

## Response to the Reviewer's Comments

### Reviewer's comments:

This manuscript reconstructs the first global, long-term (1955 - 2018), gap-free, gridded surface solar radiation (SSR) dataset by integrating nine SSR datasets using a CNN method. The main inputs include long-term ground networks, regional homogenized products, ERA reanalysis, etc. Overall, the proposed dataset is significant to the research community and will play an important role in evaluating climate modeling and analyzing global dimming and brightening. In addition, the authors did very comprehensive work in data processing. I would recommend the author include more product inter-comparison with other long-term SSR products to demonstrate the reliability and superiority of the proposed dataset. Some details of the data processing are still required.

### Response:

Thank you for the positive comments and your suggestions concerning our manuscript (essd-2023-178).

These comments /suggestions are all valuable and very helpful for revising and improving our manuscript, as well as the important guiding significance to our research. We have studied the comments carefully and have made corrections which we hope meet with approval. As suggested by the reviewer, we have compared our dataset with the ERA5 and CERES data, and the results will be shown in the SI. The main corrections to the manuscript and responses to the reviewer's comments are as follows.

### Major:

1. The manuscript adequately emphasizes the importance of a long-term SSR dataset for global dimming analysis but lacks a review of existing SSR datasets. Since ESSD is a data journal, I recommend including a paragraph about existing SSR datasets and their limitations in the literature review section.

**Response:** Thank you for the above suggestions.

We have systematically reviewed the limitations of existing SSR datasets in the previous paper (Jiao et al., 2022).

In the Introduction section of this manuscript, the second paragraph includes a description of some existing SSR datasets and their limitations but omits a review of the reanalysis and modal data. We have revised the Introduction section and added more references.

Specifically:

Lines 55-72 included a review of station datasets and their limitations. We present SSR datasets with global and regional coverage and point out their inhomogeneity and limited coverage.

Lines 76-80 provided a review of existing SSR satellite datasets and their limitations. We have made some additional revisions to this section. The new additions are as follows: “The spatial, temporal, and spectral coverage of a single satellite is limited, and multiple satellite data are therefore often used in tandem with each other; however, such a discontinuity in time and space can introduce inhomogeneity into a dataset (Evan et al., 2007; Feng and Wang, 2021; Shao et al., 2022).”

We have included a paragraph about existing reanalysis and model SSR datasets and their limitations in the literature review section. The new additions are as follows: “Reanalysis products are an important complement containing long-term SSR data, therefore have been widely used in climate studies (Zhou et al. 2017; Huang et al. 2018; Urraca et al. 2018a; Zhou et al. 2018; Jiao et al., 2022) due to the dynamically consistent and spatiotemporally complete atmospheric fields with high resolution and open access to data. However, existing studies have shown that reanalysis products generally overestimate multi-year mean SSR values compared to observations over land. With the continuous development of climate system simulations, model data from the Coupled Model International Program (CMIP) have become an important resource for conducting climate change research (Gates et al., 1999; Zhou et al., 2019). Previous studies have shown that the models used in CMIP6 overestimate the global mean SSR (Wild, 2020; Jiao et al., 2022; He, et al., 2023;).”

Lines 81-88 presented a brief review of SSR reconstruction using a machine learning approach.

#### Reference

Evan, A.T., Heidinger, A.K., Vimont, D.J., 2007. Arguments against a physical long-term trend in global ISCCP cloud amounts. *Geophys. Res. Lett.* 34 (4), L04701.

Feng, F., Wang, K., 2021. Merging high-resolution satellite surface radiation data with meteorological sunshine duration observations over China from 1983 to 2017. *Remote Sens.* 13 (4), 602.

Shao, C., Yang, K., Tang, W., He, Y., Jiang, Y., Lu, H., Fu, H., and Zheng, J.: Convolutional neural network-based homogenization for constructing a long-term global surface solar radiation dataset, *Renewable and Sustainable Energy Reviews*, 169, 10.1016/j.rser.2022.112952, 2022.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E., and Williams, D. N.: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I), *Bulletin of the American Meteorological Society*, 80, 29-55, 10.1175/15200477(1999)080<0029:Aootro>2.0.Co;2, 1999.

He, J., Hong, L., Shao, C., and Tang, W.: Global evaluation of simulated surface shortwave radiation in CMIP6 models, *Atmospheric Research*, 292, 10.1016/j.atmosres.2023.106896, 2023.

Huang, J., L. J. Rikus, Y. Qin, and J. Katzfey, 2018: Assessing model performance of daily solar irradiance forecasts over Australia. *Sol. Energy*, 176, 615–626, <https://doi.org/10.1016/j.solener.2018.10.080>.

Jiao, B., Li, Q., Sun, W., and Martin, W.: Uncertainties in the global and continental surface solar radiation variations: inter-comparison of in-situ observations, reanalyses, and model simulations, *Climate Dynamics*, 1-18, doi:10.1007/s00382-022-06222-3, 2022.

Urraca, R., T. Huld, F. J. Martinez-de-Pison, and A. Sanz-Garcia, 2018a: Sources of uncertainty in annual global horizontal irradiance data. *Sol. Energy*, 170, 873–884, <https://doi.org/10.1016/j.solener.2018.06.005>.

Wild, M.: The global energy balance as represented in CMIP6 climate models, *Clim Dyn*, 55, 553-577, 10.1007/s00382-020-05282-7, 2020.

Zhou, C., and Q. Ma, 2017: Evaluation of eight current reanalyses in simulating land surface temperature from 1979 to 2003 in China. *J. Climate*, 30, 7379–7398, <https://doi.org/10.1175/JCLI16-0903.1>.

Zhou, C., Y. He, and K. Wang, 2018: On the suitability of current atmospheric reanalyses for regional warming studies over China. *Atmos. Chem. Phys.*, 18, 8113–8136, <https://doi.org/10.5194/acp-18-8113-2018>.

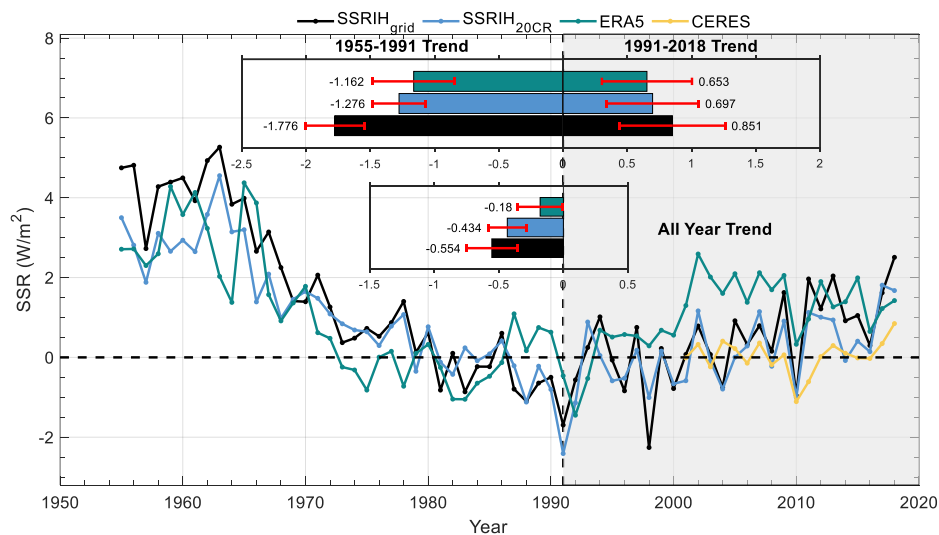
Zhou, W., Gong, L., Wu, Q., Xing, C., Wei, B., Chen, T., Zhou, Y., Yin, S., Jiang, B., Xie, H., Zhou, L., and Zheng, S.: Correction to: PHF8 upregulation contributes to autophagic degradation of E-cadherin, epithelial-mesenchymal transition and metastasis in hepatocellular carcinoma, *J Exp Clin Cancer Res*, 38, 445, 10.1186/s13046-019-1452-0, 2019.

2. The evaluation methods utilized in this work primarily compared ground-measured series and CMIP6 simulations, which may have ignored uncertainty due to the latter's limitations. To strengthen the paper, I suggest providing additional comparisons with independent global datasets, such as short-term remote sensing data or long-term reanalysis datasets, to demonstrate the proposed data's temporal stability, long period, and high accuracy.

**Response:**

In fact, we did not compare ground-measured series and CMIP6 simulations, but only used the CMIP6 SSR data as a training set to develop our CNN model used in this manuscript.

As suggested by the reviewer, we have compared our dataset with ERA5 and CERES data, and the results are shown below. We will include these results in the SI.



**Figure S8: Global land (except for Antarctica) annual SSR anomaly variations (relative to 1971-2000) before/after reconstruction. The Black solid line represents the SSRIH<sub>grid</sub> annual anomalies. The solid blue line represents the SSRIH<sub>20CR</sub> annual anomalies. The solid green line represents the ERA5 annual anomalies.**

The solid yellow line represents the CERES annual anomalies. The histograms represent the decadal trends of the SSRIH<sub>grid</sub> /SSRIH<sub>20CR</sub> / ERA5 (unit: W/m<sup>2</sup> per decade) and their 95% uncertainty range from 1955 to 1991, 1991-2018 and 1955-2018.

3. The description of the CNN modelling in section 3.2 needs clarity. Please provide details about the sampling of input data and the measures taken to prevent overfitting.

**Response:** Thank you for the comments.

A description of the input data (no sampling) is given in Section 5.1, lines 380-385, and a more detailed description of the CNN is given in the SI.

In this manuscript (SI), we will add descriptions of the measures taken to prevent overfitting of the CNN modelling. “We set the batch size to 16 in the first 500000 iterations and fine-tuned it to 18 in the last 10000000 iterations, for a total of 1500000 iterations, to suppress the overfitting phenomenon generated during the training process, and validate the model every 10000 times and early stopping if the validation shows a decreasing trend, the final number of training times used is 1100000. Second, L2 regularization is also added to regulate the loss function.

$$\mathcal{L}_{total} = \mathcal{L}_{valid} + 6\mathcal{L}_{hole} + 0.05\mathcal{L}_{perceptual} + 120 \left( \mathcal{L}_{style_{out}} + \mathcal{L}_{style_{comp}} \right) + 0.1\mathcal{L}_{tv} + \alpha \|\omega\|_2^2,$$

4. Additionally, Figure 4 clarifies how one CMIP6/20CR model was selected from their ensembles. Randomly selecting one model as input may lead to errors due to biases among different CMIP6 models.

**Response:** Thank you for your comments.

Rather than randomly selecting one model as input, we selected all 80 members of the 20CR as input (1 for evaluation and to test reconstruction, the other 79 for training the CNN model). Similarly, we selected 125 members out of a total of 507 members from several CMIP6 large ensemble models (with more than 10 realizations/runs) with high correlation coefficients with observations as input to train and validate the CNN model (1 for evaluation and to test reconstruction, the other 124 for training the CNN

model).

We have slightly revised Figure 4 to avoid ambiguity.

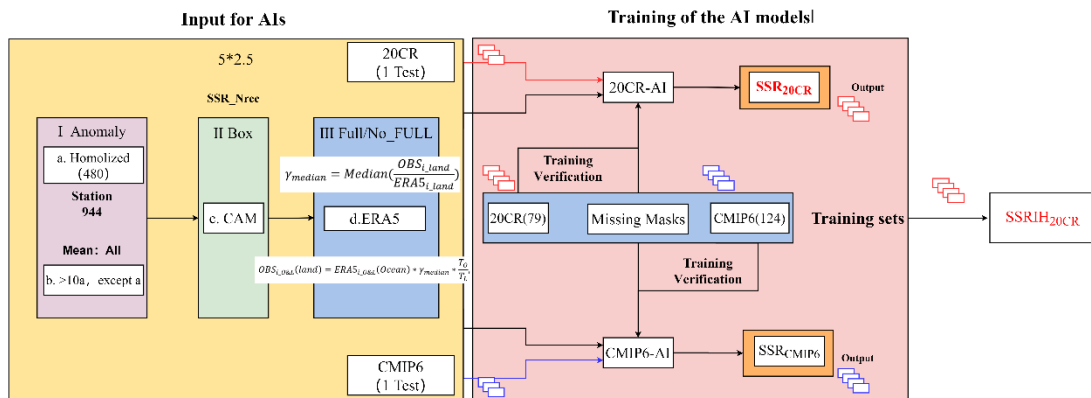


Figure 4: Flowchart of AI reconstruction.

5. In Figure S1, there appears to be a bias between the homogenized series and the original observation series for several years, followed by a good match. While I understand that the homogenization algorithm revised the series, I am concerned that this processing may introduce discontinuities in the time dimension. Please address this issue, as it is observed in nearly all sites.

**Response:** We totally understand your concern.

Figure S1 shows a comparison of the interannual variability of the station series before and after homogenization. To succeed in future observations, it is generally assumed that the most recent series are correct, while only the previous series are adjusted. This situation is therefore exactly the phenomenon caused by homogenization adjustments.

6. Considering that the proposed dataset covers 1955-2018 after reconstruction, it would be valuable to discuss the benefits of the product compared to long-term reanalysis data. It is important to acknowledge that reanalysis data also assimilate actual observations globally.

**Response:** Thank you for this comment.

This manuscript discusses the adjustment and reconstruction of *in situ* observational data, which serves as a benchmark for other comprehensive datasets, such as satellites, reanalyzes and model simulations.

As the reviewer points out, the reanalysis data assimilates some observations, but it is based on a state-of-the-art model and assimilation system. It does not contain a time function and is therefore affected by the data numbers, types, or quality of the assimilated observations.

7. 'data quality' or 'quality check' information should be given in the data file, which is required by the ESSD.

**Response:** Thank you for your advice.

The quality control procedure for the observations used in this manuscript includes extreme value checking, internal and spatial consistencies, etc. However, as our data sources (including the GEBA dataset, WRDC, CMA, etc.) have been systematically quality controlled by the data providers, the quality control of the raw data sources is not the main focus of this manuscript.

**Other:**

1. Title: Change "an artificial intelligence method" to "convolutional neural network" for more precise terminology. Additionally, please note that "AI" is often associated with models that can perform or think like human beings, which differs from "machine learning" or even "CNN."

**Response:** Thank you for your rigorous consideration. We changed the title to “An integrated and homogenized global surface solar radiation dataset and its reconstruction based on a convolutional neural network approach”

2. Line 105: Specify that there are 125 total inputs.

**Response:** Thank you very much for your comments. We selected 125 members out of a total of 507 members from several CMIP6 large ensemble models (with more than 10 realizations /runs) with high correlation coefficients with observations as input to train and validate the CNN model (1 for evaluation and to test reconstruction, the other 124 for training the CNN model).

3. Line 183: Revise the phrases "much better" and "excellent resource" or provide evidence to support these claims. Carefully review the entire context and be mindful of similar words.

**Response:** Thank you for pointing out this error.

Changed “Compared to previous model comparison projects, the CMIP6 project has a much better experimental design and more model development centres involved, as well as providing a much more significant amount of data.” to “Specifically, CMIP6 is considered as the current state of the art way of producing future climate simulations, including predicting future SSR based on different climate scenarios (Zhou et al, 2019).”

Changed “excellent resource” to “important resource”

Reference

Zhou, W., Gong, L., Wu, Q., Xing, C., Wei, B., Chen, T., Zhou, Y., Yin, S., Jiang, B., Xie, H., Zhou, L., and Zheng,

S.: Correction to: PHF8 upregulation contributes to autophagic degradation of E-cadherin, epithelial-



mesenchymal transition and metastasis in hepatocellular carcinoma, *J Exp Clin Cancer Res*, 38, 445, 10.1186/s13046-019-1452-0, 2019.

4. Line 206: Explain the rationale behind "five times." Consider removing years associated with major global volcanic eruptions (e.g., 1992) if they might impact the analysis.

**Response:**

We are very sorry for our negligence of the clerical error. It should be three times the standard deviation. The  $3\sigma$  criterion is also called PauTa criterion, which assumes that a group of data obeys or approximately obeys the normal distribution and only contains random errors. The standard deviation of this set of data is calculated and an interval is determined according to a certain probability. It is considered that the error outside this interval is a gross error rather than a random error, which should be eliminated (Olanow et al ,1998). Based on this criterion, 247 records were deleted. This represents approximately 0.4% of all station records.

In this manuscript, since we reconstruct the monthly SSR data through a CNN approach (image inpainting without time function as mentioned above), the extreme values associated with global volcanic eruptions (which may be spatially responded to) do not have a significant effect on the reconstruction.

Reference

Olanow C W, Koller W C. 1998 An algorithm (decision tree) for the management of Parkinson's disease: Treatment guidelines vol 50 no 3 (*Neurology*).

5. Line 223: Clarify the meaning of "potential reference pool" in this context.

**Response:** Thanks for your question. The potential reference pool contains all stations that can be used as reference series (Xu et al, 2013).

Reference

Xu, W., Li, Q., Wang, X. L., Yang, S., Cao, L., and Feng, Y.: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices, *Journal of Geophysical Research: Atmospheres*, 118, 9708-9720, doi:10.1002/jgrd.50791, 2013.

6. Line 274: Change "non-hole" to "gap-free."

**Response:** Changed. Thanks.

7. Line 294: Address the concern that taking a simple average between two sites with different time spans (e.g., 1950-1970 and 1960-1980) may result in discontinuity.

**Response:**

Thank you so much for your careful check.

In this manuscript, we followed the climate anomaly method (CAM) to calculate the global, regional and grid box average SSR change (Jones et al, 2001; Sun et al, 2021; Li et al, 2021). In a single 5\*5 grid box, we also calculate the average climate anomalies among all stations, which avoids the problems you mention by calculating the simple average of the absolute values (Li et al., 2009).

Reference

Jones, P., Osborn, T., Briffa, K., Folland, C., Horton, E., Alexander, L., Parker, D., and Rayner, N.: Adjusting for sampling density in grid box land and ocean surface temperature time series, *Journal of Geophysical Research: Atmospheres*, 106, 3371-3380, doi:10.1029/2000JD900564, 2001.

Sun, W., Li, Q., Huang, B., Cheng, J., Song, Z., Li, H., Dong, W., Zhai, P., and Jones, P.: The Assessment of Global Surface Temperature Change from 1850s: The C-LSAT2.0 Ensemble and the CMST-Interim Datasets, *Advances in Atmospheric Sciences*, 38, 875-888, 10.1007/s00376-021-1012-3, 2021.

Li Q, Sun W, Yun X, Huang B, Dong W, Wang X, Zhai P and Phil Jones: An updated evaluation of the global mean Land Surface Air Temperature and Surface Temperature trends based on CLSAT and CMST, *Climate Dynamics*, 56:635-650, DOI: 10.1007/s00382-020-05502-0, 2021.

Li W, Li Q, Jiang Z: Discussion on Feasibility of Gridding the Historic Temperature Data in China with Kriging Method, *Journal of Nanjing Institute of Meteorology*, 30(2): 246-252, 2009.

8. Line 216: Specify the extrapolation methods used.

**Response:** No extrapolation is used in this manuscript.

9. All trend statistics should include a significance test.

**Response:** Thanks for the reminder. A table of trends (including a significance test) and their uncertainties for each region is presented below and attached to the SI.

Table S3 Trends evaluation in Continental and hemispheric SSRIH<sub>20CR</sub> change from different scales

(Units: W/m<sup>2</sup> per decade).

Continental	Time period /Trend	Time period /Trend
North America	1955-1973	1973-2018
	-3.588±1.290	1.074±0.278
South America	1955-1990	1990-2018
	-0.408±0.619	0.049±0.768
Europe	1963-1978	1978-2018
	-2.180±1.866	1.081±0.312
Africa	1955-1991	1991-2018
	-1.506±0.496	0.340±0.998
Asia	1955-1990	1990-2018
	-1.633±0.473	0.435±0.505
North Hemisphere	1955-1991	1991-2018
	-1.457±0.246	0.887±0.415
South Hemisphere	1955-1991	1991-2018
	-0.708±0.330	-0.076±0.656

Table S4 Trend assessment in various data sources Global SSR change from different scales (units: W/m<sup>2</sup> per decade).

Type	1955-1991	1991-2018	1955-2018
SSRI <sub>grid</sub>	-1.995±0.251	0.999±0.504	-0.494±0.228
SSRIH <sub>grid</sub>	-1.776±0.230	0.851±0.410	-0.554±0.197
SSRIH <sub>20CR</sub>	-1.276±0.205	0.697±0.359	-0.434±0.148
ERA5	-1.162±0.319	0.653±0.350	-0.180±0.176