Reviewer 1. Review Report (Round 1)

Review of “Uncertainty assessment in drought severities for the Cheongmicheon watershed using multiple GCMs and reliability ensemble averaging method” by Abdulai and Chung.

In this paper, the authors investigated the approach of quantitatively assessing uncertainties on drought pattern using REA based on model performances on RMSE. Uncertainty assessment on GCM projections is an important topic and attracts attentions from both climate community and inter-discipline community, especially for the stakeholders. The analyses in this study are meaningful, and the manuscript is well-written. However, I still have some comments regarding to the methods used in this manuscript. Therefore, I would recommend that the present manuscript may be accepted for publication after some revisions.

Thank you for your input and contribution to this paper. All comments were taken in due consideration. However, below are responses to your concerns about the methodology used in this study.

1. One of my concerns is that the correlation coefficient and RMSE are not suitable to evaluate the model performance, because the inter-annual variability is mainly determined by the model internal variability rather than external forcing, therefore there is no way to expect the models can reproduce flood or drought on the exactly same year with observation. That is why the $R^2$ are all negative and small in Fig. 6. The authors need seriously think about this issue, since this is the base of the whole study.

According to the IPCC, “climate variability refers to variations in the mean state and other statistics (such as standard deviations, the occurrence of extremes, etc.) of the climate on all spatial and temporal scales beyond that of individual weather events. Variability may be due to natural internal processes within the climate system (internal variability), or to variations in natural or anthropogenic external forcing (external variability).” Based on this fundamental concept, model internal variability can as well be determined by external forcing which causes climate variability. Moreover, our study focused on multi-model averaged changes based on the Reliability Ensemble Averaging (REA) method with some concept of Giorgi and Means (2002). We analyse differences or changes mean fields between three 30-year period of the 21st century (2011-2040, 2041-2070, 2070-2100) and the reference period 1976-2005, that is a recent climate period characterized by relatively small anthropogenic component (IPCC, 2001). We then calculate the bias using quantile mapping and the values where then calculated for RMSE and $R^2$ with ensemble average changes across models using REA method which in this study reliability parameter depends only on the model bias using observational reference period and three 30-year period of each predicted value. To this point model performances were evaluated based on the RMSE, NSE and $R^2$ wherein the inverse of the bias, RMSE and correlation coefficient was proportionately used as weights with equal weights derived from RMSE. We use the weights to describe our uncertainty of intermodal variability.
Moreover, I am curious that for the future projection, when calculating the RMSE, what the references are since there is no observation as in historical period?

The weights of the model performance of the ensemble mean and each GCM was compared to the three 30-year periods to predict uncertainty of climate change as shown in Figure 1 below with some references. As mentioned previously bias corrected values by quantile mapping were used to calculate RMSE and R² in this study. However, some references on similar studies of RMSE significantly improving model performances and projections are as follows:


Ensemble Mean and Uncertainty for Future Duration

Projections hinge on climate model ensembles often reckon that each individual model simulation is of equal value. The uncertainty considered in this study is intermodal variability. Future projections were analyzed in terms of weight uncertainty and the ensemble mean for the downscaling methods using 27 GCMs and three future periods
(2011-2040, 2041-2070, 2071-2100) under RCP 4.5. The weighting methodology is usually hinge on model performance differences, may be applied, and lead to some improvement in the projected mean. The REA uncertainty shows similar ensemble average of the future projected annual mean as shown in Figure 1a, b and c. The changes projected by the three future periods varies according to the model and projected years. Here, the weights accounts for model performance. The performance weights are inversely proportional to each simulated RMSE. In Figure 5, it was established that the annual averaged RMSE of nearly all climate models did not increase substantially during the 30-year long period of simulations. Applications of these weights to the ensemble mean is forthright. Prediction is made by comparing the ensemble mean weights for the three-30-year periods in Figure 1a, b and c it is established that the model weights are homogeneous but not exact. This is an interesting phenomenon that the simpler approach and Reliability Ensemble Averaging achieve related results even though the ensembles derive various weights across these two-ensemble approach.

Moreover, in Figure 1a, b and c it shows that annual ensemble mean approach can improve the uncertainty results compared to a single individual model. It stipulate that the uncertainty spawn from the individual models are smaller compared to the ensemble as thus: In figure 1a, the following models produces the smallest uncertainty when compared to the ensemble, IPSL-CM5A-MR, CCSM4, CanESM2, CSIRO-MK3-6-0 while in Figure 1b, inmcm4, CCSM4, bcc-CSM1-1, CESM-BGC, MICRO5 and in Figure 1c, MICRO-ESM-CHEM, CSIRO-MK3-6-0, CMCC-CAM5 and CCSM4 respectively. Furthermore, the study confirms that uncertainty performances of the ensembles are preferable to that of any single model. Therefore, a single model’s output is not adequate accuracy which again rationalize the expansion of multi-model ensemble scenario. Hence, the evaluation of the uncertainties from climate models ensemble gives an essential evaluation of overall climate projection uncertainties. In summary the ensemble mean tends to perform better than any individual model as averaging across models reduces errors in both mean climate and when variability is considered.
Figure 1a. Ensemble uncertainty of annual climate results with 30-year mean (2011-2040)
Figure 1b. Ensemble uncertainty of annual climate results with 30-year mean (2041-2070)
3. Another important point is that, the authors also need to compare their results with simple averaging to show the improvements of their results.

Prior to explore this result further and as a base for the REA method a simpler approach (is the simplest multi-model technique that averages output of all climate models with equal weights) of ensemble mean and the related uncertainty weights was conducted in this study. (Equation 1 & 2). However, the ensemble mean does not distinctly take into consideration the reliability criteria and weights equally all models. The Simpler Approach was compared to the Reliability Ensemble Averaging approach to show the improvement of model uncertainty of the study. It is worthwhile to note that, the uncertainty herein is only as a result of various projections of the ensemble models and other sources of uncertainty are not influenced by the predictions used herein. In Figure 2 the Reliability Ensemble Averaging projected higher ensemble of uncertainty than that of the Simpler Approach and projection in uncertainty can be seen to increases towards the 21st century.
4. For the introduction section, I believe some recent literatures about downscaling and uncertainty assessment, would be helpful to strengthen the motivation and importance of this manuscript, such as:

Ning, L., E. Riddle, and R. S. Bradley, 2015: Projected changes in climate extremes over the northeastern United States. J. Climate, 28, 3289–3310
Maurer, E. P., 2007: Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios. Climatic Change, 82, 309–325

Thank you for recommending some articles, the introduction has been strengthened in the manuscript.

5. Explanations of the variables shown in the Eqs 1-4 are needed.

Explanation of the variables form Equations (1)-(4) have been added in the text and are as follows: The variables have been explained in the text as thus: The $\Delta p_i$ is the average bias of the model precipitation change. Operator $A$ denotes REA averaging, $R_i$ is the overall reliability of individual model as function of model bias and $p$ is the simulated precipitation of the ensemble member.

6. More details about how to calculate the SPI, SDI, and SPEI are needed.

More details of how the SPEI, SPI and SDI are calculated based on a coded software packages have been explained and added in the text. Thank you.

The approach focusses to have an understanding regarding drought deficit by assessing meteorological and hydrological drought in the Cheongmicheon watershed through...
comparing SPI, SDI, SPEI in order to evaluate drought model performances. Drought index helps to understand the development and dynamics of droughts revealed through their severities, duration and intensity. The SPEI and SPI were produced using the SPEI package coded in the R software, it is a free software used for environmental, statistical calculation and illustrations, this package yields various alternatives for computing SPI and SPEI.

The Standardized Precipitation Evapotranspiration Index (SPEI) is an extension of the widely used Standardized Precipitation Index (SPI). The SPEI is designed to take into account both precipitation and potential evapotranspiration (PET) in determining drought. Thus, unlike the SPI, the SPEI captures the main impact of increased temperatures on water demand. Like the SPI, the SPEI can be calculated on a range of timescales from 1-48 months. As a result, this study calculated the SPIEs for 3-, 6- and 9-month intervals are computed with special focus on severe and extreme drought. The procedure of the SPEI calculation depends on the actual SPI computation but uses the monthly variation linking precipitation (P) and PET. SPEI is the most suitable water deficit index in drought identification, supervision and evaluation in relation to climate variation events. If only limited data are available, say temperature and precipitation, PET can be estimated with the simple Thornthwaite method. In this simplified approach, variables that can affect PET such as wind speed, surface humidity and solar radiation are not accounted for. In cases where more data are available, a more sophisticated method to calculate PET is often preferred in order to make a more complete accounting of drought variability. However, these additional variables can have large uncertainties. Therefore, this study used an R package software for calculating the SPEI and SPI from user-selected input precipitation and temperature data using the Hargreaves method. Calculation of the SPEI and SPI is implemented in the R package SPEI (http://cran.r-project.org/web/packages/SPEI). This package is preferred over previous implementation in C language (http://digital.csic.es/handle/10261/10002). This latter implementation only allows computation of the original formulation of the SPEI (based on the Thornthwaite ETo equation). The SPEI R package allows three ETo equations (Thornthwaite, Hargreaves and FAO-56 Penman-Monteith).

The SPI index was developed by [20] in order to quantitatively study precipitation shortage. It is the major water shortage index to divulge the possible severity of water based on the idea of hydrological drought, agricultural and socio-economical drought. The study procedure and time scale calculation of SPI is same as described above. Most importantly continuous long-term data of at least 30 years is required to compute SPI and does not allow missing data. This versatility allows SPI to assess short, medium- and long-
term water supplies and drought severity. The SPI index helps to distinguish dry years from wet years and a drought occurs when the SPI is consecutively negative, and its value reaches an intensity of -1 or less and ends when SPI becomes positive. Hence, the SPI for any place is calculated based on the long-term rainfall recorded at desired station and is then first fitted to a probability distribution (example, Gamma distribution) which is modified into a normal distribution so that the average SPI is zero.

Peculiar to this evolving described above, Streamflow Drought Index (SDI) was employed for distinguishing hydrological drought. SDI is interpreted depending on cumulative streamflow volumes. The SDI has calculation techniques almost like that of the SPI. The variation linking the SDI and SPI is that the SDI uses observed streamflow data, while the SPI uses rainfall data. The calculation of the drought indices was carried out using the software package DrinC [21], which was strengthened for the needs of this study. The DrinC (drought indices calculator) is an MS Windows based software through which various drought indices can be calculated. It functions on a graphical user interface (GUI) and embodied various mechanisms that promote data handling analysis of the results, drought auditing, spatial calculation of the indices etc. Therefore, this study fits streamflow based on gamma distribution using various intervals of 3-, 6-, and 9-month.