Anonymous Referee #3

The authors proposed a high-spatial-resolution monthly temperature and precipitation dataset for China by Delta downscaling of CRU dataset. The original CRU at 30' resolution is downscaled to 1km grid. The new downscaled data set include four common climate elements that are always the driven data for various models. This topic is quite interesting and would be useful for the climate change community. However, there are some obvious flaws in the downscaling procedure and the evaluation part. More interpretation and discussion should be improved. Therefore, I do not support this publication in ESSD at current version.

General comments:

1. Downscaling is a complicated procedure, especially for precipitation from 30' to 1km grid. I do not agree that the downscaled data set represents the local physical process. Actually, Delta downscaling is an interpolation method. CRU data set is also actually produced by interpolation method. The final downscaling result is the sum of "raw" CRU and interpolated anomaly. For my understanding, there is not any physical process involved. Conventionally, for a better local representation, local topography features should be considered such as aspect, slope and elevation.

Response: You are right. The methods that CRU used to construct 30' time series dataset are similar to the Delta downscaling framework. However, this study further improved the CRU data to higher spatial resolution by combining CRU and WorldClim datasets. The WorldClim dataset have four spatial resolutions (i.e., 10', 5', 2.5', and 0.5'), which consider the effects of local topography features, distance to the nearest coast, and three satellite-derived covariates. Thus, the downscaled datasets considered the above physical process. Our manuscript did not present this detailed information. We will add them to the revised version.

For the CRU time series data, we will describe it as follows.

Monthly mean, maximum, and minimum TMPs, as well as PRE, with a spatial resolution of 30' and covering a period from January 1901 to December 2017, were obtained from the CRU TS v. 4.02 dataset (http://www.cru.uea.ac.uk) (Harris et al., 2014). Methodologies used by CRU group to construct 30' time series dataset are similar to the Delta downscaling framework. First, more than 5000 weather stations are employed, and each station series is converted to anomalies by subtracting (for temperatures) or dividing (for precipitation) the 1961–1990 normal from all that station's data. Then, the station anomaly time series are linearly interpolated into 30' grids covering the global land surface. Finally, the grid anomaly time series were transformed back to absolute monthly values by a 30' climatology during 1961–1990. Specifically, the 30' climatology used by the CRU group contain the climatology for each month and are obtained from New et al. (1999). This climatology were generated by a function considering the latitude, longitude, and elevation, base on global 3615–19800 weather stations. Elevation data used in this climatology had a spatial resolution of 30', which was a mean result of global 5' digital elevation model. Specifically, elevation at each 30' grid was the mean of 36 grids of 5' digital elevation model (New et al., 1999). Therefore, the CRU dataset could well represent orographic effects on climate variation at the 30' spatial resolution; compared with similar gridded products, the CRU dataset had better performance. In addition, 323 weather stations across China region were employed by CRU group to generate the CRU time series data (Harris et al., 2014) (Figure 1).

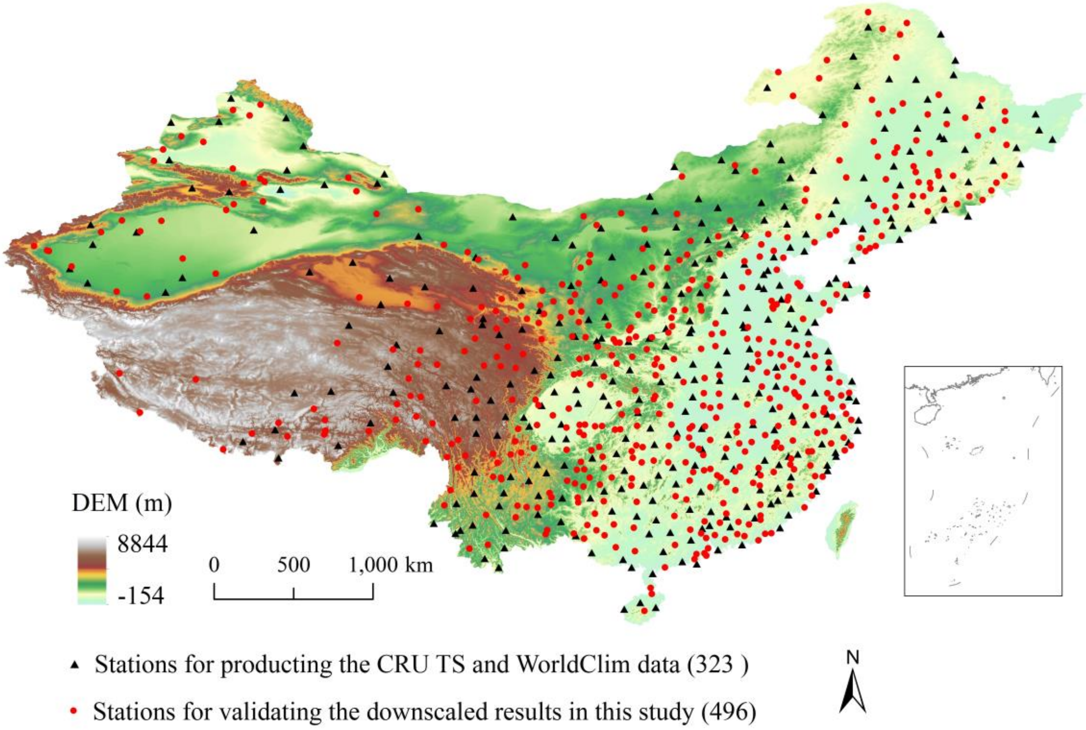

Figure 1. Spatial distribution of the national weather stations across China.

For the WorldClim data, we will describe it as follows.

To downscale CRU TMPs and PRE time series to higher spatial resolutions, we obtained four high-resolution reference datasets at spatial resolutions of 10', 5', 2.5', and 0.5' from the WorldClim v. 2.0 (http://worldclim.org) (Fick and Hijmans, 2017). The reference datasets contained monthly averages of climatic variables (mean, maximum, and minimum TMPs, as well as PRE) for the period 1970–2000, generated based on global 9000–60000 weather stations, using the thin-plate splines interpolation method. Thus, for each climatic variable, it has 12 climatology layers, representing the climatology ranging from January to December. Remarkably, the interpolation considered covariation with the latitude, longitude, elevation, distance to the nearest coast, and three satellite-derived covariates: the maximum and minimum land surface temperature and cloud cover, obtained from the MODIS satellite platform. Thus, these reference data reflect orographic effects and observed climate information for each month. Moreover, cross-validation correlations indicated that these reference data held good performance around the world, because of the introduction of satellite-derived and distance to the nearest coast covariates. In addition, weather stations over China region used in the WorldClim were the same as the CRU group (Fick and Hijmans, 2017) (Figure 1). For independent evaluation of downscaled dataset in this study, these weather stations would be excluded.

For the Delta downscaling method, we will describe it as follows.

Delta downscaling was employed to generate monthly TMPs and PRE for the period 1901–2017 at spatial resolutions of 10', 5', 2.5', and 0.5'. The Delta downscaling procedure contains four steps (Peng et al., 2018).

The first step constructs the climatology for each month and each climatic variable based on the 30' CRU time series. In specific, the long-term average of TMPs and PRE were calculated for each month using CRU TMPs and PRE time series. This step keeps the spatial resolution of 30' from CRU. To match the data period of the WorldClim, the period 1970–2000 was selected.

The second calculates the anomaly time series for each climatic variable using the 30' CRU time series and the calculated monthly averages. The TMP anomaly was calculated as the difference between the TMP time series and their long-term average in each month, while the PRE anomaly was calculated as the ratio of the PRE time series to their long-term average in each month.

$$\mathrm{An\_TMP}(yr, m) = \mathrm{TMP}(yr, m) - \mathrm{CRUClim\_TMP}(m) \quad (1)$$

$$\mathrm{An\_PRE}(yr, m) = \mathrm{PRE}(yr, m) \,/\, \mathrm{CRUClim\_PRE}(m) \quad (2)$$

where An_TMP(*yr*, *m*) and An_PRE(*yr*, *m*) are the anomaly for temperatures and precipitation, respectively, at *m* month and *yr* year; TMP(*yr*, *m*) and PRE(*yr*, *m*) are the absolute temperatures and precipitation values, respectively, at *m* month and *yr* year; CRUClim_TMP(*m*) and CRUClim_PRE(*m*) are the 30' climatology for temperatures and precipitation, respectively, at *m* month. *m* ranges from 1 to 12, representing January to December.

The third step spatially interpolates the 30' anomaly time series to higher spatial resolution. In this study, the 30' anomaly was interpolated to four spatial resolutions (i.e., 10', 5', 2.5', and 0.5') to match those of reference datasets from the WorldClim. To optimize the interpolation, we compared the performance of three methods, including bicubic interpolation, bilinear interpolation, and nearest-neighbor interpolation methods, to select the appropriate approach.

The last step reversely transforms the time series of anomaly to those of absolute values. Contrary to the steps for anomaly calculation, addition was used for TMPs, while multiplication for PRE. It should be noted that this step was carried out for each high-spatial-resolution anomaly time series.

$$\mathrm{TMP}(yr, m, res) = \mathrm{An\_TMP}(yr, m, res) + \mathrm{WorldClim\_TMP}(m, res) \quad (3)$$

$$\mathrm{PRE}(yr, m, res) = \mathrm{An\_PRE}(yr, m, res) \times \mathrm{WorldClim\_PRE}(m, res) \quad (4)$$

where *res* represents the spatial resolution, i.e., 10', 5', 2.5', and 0.5'; TMP(*yr*, *m*, *res*) and PRE(*yr*, *m*, *res*) are the absolute temperatures and precipitation values with a spatial resolution of *res*, respectively, at *m* month and *yr* year; An_TMP(*yr*, *m*, *res*) and An_PRE(*yr*, *m*, *res*) are the anomaly with a spatial resolution of *res* for temperatures and precipitation, respectively, at *m* month and *yr* year; WorldClim_TMP(*m*, *res*) and WorldClim_PRE(*m*, *res*) are the climatology from the WorldClim with a spatial resolution of *res* for temperatures and precipitation, respectively, at *m* month.

Figure 2 presented the steps of Delta downscaling for mean TMP using the CRU 30' time series and WorldClim 0.5' climatology datasets.

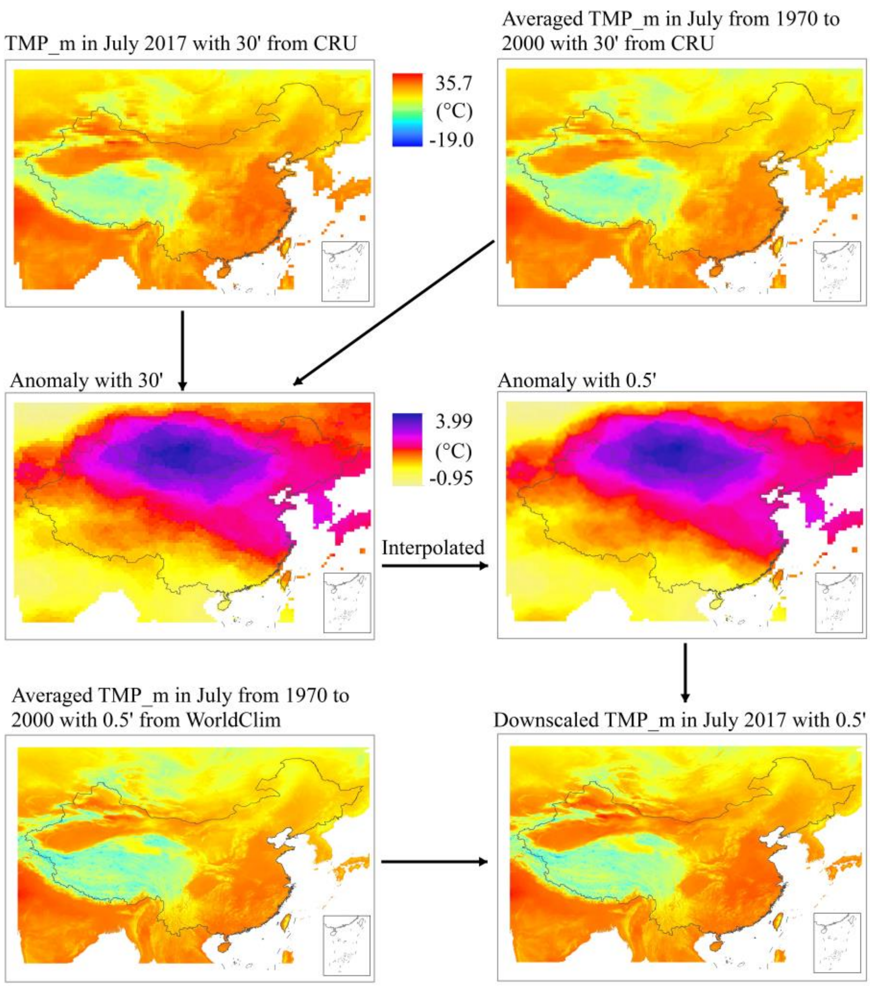

Figure 2. Schematic illustration of the Delta downscaling procedure. The mean TMP (TMP_m) in July 2017 obtained from CRU is used as an example.

2. WorldClim data set is used as the reference data in downscaling. However, how well does WorldClim represent the climatology over China? I did not find this information in the current version. The bias of WorldClim could be transferred into the final results. Therefore, it is not easy to understand why the downscaled data has a better performance. If the authors use other reference data, how will the downscaling result be?

Response: There have several overlapped weather stations for the creations of CRU and WorldClim data as well as the evaluation in the original manuscript. Now, based on the independent stations, we evaluate the monthly WorldClim data for each climatic variables (Tables 1-2). The evaluation indicates that (1) the WorldClim datasets have high performance to represent the monthly climatology over China region; and (2) the performance is better almost along the higher spatial resolution. Therefore, the accuracy representation of monthly WorldClim data result in the better performance of the downscaled dataset, especially at the spatial resolution of 0.5’. Considering the advantage of the monthly WorldClim data (e.g., the introduction of satellite-derived and distance to the nearest coast covariates), we just used this reference data. Besides, there are no other reference data with different resolutions for this study. We will added above evaluation in the revision.

Table 1. The mean absolute errors between the observed and WorldClim climatology at different

spatial resolutions over the 496 weather stations. The period ranges from 1970 to 2000.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Minimum	10’	0.726	0.675	0.615	0.533	0.515	0.533	0.789	0.759	0.719	0.639	0.643	0.656
TMP (°C)	5’	0.653	0.596	0.521	0.467	0.450	0.429	0.660	0.633	0.607	0.523	0.514	0.550
	2.5’	0.632	0.563	0.484	0.433	0.411	0.372	0.602	0.574	0.543	0.459	0.449	0.503
	0.5’	0.622	0.549	0.474	0.430	0.408	0.354	0.567	0.541	0.513	0.428	0.420	0.484
Mean	10’	0.450	0.481	0.470	0.482	0.487	0.478	0.455	0.445	0.427	0.425	0.425	0.427
TMP (°C)	5’	0.401	0.426	0.385	0.390	0.400	0.391	0.379	0.387	0.380	0.367	0.362	0.377
	2.5’	0.365	0.378	0.338	0.332	0.351	0.342	0.338	0.356	0.348	0.333	0.331	0.349
	0.5’	0.355	0.366	0.328	0.322	0.337	0.330	0.334	0.351	0.343	0.331	0.324	0.342
Maximum	10’	0.832	0.821	0.809	0.909	0.827	0.678	0.718	0.734	0.644	0.658	0.630	0.687
TMP (°C)	5’	0.727	0.711	0.666	0.760	0.687	0.560	0.645	0.658	0.568	0.561	0.511	0.576
	2.5’	0.664	0.637	0.591	0.670	0.597	0.485	0.589	0.600	0.531	0.509	0.447	0.517
	0.5’	0.631	0.596	0.544	0.611	0.544	0.445	0.574	0.578	0.516	0.484	0.405	0.479
PRE	10’	2.165	1.869	3.476	4.662	5.651	8.416	9.716	7.993	5.825	3.968	2.202	1.378
(mm)	5’	2.077	1.834	3.407	4.641	5.637	8.291	9.702	7.841	5.805	3.908	2.183	1.348
	2.5’	2.074	1.813	3.404	4.603	5.594	8.268	9.664	7.705	5.742	3.904	2.182	1.334
	0.5’	2.072	1.797	3.360	4.495	5.564	8.190	9.630	7.651	5.699	3.895	2.170	1.300

Table 2. The correlation coefficients between the observed and WorldClim climatology at different spatial resolutions over the 496 weather stations. The period ranges from 1970 to 2000.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Minimum	10’	0.987	0.984	0.977	0.969	0.963	0.962	0.955	0.957	0.956	0.971	0.984	0.987
TMP (°C)	5’	0.989	0.987	0.983	0.977	0.973	0.973	0.964	0.966	0.968	0.980	0.990	0.991
	2.5’	0.989	0.988	0.985	0.981	0.978	0.977	0.968	0.971	0.974	0.985	0.992	0.992
	0.5’	0.989	0.989	0.986	0.983	0.981	0.980	0.972	0.974	0.977	0.988	0.993	0.993
Mean	10’	0.986	0.979	0.968	0.955	0.949	0.949	0.956	0.958	0.966	0.974	0.982	0.987
TMP (°C)	5’	0.991	0.986	0.980	0.969	0.962	0.959	0.963	0.965	0.973	0.983	0.989	0.991
	2.5’	0.993	0.990	0.986	0.977	0.970	0.965	0.968	0.970	0.978	0.986	0.992	0.993
	0.5’	0.994	0.992	0.989	0.981	0.973	0.968	0.970	0.972	0.980	0.988	0.993	0.995
Maximum	10’	0.958	0.946	0.920	0.892	0.889	0.899	0.893	0.890	0.935	0.957	0.968	0.974
TMP (°C)	5’	0.969	0.961	0.946	0.921	0.912	0.912	0.898	0.896	0.939	0.965	0.978	0.982
	2.5’	0.976	0.971	0.960	0.941	0.930	0.925	0.910	0.909	0.945	0.971	0.984	0.986
	0.5’	0.979	0.976	0.968	0.951	0.940	0.932	0.913	0.912	0.946	0.973	0.988	0.989
PRE	10’	0.976	0.980	0.978	0.979	0.974	0.961	0.903	0.920	0.941	0.908	0.939	0.965
(mm)	5’	0.976	0.980	0.979	0.979	0.974	0.961	0.905	0.924	0.943	0.911	0.940	0.966
	2.5’	0.976	0.981	0.980	0.979	0.974	0.962	0.908	0.930	0.943	0.913	0.941	0.967
	0.5’	0.977	0.981	0.981	0.980	0.975	0.962	0.909	0.930	0.944	0.914	0.941	0.968

3. In addition, the "Direct evaluation" is not adequate. The time series are different for CRU, WorldClim, and observation. How do the authors guarantee the consistency of time series, in particular the period 1901-1950? Meanwhile, the mean climatology is calculated from 1970-2000. Is this time period appropriate for representation? For precipitation, the observation has shown significant nonstationary features after 1980s in China under the global warming. Unfortunately, Delta downscaling method does not consider the nonstationary.

Response: We will revised the "Direct evaluation" in the revision.

The Delta downscaling method involved the monthly reference climatology during a period. It is necessary to guarantee the accuracy representation of monthly reference climatology during this period. However, it is not necessary to guarantee the consistency of time series between WorldClim and other datasets. The details on the application of WorldClim data could be found in the responses on General comment #1. Now, we have evaluated the monthly WorldClim data for each climatic variables (above Tables 1-2), and the WorldClim data can represent the monthly climatology over China region for each climatic variable.

Indeed, the Delta downscaling method does not consider the "nonstationary", while it has the ability to enhance the spatial resolution and reliability for the 30' CRU time series. Thus, whether the precipitation over China has significant nonstationary features after 1980s mainly depends on the original CRU time series.

4. The authors evaluated the new data set using 745 observations over China. I think it is not enough, especially for the west of China, such as the high mountains areas and Tibet Plateau. Meanwhile, most observations begin after 1950, how about the pre-1950? Therefore, it is hard to conclude the data set is "sufficiently reliable".

Response: Many thanks for your queries. For the independent evaluation in this study, we will use 496 independent weather stations (above Figure 1) to evaluate the original CRU time series, WorldClim, and downscaled time series. Further, we have analyzed the representativeness of the 496 independent stations over China region (Figure 3). Figure 3 shows the orographic statistic information (e.g., elevation, slope, and aspect) of China and the stations. The results presented that the proportions of the weather station numbers in different orographic gradients almost correspond to those in China excepting the areas with elevations exceeding 4500 m, which indicated that these weather stations could represent the climate variation over China and be used for validating the downscaled dataset. This exception is inevitable, because of the observability, installation, and maintenance of the weather stations in those areas. We will revised the related contents in the revision.

Although China has some weather stations with data starting from 1901, all of them have been used to generate the CRU time series. We thus cannot verify the data quality before 1950 because of data availability. However, our downscaling procedure used data from CRU and WordClim and did not incorporate the observation, the data quality thus mainly depends on that of the data center. Our evaluation showed that the data quality of the data centers is overall satisfactory, and our downscaling procedure can further improve the data quality. We will revised the related contents in the revision.

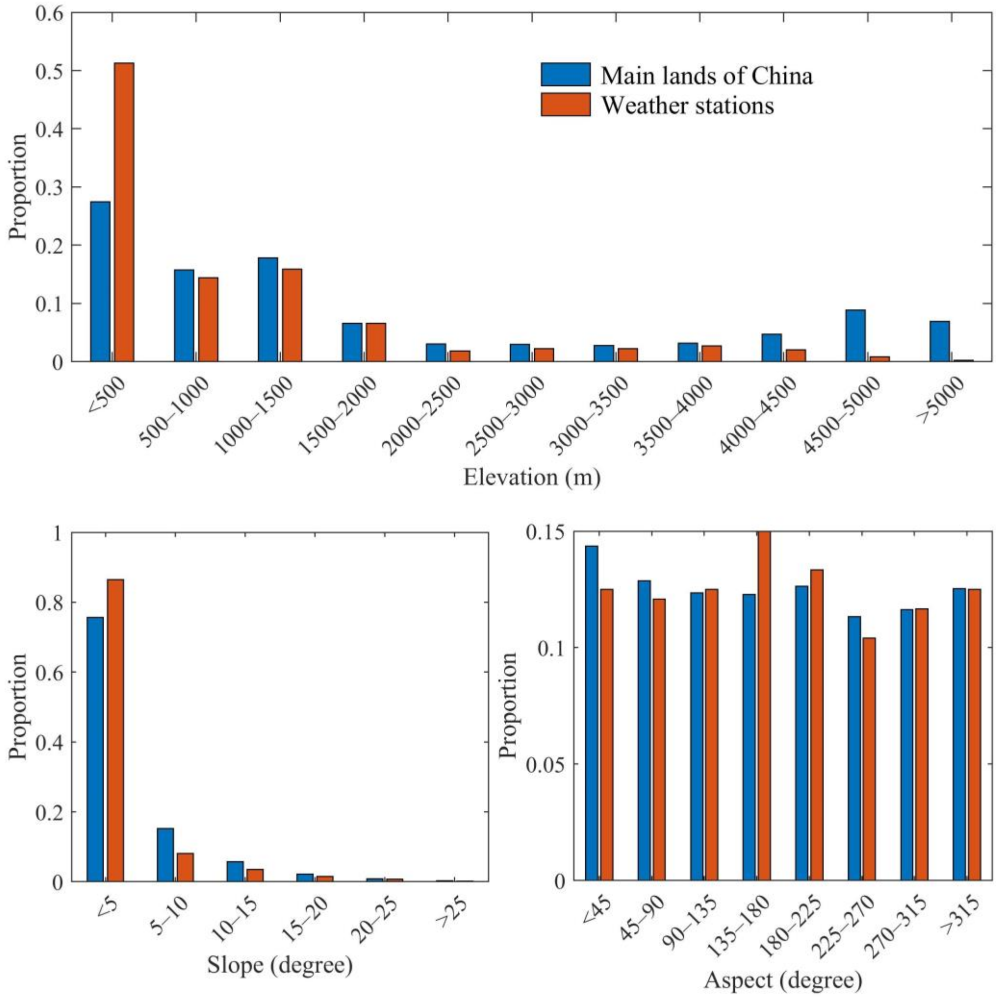

Figure 3. Orographic statistic information at different gradients for China and weather stations used in this study.

5. How many observations have been used in CRU and WorldClim? These sites should be excluded since they destroy the independence of evaluation.

Response: 323 weather stations across China region were employed by CRU group to generate the CRU time series data (Harris et al., 2014) (above Figure 1), and the WorldClim monthly datasets were generated using all of these stations obtained from the CRU group (Fick and Hijmans, 2017). For independent evaluation of downscaled dataset in this study, these weather stations will be excluded.

We appreciate you very much to point out the issues on the independent evaluation. Now, we have used the 496 independent weather stations to evaluate the original CRU time series (Table 3), WorldClim (above Tables 1-2), and downscaled time series (Table 3). The results in Table 3 indicated that downscaled datasets had better performance than the original CRU dataset, especially for the 0.5’ dataset. We will revised the related contents in the revision.

Table 3. Statistical characteristics between original/downscaled and observed monthly TMPs and PRE in the time series (1951–2016). The values shown here are the averaged evaluation results at all 496 weather stations.

	Res	MAE_c	MAE_l	MAE_n	$RMSE_c$	$RMSE_l$	$RMSE_n$	NSE_c	NSE_l	NSE_n	Cor_c	Cor_l	Cor_n
Minimum	30’	1.766			1.947			0.887			0.994		
TMP (°C)	10’	1.673	1.515	1.558	1.802	1.726	1.793	0.896	0.902	0.899	0.995	0.995	0.995
	5’	1.338	1.292	1.325	1.666	1.503	1.582	0.904	0.937	0.923	0.995	0.995	0.995
	2.5’	1.233	1.142	1.211	1.401	1.349	1.384	0.946	0.951	0.949	0.995	0.997	0.996
	0.5’	1.140	1.050	1.137	1.322	1.248	1.271	0.955	0.972	0.963	0.997	0.998	0.997
Mean	30’	1.598			1.759			0.888			0.996		
TMP (°C)	10’	1.277	1.140	1.188	1.433	1.293	1.358	0.899	0.914	0.904	0.997	0.997	0.997
	5’	1.117	0.980	1.003	1.222	1.133	1.197	0.926	0.950	0.933	0.997	0.997	0.997
	2.5’	0.977	0.836	0.859	1.157	0.988	0.993	0.966	0.976	0.973	0.997	0.998	0.997
	0.5’	0.826	0.820	0.822	0.974	0.969	0.970	0.977	0.981	0.980	0.998	0.998	0.998
Maximum	30’	2.034			2.206			0.800			0.995		
TMP (°C)	10’	1.800	1.672	1.755	2.044	1.886	1.968	0.811	0.832	0.824	0.995	0.996	0.996
	5’	1.649	1.487	1.548	1.864	1.700	1.756	0.843	0.856	0.850	0.996	0.996	0.996
	2.5’	1.455	1.310	1.387	1.666	1.523	1.632	0.875	0.909	0.887	0.996	0.997	0.996
	0.5’	1.296	1.282	1.291	1.511	1.491	1.500	0.909	0.910	0.910	0.997	0.997	0.997
PRE	30’	17.850			29.559			0.614			0.885		
(mm)	10’	16.884	16.647	16.741	28.022	27.559	27.946	0.675	0.735	0.700	0.887	0.890	0.890
	5’	16.134	15.223	15.942	26.222	25.185	25.888	0.764	0.791	0.773	0.892	0.900	0.894
	2.5’	14.867	14.024	14.557	24.374	23.191	23.867	0.791	0.792	0.791	0.914	0.920	0.919
	0.5’	13.772	13.269	13.443	22.655	21.941	22.213	0.794	0.808	0.802	0.920	0.929	0.926

Notes: Res indicates the spatial resolution. The subscripts *c*, *l*, and *n* indicate bicubic, bilinear, and nearest-neighbor interpolations, respectively. The original TMPs and PRE are the 30’ CRU data and directly compared with the observed data. Evaluations at 10’, 5’, 2.5’, and 0.5’ are the evaluations for the downscaled datasets. MAE, RMSE, NSE, and Cor indicate the mean absolute error, root-mean-square error, Nash–Sutcliffe efficiency coefficient, and correlation coefficient.

Specific comments:

1. Figure 1, the range of DEM from 0 to 8848 is wrong. The Turpan Basin is for sure below the mean sea level. What is the spatial resolution of DEM in this figure?

Response: We appreciate you very much to point out this issue. The spatial resolution of DEM in this figure is approximately 1 km. We will revised the Figure 1 in the revision.

2. Figure 2, it is hard to follow this downscaling framework. There is no legend for all figures, which is the mandatory element. The color scales should be the same for a better comparison.

Response: Many thanks for your concerns. We will revised the contents regard to the Delta downscaling framework as well as the schematic illustration. The detailed revision could be found in the responses of General comment #1.

3. More interpretation should be given for the Delta downscaling method. For example, how to calculate the “ratio” for PRE anomaly? Is there a simple mathematical formula?

Response: Many thanks for your suggestion. We will revised these issues in the revision. The detailed revision could be found in the responses of General comment #1.

4. Once again, "Direct evaluation" is not sufficient. More details about the bias or errors should be supplemented.

Response: Many thanks for your comment and suggestion. In the revision, we will revise the "Direct evaluation" and add more details about the bias based on the independent and representative stations. Especially, the spatial distributions of bias at the station scale of original and downscaled CRU time series will be mapped; further, in the geographic space, where have much improvement and what result in the improvement will be analyzed and discussed.