Reviewer #2:

This manuscript is interesting and convincing. The Manuscript develops the first, long-term (1955-2018), homogenized, gap-free global land SSR anomalies dataset by training improved partial convolutional neural network deep learning methods. Authors analyzed the global land (except for Antarctica) /regional scale SSR trends and spatio-temporal variations. Comparative validations /evaluations show that the $SSRIH_{20CR}$ provides a reliable benchmark for global SSR variations. Therefore, this manuscript may be considered for formal publication with minor modifications after addressing the following issues:

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 made corrections (please refer to the detailed revision after each comment) which we hope to get approval.

1. The resolution of the SSR data in this paper is only 5°×5°. Why not develop a product with higher resolution? What's the difficulty? Is it necessary?

Response: Thanks for your question.

Firstly, the reason why we did not develop a higher-resolution SSR dataset is the scarcity of the *in situ* observations. There are only about 1,000 compliant SSR sites worldwide. Even if we obtain higher resolution SSR datasets through interpolation techniques, they are only based on objective analysis in mathematical methods and do not add more SSR observation information, nor can they represent local scale SSR changes. Secondly, the SSR anomaly dataset in this manuscript is a benchmark dataset, which is designed to reflect large-scale SSR climate change. The global SSR data in 5° × 5° resolution already represents the long-term changes in SSR, as SSR and

temperature are similar in that they are both highly spatially representative.

2. Remote sensing inversion based on satellite measurements or some fusion products can provide space-time continuous SSR data, whether global or regional. I suggest that the authors clarify the reason why is the long-term trends of SSR data in this paper quite different from the current high-resolution satellite fusion data?

Response: Thanks for your question.

Firstly, the high-resolution satellite fusion datasets cover a too short period to investigate their decadal and multi-decadal variations.

Secondly, note that satellite fusion SSR datasets is also largely a modelled product, since satellites can only accurately measure the TOA fluxes, but not at the surface, since the atmosphere perturbs the surface signal received at the satellite sensor. Therefore, although it is a good estimation, it can still deviate from the real world (Zhang et al., 2016).

Thirdly, 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 (Tang et al., 2019; Feng and Wang, 2021; Shao et al., 2022).

Finally, the purpose of the application differs between the two datasets. The spatial resolution of the satellite fusion datasets are higher than those of the products of our dataset and will contribute to high-resolution photovoltaic applications. The SSR dataset in this manuscript is a benchmark dataset, which is designed to reflect large-scale SSR climate change.

Therefore, many aspects (including long-term trends) of SSR data in this manuscript are quite different from the current high-resolution satellite fusion data.

Reference:

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.

Tang W., Yang, K., Qin, J, Li, X., Niu, X.: A 16-year dataset (2000–2015) of high-resolution (3 h, 10 km) global surface solar radiation. Earth System Science Data, 11, 1905-1915, 2019.

Zhang X, Liang S, Wang G, Yao Y, Jiang B, Cheng J. Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. Rem Sens 2016;8(3):225.

3. The introduction provides a detailed overview of existing SSR datasets. However, the limitations of existing datasets are described rather briefly.

Response: Thank you for the above suggestions. It is really true as the first Reviewer suggested that we need to include a paragraph about existing SSR datasets and their limitations in the literature review section.

In the Introduction section of this manuscript, the second paragraph includes a description of some existing SSR datasets and their limitations but the limitations of existing datasets are described rather briefly. We have revised the Introduction section and added more references.

We provided a review of existing SSR satellite datasets and their limitations (**Page 4, Lines 78-83**). 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)." (**Page 4, Lines 83-86**)

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 (He, et al., 2022). 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).” (**Page 4, Lines 86-95**)

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.

He, Y., Wang, K., and Feng, F.: Improvement of ERA5 over ERA-Interim in simulating surface incident solar radiation throughout China, Journal of Climate, 34, 3853-3867, 2021.

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, ClimDyn, 55, 553-577, 10.1007/s00382-020-05282-7, 2020.

4. It is proposed to provide a more detailed description of the CNN method. Better to provide details about the measures taken to prevent overfitting.

Response: Thank you for your comments.

In this manuscript (SM), 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 (**Page 14, Lines 362-367**).

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

" (**Page 3, Lines 54-62 in the SM**).

5. Trends for the regional scales also need to be tested for significance.

Response: Thanks for your reminder. A table of trends (including a significance test) and their uncertainties for each region is presented below and attached to the SM (**Pages 5-6, Lines 65-70 in the SM**).

Table S3 Trend assessment in various data sources Global SSR change from different scales (units: W/m^2 per decade).

Type	1955-1991	1991-2018	1955-2018
$SSRI_{grid}$	-1.995±0.251	0.999±0.504	-0.494±0.228
$SSRIH_{grid}$	-1.776±0.230	0.851±0.410	-0.554±0.197
$SSRIH_{20CR}$	-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

Table S4 Trends evaluation in Continental and hemispheric $SSRIH_{20CR}$ change from different scales (Units: W/m^2 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

6. Figure 1 &4: The font size should be bigger.

Response: Thank you for pointing out this problem in the manuscript. We have redrawn Figures 1 (**Page 32, Line 813**) &4 (**Page 36, Line 822**) and enlarged the font size.

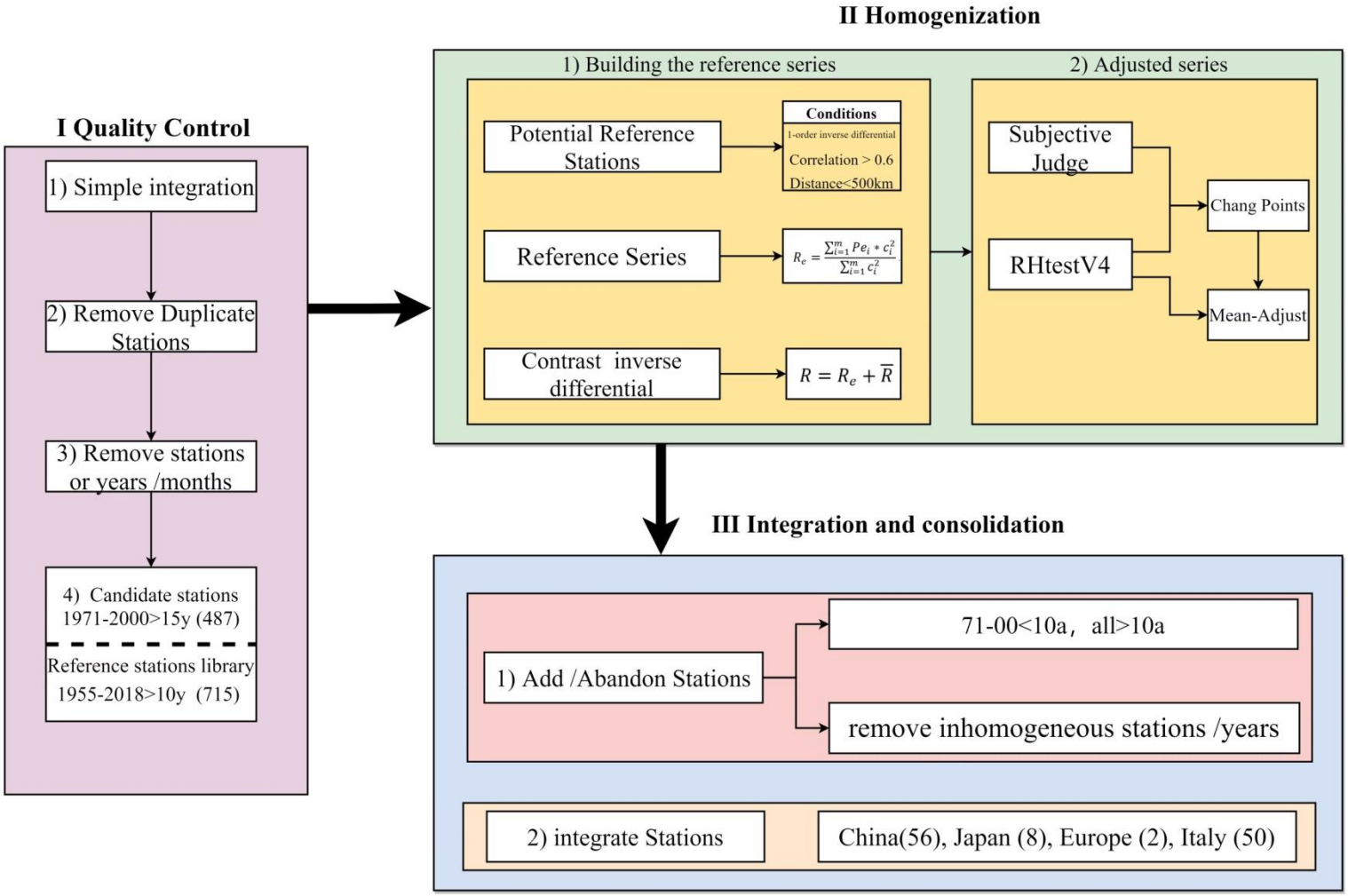

Figure 1: Flowchart of quality control (QC) (first step), homogenization (second step) and integration (third step).

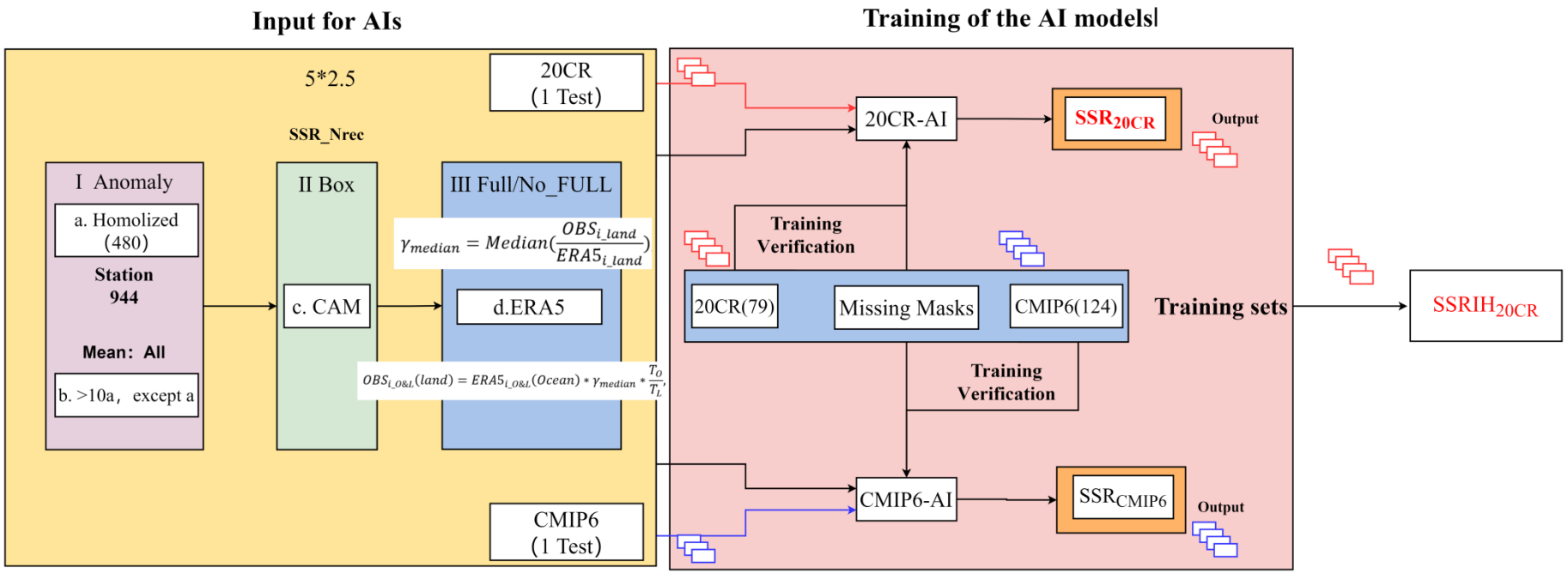

Figure 4: Flowchart of AI reconstruction.

7. The number of decimals should be consistent throughout. for example: Figure 9 and Line 441.

Response： We gratefully appreciate for your valuable suggestion. Considering the Reviewer's suggestion, we have redrawn Figure 6 (**Page 39, Lines 831-832**), Figure 9 (**Page 43, Line 852**), Figure S5b (**Page 22, Lines 119-124 in the SM**) and S7b (**Page 41, Lines 239-244 in the SM**).

We have revised the number of decimals (three valid decimals) throughout the manuscript (**Throughout**).

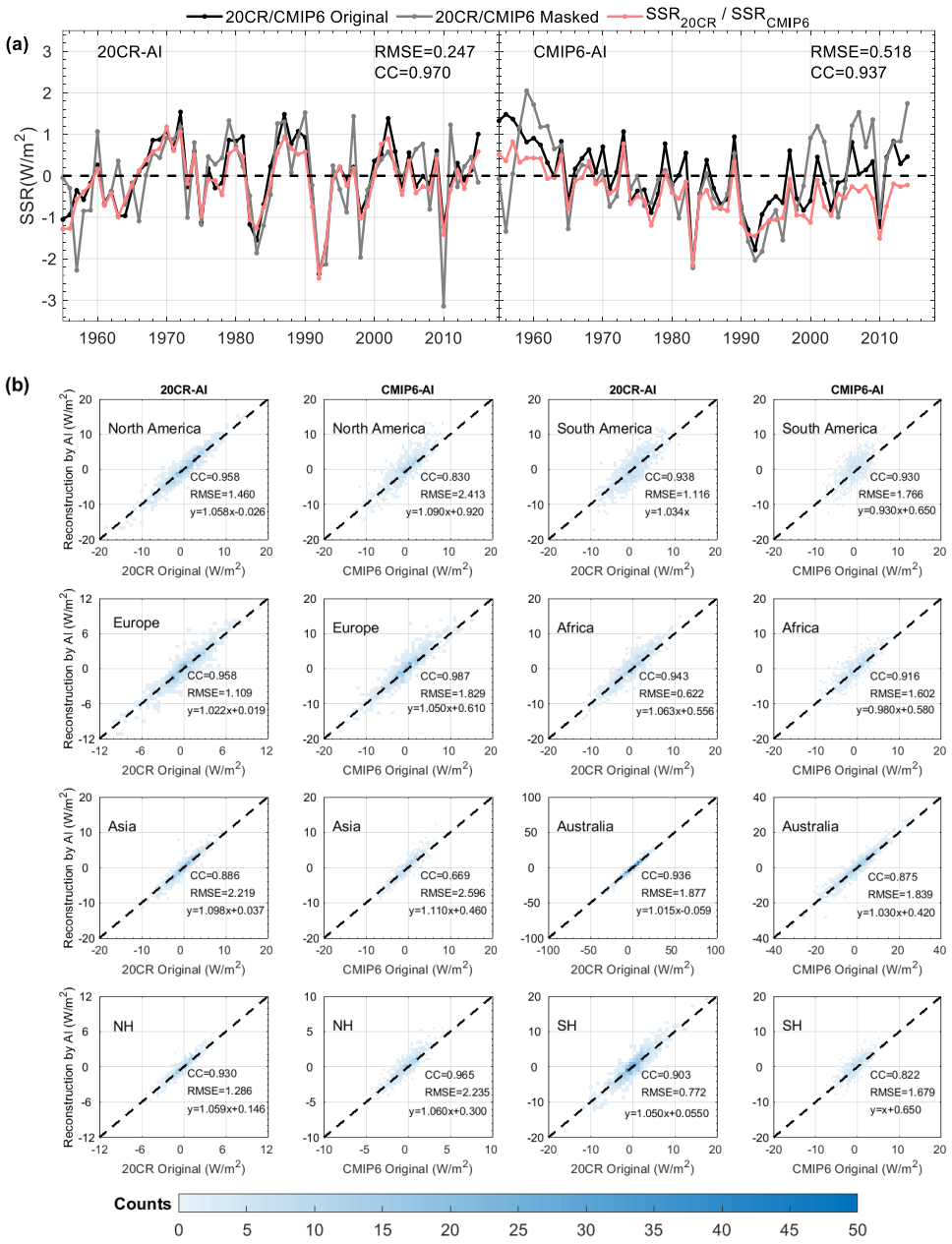

Figure 6: Reconstruction capabilities of the AI model. (a) Global land (except for Antarctica) means time-series analysis and AI model reconstruction evaluation. The red line is the SSR of the reconstruction based on the 20CR-AI /CMIP6-AI model (SSR_{20CR} /SSR_{CMIP6}); The grey line is the masked datasets with missing values of the $SSRIH_{grid}$. The solid black line is the 20CR and CMIP6 validation set (the SSR from the 1th member of 20CRv3 /CMIP6). (b) Comparisons of the SSR_{20CR} (columns 1, 3) /SSR_{CMIP6} (columns 2, 4) with the SSR from the 20CR and CMIP6 validation set. Colour bars represent counts with the same values for both. Figures also show the SSR_{20CR} (SSR_{CMIP6}) correlation coefficient (CC), root mean squared error (RMSE) and fitting equation compared to the original dataset in different regions.

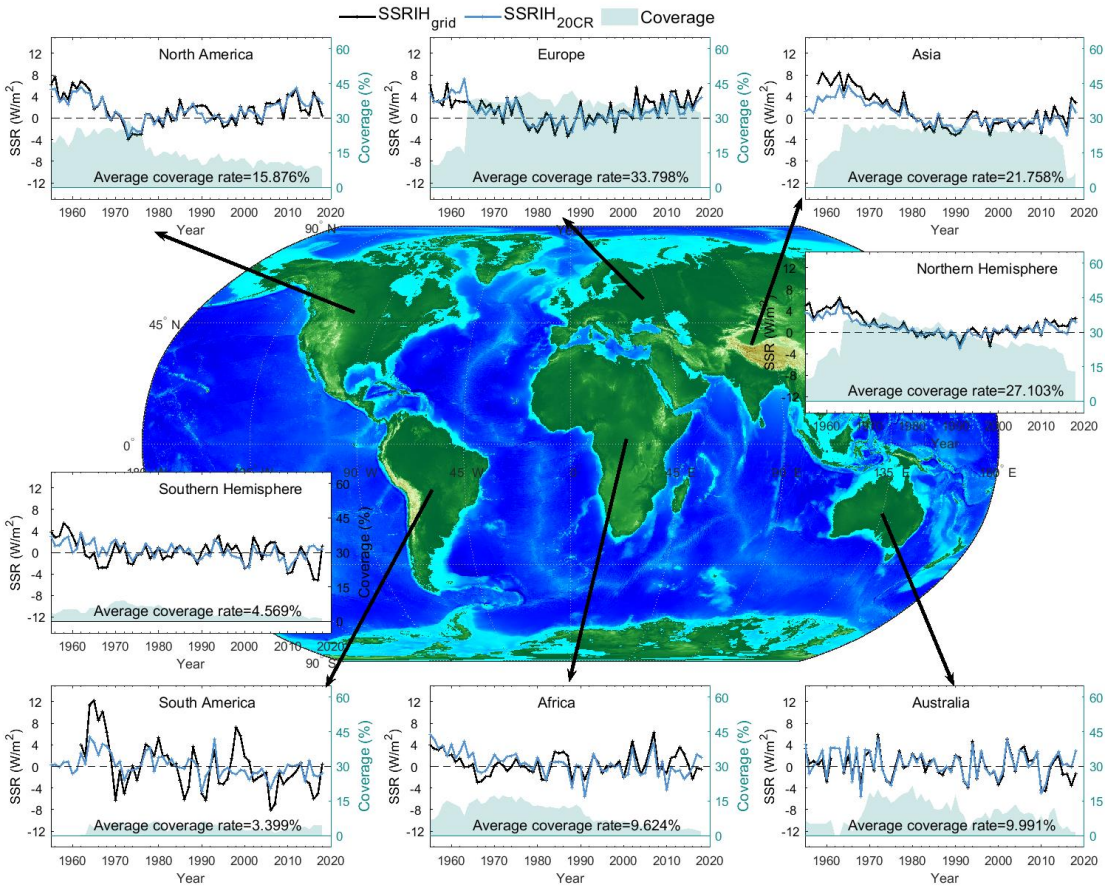

Figure 9: Same as Figure 8, but for regional annual anomaly variations. The green colour filling diagram represents the variation in grid box coverage (before reconstruction).

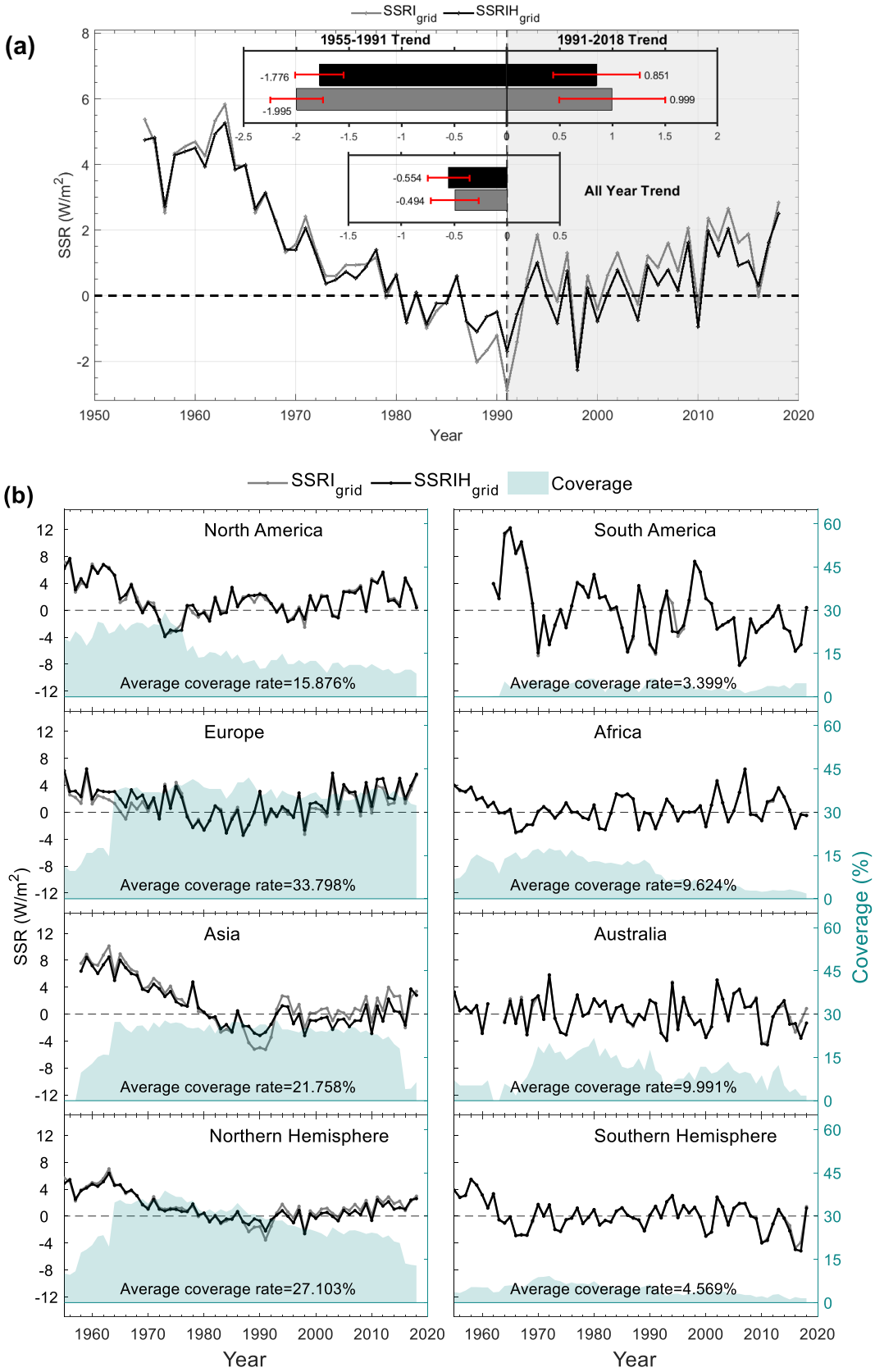

Figure S5: Time series of the annual global (a) /regional (b) SSR anomaly variations (relative to 1971-2000) before /after homogenization.

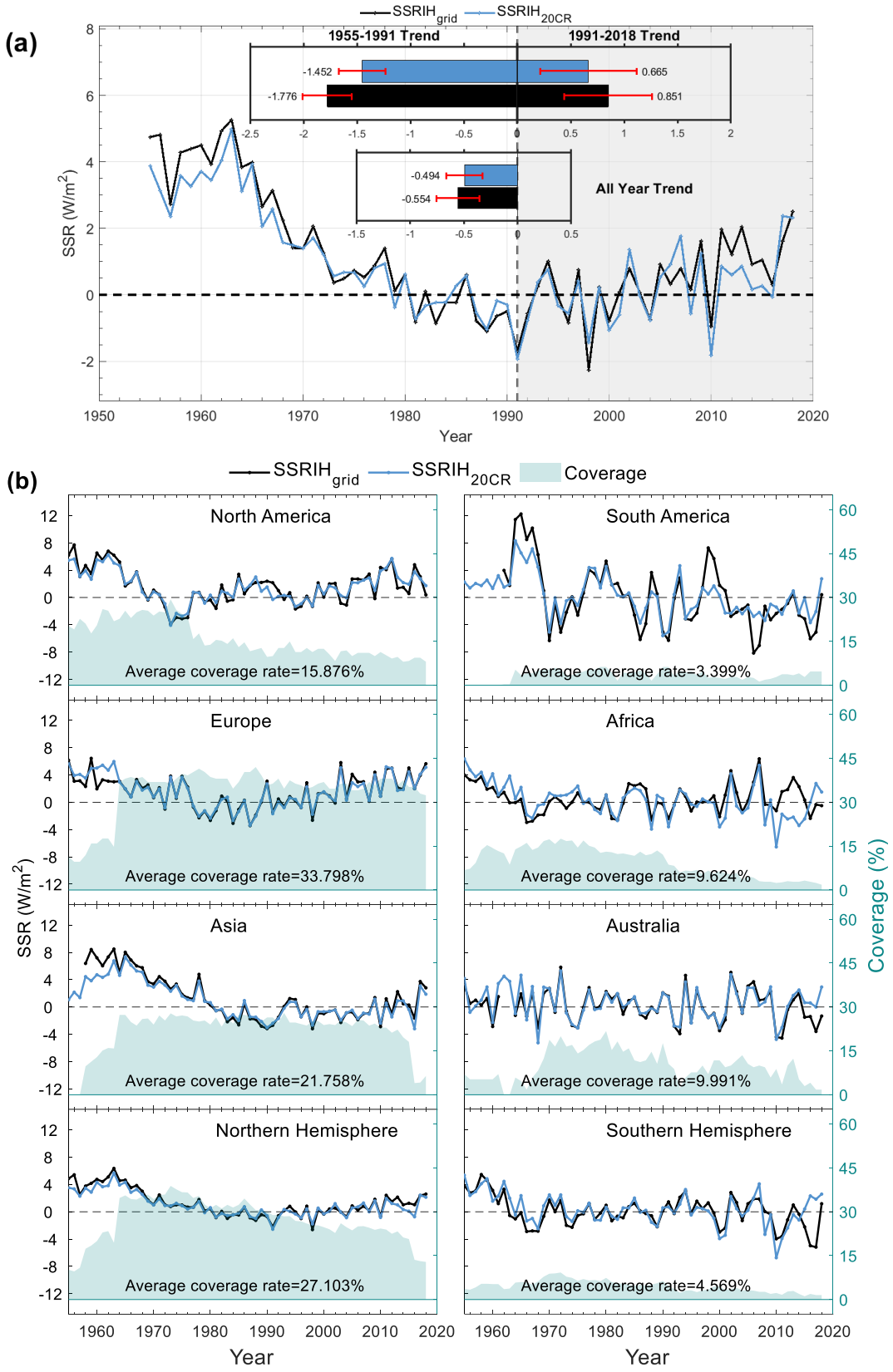

Figure S7: Same as Figure 8 and Figure 9, but the $SSRIH_{20CR}$ is reduced to the grid boxes with in situ observations.

8. Some sentences need to be polished and/or improved.

For example:

Lines 50-54: They allowed for the first time the detection of decadal changes in SSR known as “dimming and brightening” (Wild et al., 2005), especially considering that they cover a longer period concerning another type of data like for example satellite data (Pfeifroth et al., 2018) even if observational data often have uneven distribution and missing data with respect to the satellite data, especially in areas with complex orography (Manara et al., 2020).

Response： We agree with the comments. We have split this long sentence into two simple sentences.

They allowed for the first time the detection of decadal changes in SSR known as “dimming and brightening” (Wild et al., 2005), especially considering that they cover a longer period concerning another type of data like for example satellite data (Pfeifroth et al., 2018). Even observational data often have uneven distribution and missing data with respect to satellite data, especially in areas with complex orography (Manara et al., 2020) (**Page 3, Line 52**).

Lines 353-355: At the regional scale, the $SSRIH_{grid}$ has a generally similar variation to the $SSRI_{grid}$, and the $SSRIH_{grid}$ is usually more representative of climate change than $SSRI_{grid}$ at individual 355 stations. Remove “is”

Response： Remove “is” (**Page 15, Line 380**).