Response to Reviewer 2

We would like to thank the anonymous referee for reviewing our manuscript. These constructive comments are very important for us to improve the present manuscript. In the following, we address all comments point-by-point according to the comments. All revisions are highlighted in the context.

The authors present a high-resolution (1km, 6h) air temperature data set for the Chinese Tianshan Mountains from 1979 to 2016 based on a downscaling method. This topic is quite interesting and the data set would be useful for the potential end-users who focus on the alpine climate and cryosphere issues in the Tianshan Mountains. In general, this paper is well-written for most parts. However, a major revision is needed before it is published in ESSD.

General Comments:

1. The mentioned downscaling method in the paper has been validated in the Alps Mountains and the Tibet Plateau. However, the relative references (Gao et al., 2012, 2017) named this method as "elevation correction" rather than "downscaling". What is the difference between these two methods or terms? For me, they are the same. Thus, which one is more appropriate?

-Answer: Thanks a lot for the comments. In this study we used the term of "downscaling" to emphasize the spatial resolution enhances from 0.25° to 1 km grid. But the reviewer is right that the correction has the same meaning with downscaling in this study. The elevation correction can be considered a form of downscaling. The reviewer 1 (Dr. Gerlitz) also pointed out this issue. He thought that correction is more intuitive because we only used the single factor (elevation) rather than local-scale processes. We revised this term in the revision version.

2. 24 meteorological stations are not enough for validation for such a large

area (more than 80000 points). Is there any other data resources could be used for further validation?

-Answer: Thanks a lot for the comments. Validation is the necessary process for a new data set. Unfortunately, the observation sites in the CTM are really sparse. This is the motivation we want to product a high-resolution spatial continuous and long-term data set for the CTM. There are around 760 meteorological sites with good quality (daily, more than 40 years) can be used over the whole China. Among them around 35 sites in Xinjiang Province with long-term records can be used. For the CTM, only 24 sites are available for validation. This is the best data resource we have. For sure, there are some global grid data sets which cover the CTM. However, most of them are produced by interpolated or modeled (e.g. CRUTEM3, E-OBS). The remote sensing has shorter time series and large bias. We believe that the observation from the meteorological stations is the best data set for validation. It is unscientific to validate the new data set using the data that contains bias. Although, 24 sites are not enough for validation, these sites can generally test the quality of the new data set. We welcome the potential users to evaluate the new data set based on different data resources that we do not have. More assessments can help us improve the product.

3. The authors pointed out that about 24% of RMSE was reduced by the downscaling method compared to the original ERA-Interim. Is it good enough? How to evaluate the data set (or any reference/standard) is good enough for end-users?

-Answer: Thanks a lot for the comments. To be frank, this is a difficult question to answer. Normally, there is no standard to measure the significance of error reduction. However, we thought that 24% of RMSE (around 1°C) is quite good for the CTM. For mountainous, the original ERA-Interim always has a large bias more than 3°C, for example 3.7°C in the Tibetan Plateau (Gao et al., 2014). Gao et al. (2017) used the similar method reduced 62% of RMSE for ERA-Interim (from 4.31 to 1.64°C) in the Tibetan Plateau. But in that study, 80 meteorological stations are used for validation. This is another potential cause for the "insignificant" bias reduction in the present study. Thus, we need more observations and other data resources as well as other users to evaluate whether the new data is good enough.

4. For my understanding, the downscaling method is mainly based on the elevation (DEM). Is it possible to get higher resolution data set if we use the 100 m DEM? The ERA-Interim product provides 3-hourly forecast data. Thus, is it possible to obtain 3-hourly data set for whole Tianshan Mountains?

-Answer: Thanks a lot for the comments. The reviewer raised a very interesting question. In fact, at the very beginning of this study, we planned to produce a 100m 3-hourly temperature data set for the CTM. However, we met two challenges: 1) China is not the member country of ECMWF, which means we cannot gain the 3-hourly forecast data in a direct way. We only can download the 6-hourly analysis data from the public data set archive; 2) 100 m resolution means the total number of grid reaches more than 8 millions. For an ordinary computer even a computer workstation, it is too large to process. For the end-users, such a large amount of data is not convenient. We plan to produce 100m 3-hourly data sets in the future but only for selected area such as basins or valleys that the end-user interested in.

5. How to evaluate the lapse rate is correct or appropriate for the downscaling? The lapse rate varies significant in different topographical situations and time period.

-Answer: Thanks a lot for the comments. This is an important issue. The lapse rate has a large spatial-temporal variability in the mountain areas. The reviewer 1 also pointed this question out. We added the evaluation of the lapse rates between ERA-Interim and observations in the section 4.2 in the revision (Figure 1). Figure 1 shows the temporal variation of monthly lapse rates for observation and ERA-Interim (Γ_{700_925}) over the 24 sites. In general, the ERA-Interim has a higher temperature gradient than observation. However, ERA-Interim captures the variability of observed lapse rate very well, especially in the warmer months (May to August). The inter-monthly variability of observed lapse rate is much higher than ERA-Interim, especially from September to January. Table 1 shows the monthly lapse rates over all sites in 1979-2013. The lapse rate differences are small (less than 0.5 °C km^{-1}) from May to August, while the differences are larger than 1 °C km^{-1} from September to December and January. More details please see the section 4.2 in the revision (**P8 L29-31, P9 L1-20**).

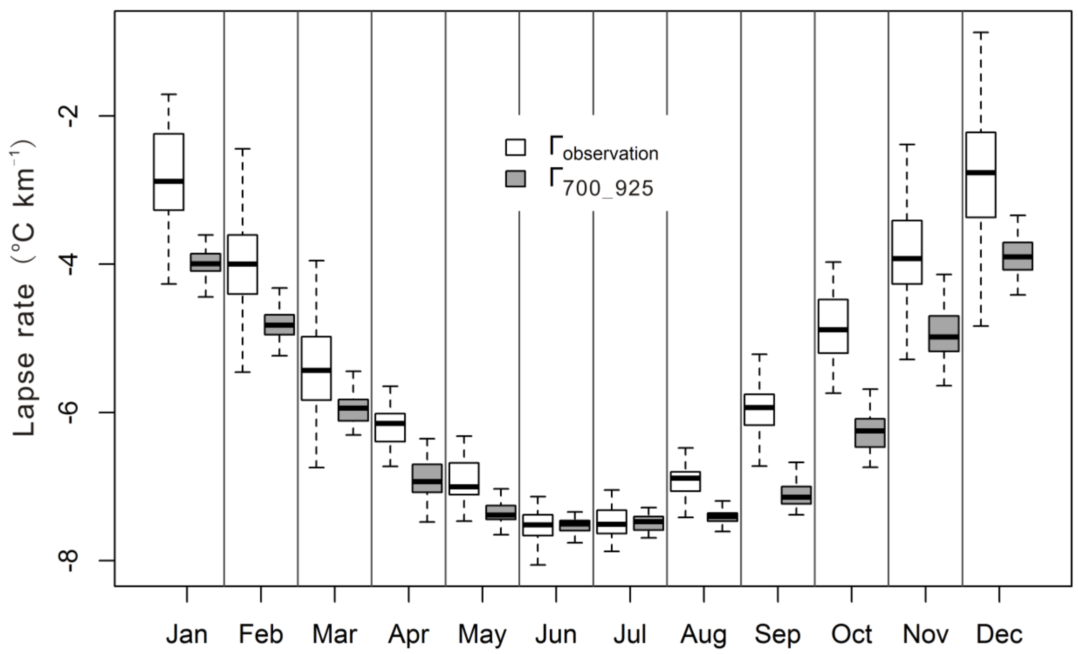

Figure 1 Boxplots of monthly lapse rates for observation and ERA-Interim (Γ_{700_925}). Thick horizontal linesin boxes show the median values. Boxes indicate the inner-quantile range (25% to 75 %) and the whiskers show the full range of the values.

Table 1 Monthly lapse rate (°C km^{-1}) over all sites in 1979-2013.

Month	observation	Γ_{700_925}
January	-2.79	-4.00
February	-4.01	-4.81
March	-5.42	-5.96
April	-6.14	-6.90
May	-6.92	-7.35
June	-7.55	-7.52
July	-7.48	-7.49
August	-6.95	-7.40
September	-5.93	-7.10
October	-4.86	-6.27
November	-3.94	-4.95
December	-2.88	-3.88

6. Precipitation is another basic and important variable for climate and environmental models. Can you produce any high resolution precipitation data set using some similar methods for this region?

-Answer: Thanks a lot for the comments. This is another interesting issue. Hundreds methods/models for precipitation downscaling were developed in the past two decades. However, at present, there is no universal method. Meanwhile, precipitation is more complex and difficult to downscale for a finer grid, especially independent of observations. To our knowledge, PRISM (Precipitation-elevation Regression on Independent Slope Model) is one of the possible methods to downscale the coarse grids to a finer grid. PRISM takes the elevation and other microtopographic factors (slope, aspect) into account. Until now, we did not test this method in the CTM. But it is our future research plan. Surely, PRISM should be adjusted and further developed for the CTM because the snow is the main form of precipitation in winter.

Specific comments:

1. Although the authors listed many references about the downscaling method, I believe it is necessary to clarify the method specific for the readers

who are not familiar with downscaling method.

-Answer: Thanks for pointing this out. We added more information on the correction methods in the section 3.3. We also gave an example to calculate the lapse rate (**P6 L2-11**).

2. The downscaling method is more appropriate named "elevation correction". Since only elevation is involved. The conventional circulation variables such as wind, sea level pressure, humidity are not considered in the downscaling method.

-Answer: Thanks for pointing this out. Yes, we agree that elevation correction is more appropriate for this study. We have changed it for the whole context in the revision.

3. The data set is not friendly to download and use. The data set is divided into so many sub-files. Is it possible to find a more easy way for users?

-Answer: Thanks. Other reviewers also pointed this problem out. We also realize that the data set is not very friendly. We have tried many ways to make it easier for end-user. For example, we put all points together for a single year in a signal NetCDF file, but it was more than 5 GB. A normal desktop cannot read it. The Matlab (we process the data in Matlab) always says out of memory. Thus, we prefer to provide the small parts. The advantage is that the potential users can download the data points according to the coordinates of study area, rather than download the whole data points. We are working on the version 2.0, which is friendlier to users. The accessibility of data set also will be improved in the version 2.0 (**P15 L12-18**).

4. The downscaled data at some sites are worse than the original ERA-Interim data. Why? The authors should discuss this issue. It is very important because only 24 sites are available for validation.

-Answer: Thanks a lot for the comments. It is true that the elevation correction method did not work very well in some sites. We thought that the micro-topographical features such as aspect and slope of the mountain are the main reason. For example, for station No. 16 (Baicheng) located in the valley, the lapse rate might changes sharply even inverse in the cold winter. We added some discussion in the revision (**P14 L19-23, 28-29**).

5. If someone plans to run a hydrology model in a small catchment in the Tianshan Mountains. How to adjust the data set points to match the model grids?

-Answer: Thanks a lot for the comments. Firstly, the end-user should confirm the boundary (coordinates of the four directions) of the basin. Secondly, confirm the center points of hydrological model grids, for example 3 km grid. Each grid has a center point with the coordinate. Third, download all the grids (818126) information (not the data, only a small txt file) from the data set. This file contains the coordinates of all points and their elevation. Last, the data point can be selected by calculating the distance (define a threshold according to the hydrology model, for example 3 km) from the data point to the model point. There is a more intuitive way using ArcMap software. Firstly, create two grids layers in ArcMap, and then overlay or interact two layers. The overlap part is the matching points for data points and model points. Then, export the coordinates of matching grids from ArcMap. According to the ID of data points, download the relative data from PANGAEA for the time series the users want.

6. I found that the data amount is around 187G. How to process such large data set? What is software or platform to process it? Maybe, the authors could provide some codes for data processing.

-Answer: Thanks for the question. We have spent much time on the data processing and calculation. Matlab is the tool we used. Because of the large data amount, we used the VSC-3 super computer system in Vienna of Austria. It took almost one month to finish the calculation. But now, the end-users can process and analyze data on the laptop or desktop because we divided the grids and time series into smaller ones (the amount of single file is less than 2 GB). The Matlab code is simple to follow. We could provide it if some end-users need it.

7. I am not sure the data set could capture the temperature changes in the micro-topography since the original data is 0.25 degree. The slope and aspect of mountains also affect the temperature significant, especially in the night.

-Answer: Thanks a lot for the comments. The reviewer raised a very important issue. The original spatial resolution of ERA-Interim is reduced Gaussian grid N128 (around 0.75 degree). In general, ERA-Interim has a relative small large-scale bias. However, we do not think it can capture the temperature changes in the micro-topography. Other reviewers also mentioned this question. We added more analysis on the maximum temperature, minimum temperature and diurnal temperature ranges as well as the warming trends. This analysis could generally illustrate the ability of the new data set on the temperature changes. In the future, we plan to adjust the temperature according to the micro-topography features, especially for local scales such as basins. We added some discussion on this issue in the revision (**Section 4.5 P12, P13 L1-18**).

8. Can other temperature downscaling methods be used for the high-resolution data set? And why?

-Answer: Thanks a lot for the comments. This is an interesting open question. The downscaling based on the lapse rate is the most common used method for temperature. However, most cases used the fixed lapse rate (-6.0 or -6.5 °C km^{-1}) or monthly lapse rate from Kunkel (1989). Fewer studies focused on the lapse rate variability. We do not think the conventional method could be used for high-resolution data set, especially for the high mountain areas. These methods rely on a high density of observations. However, the sites are very sparse in the high mountain areas. The present method is independent of observations, which could be extended to any other high mountain areas over the world.

9. Some expression and description of language is not clear. A native speaker would be helpful for the improvement of readability for the whole context.

-Answer: Thanks a lot for the comments. It is true that there are some language problems. The English expression in the revision is corrected by the Elsevier publishing group (https://webshop.elsevier.com/)

References:

Gao, L., Hao, L., and Chen, X. W.: Evaluation of ERA-Interim monthly temperature data over the Tibetan Plateau, J. Mount. Sci., 11(5), 1154-1168, doi : 10.1007/s11629-014-3013-5, 2014.

Gao, L., Bernhardt, M., Schulz, K., and Chen, X. W.: Elevation correction of ERA-Interim temperature data in the Tibetan Plateau, Int. J. Climatol., 37, 3540–3552, doi: 10.1002/joc.4935, 2017.

Kunkel, E. K.: Simple Procedures for Extrapolation of Humidity Variables in the Mountainous Western United States, J. Climate, 2, 656–669, 1989.