Details of the revisions and responses to **Reviewer 3** comments on the manuscript entitled '*GSDM-WBT: Global station-based daily maximum wet-bulb temperature data for 1981–2020*' (essd-2022-309)

We would like to thank the reviewer for your insightful and constructive comments that help to enhance the overall quality of our manuscript. Our responses (on a comment-by-comment basis) are indicated in blue text, and all updates to the original submission will be highlighted in the revised manuscript.

Comments: Wet bulb temperature is of great significance for the study of global or regional extreme heat events and humid heat, and this study realizes the calculation and homogenization of wet bulb temperatures at 1834 sites in a global long-term sequence, which is very meaningful work. Overall, this manuscript is clear and well-written and presents interesting research, but some concerns need to address before acceptance and publication. (Minor revision)

Response: Thank you for recognizing our work and providing the comments. All responses to your comments are as follows.

1. Figure S2 is difficult for readers to distinguish different zones. I suggest using different colors and symbols to distinguish different zones

Response: Thank you for the suggestion. We redrew the figure using different colors and symbols.

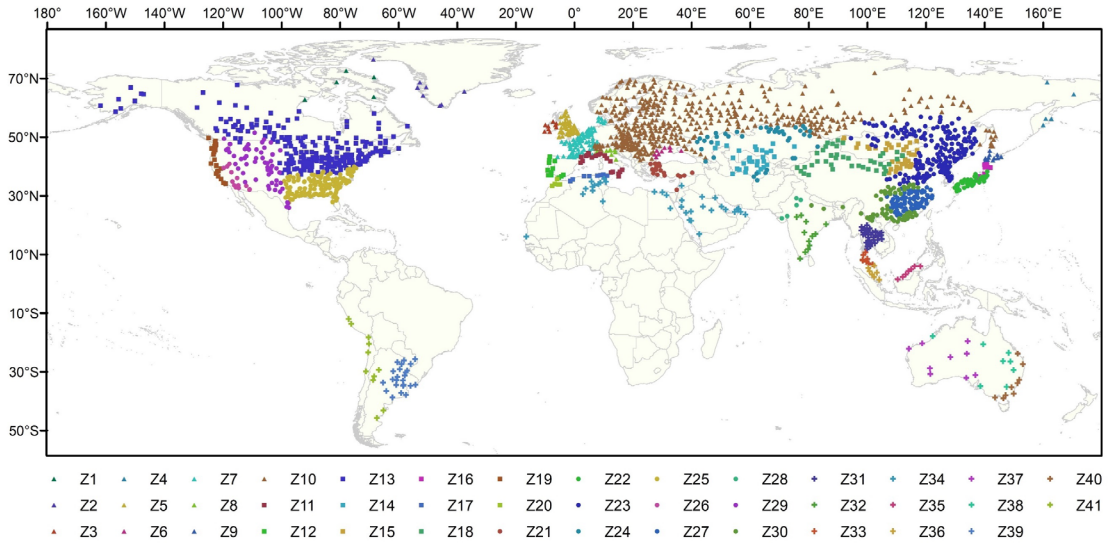

Figure S2. Spatial patterns of 41 station zones (total 1834 stations) based on Koppen-Geiger climate classifications. Each station zone contains at least 5 stations for better homogenization.

2. As mentioned by the author, the air temperature, specific humidity, and surface pressure data of each site are reanalysis meteorological data of the nearest grid point directly extracted, but the spatial resolution of NCEP-DOE data is relatively low. When using the nearest neighbor algorithm to extract the temperature, humidity, and surface pressure data of multiple adjacent sites may be the same. Will this affect WBT calculation? If the bilinear interpolation algorithm is used for extraction, does the result of WBT change greatly?

Response: Thank you for the comments. We are sorry for the unclear descriptions on extracting reanalysis series based on NCEP-DOE dataset. Due to the relatively coarse resolution of reanalysis data, one grid might involve two or more stations spatially. Therefore, to remove the effect caused by the same reanalysis series, we deleted the duplicate series and paired it with the station-based series with highest correlation coefficients for further bias correction. We will add related descriptions in the Methods in the next version of manuscript.

We did not apply the interpolation algorithm to improve the spatial resolutions because it is hard to determine the best scales of grids due to the unevenly distributed stations. The bias correction was also introduced to get the eventual complementary series so as to reduce the uncertainties from reanalysis dataset. In addition, it is notable that the process of complementing reanalysis series was essential but with relative low impact on the whole results. As shown in Table R1, the percentages of void time steps in all stations (0.04%-2.59%) relative to 14610 total time steps were low. The analysis about the effect of complementary series will be added in the Results.

Table R1. The effect of complementary series in different station zones.

Station zone	Number of complementary series	Number of all stations	Number of void time steps in all stations	Percentage of void time steps (%) in all stations
Z1	1	5	27	0.18
Z2	1	8	8	0.05
Z3	1	9	14	0.10
Z4	1	5	36	0.25
Z5	5	54	12	0.08
Z6	1	9	10	0.07
Z7	9	87	5	0.03
Z8	1	11	8	0.05
Z9	1	12	378	2.59
Z10	45	451	6	0.04
Z11	2	20	8	0.05
Z12	1	12	6	0.04
Z14	3	31	15	0.10
Z15	4	35	12	0.08
Z16	1	8	24	0.16
Z17	1	9	27	0.18
Z18	4	41	7	0.05
Z20	1	5	169	1.16
Z21	2	18	9	0.06
Z22	3	34	13	0.09
Z23	19	187	6	0.04
Z24	4	41	7	0.05
Z27	6	64	12	0.08
Z28	1	5	67	0.46
Z30	6	56	7	0.05
Z31	4	38	35	0.24
Z32	1	12	46	0.31
Z33	1	8	41	0.28
Z34	4	36	5	0.03
Z35	1	7	145	0.99
Z36	1	7	123	0.84
Z37	1	9	21	0.14
Z38	1	8	21	0.14
Z39	2	24	41	0.28
Z40	1	9	32	0.22
Z41	1	11	17	0.12

3. Line 168: How is the initial daily maximum WBT calculated? The time resolution of NCEP-DOE reanalysis data is 6h instead of 1h. More detailed description is required here.

Response: Thank you for the comment and suggestion. We at first calculated sub-daily (6h intervals) wet-bulb temperature by the same algorithm as demonstrated in section 2.2 and computed the daily maximum wet-bulb temperature, which is the initial wet-bulb temperature before bias correction. The six-hour interval of NCEP-DOP was also consistence with the criteria of our data quality control in section 2.3. We added more details here for better understanding, and also emphasized the time resolution again in the manuscript. The revised sentences are as follows "*First, the air temperature, specific humidity and surface pressure of the grid point nearest to each station were extracted, and the sub-daily (six-hour interval) TW was calculated (see section 2.2). Then the initial series of daily maximum TW and monthly mean were computed before bias correction.*".

4. More detailed description of Figure 4 (a) is helpful for readers to compare the results before and after homogenization. In addition, "Before homogeization" should be "Before homogenization" in Figure 4 (a).

Response: Thank you for the suggestions. Figure 4 (a) showed the correlation coefficients of series between paired stations before and after homogenization, and the sub-plot showed for the paired stations of which distances lower than the first quarter. When the coefficients were more than 0, the dots in the upper areas of black diagonal indicated the higher coefficients after homogenization. The maximum increment of coefficients was 0.28. There was also an obvious increase in the coefficients between closer stations as shown in the blue dots. In the sub-plot of Figure 4 (a), about 80.23% of paired stations had larger coefficients after homogenizations. We have added more descriptions in the Results as "…*Overall, the correlation coefficients after correction were higher and the maximum increment of coefficients was 0.28. It is also notable that there was a significant increase in correlation between stations that were closer together as shown in the blue dots. In the sub-plot of Figure 4 (a), about 80.23% of paired stations had larger coefficients after homogenizations…*", and in the caption of figure as "*When the coefficients were more than 0, the dots in the upper areas of black diagonal indicated the higher coefficients after homogenization*". We also revised "Before homogeization" to "Before homogenization" in Figure 4(a).

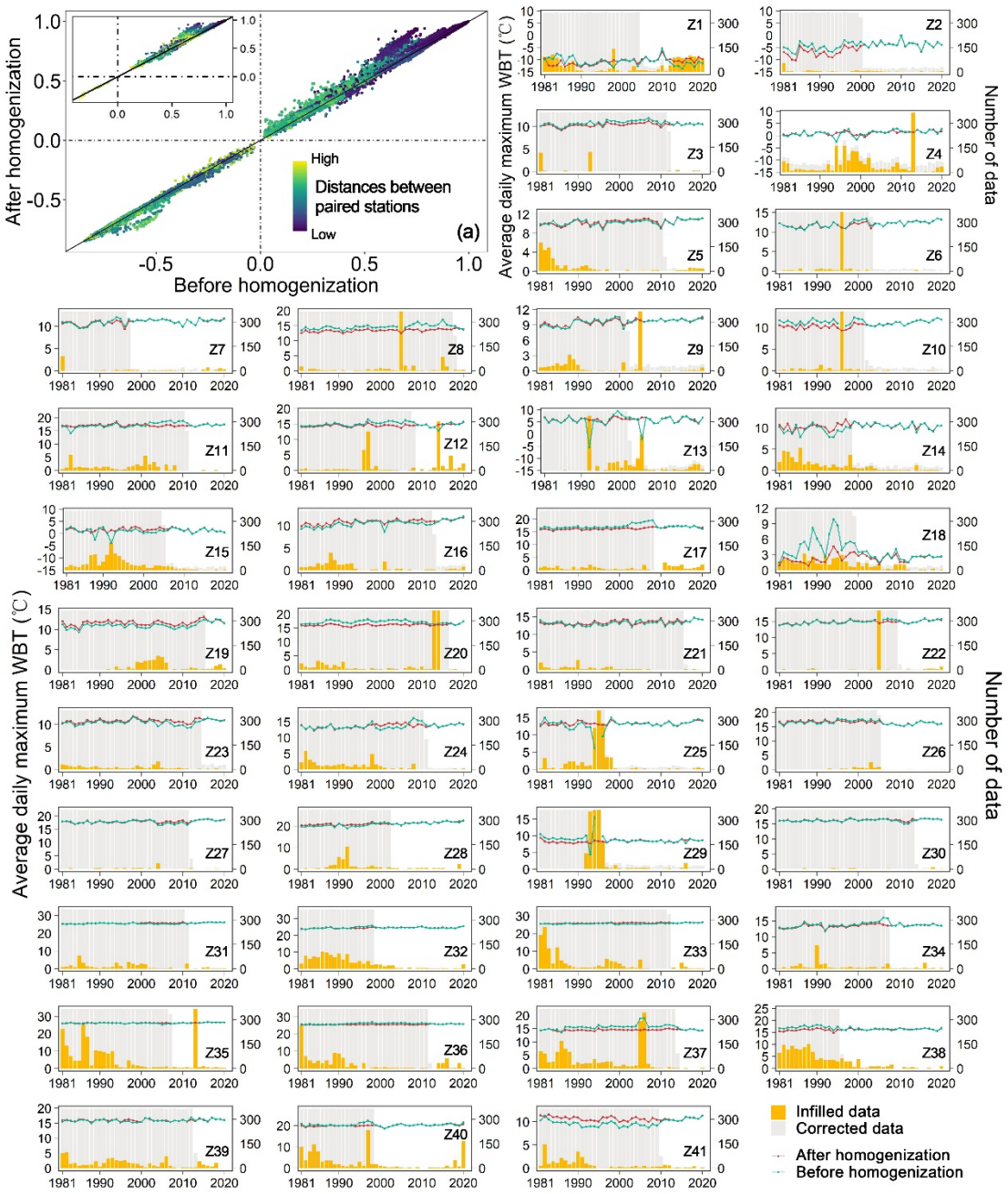

Figure 4. Correlation coefficients ($p<0.05$) between paired series before and after homogenization (a), annual average daily maximum TW (℃) and the number of infilled or corrected data for one typical station in each station zone (Z1-Z41). Note that sub-plot of (a) showed the correlation coefficients between paired stations of which distances lower than the first quarter. When the coefficients were more than 0, the dots in the upper areas of black diagonal indicated the higher coefficients after homogenization. Detailed information of all typical stations was shown in Table S2.

5. Line 243: What does higher SNHT value represent?

Response: Higher SNHT values mean higher probability of such stations to be detected the break points (also the inhomogeneous series). In the Climatol, SNHT test is applied to the series of anomalies between the actual values and the reference values and the SNHT values are used

to identify the break points. We have further added the related explanation in the Methods as "*Higher standard deviations and SNHT values mean higher probability of such stations to be detected as the outliers and break points.*".

6. In Lines 264-265: The WBT is calculated site by site and day by day. The statistical results do show that there are many missing data of WBT in the HadISD-Humidity data, but the author believes that HadISD-Humidity has relatively low accuracy and higher uncertainties. From my understanding, the existing description is not enough to prove that HadISD-Humidity has relatively low accuracy and larger uncertainties.

Response: Thank you for the comments. In this section, we aimed to explain two main problems in HadISD-Humidity, and the first one is about missing values. Particularly, since the daily maximum TW is the main measurement for characterizing extreme humid-heat in the warm seasons, the number of missing days in different months was shown in the manuscript. It was found that there were more missing values during the warm season, especially in the Northern Hemisphere. Because the extremely humid heat events are generally identified based on daily TW and the daily thresholds in the historical baselines, more missing values could lead us to detect inaccurate thresholds or identify insufficient events. So, the probable uncertainties may exist when directly using HadISD-Humidity to characterize humid heat.

We are sorry for the misunderstanding on the description, and according to the above explanation we revised the related contents as "*In terms of seasonality, there are evidently more missing days in the warm season (May-September) in the Northern Hemisphere, especially in summer (June-August). Because the extremely humid heat events are generally identified based on daily TW and the daily thresholds in the historical baselines, more missing values could cause inaccurate thresholds or insufficient events to be detected. Therefore, it needs to be noticed the probable uncertainties when directly using HadISD-Humidity to characterize humid heat*".

7. As described by the author, the results of HadISD-Humidity and HiTiSEA are overestimated or underestimated, and the author uses HadISD-Humidity and HiTiSEA data to compare and analyze the results of GSDM-WBT, then the verification results do not represent the true accuracy of GSDM-WBT, but only the relative accuracy.

Response: Thank you for the comments. To get the true accuracy of GSDM-WBT, the best approach is to directly compare our dataset with the long-term homogenous observations of wet-bulb temperature. However, to our knowledge, there is no global observation-based dataset

which could be used to validate the absolute accuracy of GSDM-WBT until now. According to your suggestion, we would emphasize the “relative accuracy” of the evaluations in the revised manuscript.

As for the dataset HiTiSEA, its underestimation of average 0.4°C was also found by their producers. When comparing the HiTiSEA to GSDM-WBT, the average bias was about 0.34°C which proved its underestimation. In previous regional studies on TW, the underestimation caused by using reanalysis dataset has also been demonstrated (Freychet et al., 2020; Raymond et al., 2020). When comparing with the HadISD-Humidity, the daily maximum TW of GSDM-WBT is overall lower. Because of the same data source used in the GSDM-WBT and HadISD-Humidity, the existing missing values might increase the potential inhomogeneity of HadISD-Humidity. So, in the section of comparing with station-based dataset, we would like to illustrate the possible reasons for the bias between GSDM-WBT and HadISD-Humidity, but not determine whether the HadISD-Humidity was overestimated. We have checked all inappropriate descriptions and revised the contents.

References:

Freychet, N., Tett, S. F. B., Yan, Z., and Li, Z.: Underestimated change of wet-bulb temperatures over east and south China, Geophys. Res. Lett., 47, e2019GL086140, https://doi.org/10.1029/2019GL086140, 2020.

Raymond, C., Matthews, T., and Horton, R. M.: The emergence of heat and humidity too severe for human tolerance, Sci. Adv., 6, eaaw1838, https://doi.org/10.1126/sciadv.aaw1838, 2020.