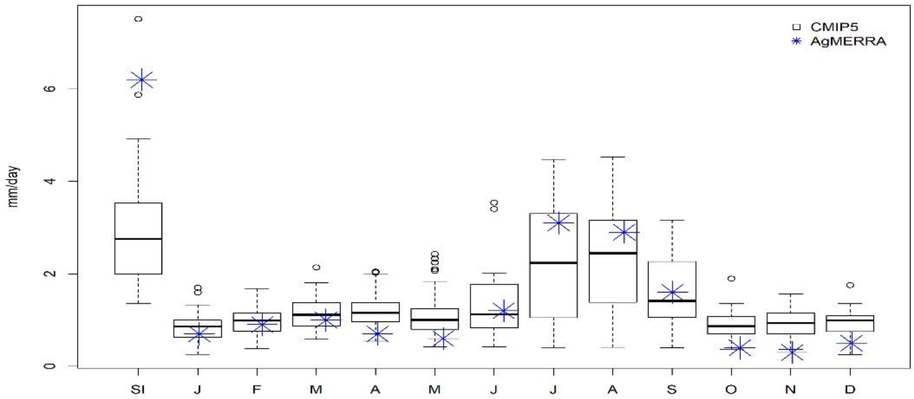

Fig. 5. Box and whisker plot of precipitation and seasonality indices for the CMIP5 multimodel ensemble. Blue asterisks represent AgMERRA dataset whereas boxes represent CMIP5 multimodel ensemble

Fig. 5 shows analysis for precipitation climatology and seasonality index. There is large variability in the CMIP5 multimodel ensemble shown in the months of June, July, August and September as well as for the seasonality index. There are few outlier models which are represented in the upper quartile. The median values of CMIP5 models coincide with those of AgMERRA for the months of January, February, March and June however, there is large difference between the model ensemble and the observations in the rest of the months (up to 1.5 mm/day). The seasonality index also show large variation between the models and the observed dataset. AgMERRA shows up to 3mm/day higher value in the seasonality index as compared to the CMIP5 multimodel ensemble median. The large variation could be result of models which are unable to simulate the annual seasonal cycle correctly.

Conclusion

The CMIP5 multimodel ensemble has been analyzed in comparison to observed dataset AgMERRA for the period 1981-2005. The analysis have been done at monthly time scales for the precipitation. Spatial and time averaging has been applied to assess the overall performance of the models. A seasonality index is also defined for precipitation. In addition to CMIP5 multimodel ensemble, we have also used ensmean and ensmedian models. ENSMEAN, ENSMEDIAN models outperform other individual models for all variables consistent with most of the performance metrics studies on CMIP5 models. For winters, EC-Earth show better performance for precipitation, and for summer months GFDL-CM3, GFDL-ESM2G and GFDL-ESM2M show smaller $RMSE_r$ over the domain. There is large variability in CMIP5 models in simulating the precipitation. For peak monsoon months July, August and September, MRI-CGCM3, GFDL-ESM-2M, GFDL-ESM2G and GFDL-CM3 show better results than the rest of the ensemble. For precipitation seasonality index MIROC5 and EC-Earth show better results over the selected domain. The boxplots are also included in the analysis for the overall performance of CMIP5 ensemble. The results of analysis are consistent with the results of studies based on upper Indus basin, and full domain of Pakistan (Najeeb Ullah Khan et al., 2018: 1793; Kamal Ahmend et al, 2019, Asim Jehangir Khan et al. 2018: 1793; Nadia Rehman et al., 2018: 381-415; etc). Precipitation is under estimated in the ensemble especially in JJAS months with large variability between the models. The precipitation seasonal cycle is under estimated in the CMIP5 ensemble. For winter months (JFM), the models perform relatively better.

Acknowledgment

The author would like to thank Earth System Physics Section of The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy for providing the CMIP5 GCMs data to conduct this study.

References

Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C., and Chung, E.-S. (2019). Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics. Hydrol. Earth Syst. Sci. Discuss, 23, 4803-4824. https://doi.org/10.5194/hess-2018-585.

Bull, W. B. (2009). Tectonically active landscapes. Hoboken: Wiley-Blackwell.

Boos, W. R., Hurley, J. V. (2013). Thermodynamic bias in the multimodel mean boreal summer monsoon. Journal of Climate, 26(7), 2279-2287. https://doi.org/10.1175/JCLI-D-12-00493.1

Chen, L., Frauenfeld, O. W. (2014a). A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. Journal of Geophysical Research: Atmospheres, 119(10), 5767-5786. https://doi.org/10.1002/2013JD021190

Chen, L., Frauenfeld, O. W. (2014b). Surface air temperature changes over the twentieth and twenty-first centuries in China simulated by 20 CMIP5 models. Journal of Climate, 27(11), 3920-3937. https://doi.org/10.1175/JCLI-D-13-00465.1

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C. (2014). Evaluation of climate models. In Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 741-866). Cambridge: Cambridge University Press.

Giorgi, F., Francisco, R. (2000). Evaluating uncertainties in the prediction of regional climate change. Geophysical Research Letters, 27(9), 1295-1298. https://doi.org/10.1029/1999GL011016

Giorgi, F., Hewitson, B., Christensen, J., Hulme, M., Von Storch, H., Whetton, P., Jones, R., Mearns, L., Fu, C., Arritt, R. and Bates, B. (2001). Regional climate information-evaluation and projections, 110. Available at: https://lib.dr.iastate.edu/ge_at_pubs/110/

Giorgi, F., Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging”(REA) method. Journal of Climate, 15(10), 1141-1158. https://doi.org/10.1175/1520-0442(2002)015%3C1141:COAURA%3E2.0.CO;2

Gleckler, P. J., Taylor, K. E., Doutriaux, C. (2008). Performance metrics for climate models. Journal of Geophysical Research: Atmospheres, 113(D6). https://doi.org/10.1029/2007JD008972

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R. and Wargan, K. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of Climate, 30(14), 5419-5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Hawkins, E., Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society, 90(8), 1095-1108. https://doi.org/10.1175/2009BAMS2607.1

Hawkins, E., Sutton, R. (2012). Time of emergence of climate signals. Geophysical Research Letters, 39(1), L01702. https://doi.org/10.1029/2011GL050087

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. Journal of Climate, 23(10), 2739-2758. https://doi.org/10.1175/2009JCLI3361.1

Khan, N., Shahid, S., Ahmed, K., Ismail, T., Nawaz, N., Son, M. (2018). Performance assessment of general circulation model in simulating daily precipitation and temperature using multiple gridded datasets. Water, 10(12), 1793. https://doi.org/10.3390/w10121793

Khan, A., Koch, M. (2018). Selecting and Downscaling a Set of Climate Models for Projecting Climatic Change for Impact Assessment in the Upper Indus Basin (UIB). Climate, 6(4), 89. https://doi.org/10.3390/cli6040089

Li, Q., Wang, S., Lee, D., Tang, J., Niu, X., Hui, P., Gutowski, W.J., Dairaku, K., McGregor, J.L., Katzfey, J., Gao, X., Wu, J., Hong, S., Wang, Y., & Sasaki, H. (2016). Building Asian climate change scenario by multi-regional climate models ensemble. Part II: mean precipitation. International Journal of Climatology, 36, 4253-4264. https://doi.org/10.1002/JOC.4633

Lutz, A.F., ter Maat, H.W., Biemans, H., Shrestha, A.B., Wester, P., Immerzeel, W.W. (2016). Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. International Journal of Climatology, 36(12), 3988-4005. https://doi.org/10.3390/cli6040089

McSweeney, C. F., Jones, R. G., Lee, R. W., Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. Climate Dynamics, 44(11-12), 3237-3260. https://doi.org/10.1007/s00382-014-2418-8

Meehl, G.A. (2007). Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

Meher, J. K., Das, L., Akhter, J., Benestad, R. E., Mezghani, A. (2017). Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the Western Himalayan region. Journal of Climate, 30(19), 7777-7799. https://doi.org/10.1175/JCLI-D-16-0774.1

Rehman, N., Adnan, M. Ali, S. (2018). Assessment of CMIP5 climate models over South Asia and climate change projections over Pakistan under representative concentration pathways. International Journal of Global Warming, 16(4), 381-415. http://dx.doi.org/10.1504/IJGW.2018.095994

Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. Journal of Geophysical Research: Atmospheres, 118(19), 884-906. https://doi.org/10.1002/jgrd.50843

Sperber, K.R., Annamalai, H., Kang, I.S., Kitoh, A., Moise, A., Turner, A., Wang, B. Zhou, T. (2012). The Asian Summer Monsoon: An Intercomparison of CMIP-5 vs. CMIP-3 Simulations of the 20th Century (No. LLNL-JRNL-563734). Livermore: Lawrence Livermore National Lab.(LLNL).

Ta, Z., Yu, Y., Sun, L., Chen, X., Mu, G., Yu, R. (2018). Assessment of Precipitation Simulations in Central Asia by CMIP5 Climate Models. Water, 10(11), 1516. https://doi.org/10.3390/w10111516

Taylor, K.E., Stouffer, R.J. Meehl, G.A. (2012). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1

Wu, J., Xu, Y., Gao, X. J. (2017). Projected changes in mean and extreme climates over Hindu Kush Himalayan region by 21 CMIP5 models. Advances in Climate Change Research, 8(3), 176-184. https://doi.org/10.1016/j.accre.2017.03.001