Anonymous Referee #2

This study by Peng et al. developed a high-resolution and long-term climate dataset over China. The CRU data was downscaled to 1km using the Delta downscaling framework. The topic is interesting, and the product would be useful in climate-related studies for the nation. However, I think the paper needs some improvement and further discussion before it can be published.

My major concerns include:

1. The Delta downscaling improves the spatial representation of temperature/ precipitation climatology using high-resolution WorldClim as the reference climatology. However, it is hard to understand how this downscaling method improves temporal variability (or trend), because the temporal change is simply based on the interpolated anomalies from low-resolution CRU. This limitation should be at least explained and discussed in the manuscript.

Response: Thank you for this point. The Delta downscaling not only improved the spatial resolution of CRU time series, but also corrected the bias of CRU time series. Table 1 in the original manuscript clearly showed the good performances on bias correction.

Now, we have analyzed the correlations between original/downscaled and observed data in the time series (Table 1), based on the 496 independent stations (Figure 1), which were not taken part in the creation of the CRU time series and WorldClim climatology. The results indicated that downscaled datasets have better performance than the original CRU dataset, especially for the 0.5' dataset. Specifically, the bias has been improved very much, and the temporal variability has been slightly improved. These imply that the Delta downscaling used in this study can improve the temporal variability of original CRU time series. It should be attributed to the introduction of monthly WorldClim climatology for each climatic variable.

I understand your query very much. If one climatology was used as the reference climatology for the downscaling, the temporal variability of downscaled dataset is the same as the original dataset. However, this study employed 12 climatology layers, representing the climatology from Jan to Dec, to downscale the CRU data. Indeed, the description of downscaling method in the original manuscript is ambiguous, and we will revised this part in the revision.

Table 1. 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.

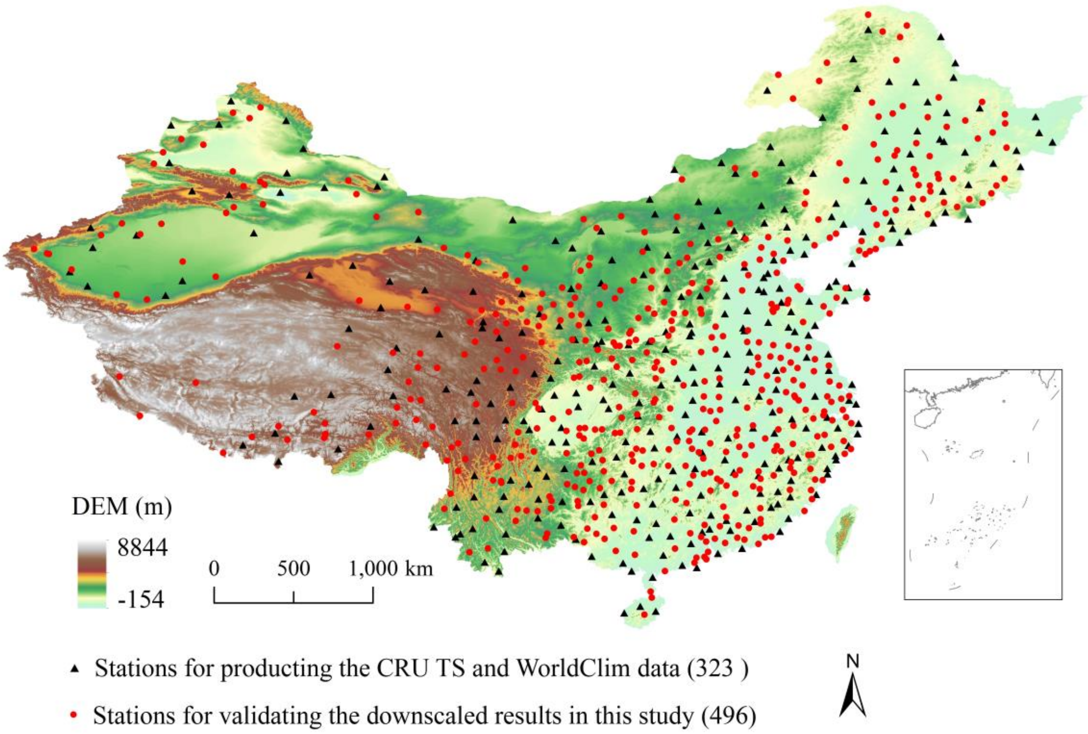

Figure 1. Spatial distribution of the national weather stations across China.

2. The downscaled dataset is developed based on the WorldClim reference climatology from 1970 to 2000, and the data evaluation is performed for the period 1951-2016. Can the authors first evaluate the reference data (WorldClim) at different resolutions? Also, because there is an overlapping period for data training and evaluation, is it possible to use two separate periods, in which one is for downscaling and the other one is for data evaluation?

Response: Thanks for the suggestion. Your proposal of two separate periods (one period for downscaling and the other period for evaluation) was substituted by another method.

Some of the weather stations used in our original manuscript have been involved in the data assimilation of CRU and WorldClim data. They were also used for the data evaluation. Now, we kicked out those stations and evaluated the WorldClim, CRU and downscaled data only with 496 stations (above Figure 1). The modified procedure can improve the reliability of the evaluation (Tables 1-3). Overall, the WorldClim datasets have high performance to represent the monthly climatology over China region, and the dataset performs better for higher spatial resolution. In specific, the absolute errors become smaller with increasing spatial resolution (Table 2) and the correlations get greater with increasing spatial resolution (Table 3).

Table 2. 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 3. 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. The authors need to discuss the possible reasons why CRU temperatures have systematic cold biases.

Response: The CRU group introduced the averaged 30’ DEM to generate the CRU data, which weakens the representation of temperatures in the actual land surface, especially in the regions with complex terrain. Moreover, the evaluation for the original CRU data was carried out at the station scale. For instance, it may present a warm bias if the averaged grid elevation involved by the CRU data is lower than the station elevation, while a cold bias if the averaged grid elevation involved by the CRU data is greater than the station elevation. The weather stations used in this study belong to the national weather station, which were often established in the valley near to the county or city. Thus, it is very likely that the averaged grid elevation involved by the CRU data is greater than the station elevation at most of the weather station, and presenting the “cold bias”.

In the revision, we will enhance the related discussion. In the Data section, we will add additional information to describe how CRU generated their data as follows.

Methodologies used by CRU group to construct 30’ time series dataset are similar to the Delta downscaling framework employed herein. 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) (above Figure 1).

4. The dataset covers from 1901 to 2017, but most of the evaluations and discussion are about post-1950. Data quality or uncertainties before 1950 need more discussion.

Response: 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.

Specific Comments:

1. P3, L10-L15: Can we get the information about how many stations in China were used for CRU TS and WorldClim? How different are they? Are they comparable to the 745 weather stations used in this study?

Response: There have several overlapped weather stations for the creations of CRU and WorldClim data as well as the evaluation in the original manuscript. 323 weather stations across China region (above Figure 1) were employed by CRU group to generate the CRU time series data (Harris et al., 2014), and the WorldClim monthly datasets were generated using all of these stations obtained from the CRU group (Fick and Hijmans, 2017). For the independent evaluations in this study, we have used the 496 independent weather stations to evaluate the original CRU time series, WorldClim, and downscaled time series (above Figure 1 and Tables 1-3). We will revised the related contents in the revision.

Further, we have analyzed the representativeness of the 496 independent stations over China region (Figure 2). Figure 2 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.

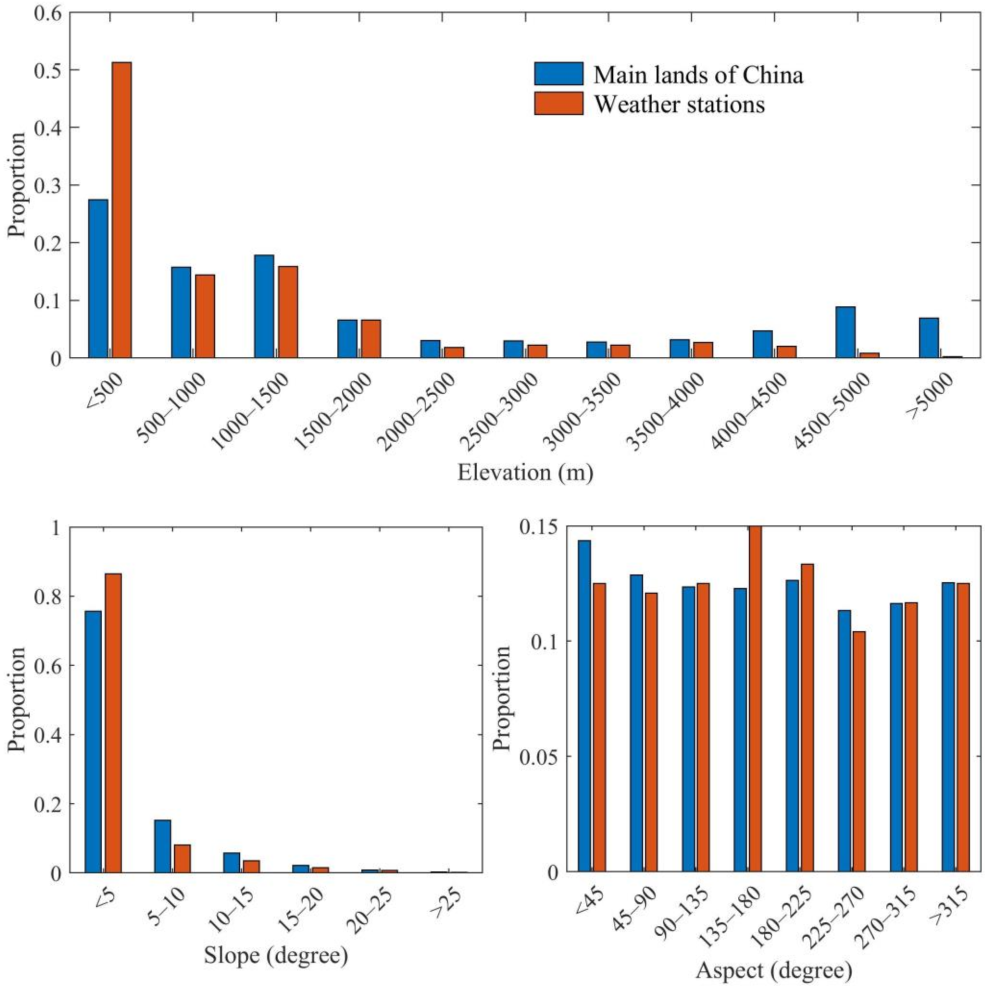

Figure 2. Orographic statistic information at different gradients for China and weather stations used in this study.

2. P4, L17-19: I assume the final product is generated using the bilinear interpolation method? This should be mentioned in conclusion and abstract as well.

Response: Yes. The bilinear method was finally used to interpolation. We will add it in the revision.

3. P5, section 4.3: Trend is one aspect of the temporal variations. It would be better to also calculate the correlation of the time series.

Response: Thank you for this suggestion. We calculated the correlation and presented the results in Table 1.

4. Table 1: These metrics are applied to the climatology of TMP and PRE for 1951-2016? Or applied to time series of monthly TMP and PRE, then averaged over the 745 stations? Or any other way? This should be clarified in the main text or table caption. Same issue for Table S1.

Response: These metrics are applied to time series of monthly TMP and PRE, then averaged over the 745 stations. Based on the 496 independent stations, we have calculated the results, as well as the correlation coefficient of time series (above Table 1). We will revise the related contents as your suggestions.

5. Figure 4: Because the climatology is "corrected" using high-resolution reference data, it is not surprising that the downscaled data outperformed the CRU data in terms of the climatology. As suggested in Comment #3, it would be better to have a similar figure to show the time series (or anomalies).

Response: We will revise the related contents as your suggestions, based on the independent stations.

6. Figures 6-9: These figures are not quite informative. It would be better to add the trends as text on the figures. For figure 9, it is really hard to distinguish those three lines.

Response: We will revise the related contents as your suggestions, based on the independent stations.