Replies to Reviewer 3

Thanks very much for anonymous Referee #2 to make comments. We will make corresponding changes in our manuscript according to these comments.

Comment 1 (General comments):

Comment 1.1

"The paper attempts on quantifying monotonic trends in "heavy snow" indices in the northern region of China, Xinjiang, and generate spatial fields of results by using a standard interpolation method "universal krigging". While the purpose of the research could be a-priori interesting for the reader of Natural Hazards and Earth System Science, the outcome is certainly disappointing, and it does not reach the standards for an international journal such as NHESS. My recomendation is to reject the manuscript; however I encourage the authors to re-think about their research, and how they could make it suitable for an international publication. For this I specify the main drawbacks of the manuscript, as well as some of the (countless) technical errors that I found."

Authors' answer:

Thanks for anonymous Referee #2 to make comments. We will further improve our paper.

Comment 2 (Specific comments):

Comment 2.1 Presentation/quality:

Comment 2.1.1:

"The first impression when one reads the paper is confirming (as Referee #1 pointed out) the poor language/English that it contains, that even non-English speakers as myself can find countless grammatical, wording, and spelling errors along the manuscript. Sentences such: "Actually, there is a bunch of researches devoted to the trends of snowstorms over a variety of regions" or, "Mentioned above make it easy to identify snow days and their snowfall amounts. What needs to be mentioned is that the precipitation is usually measured by rain gauge in China. In case of snowfall, the snow amount captured by the rain gauge is taken from observation site to room. When the snow in rain gauge melted at room temperature, the melted-water amount is measured in unit millimeter (mm)", are clear examples of very bad use of English, that alone could be a reason for rejecting the manuscript."

Authors' answer:

We will improve the linguistic expression in the revised manuscript.

Comment 2.1.2:

"The figures are as well of very poor quality, lacking any kind of edition (such as Fig.

2), or entirely uninformative (such as Fig. 4).”

Authors' answer:

Thanks for the advice of anonymous Referee #2. Figure 2's resolution is 600 dpi, and its format is EPS, so its quality is not bad. Figure 4 showed the hot spots and cold spots of ESE variation in northern Xinjiang. The capital of Figure 4 is concise but not uninformative. The meanings of hot spots and cold spots are explained in the main text, so we will not show it in the figure capital.

Comment 2.1.2:

“Some paragraphs, especially those in the methodology section, are dispensable. For example, the statistical explanation of the Mann-Kendall tests is given in thousands of previous papers, and more important, in the original references, so there is no need to write it down, maybe with some references it would be enough. Also the whole explanation of the universal krigging interpolation method seems to me disproportioned, especially when the objective of the paper is not comparing interpolation methods, or measuring the goodness and suitability of the used one. Moreover, as I argue in the next paragraph, the spatial interpolation of results in this study seems to me inadequate, given the small amount of cases (stations) for such a wide and topographically complex study area.”

Authors' answer:

Thanks for the advice of anonymous Referee #2. We will delete some contents, in the methodology section, about the Mann-Kendall test and kriging interpolation method. And the results of spatial interpolation will be explained in the answers to Comment 2.2.1.

Comment 2.2 Lack of representativeness/lack of knowledge of climatological basics:

Comment 2.2.1:

“In the introduction, the authors make a quick description of the study area, but they forget to talk about the extension and other geographic features, which are essential for the further interpretation of results and suitability of methodology. At a guess I’d say that the area of the region is about 0.5 million Km2, which is about the size of countries such as Spain or Sweden. Two mountain chains are observed in the northeast and southwest of the region, with a wide basin in between. Nothing is said about the well-known gradients of precipitation (orographic precipitation) and temperature (adiabatic lapse rate) of mountain areas, which accumulate much larger amounts of snow than the plains. I’m afraid that 18 pluviometric stations are not representative of such a wide and diverse area, and this invalidates any spatial interpolation to be made (none can imagine an interpolation of snow indices in, let’s say, Spain with only 18 meteorological stations). None of the 18 stations are located in the mountains, only a few are in the foothills, and the majority are located in the inner basin, therefore the tendency surfaces show in Fig 3 are totally arbitrary. There are more suitable interpolation methods when orography is present such as the co-krigging or the multiple regression method (Ninyerola et al, 2000), that enable

introducing the elevation as a co-variable, thus obtaining more reliable results."

Authors' answer:

The main idea of anonymous Referee #2 is that the stations in northern Xinjiang are sparse and most of them distribute in the plain, and this invalidates any spatial interpolation to be made. Additionally, anonymous Referee #2 suggests that co-kriging or the multiple regression method is more suitable when orography is present. Our answer is as follows.

The trends of ESE indices for every station have been studied in detail (Table 6) and we describe the trends exhaustively in language. So we don't think it's necessary to show the trends of every ESE indices of 18 selected stations in a spatial figure. We tend to show the rough spatial distribution for the trends of every ESE index over the entire northern Xinjiang using interpolation method. Because the spatial distributions of trends of ESE indices over northern Xinjiang can be observed more intuitively, and it can help us bring to light the large-scale causes that lead to the increasing ESEs, though the spatial interpolation certainly has the problem of the accuracy.

In this paper the method of spatial interpolation used is universal kriging. We select this method because of the following reasons.

Firstly, universal kriging method can fit the dominant trends in space well and can test the accuracy of the interpolation in the way of cross-validation (the detail is showed in this paper). It is the optimal unbiased estimator. By the data explore tools of geostatistical, we found that although the topography of northern Xinjiang is a little complicate, there are dominant trends in space for the trends of the ESE indices. So universal kriging can be used to interpolate and can well fit the trends which are influenced by topography.

Secondly, the results showed in this paper about the spatial distribution of the trends of the ESE indices are consistent with previous studies, such as Zhao et al. (2010) and Wang et al. (2012). So the results of the spatial interpolation can be acceptable.

In summary, in this paper, spatial interpolation by using universal kriging is satisfactory. Because the precipitation in northern Xinjiang is related to a lot of factors such as moisture sources, slope aspect, elevation, and so on, the co-kriging or multiple regression method (when orography is presented) may not produce more reliable interpolation results than universal kriging method.

Comment 2.3 Mislead of scientific concepts:

Comment 2.3.1:

"From the scientific perspective there are as well various inaccuracies that the authors should take into consideration for any further investigation. For example, they use the term "trend" inadequately, as it only should be used when the Mann-Kendall coefficients are statistically significant. Example: Page 7069 line 4: "Upward trends were observed at 17 out of 18 stations, while only one station exhibited a downward trend (Table 4). The highest upward trend occurred at Urumqi station, and the downward trend occurred at Altay station (Table 6). MK significance test for the trends in the time series of the SX1day showed that 10 out of 18 stations

had significant upward trends (at p <0.05), accounting for 55.6% of the total stations (Table 5)."

If only 10 out of 18 stations showed significant trends, they shouldn't say "upward trends were observed at 17 out of 18 stations". They should rather use the term coefficient, i.e.: positive coefficients were observed in 17 out 18, 10 of which were statistically significant, indicating upward trends…"

Authors' answer: We will do the corresponding modifications according to the advice of anonymous Referee #2.

Comment 2.3.2:

"The authors use the non-parametric Mann-Kendall test to search for the significance of trends, and they justify it because you don't have to assume any distribution of data, and because it is not influenced by outliers; however they use the parametric Pearson test to compute for the magnitude of trends, and this test it's influenced both by the presence of outliers and by the distribution of the sample. The use of both tests is contradictory, and the results of table 6 are confusing. What does the bold and italics refer to, p-level of the Mann-Kendall test or p-level of the Pearson test? If the amount of change is given by the Pearson test, it is erroneous to use the p-level of MK test, as both tests have different sensitivity. The Thiel-Sen slope estimator (Yue et al., 2002) should be used instead of the Pearson test, as it complements the MK test, giving the value of the slope and thus the magnitude of any existent trend.

Authors' answer:

Thanks for the advice of anonymous Referee #2. We have used the Thiel-Sen slope estimator instead of the Pearson test, to give the value of the slope and thus the magnitude of any existent trend. We will show the modified Table 6 in the revised paper.

Table 6. Slopes of 5 extreme indices in MK test for each station. * and ** respectively represent the slopes which are significant at the 0.05 level and the 0.01 level.

		Slopes				
No.	Stations	SX1day (mm/10a)	SX1process (mm/10a)	DSb (day/10a)	DSc (day/10a)	PSb (times/10a)
1	Alashankou	0.061*	0.110**	0	0.085	0.047
2	Altay	-0.128	-0.143	-0.029	0.172*	-0.057
3	Beitashan	0.125	0.045	4.95E-08**	-0.040	0.084
4	Caijiahu	0.046	0.158*	-12.953**	0.309**	0.081
5	Fuhai	-0.008	0.064	-5.11E-09	-0.041	0.057
6	Habahe	0.018	-0.003	-0.024	0.195*	0.032
7	Hoboksar	0.088	0.053	7.35E-08**	0.087	0.032

8	Jinghe	0.028	0.041	8.59E-08**	-0.013	0.006
9	Karamay	-0.011	-0.072	-0.031	0.191*	0.003
10	Qinghe	0.016	-0.055	0	-0.191**	0.014
11	Qitai	0.005	0.082	0	0.082	-0.035
12	Shihezi	0.087	0.094	0.048	0.207*	0.088
13	Tacheng	0.194*	0.262**	0.089**	0.006	0.108
14	Tory	-0.059	-0.017	0.025*	-0.104	-0.031
15	Urumqi	0.170*	0.153*	0.005	0.105	0.063
16	Wenquan	0.286**	0.197*	0.052*	0.021	0.080
17	Wusu	0	0.099	0	0.041	0.027
18	Yining	0.153*	0.048	0.070*	0.080	0.191*

Comment 2.3.3:

"Moreover, data series should be checked for autocorrelation (Yue et al., 2002). The existence of autocorrelation can lead to erroneous rejection of the null hypothesis in the MK test, thus removing any serial correlation is highly advisable prior to run MK test."

Authors' answer:

This advice has been improved by Reviewer 2, and we have answered this question in detail in the replies to Reviewer 2 (AC C2807).

Comment 2.3.4:

"I have serious concerns as well about the suitability of the data used for this kind of analysis. Firstly, the authors are assuming (should the readers assume it too?) that any precipitation recorded during the December-February period is below 0_C. This should be demonstrated in the manuscript. Secondly, the authors indicate that the daily precipitation types were discriminated, but then, for the calculation of the indices, we don't know if they are using daily precipitation, or daily amounts of snow… Thirdly, as indicated by Referee #2, there is no explanation on how the indices were calculated. This whole methodology part is full of assumptions and rather obscure and needs a lot more detail to be reliable. Finally, the authors use the concept of "extreme" without a real comprehension of its meaning. From a statistical perspective extreme refer to unusual, or very little frequent (in the extremes, or tails of the distribution), and from a meteorological perspective it includes as well an exceptional magnitude of the event. As we don't know how the indices were calculated we cannot really appreciate if they refer to extreme events."

Authors' answer:

Anonymous Referee #2 asked four questions here. We will answer these questions one by one.

Firstly, the winter in northern Xinjiang is from November to next March. In order to make sure the form of precipitation is snow, we only selected the daily precipitation data from December to next February, in this period, all the daily average temperature of each selected station were below 0 ℃. As we could not list all the records in recent 49 years (from 1959/1960 to 2008/2009) of daily temperature for all selected station, so we just made a linguistic explain to the situation.

Secondly, for the calculation of the indices, we used daily precipitation (=snow water equivalent) but daily amounts of snow.

Thirdly, Table 3 shows the definitions of the ESE indices and it is easy to understand. In order to show how the indices were calculated we just add a sentence after the last sentence of the section 2.2, and the sentence is "These indices during recent 49 years (1959/1960-2008/2009) were calculated for each selected station in northern Xinjiang, China, at yearly level."

Finally, the indices defined in this paper (Table 3) are exactly extreme ones. These ESE indices are similar with the extreme precipitation indices defined by ETCCDI (Expert Team on Climate Change Detection and Indices). The extreme precipitation indices defined by ETCCDI are well introduced in RClimDex User's Guide, on Page 19-20 (http://etccdi.pacificclimate.org/software.shtml).

Comment 2.3.5:

"The terminology used by authors is as well confusing. What does "snowfall" refer to? Is it amount of precipitation? Once again, how do we know that all this precipitation was in the form of snow? They should rather use the concept of Snow Water Equivalent (SWE). The values shown in table 2 are hardly representative of an extreme amount of snow. Is really 10 mm of SWE in 24h considered a heavy snowfall in China?"

Ninyerola, M., Pons, X. and Roure, J. M. (2000), A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. Int. J. Climatol., 20: 1823–1841. doi: 10.1002/1097-0088(20001130)20:14<1823::AIDJOC566> 3.0.CO;2-B

Yue, S., P. Pilon, B. Phinney, and G. Cavadias. 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes 16:1807-1829."

Authors' answer:

Thanks for anonymous Referee #2. This question was answered in detail in the replies to the Review 1 (AC C2220).