Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-585-AC1, 2019

Hydrology and Earth System Sciences Discussions

Interactive comment on "Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics" *by* K. Ahmed et al.

K. Ahmed et al.

eschung@seoultech.ac.kr

Received and published: 26 August 2019

General Comments This manuscript was written fairly well in terms of its academic originality, and scientific descriptions for introduction, methodology, results and conclusions. Especially, it developed a systematic selection framework for many GCMs from CMIP5 based on various state-of-the-art spatial performance metrics. The selected GCMs showed their capability to mimic the spatial patterns of annual and seasonal precipitation. The most impressive point is to summarize so many relevant articles which were published in recent years. This manuscript is worthwhile to be published in this journal. Nevertheless, the following point should be thought carefully in my opinion.

This manuscript focuses on the simulated data for the very past period (1961-2005). However, the superiority of performances of GCMs for the past period doesn't guarantee the exactness of projection for the future period. In addition, the main objective of GCMs is to support the forecasted future data for 2010-2100. Of course, a part of them can be evaluated using the recent data (2010-2018). If you cannot quantify GCMs' performances for the recent data, you can mention that point in the manuscript. Reply Thank you for your constructive comments on our manuscript. Your suggestions helped us to improve the quality of our manuscript. In response to your above comment we have added following paragraph as a limitation in the discussion of the manuscript. "In this study performance of GCMs was assessed based on their ability to simulate past observed P, Tmax and Tmin 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. . Comment 1 In section 3.3, the four top ranked GCMs were used to generate the most appropriate ensemble of GCMs. Is there any reason why four is used? You can compare your results with those form different numbers of GCMs. This number can affect the results.

Reply Thank you very much for your very interesting comment. As seen in the literature the number of GCMs selected for the MME is an arbitrary choice. The choice for the selection of four top GCMs in this study is also arbitrary. We have added the following text to section "3.4 Identification of Ensemble Members" of the manuscript as below: "3.4

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, four top ranked GCMs were considered for the development of MMEs for P, Tmax and Tmin. The review of the literature revealed that there is no well-defined guideline on the selection of the optimum number of GCMs for the MME and most of the studies considered the first three to ten GCMs ranked according to the descending order of their performance for the MME. For instance, in the study by Xuan et al. (2017) over Zhejiang, China, ten top-ranked GCMs for an MME for precipitation were used. In another study over China, Jiang et al. (2015) developed MMEs for daily temperature extremes using the five top-ranked GCMs. In a study over Pakistan, Khan et al. (2018) considered six common GCMs that appeared in the lists of ten top-ranked GCMs for daily temperature and precipitation. Ahmadalipour et al. (2015) used the four top-ranked GCMs for simulating daily precipitation and temperature over the Columbia River Basin in the Pacific Northwest USA. In the study by Hussain et al. (2018) the three top-ranked GCMs for the development of an MME for precipitation over Bornean tropical rainforests in Malaysia were used. In the present study, the ensemble of GCMs was identified in two steps: (1) RM values of GCMs for annual, monsoon, winter, pre-monsoon and post-monsoon P, Tmax and Tmin were individually used to derive an overall rank for each GCM, and (2) four top ranked GCMs based on RM values for all variables 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."

Following lines are added as the limitation of the study in discussion section. "In the

present study, the MME of P, Tmax and Tmin 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. It can be remarked that the performance of an MME is probably sensitive to the choice of the 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 on its performance."

Comment 2 In section 3.4, you mentioned numerous approaches have been used to calculate the mean time series from an ensemble of better performing GCMs. Thus, it is better to add the reason why two representative methods should be used. What is the improved one? If the method is very critical to the results, you should add the descriptions on simple mean and random forest methods.

Reply Thanks for your comment. We have already addressed the above comment partly in the introduction and section 3.4 of the manuscript as given below. In addition to that, we have now added details on simple mean and random forest to sections 3.4 of the revised manuscript. In introduction section: "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., 2018). 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., 2018;Oh and Suh, 2017;Giorgi and Mearns, 2002). The SCM is relatively simple to apply and found to perform better than individual GCMs (Weigel et al., 2010;Acharya et al., 2013;Wang et al., 2018). However, WEM is preferred as it has the capability to remove the systematic biases and improve the prediction capability since higher weights are assigned to better GCMs (Krishnamurti et al., 1999;Krishnamurti et al., 2000). Salman et al. (2018) 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., 2018). Therefore, selection of a sub-ensemble is a major challenge in climate change modelling." We revised section 3.4 as below: "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, four top ranked GCMs were considered for the development of MMEs for P, Tmax and Tmin. The review of the literature revealed that there is no well-defined guideline on the selection of the optimum number of GCMs for the MME and most of the studies considered the first three to ten GCMs ranked according to the descending order of their performance for the MME. For instance, in the study by Xuan et al. (2017) over Zhejiang, China, ten top-ranked GCMs for an MME for precipitation were used. In another study over China, Jiang et al. (2015) developed MMEs for daily temperature extremes using the five top-ranked GCMs. In a study over Pakistan, Khan et al. (2018) considered six common GCMs that appeared in the lists of ten top-ranked GCMs for daily temperature and precipitation. Ahmadalipour et al. (2015) used the four top-ranked GCMs for simulating daily precipitation and temperature over the Columbia River Basin in the Pacific Northwest USA. In the study by Hussain et al. (2018) the three top-ranked GCMs for the development of an MME for precipitation over Bornean tropical rainforests in Malaysia were used. In the present study, the ensemble of GCMs was identified in two steps: (1) RM values of GCMs for annual, monsoon, winter, pre-monsoon and post-monsoon P, Tmax and Tmin were individually used to derive an overall rank for each GCM, and (2) four top ranked GCMs based on RM values for all variables 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.

Details of both Simple Mean and Random Forest used in developing MMEs are given below.

3.5.1 Simple Mean (SM) Simple Mean (SM)-based MMEs were developed by simply averaging the individual P, Tmax and Tmin simulations of the four top-ranked GCMs using Eq.16.

SM=1/n $\sum_{(i=1)}^{n} GCM_i (16)$

In Eq. 16, n refers to the number of GCMs considered for the development of MMEs which is four in the present study and GCMi refers to the P, Tmax and Tmin simulation of the ith GCM.

3.5.2 Random Forest (RF) Random Forest (RF) algorithm (Breiman, 2001) was used in the calculation of the mean time series of P, Tmax and Tmin 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. A bootstrap resampling method is used to select sample sets from training data. Classification And Regression Tree (CART) technique is used to develop unpruned trees using the bootstrap sample. 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. A regression model is fitted for all the trees and the performance of each tree is assessed. 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, Tmax and Tmin simulations of GCMs as independent variable and historical observed P, Tmax and Tmin as dependent variable provide weights to the GCMs according to their ability to simulate historical observed P, Tmax and Tmin. 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 3 When the abbreviation was defined, it should be done at the first appearance. E.g.) p3 L13, root mean square error, p2 L17 multi-model ensemble. Check the abbreviation. When any was defined, abbreviation should be used afterwards. E.g. P4 L12, P5 L31 MME, P6 L24 SPAtial EFficiency metric; P6 L8, P10 L7, P16 L24 Rating metric, P11 L5, P15 L19 simple mean and random forest.

Reply Thanks. Corrected as suggested.

Comment 4 P7 L12 "lamba"? Check the name of variables in all equations. "N"s in Eq. (5) and (6) are the same? Check the other variables. "(MME)" in the sub-title can be removed.

Reply Thanks, "lamba" was corrected as "lambda". The "N" in Eq. 5 and 6 are different and defined differently. "MME" in the subtitle was removed.

Comment 5 P2 L17 Check Pour et al.(2018b) which is not included in the reference.

Reply Thanks, we have updated reference list. It is corrected now.

Comment 6 P2 L22 Check Wang et al.(2017b) which is not included in the reference list. Reply Thanks, we have updated reference list. It is corrected now.

Comment 7 P2 L23 Check Wang et al.(2017a) which is not included in the reference list. Reply Thanks, we have updated reference list. It is corrected now.

Comment 8 P21 L15-16 Salman et al.(2018a) is the same to Salman et al. (2018b), Reply Thanks, we have updated reference list. It is corrected now.

Comment 9 P2 L30 Check Pour et al.(2018b) which is not included in the reference. Reply Thanks, we have updated reference list. It is corrected now.

Comment 10 P3 L8 Are Tebaldi et al. (2005) and Chandler (2013) included in the reference list? Reply Yes, they are also included in revised reference list.

Comment 11 P4 L9 "and" should be added at the end of this sentence. Reply Thanks, we have added "and" at the end of the sentence.

Comment 12 P7, P8, P10 Variables "m" and "n" were used in the different equations. Check their consistency. Reply Thanks for the comment, we have re-checked, "m" and "n" were used in the different equations and they are defined in accordingly.

Comment 13 P7 L1 in equation1, is "KGE" correct? SPAEF? Reply Thanks, it was corrected.

Comment 14 P9 L15-16: it should be moved below Equation 8. Reply Thanks for this comment. P9 L15-16 are moved below equation 8.

Comment 15 P9 L17-18: it should be moved below equation 9 and 10. In the conclusions, abbreviations were defined again. Is it correct in this journal? Check it. Reply Thanks for this comment. P9 L17-18 are moved below equation 9 and 10. The journal does not have any restriction on the use of long-terms of abbreviations in the conclusions. We have defined some of the abbreviations in the conclusion as it may assist a reader who wants to have a glance at the conclusions and understand the main outcomes of the work.

Comment 16 P22 L18 Wang et al. (2016) is not cited in the manuscript. Check the reference format. Reply Thanks, we have updated reference list. It is corrected now.

Comment 17 P2 L24 “to use” is right? Reply We have changed the phrase “to use” to “ to apply”

Comment 18 P2 L30, “selection âĹij modelling.” is right? Reply The word “selection” is more appropriate here as we are referring to the selection of GCMs.

Comment 19 P3 L7 “such as” was repeated. Reply Thanks. Corrected as suggested.

Comment 20 P3 L25 “scale” or “scales”? Reply Thanks, we changed scale to scales.

Comment 21 P3 L29 “should able” or “should be able”? Reply Thanks, “should able” was replaced with “should be able”.

Comment 22 P5 L16 the second “20” is not necessary. It was already mentioned at the previous sentence. Reply Thanks, the second “20” was removed from the next sentence.

Comment 23 P5 L30 “are” or “is”? Reply Thanks, We have changed “are” to “is”.

Comment 24 P14 L14 “point” or “points”? Reply Thanks, we changed point to points.

Comment 25 P14 L18 “scatter” is right? Reply Yes, it is correct.

Comment 26 P14 L16 “skillful”? Reply Thanks, skilfull is changed to skillful.

Comment 27 P14 L17 Check the location of”also”. Reply Thanks, we revised sentence as shown below. “Over and underestimation of precipitation can also be seen in the scatter. . ..”

References

Acharya, N., Singh, A., Mohanty, U. C., Nair, A., and Chattopadhyay, S.: Performance of general circulation models and their ensembles for the prediction of drought indices over India during summer monsoon, Nat. Hazards, 66, 851-871, 10.1007/s11069-012-0531-8, 2013.

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

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. Folberth, C., Baklanov, A., Balkovič, J., SkalskÃ¡, 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.

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.

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.

Hussain, M., Yusof, K. W., Mustafa, M. R. U., Mahmood, R., and Jia, S.: Evaluation of CMIP5 models for projection of future precipitation change in Bornean tropical rainforests, Theor. Appl. Climatol., 134, 423-440, 10.1007/s00704-017-2284-5, 2018.

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.

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.

Kim, J., Ivanov, V. Y., and Fatichi, S.: Climate change and uncertainty assessment over a hydroclimatic transect of Michigan, Stoch. Environ. Res. Risk Assess., 30, 923-944, 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.

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.

Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P., and Immerzeel, W. W.: Selecting representative climate models for climate change impact studies: an advanced envelopeâĂŘbased selection approach, Int. J. Climatol., 36, 3988-4005, 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.

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

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, Atmos. Res., 213, 149-162, https://doi.org/10.1016/j.atmosres.2018.06.006, 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.

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

Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A., and Yu, Q.: Using multiâĂŘmodel ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia, Int. J. Climatol., 38, 4891-4902, 2018.

Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C.: Risks of model weighting in multimodel climate projections, J. Clim., 23, 4175-4191, 2010.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-585, 2019.