Earth Syst. Sci. Data Discuss., community comment CC1
https://doi.org/10.5194/essd-2021-460-CC1, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2021-460

Nancy Swift

Community comment on "A 148-year precipitation oxygen isoscape for China generated based on data fusion and bias correction of iGCMs simulations" by Jiacheng Chen et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-460-CC1, 2022

The authors create a high-resolution precipitation oxygen isoscape dataset for China by fusing eight iGCMs simulations and in-situ observations based on data fusion and bias correction techniques. I appreciate the authors' great efforts to develop such a dataset, but I am quite dubious about the general content, and have large concerns about the novelty and quality of the dataset.

My main concerns are:

- Data-quality and novelty:

It feels like a direct comparison of five commonly used fusion methods for developing a high-resolution dataset in China, without any new advanced fusion methods. Furthermore, the quality of the developed dataset is still questionable and unreliable due to its poor and insufficient present form. It seems to me that the intended novelty might be a high-resolution dataset.

The methods used to develop the dataset are unclear and not robust. For example, the sensitivity of model parameters are not evaluated and discussed; thus, the results are not robust. Why does CNN perform much better? Why is it set to a three-layer structure model? The observed data covers a short period and is not sufficient to train the model.

Using the interpretations of spatial pattern at the seasonally averaged scale (1969-2007) and the temporal pattern at the regional scale to validate the effectiveness and reliability of the data is not persuasive. Why not give us a comprehensive assessment at finer scale, such as time-series comparisons between the gridded simulations (50km) and in-situ observations at each station, and spatial patterns for each month. Without these comprehensive evaluations at finer scale, I am quite dubious about the data-quality and usefulness of this data set.

- The presentation of dataset.

Introduction: The section of Introduction is not well written, and lacks to interconnect of the data it shares to and to show how it is valuable in relation to the Earth's system. For example, readers should have a clear understanding of the motivation of this study, the purpose of creating such a dataset, which can be seen from a literature review of

precipitation oxygen isoscape in hydrological and biogeochemical cycles. Then, followed by a detailed description of the available datasets, we could not find any description of the previously evaluated performance of iGCMs simulation in the Introduction. A description of the data-fusion method should also be added.

Data set and study area: Please add more details about the in-situ data (e.g., time-series of available in-situ data, number of data points at each station) in the supplementary material to justify the machine learning. For better visualization, a mesh of the iCGMs could be added in Fig.1

Methodology: This section is not well written, and should be clarified. For example, did you correct for iGCMs simulation bias at grid scale (grid by grid) or regional scale (using all gauges and all iGCMs within a specific region). How do these methods generate 50 km simulations from various iGCMs with different spatial resolutions?

Results and discussion: see above comments. Many published ESSD papers have demonstrated the uncertainties, limitations, advantages and 1-3 specific applications (for validation, evaluation, and analysis) of their dataset, which are also highly recommended for your paper.

For above reasons, I do not support its publication in the ESSD, without advanced approach or comprehensive evaluations of dataset at finer scale. Anyway, I still look forward to seeing the reviewers' comments and editor's decision.