

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2022-235-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-235

Anonymous Referee #2

Referee comment on "Machine-learning ensembled CMIP6 projection reveals socio-economic pathways will aggravate global warming and precipitation extreme" by Piaoyin Zhang et al., Hydrol. Earth Syst. Sci. Discuss.,
<https://doi.org/10.5194/hess-2022-235-RC2>, 2022

Comments on the manuscript entitled "Machine-learning ensembled CMIP6 projection reveals socio-economic pathways will aggravate global warming and precipitation extreme" by Zhang et al.

Recommendation: Major revision

The manuscript examined three ensemble mean methods, which are based on machine learning (ML), using 16 CMIP6 models and observed temperature and precipitation data. Their evaluation results suggested that three ML-based ensemble methods can generate more reasonable simulations relative to individual CMIP6 models. The Deep Neural Networks method shows an overall better performance than the Ordinary Least Squares regression and Decision Tree methods in terms of the ensemble mean precipitation and temperature of CMIP6 models. The authors also projected the changes in temperature and precipitation based on the proposed multi-model ensemble mean method for SSP126, SSP245, and SSP585. The ensemble mean methods presented in the manuscript are interesting and the study fit the scope of HESS. However, the methods were not well explained and many important details are missing. In addition, comparing the ML-based ensemble mean against individual CMIP6 models cannot well define the merit of the method. At least, the uniformly weighted multi-model ensemble mean should be included in the comparison. The main findings of the study were not well summarized, either. Given these issues, substantial revision is required before the manuscript can be considered for publication with HESS. Detailed comments are as follows:

Major comments:

- The authors evaluated the performance of ensemble mean derived from three machine-

learning (ML) methods as well as individual CMIP6 model performance against observations. They concluded that the ML-based ensemble-mean precipitation and temperature were more credible than that simulated by individual GCMs. Previous studies suggested that multi-model ensemble mean with the uniform weight (MME) also shows generally good performance than individual GCMs in terms of various variables or indices (e.g. Massoud et al, 2019; Zhang et al, 2022). It is not enough to justify the merit of ML-based ensemble mean by simply comparing it with individual GCMs. I strongly suggest the authors include MME in the evaluation and compared it with the ML-based ensemble mean.

- The methods were not explained in enough details. For example, how was the weight determined in Eq. (1)? What does p mean in Eq. (2)? Does the OLS method require a significant correlation between model and observation? Note that no significant correlation is very common if one compares the year-to-year variation of precipitation (or temperature) between CMIP6 output and observation. What does “ s ” refer to in Eq. (3)? How was the weight determined in Eq. (7)? Without detailed method description, It is hard for readers to follow the authors’ method and repeat their results.
- The evaluation of three ML-based ensemble mean and individual CMIP6 models is not explained clearly. For example, it is not clear that the authors evaluated the ‘climatological mean’ precipitation (temperature) or their temporal variation against observation. It seems the authors evaluated the climatological means. Note that a model or multi-model ensemble mean can better capture the climatological means (section 3.1) do not necessarily suggest it can also generate a more reliable projection of future climate (section 3.2). No relationship between historical simulation and future projection was established in the evaluation, either. Moreover, the authors did not consider the uncertainty range of climate projection, which is also crucial for climate projection. These caveats should be discussed in the Discussion section at least.
- The conclusion section did not well summarize the main findings of this study. Some conclusions are not new. For example, it is not a very new finding that CMIP6 models tend to project a warmer climate relative to CMIP5 models (Section 4.2). The “hot model” problem was also reported in recent studies (e.g. Hausfather et al, 2022; Voosen, 2022). Also, it is not surprising that “the intensity order of temperature rising is SSP5-8.5 > SSP2-4.5 > SSP1-2.6 over a global scale” (L553-555).

Reference:

Hausfather Zeke, et al, 2022: Climate simulations: recognize the ‘hot model’ problem. *Nature*, 605, 26-29

Voosen Paul, 2022: “Hot” climate models exaggerate Earth impacts. *Science*, 376

Massoud, et al, 2019: Global Climate Model Ensemble Approaches for Future Projections of Atmospheric Rivers. *Earth’s Future*

Zhang et al., 2022: Evaluation of CMIP6 models toward dynamical downscaling over 14 CORDEX domains. CD

Minor comments:

L17: "precious studies", typo?

L24,25: "The new ensemble precipitation (temperature) data with the $R=0.81(0.99)$ is more accurate". It is not clear to me that the authors refer to climatological mean precipitation or interannual variation of precipitation. What data is used as the reference data when calculating the correlation coefficient?

L32: "The proposed analysis provides *credible* opportunities...". Here and elsewhere, you may say "improve the credibility" rather than "credible projection of future climate". Given the large uncertainties, no climate projection is credible.

L45: "anthropogenic activities" -> "human activities"

L50-51: delete the sentence before "However"

L52: "GCMs (General Circulation Models)" -> "General Circulation Models (GCMs)"

L53: "catch" -> "project"

L63-64: "However, the findings generated by *new* ensemble climate global dataset are *rarely* reported under CMIP6 with the new emission strategy." What do you mean by "new ensemble climate global dataset". Many papers have been published in terms of climate projection with the CMIP6 dataset.

L67: "physical parameters sensitivity" is not the only factor that affects the model performance.

L68-69: "Climate change projection ignoring the temporal and spatial heterogeneity leads to the incredibility of the estimation." This is not true. Climate change projections do consider the temporal and spatial differences. Moreover, a projection of climate change at a global or continental scale is usually *more reliable (rather than unreliable)* than that at a regional or local scale.

L70: "Utilizing only one model will 'improve' the uncertainty of climate projection"?

L90: "precious regional studies" -> "previous regional studies"

Table 1: Please double-check the model info presented in Table 1. The grid spacing of MPI-ESM1-2-LR seems incorrect.

L137: Why are there 540 observation images? How many years of observation data are used for input?

Fig.1: Which month is used, 01 or 02 (YYYY01 or YYYY02), in the second group of observation data?

L143: "a widely technique applied for" -> "a technique widely applied for"

L146: There are two "Lee et al., 2022" in the reference list. Please clarify which one you referred to here.

- L172-173: "following equation 4"?
- L191: delete "neural network"
- L210: delete "ranging from -1 to 1"
- L211-214: On what basis were correlation coefficients divided into five categories? The range of correlation coefficients is also affected by the size of the samples. A significance test is more desirable than a specified range of correlation coefficients. Is it appropriate to define no correlation as $R=0$ strictly?
- L226: Please double-check the citation of equations here.
- L236-243, L329-331: The authors define the p^{th} index with R , CRMSE, SD, and MAE. Note that these statistics are not independent of each other. For example, CRMSE is a function of R and SD (Taylor 2001; Xu et al., 2019). In another word, CRMSE already includes R and SD. Thus, it is not necessary to combine these statistics and compute CRI. However, the authors may consider computing CRI using the temperature and precipitation indices and compute CRI, which represents the model's overall ability to simulate both variables.

- Reference:
- Taylor, 2001: Summarizing multiple aspects of model performance in a single diagram. *JGR-Atmos*, 7183-7192
- Xu et al, 2019: Comments on 'DISO: A rethink of Taylor diagram'. *IJoC*.40, 2506-2510
-
- 3: The figure appears of poor quality. It is extremely hard to distinguish different models! The authors should also clarify SD and cRMSE were normalized by observed SD. Are the authors showing climatological mean temperature and precipitation? Over which region? Which season? Only for the land area? Please clarify.
- 4: Similar to Fig. 3, the figure caption failed to provide enough information. Does the figure show annual mean or seasonal mean precipitation (temperature)? Which region? Are the MAE and percentile calculated based on a climatological mean spatial field or a time series?
- L293-295, L207-209: why? Could you please explain the reason?
- L343: "overall pattern" -> "overall spatial variability"
- L345: This is not true.
- 8: Please use thick lines for the global mean temperature anomalies.
- L435-436: Please double check the citation of Fig. 10a and 10b.
- L441: "What's more" -> "Moreover"
- L442: delete "with"
- L443-444: The sentence does not read well.
- L469: "decrease" -> "reduce"
- L493: "quicker" -> "slower"
- L512: "aggravate" -> "generate a stronger"
- L522: "intervention" -> "mitigation"
- L535, 538: "expected to *accurately* project climate change", "high credible findings". Climate projection is strongly affected by scenario uncertainty, model uncertainty, and internal climate variability. Given these uncertainties, it is impossible to accurately project climate change using a multi-model ensemble mean. Thus, uncertainty range estimation is also of great importance.