

Earth Syst. Sci. Data Discuss., author comment AC1  
<https://doi.org/10.5194/essd-2022-45-AC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

Haibo Gong et al.

---

Author comment on "1km Monthly Precipitation and Temperatures Dataset for China from 1952 to 2019 based on a Brand-New and High-Quality Baseline Climatology Surface" by Haibo Gong et al., Earth Syst. Sci. Data Discuss.,  
<https://doi.org/10.5194/essd-2022-45-AC1>, 2022

---

We have answered point-by-point to all the comments raised by the referee. The answers are provided below each comment. We would like to thank the reviewer for the quality and detailed reviewing provided.

- Referee Comment 1:

This manuscript reported on one dataset of 1km Monthly Precipitation and Temperatures Dataset for China from 1952 to 2019. Basically, the dataset of temperature and precipitation at monthly scale are believed to be not able make any new contributions to the field of weather and climate community. China Meteorological Administration has compiled and archived a wide range of temperature and precipitation dataset at finer scales, in which a series of strict and physically- based quality-control measures have been implemented. Technically, a variety of data have been used to compile the dataset. Nevertheless, my big concern lies at the ambiguousness of data sources, which needs further clarification. Importantly, the selection of the optimization model was simply relying on the sorting combination of various data sources, while the rationality or physical basis behind the different combinations is not explained at all. In addition, there were a lot of still descriptive errors in this manuscript, and most of the figures were not presented in scientific ways, among others. Sorry for that I cannot be more positive. Therefore, I have to recommend the rejection of the work. The specific concerns of mine are listed as follows:

We sincerely thank the reviewer for his very insightful and effective comments and suggestions. Although the usefulness of climate datasets at a monthly scale may be weak in the field of weather and climate community, the climate datasets at monthly/annual scales are essential in the field of vegetation ecology, such as the effects of climate change on species distribution and the response of vegetation productivity/phenology to climate change. To our knowledge, the China Meteorological Administration has compiled a wide range of datasets, however, there were presently no publicly and freely available baseline climatology surface and the long-term time series datasets at high-quality spatial resolution (1km). Previous studies suggested that the release of public climate data over

China is insufficient to construct a baseline climatology dataset better than that available from WorldClim2 (Peng et al., 2019). In comparison to WorldClim/WorldClim2, ChinaClim\_baseline, was a better baseline climatology surface than WorldClim/WorldClim2, taking into account the dense weather stations (over 2000) and the useful satellite-driven TRMM3B43. We acknowledged the ambiguousness of our data sources and explained the source and processing of the data in a supplement called station\_informations in supplement2.zip. Notably, WorldClim/WorldClim2, which was one of the most popular climate datasets, was generated by the optimization model relying on the combination of various variables. Additionally, variations in local context and seasons have significant influence on climate processes, thus the model for fitting baseline climatology surface should vary from various climatic regions and different months. ChinaClim\_baseline was created by using the optimal TPS model for each climatic region and different months. This adaptive method allowed for better model fits in remote regions and specific months. Hence, our method of generation ChinaClim was feasible. Indeed, we also must acknowledge that the physical basis behind the different combinations was weak and current interpolation models hardly reflect this physical mechanism. We carefully addressed this manuscript's shortcomings in terms of descriptive errors and erroneous figures, and we rewrote several phrases with existing expression issue. We apologize for the numerous errors in this manuscript. We carefully analyze each problem in this manuscript and try all effort to improve it here. We truly hope to present a strict, meaningful, and attractive manuscript to readers.

- **Section 3.1, line 273 "After removing duplicate and invalid weather stations", based on**

**what methods and standards are invalid stations defined, and which station data have**

**been removed?**

Thanks for your suggestion!

We provided a supplement called station\_informations in supplement2.zip containing weather station information, as well as a list of the weather stations that were utilized and removed in this manuscript.

Specially, we excluded 47 weather stations with no records longer than ten years from a dataset of 2160 weather stations for the 30-year average climate (1981-2010). We excluded 143 weather stations with no records from 1981 to 2010 and less than 30 years of records from monthly climate observations (756 weather stations).

- **Line286, Table 1 introduces the data used for the interpolation. The precipitation data**

**used the TRMM3B43 data from 1998 to 2019 as a covariate. Was the data from 1952 to 1997 a default covariate? Whether the two pieces of data (1952-1997, 1998-2019) were continuous? For the same reason, temperature series (TMEAN, TMAX, TMIN), the time range of LST data was 2001-2019, then how to deal with the data from 1952-2000, Were the two pieces of data (1952 -2000, 2001-2019) continuous?**

Indeed, continuity is a serious issue since it may impact greatly the quality of time series

datasets. We made attempts in both data input and model selection to preserve the continuity of the time series dataset (ChinaClim time-series) as much as feasible. The continuous CRU dataset was always included as one of the variables to construct the TPS model across the whole time for data input. We used a segmented strategy for model selection, which first determined the basic monthly-scale TPS model, one of 7 model formulations (Table S6) constructed by using different combinations of variables (Longitude, Latitude, Elevation, Distance to the nearest coast, CRU anomaly (ratio), and the 30-Year normals). Then this model was used as the TPS model for the period without satellite-driven dataset (eg: 1952-1997 for precipitation); then, the basic monthly-scale TPS model, the satellite data variable was added as one of the covariates or spline variables to select the optimal TPS model for the rest of the period (eg: 1998-2019 for precipitation). This segmented strategy for model selection allowed us to preserve a degree of similarity between the models over the two periods, ensuring the finest potential continuity of the final dataset.

Additionally, we calculated the temporal trends for ChinaClim\_time-series datasets at a monthly scale and checked the possible mutations at the related years (eg., 1998 and 2001). Our dataset displayed good continuity, according to the results in a supplement called trend in supplement2.zip.

- **We learned that different optimization models (Table1, TableS7, TableS8) were used in**

**different months, which means that data from different sources was applied to the entire sequence. Has the 1952-2019 change been assessed, how to consider the homogeneity of the climate series (ChinaClim\_baseline)? Was this scheme reasonable?**

Thanks!

As shown in a supplement called trend in supplement2.zip, we assessed the temporal variation of ChinaClim\_time-series.

Additionally, I'm not sure what you mean about the homogeneity of the climate series; do you believe the spatial patterns of the various baseline climatology surfaces (or time-series climate data) should be assessed?

- **WorldClim data were downscaled from CRU-TS-4.03. The temperature data in this paper also used CRU as a covariate, and finally the WorldClim was compared with the sequence (ChinaClim\_baseline) in this paper. What was the significance of comparing the same data source?**

Thanks!

As far as we known, both the generation of WorldClim and WorldClim2 used the thin-plate smoothing spline algorithm for interpolation instead of downscaled from CRU-TS-4.0. It was also described in section 2.6 of this manuscript. The specific articles are:

- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
- Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.

We did not use CRU data as a covariate in the generation of the baseline climatology

surface, as shown in Table 1.

Generally, the accuracy of diverse data sets should be evaluated using their own independent samples. We were unable to find independent samples to test the accuracy of WorldClim2 and CHELSA in this manuscript, thus we just analyzed the differences in estimated temperature and precipitation values for different baseline climatology surfaces.

- **There were only 613 observation stations in this paper. The satellite data was inversion data and non-observation data, and the accuracy was 25km\*25km (0.25°). Other data (CRU) were the grid data with 613 station interpolation applied in China, and the accuracy was 25km\*25km (0.25°). The 30-year average climate (1981-2010) data belonged to a 30-year average of 366 data for each station. Precipitation events are small and medium-sized local events of the weather system. Precipitation events are small and medium-scale local events of the weather system. Was it reasonable to interpolate these data to 1km with this method of multiple statistical downscaling?**

Thanks!

In this work, 613 observation sites were used to construct the ChinaClim time-series, which was the whole amount of freely available weather station. Given the limited set of observation stations, attempting to interpolate directly to ~1km spatial resolution from these stations may lead to significant errors. Here, we employed climatologically aided interpolation (CAI) approach to obtain the final climatic surface by superimposing the baseline climatology surface and the monthly anomaly surface. To begin, 12 baseline climatology surfaces were generated using the multi-year monthly average observation data from more than 2,000 weather stations. These baseline climatology surfaces represent the local average monthly and annual conditions. Then, based on the anomaly time series from 613 observation stations, the monthly anomalies were interpolated to create monthly anomaly surfaces.

Notably, we downscaled the TRMM3B43 of 0.25° by applying the Cubist algorithm, which was believed to be effective for precipitation downscaling. The related references were as follows:

- Improving TMPA 3B43 V7 Data Sets Using Land-Surface Characteristics and Ground Observations on the Qinghai–Tibet Plateau.
- A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai–Tibet Plateau with the effects of systematic anomalies removed.

We used TPS interpolation to interpolate the CRU data to 1km spatial resolution, the reason for this was that we referred to the processing of some datasets such as CHELSA and KGclim. We realized that this method would increase the uncertainty of our dataset, and future work needs to strive to further improve the algorithm to improve the accuracy.

CHELSA: Climatologies at high resolution for the earth's land surface areas

KGclim: A 1 km global dataset of historical (1979–2013) and future (2020–2100) Köppen–Geiger climate classification and bioclimatic variables

- **Section 2.1, lines 164-165: "Dataset of 30-year average climate (1981-2010) was**

**obtained from two sources, 2438 weather stations from CMD and 25 weather stations**

**from Central Weather Bureau." It was inconsistent with the description in the legend of**

**Figure 1.**

Thanks!

We revised the description in the legend and attached specific information about weather stations used in this manuscript. We provided new Figure 1 in a supplement called figures in supplement2.zip.

For the generation of ChinaClim\_baseline, the number of weather stations used in this study was 2138, containing 2113 weather stations from CMD and 25 weather stations from Central Weather Bureau.

- **Section 2.1 Lines 167-168: "from the China Meteorological Data Service Center (CMD: <http://cdc.nmic.cn>)." the website cannot be accessible.**

Thanks!

We modified the inaccessible website (<http://cdc.nmic.cn>) to the correct link (<http://data.cma.cn>).

- **Figure 2 and Figure 4-7 were not the standardized maps of China, and the location of the nine-dash line and the Diaoyu Islands and Huangyan Islands were not marked.**

Thanks!

We revised the maps of China based on the standardized maps and a supplement called figures in supplement2.zip.

Given the drawing space limitations, as shown in Figure 2, we only marked the location of the nine-dash line and the Diaoyu Islands and Huangyan Islands for the bottom two maps.

- **The DEM data of the application and the method of how to make it into 1KM\*1KM were**

**not explained.**

Thanks!

The DEM data was aggregated by using nearest-neighbor interpolation method to 1km spatial resolution.

Please also note the supplement to this comment:

<https://essd.copernicus.org/preprints/essd-2022-45/essd-2022-45-AC1-supplement.zip>