Reply on comments from Anonymous Referee #1

This paper provides an interesting overview of station-based soil freeze/thaw processes in China over recent 51 years. As such, it provides useful information, especially because this information is difficulty to obtain observationally. Overall it is a good paper, and below I provide some suggestions which may improve and strengthen the paper.

Response: We appreciate the referee's insightful and constructive comments on the manuscript. All comments are very helpful for improving the manuscript. We have studied all comments thoroughly and made necessary changes and corrections.

It isn't stated anywhere what type of data actually form the basis for this entire analysis. The paper only says "station data" of "near-surface freeze/thaw," and also refers to "ground-surface temperature" (GST). The "CMA, 2007" data citation is not in the reference list. If they are GST, measured in the uppermost layer (centimeters) of the ground, what types of sensors did they come from? What is the quality of these data? Were there sensor changes over 1956-2006? Were the data homogenized to remove certain (potential) artifacts such as station moves, location biases, etc.? Please comment on these issues.

Response: We added two paragraphs in text to clarify these issues (insert into ***section 2, p3790, line 16***):

"Ground surface temperatures were measured by using a thermometer. The thermometer sensor has mercury ball on one end with diameter of 5 mm. It is required by the measurement standard that half of the thermometer sensor be buried in ground and the other half expose to the air. In practice, the sensor is usually buried more than half in the ground and it is colored in white to reduce solar heating. Daily minimum (maximum) temperatures were measured using a special minimum (maximum) temperature thermometer. The minimum (maximum) temperature thermometer records the daily minimum (maximum) temperature once a day although it cannot record the time when it occurs. Ground surface temperatures were also measured four times a day (02:00, 08:00, 14:00, and 20:00 Beijing Standard Time) and averaged as a daily mean. Daily minimum (maximum) temperature was reported at 20:00 Beijing Standard Time. The thermometer has an accuracy of ±0.1 ℃ and by requirements, these thermometers should be calibrated once a year. The thermometer sensors were used for the entire study period. The large majority of the stations have no location change over period of the records. However, information is not available for those stations with location change history. We believe that effect of station movement on overall outcome is very minimum. All of these measurements were conducted routinely each day by trained professional technicians at all meteorological station across China.

The daily surface temperature dataset was performed basic quality control that identified and excluded questionable data points. In details, we checked with the following two procedures. First, at daily time scale, we checked the consistency of temperature time series by cross-referencing temperature values with the day before and after the checking day. At annual time scale, we plotted and screened each individual time series to identify questionable data points and delete the statistical outliers of points out of the three standard deviations range from the long-term mean."

The Data and Methods section describes SRTM data, but it then isn't mentioned what this is used for. I'm sure the metadata for the station observations include elevation–so I am unclear where SRTM elevations are employed.

Response: We used only SRTM data to show the topographic features across China in Figure 1 (**p3803**).

This section 2 also mentions calculation of regression "trends" for latitude and altitude. Given that latitude and altitude probably don't change over time, how are you calculating trends for these variables?

Response: What we mean is that the relationships between the trends of the mentioned variables with latitude and altitude. We have made the changes as follows (**p3790, line 24-26**):

"We also compared the linear trends of the mentioned variables with latitude and altitude in stations to investigate the geographic characteristics of the changes."

Any mentions in the paper of "insignificant" trends and changes should be removed; a non-significant trend means there is no trend at all (it cannot be distinguished from "0"); this also includes most of the panels in figure 6. Similarly, it is also not necessary to state the p-values for all of the findings. Usually a significance threshold (like 95%) is chosen a priori, and then the results are reported as either significant, or not. But the magnitude of the p-value itself is not useful.

Response: We have deleted all figures with insignificant trends, and state only the significant level instead of p-values. Also, we have deleted all text mentioned "insignificant" in revised manuscript.

P3791, 25-27: "FD decreased, but insignificantly, before the 1970s (−0.20 day yr−1, p = 0.14) and changes during 1965–1975 are the lowest in our study period." -> "Anomalies of FD during 1965–1975 are the lowest in our study period."

P3792, 5-6: deleted "For the period before the early 1990s, FD had no significant change (0.03 day/yr, p = 0.45)."

P3793, 5-6: deleted “DR increased, but statistically insignificantly, from 1956 through 1970 (−0.27 day yr−1, P = 0.18).”

P3793, line 26- P3794, line 1: “A statistically insignificant increase in NF occurred from 1956 to 1970 (0.30 day yr−1, P = 0.10), but NF decreased after the early 1970s.” -> “A statistically significant increase in NF occurred since the early 1970s.”

P3794, line 2: “(−0.27 day yr−1, P = 0.02).”-> “(−0.27 day yr−1)”.

P3808, Figure 6: modified the figure and caption:

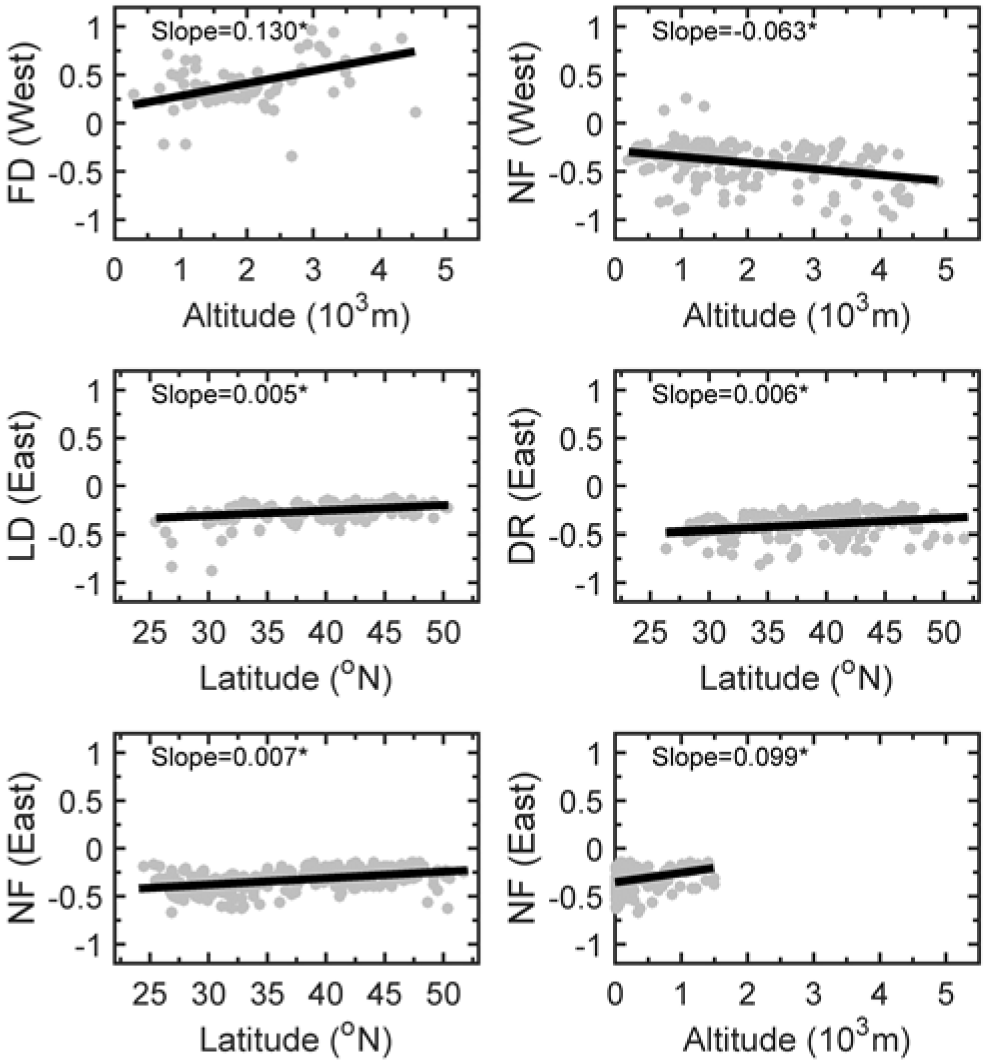

Figure 6. Trends of FD, LD, DR, and NF at west stations (the first row, longitude ≤110 ˚E) and east stations (the second and third rows, longitude >110 ˚E) against latitude (˚N) and altitude (103m a.s.l). We show only graphs passed significant test at least 95% confidential level. Solid cycles are data points, and lines are linear fitted lines. Symbol * indicates a significant trend at 95% confidential level.

How were the various "breakpoints" in the time series determined? For example, in addition to providing trends pre- and post-1970 in the text, figures 2, 3, 4, and 5 all show separate trend lines from "the early 1990s" (it looks like 1992, but this is not actually stated anywhere in the paper) onwards–how was this break-point chosen? The lead author has a previous publication where an objective change-point analysis was applied to determine breakpoints (Frauenfeld, Zhang, & McCreight, 2007, IJoC). Could this be employed here? In many cases (figures 2-5), it looks like 1995 might also (if not more so) be an appropriate break point, so using an objective method may be advisable.

> **Response**: Following the objective method used in Frauenfeld et al. (2003), we determined the break point in the time series of DR. Year 1991 is the first break point, and the year 1998 is the second break point. Here we select year 1991 as the most appropriate mutation point. In addition, we used the same break point to all variables to facilitate comparisons.
>
> *Reference*: Frauenfeld, O. W. and R. E. Davis, 2003: Northern Hemisphere Circumpolar Vortex Trends and Climate Change Implications. J. Geophys. Res., 108(D14), 4423, doi:10.1029/2002JD002958.

A more fundamental question pertains to which aspect of "climate change" the authors are attributing the observed freeze/thaw changes. The rapid and incredible urban expansion of Chinese cities is well-known. To what degree does, e.g., urbanization (and other land cover changes) factor into the findings? These surface changes are, of course, part of "climate change," so it would be useful if the authors could more precisely attribute the freeze/thaw status changes to certain aspects of climate change. E.g., you could categorize stations as rural, urban, or having transitioned from rural to urban, and then check to see if this accounts for some of the changes. Alternatively, you could explicitly state that it is not possible to distinguish between, e.g., greenhouse gas warming and land use change, and that both effects are thought to contribute.

> **Response**: Both referees gave comments about potential effect of urbanization on the changes in soil freeze/thaw cycles in this study. We agree with the reviewers' comments and have done a thorough search in literature and data. We add the following materials, including one paragraph and one figure, in the revised version of the manuscript:

“Our results indicated that urbanization may play an important role in decrease of the near-surface soil freeze days in China over the past three decades. To further explore the impact of urbanization on soil freeze, we used data and information of urban expansion in China from 1990 through 2010 (Wang et al., 2012). The urban built-up areas were manually interpreted using Landsat TM/ETM+ in the 1990s, 2000s and 2010s, which have a spatial resolution of 30 m. The interpretation processes were mainly performed by three experienced operators and revised by the high-resolution images in Google Earth. The interpreted urban areas were finally integrated by statistical data of urban areas in local official yearbooks (Wang et al., 2012).

Over the period from 1990 through 2010, three regions can be divided based upon different degree of urbanization rates, i.e., low rate (<200%), median rate (200% - 500%), and high rate (>500%) of urban expansion regions (Fig. 8). We then calculated the regional anomalies of the number of soil freeze days (Fig. 9). For all three regions, there were significant decreasing trends in the near-surface soil freeze days since 1956 (Fig. 9). For the low and median rate regions, the trends in NF were approximately -0.19 day/yr; while for the high rate regions, the trend was about -0.27 days/yr, approximately 42% larger than the other two regions. It showed a similar phenomenon to Fig.5B (spatial trend patterns of NF from 1956 through 2006 across China). Meanwhile, interannual variations were also significantly large in high rate regions (Fig. 9). However, an important issue occurred roughly before and after 1990. Here we chose 1990 as the breakpoint because (1) 1990 was the starting year of urban expansion data and information available (Wang et al., 2012), and (2) 1990 was close to the breakpoint as shown in Fig.5A.

We found NF changed insignificantly in all three regions before 1990 and significantly decreased after 1990 (Fig. 9). The NF was decreased sharply and continuously even though air temperature had a warming hiatus from approximately 1998 (Easterling et al., 2009). It showed that natural forcing (air temperature) may not be the major factor to affect NF. Further analysis indicated that after 1990, NF in the regions with the lower rate of urban expansion decreased at a rate of about –0.86 day/yr, while NF in regions with high rate of urban expansion showed a statistically insignificant change over the same period (Fig. 9).

Based on results from the above analysis, regions with large expansion rate had a significant long-term (1956-2006) decreasing trend in NF, while regions with low and median expansion rates, the decrease in NF was also significant but their magnitudes were reduced almost by one-third (Fig. 9). This is because the regions with the high urban expansion rates are large cities along the east coast of China. These regions were relatively more developed since the mid-1950s, resulting in the greater long-term impact of urban expansion over the past five decades on the near-surface soil freeze, superimposed on the long-term climate warming. Over the period from 1990 through 2006, the trend in NF was not statistically significant ($P>0.05$) probably due to the climate warming hiatus effect, while urban effect may be minimal because the urban expansion was mainly occurred around the edges of the large cities and meteorological stations were not moved. For regions with low and median expansion rates, the long-term decrease trends in NF may mainly reflect the impact of climate warming with relatively limited urban expansion effect because these regions are located far inland and less developed. Meteorological stations in these regions were installed in the 1950s and generally located away from small and median cities by several kilometers to avoid the urban effect on meteorological observations. However, over the period from 1990 through 2006, the magnitude of the decreasing trends in NF increased sharply (Fig. 9) this may be due to the urban expansion was close to and probably far beyond the meteorological stations, resulting in substantial heat island impact on the near-surface soil freeze."

Reference:

Wang, L., Li, C., Ying, Q., Cheng, X., Wang, X., Li, X., Hu, L., Liang, L., Yu, L., Huang, H., and Gong, P.: China's urban expansion from 1990 to 2010 determined with satellite remote sensing. Chin. Sci. Bull., 57, 2802-2812, doi: 10.1007/s1434-012-5235-7, 2012

Easterling, D., and Wehner, M.: Is the climate warming or cooling? Geophys. Res. Lett., 36, L08706, doi:10.1029/2009GL037810, 2009.

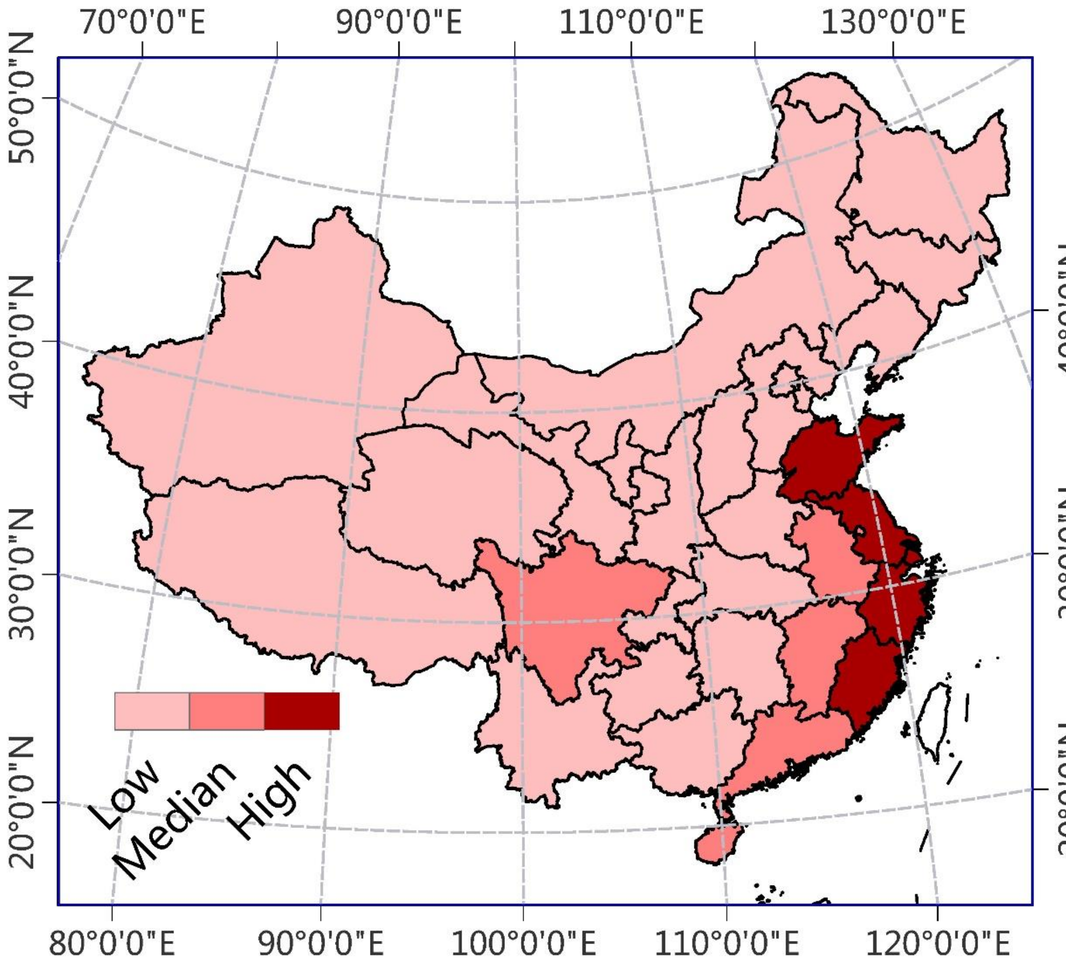

Figure 8. Rates of urban expansion from 1990s through 2010s. (Reclassified from Wang et al. (2012))

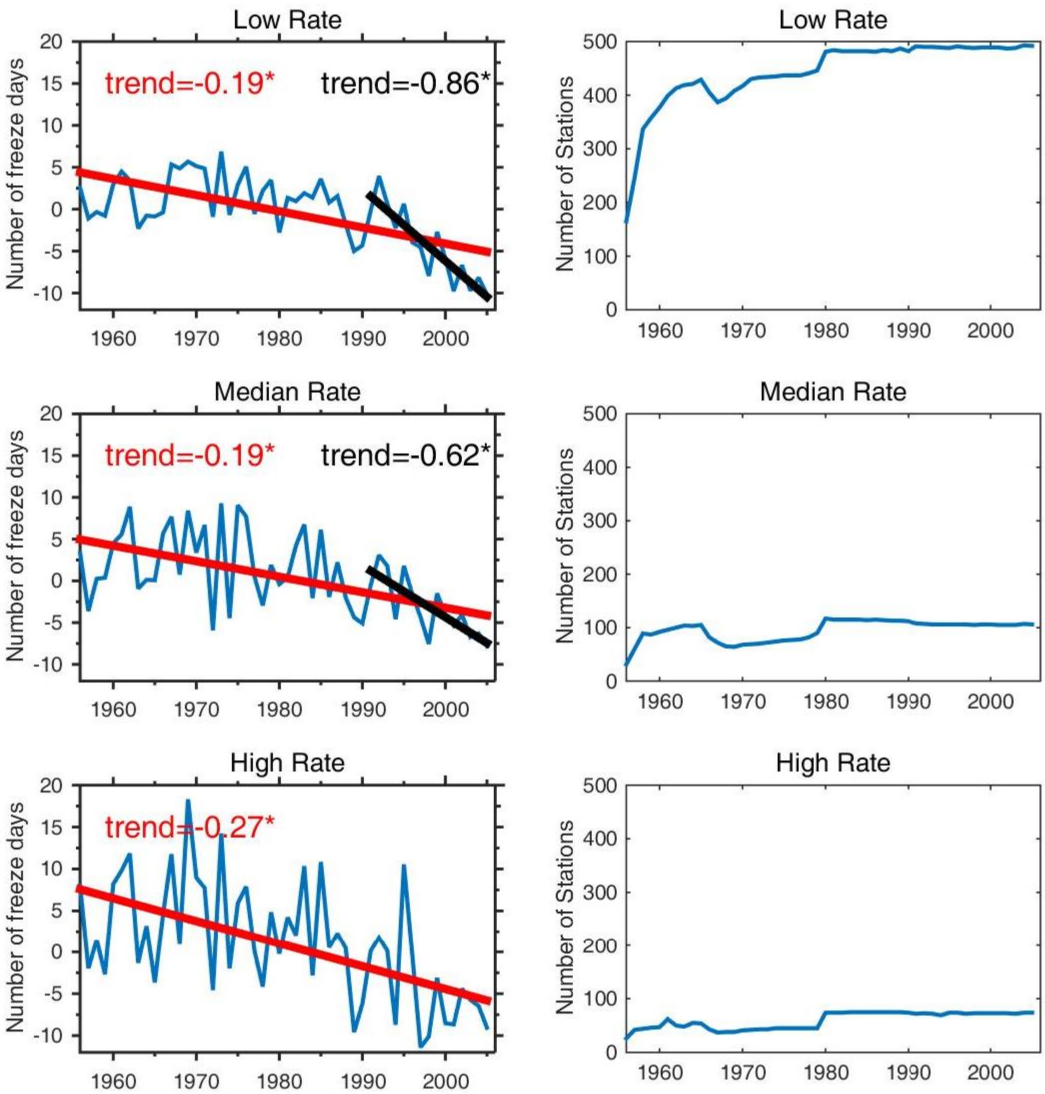

Figure 9. Regional changes of NF in regions with different urbanization rates (left). Black lines and red lines depict respectively the linear regression for the period after 1990 and the period since 1956. Symbol '*' indicates a statistically significant at 95% confidential level. Rights are number of stations used to create each time-series.

What is the explanation for some of the interdecadal variations, e.g., in regard to the "major increase in FD...after the 1970s?" In other recent work (Frauenfeld and Zhang, 2011, ERL), you suggested a strong role of the NAO in affecting soil freezing (or lack thereof) in Russia over this exact same time period–is there a similar explanation here? This is one important aspect that is currently missing from the paper: attribution of the soil freeze/thaw changes to 'something' beyond air temperature. Land surface and soil properties, vegetation, latent/sensible heat sources, snow cover (in the cold season), etc. could all be playing a role in GST variability, yet only air temperature is used. It seems a little simplistic to essentially conclude that when it is cold, the ground freezes, and when it is warm, the ground thaws.

Response: We have also examined the relationship between NAO (seasonal mean from December to February) and our indicators (not shown in the manuscript). Our results indicated barely significant in first date, last date and number of days, and only a statistically significant relationship between winter NAO and duration. It shows that effects of NAO on our indicators exist but may be limited at least in a way of statistical results.

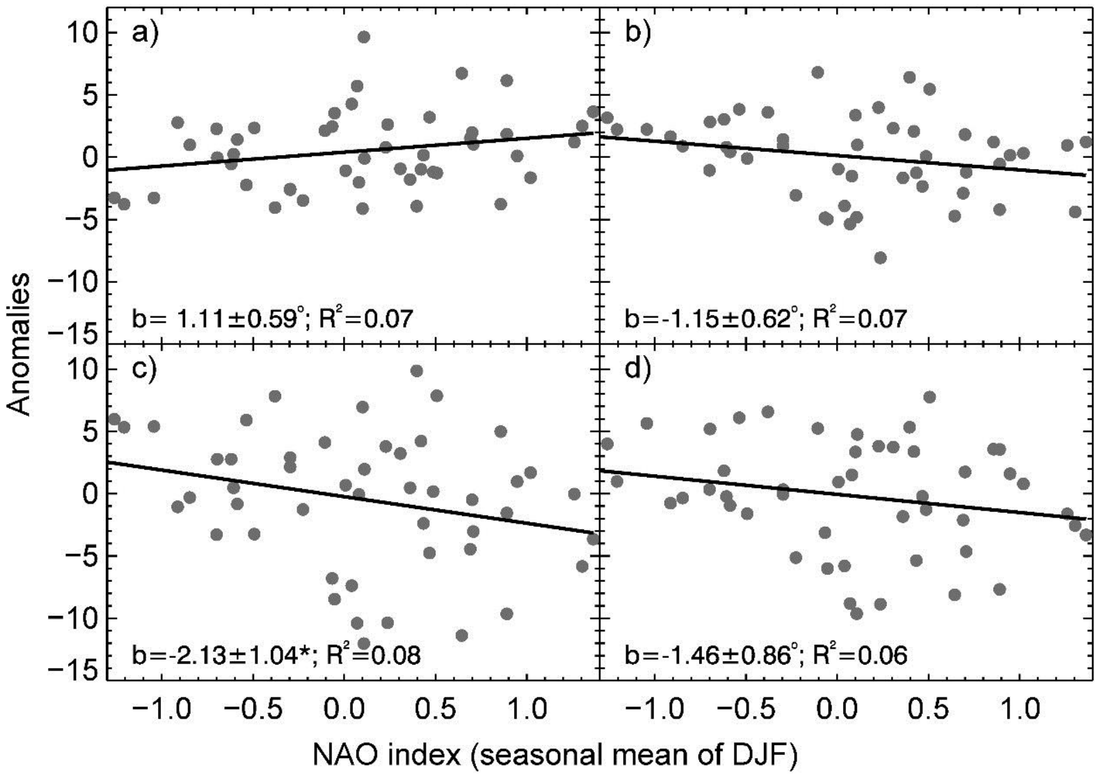

Figure 1. Relationship between of NAO in winter and (a) FD, (b) LD, (c) DR, and (d) NF from 1956 through 2006 across China. The solid circles are individual data points, and the lines are linear regression. Symbol * indicates statistically significant trend or 95% confidential level; and symbol 'o' indicates a barely significant or 90% confidential level.

In addition, soil profiles, snow cover, and other factors suggested by the Referee all might be the important issues to near-surface soil freeze/thaw states. Unfortunately, we don't have enough and valid materials of soil parameters in all stations especially the dynamics of these soil parameters. Despite we have some observations of snow depth in meteorological stations; the duration of snow cover on the ground is always short. Thus its effects on the near-surface soil freeze/thaw status also are limited.

A couple of minor, final points: why are the results of this study compared to Kansas, USA (are there some expected similarities)?

Response: We compared with previous study in Kansas, USA to indicate a common long-term trend also to show some interregional differences.

Also, please carefully check the paper for grammatical errors. There may be some PDF conversion issue, but in many places, two words are merged together (e.g., "datewas" in line 9 of the abstract, also "utilizedto" on p. 3788 line 14, "atleast" p. 3790 line 25, p. 3792 lines 20-21, and many others). P. 3792 line7 contains a mistake ("the stations 140 stations"), "observede" on p. 3794 line 15, "effectively" on p. 3796 line 13 (should be effective), and line 26 on p. 3796 is missing a word after "land-atmosphere."

Response: We have checked and corrected in revised manuscript.

P3786, line 9: "datewas"->"date was"

P3788, line 14:"utilizedto"->"utilized to"

P3789, line 25: 'atleast' -> 'at least'

P3792, line 7: "the stations 140 stations"->"the 140 stations"

P3792, line 20: 'earlier.Anomalies', add a space after '.'. (i.e., 'earlier. Anomalies').

P3792, line 21: 'appearedafter'->"appeared after"; 'alinear'-> 'a linear';

P3795, line 15: 'observede'->'observed'

P3792, line 27: 'FD and LD was' -> 'FD and LD were'

P3794, line 14: 'Qinghai-Tibet Plateau' -> 'Qinghai-Tibetan Plateau'

P3795, line 13: 'effectively' -> 'effective'

P3796, line 26: added a word "processes" after 'land-atmosphere'.

P3797, line 5: 'arctic' -> 'Arctic'

P3797, line 24: 'Qinghai-Tibet Plateau' -> 'Qinghai-Tibetan Plateau'