Reply on comments from Anonymous Referee #2

This paper examines the soil temperature data collected at numerous stations in China for the long-term trend of the first day and last day of soil frost, and the duration and number of soil frost days. A number of previous studies have reported the decreasing trend of soil frost in various parts of the world. This study is a useful exercise that adds another piece of evidence to the growing body of literature substantiating the common notion that the rising air temperature results in fewer days of soil frost. However, I feel that the paper in its present form does not contribute significant new knowledge to the scientific understanding of interaction between environmental changes and soil freeze-thaw status. That is partly because the paper is somewhat vague in the technical definition of frozen/unfrozen status of soil, as well as the statistical treatment of data set. The paper can be strengthened significantly by more creative and rigorous analysis of data and their discussions. Please see my comments below for specific suggestions.

> **Response**: We appreciate the referee's insightful and constructive comments on the manuscript. All comments are very helpful for improving the manuscript. Specifically, many questions about the fundamentals of statistical method are very important to enhance reliablibility of our statistical results. We have studied all comments thoroughly and made necessary changes and corrections.

SPECIFIC COMMENTS:

1) **P3788-3790. This section contains a large volume of texts reviewing previous literature on similar studies. I feel that the literature review is too long, and also much of the material on remote sensing is not directly relevant to this paper. On the other hand, it is fairly thin on the current understanding of the physical processes that control how the environmental changes (e.g. air temperature, precipitation, land use, urbanization etc.) affect soil freezing and thawing. I suggest that the introduction be re-written to sharpen the focus of the paper. I also suggest that the scientific question, hypothesis, or objective of the study be clearly stated in the introduction.**

> **Response:** We have rewritten the introduction through deleting few sentences not-so-close to major aim of this paper and shortening the literature reviews about advancement of study of soil freeze/thaw. The revised section is as following.

“The near-surface soil freeze/thaw state is coupled to the timing and duration of cold/warm seasons, and is an important indicator of climate change (Zhang et al., 2001). The latest assessment report from the International Panel on Climate Change (IPCC) indicated that the globally averaged combined land and ocean surface temperature rose 0.89 C over the period 1901–2012 (IPCC, 2013). During the past few decades, many studies have focused on the dynamics of the near-surface soil freeze/thaw status and the feedback between the ground and atmosphere. They have shown that changes in the near-surface soil freeze/thaw status are interrelated, and soil freeze/thaw affects hydrological processes (Cherkauer and Lettenmaier, 1999; Niu and Yang, 2006; Rempel, 2012), ecological processes (Schimel et al., 1996; Tagesson et al., 2012) and soil microbial processes (Lloyd and Taylor, 1994; Gilichinsky and Wagener, 1995; Edwards and Jefferies, 2013).

Variations in the timing and duration of the near-surface soil freeze/thaw status have been widely investigated using a range of approaches, including satellite remote sensing and in-situ observations, across spatial-temporal scales ranging from regional to global. Menzel et al. (2003) used data from 41 meteorological stations across Germany (from 1951 through 2000) to investigate soil frost dynamics and showed that the freeze-free period was extended with increasing air temperature. Henry (2008) used observations from 31 stations to examine soil freeze dynamics across Canada and found that annual soil freezing days declined from 1966 through 2004. Using long-term data from three stations in Indiana, USA, Sinha and Cherkauer (2008) analyzes found that the number of soil freeze days significantly decreased at the central and southern study sites, but the near-surface soil temperature at the northernmost site showed a significant decrease in the cold season due to the decrease in snow depth. Anandhi et al. (2013) carried out a more-detailed analysis of frost indices using data from 23 stations across Kansas, USA, and found that the first date and the last date of freezing occurred later and earlier, respectively.

Numerous studies have reported significantly improvements of monitoring soil freeze/thaw status. NASA is launching the Hydrosphere State Mission as part of the Earth System Science Pathfinder Program (ESSP) to improve satellite monitoring of global land freeze/thaw and soil moisture (Entekhabi et al., 2004). In China, a multi-scale monitoring network has been established on the Qinghai-Tibetan Plateau (Yang et al., 2013). Fifty-six (56) stations have been installed in cold and high-elevation regions to enhance monitoring of soil temperature and moisture and hence to support remote sensing data and large-scale climate modeling (Su et al., 2011; Yang et al., 2013).

In this study, we use ground-based meteorological station data to investigate the long-term spatiotemporal variation in the timing and duration of near-surface soil freeze/thaw across China over the period 1956–2006. Using data from 636 stations, we examine the first date, the last date, duration and actual number of the near-surface soil freeze, as well as their spatial characteristics over China. Finally, we briefly investigated the response of the near-surface soil freeze/thaw status to changes in climate conditions in the past few decades. ”

2) **P3790, L5. How were these stations selected? Did all of these stations register soil temperature below freezing point? Is it meaningful to include the stations from warm regions (e.g. southeastern China) in the analysis?**

   Response: Essentially, all stations with recorded soil surface temperatures are selected in this study. However, the near-surface soil temperature at some southern stations may never be below the freezing point, data from these stations were not used in the analysis.

3) **P3790, L6. How is "near-surface" defined? Please present a precise definition.**

   Response: The reviewer raised an excellent question. We often say ground surface but in real world, it is hard to define. By “near-surface” in our previous publications, we mean the top 2 to 5 cm of soils from the surface. It is hard to define the true ground surface since the surface is not a perfect flat plate, any micro-scale horizontal changes within a few centimeters may have a few millimeters or even centimeter changes in roughness. A true 0 cm surface is essentially meaningless. However, in this study, by “near-surface”, we mean the top a few millimeters of soils from the “surface” since the thermometer tip has a diameter of 5 mm, it is slightly more than half buried in soils, slightly less half exposed in the air as we defined in the text.

4) **P3790, L11. Did all stations have continuous hourly temperature data? If not, how was the daily minimum temperature determined? This needs to be explained carefully. Note that Henry (2008) used the data sets that had only two measurements per day.**

   Response: The reviewer #1 raised the same question. We have added more information on this issue as follows:

“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 designed minimum (maximum) temperature thermometer. The principle is that when the mercury ball at the tip of the thermometer cools (warms) due to the change in the near-surface soil temperature, its volume will shrink (expand). The volume change will be recorded on the graduated glass tube. For minimum (maximum) temperature thermometer, the mercury volume will not return to the normal after the temperature reaches the minimum (maximum) without the extra force. In this case, the daily minimum (maximum) temperature will be recorded and the mercury scale on the glass tube will be reset by the technician. The minimum (maximum) temperature thermometer records the daily minimum (maximum) temperature once a day but 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 recorded 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 across China. 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.”

Here is a reference.

China Meteorological Administration. 2007b. Specifications for surface meteorological observation, Part 13: Measurement of soil temperature, China Meteorological Press: Beijing, China.

5) P3790, L12. How was the temperature of ground surface (0 cm) measured? It is not trivial to measure the surface temperature. Accurate measurements would require an infrared thermometer. Was the same method used to measure soil temperature at all stations? What is the accuracy of measurements? This is a very important point, and should be discussed thoroughly in the paper.

Response: We have answered this question in the previous question. The good thing is that the same method was used for all stations across China.

6) **P3790, L23. The temporal trend analysis concerning environmental changes, such as this paper, requires rigorous statistical treatment of the data, if it is to be published in a peer-reviewed journal. The authors need to explain the statistical methods carefully and justify the assumptions used in the analysis. Why is a linear regression method used? Does the statistical distribution of data set justify the use of linear regression?**

Response: We agree with the reviewer's comments and have made thorough checks. Generally, if mean of residual approach to zero, estimated coefficient can be assured as unbiased estimator. In this case, we use a table to summary the estimated parameters of probabilistic distributions of the residual, which included mean, standard deviation, skewness and kurtosis. Then we show Q-Q plot, a quantile-quantile plot of the sample quantiles of residual versus theoretical quantiles from a normal distribution. If the distribution of residual is normal, the plot will be close to linear.

Table 1 shows the means of residual are all close to zero, thus our estimated coefficients are close to unbiased. Considering skewness and kurtosis, they stray slightly from zero. More directly, Figure 1 indicates the subplots all are close to linear. Thus, we can find the hypothesis of normal distribution for our linear fitness should be appropriate. To keep the paper short, we did not include the Table 1 and the Figure in the text. However, we added text and reference.

References:

Rice, John. Mathematical statistics and data analysis. Cengage Learning, 2006.

Freedman, David. Statistical models: theory and practice. Cambridge University Press, 2009.

Table 1. Descriptive Statistics of residual

	Minimum	**Maximum**	**Mean**	**Std. Deviation**	**Skewness**		**Kurtosis**	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
First date	-5.60	6.70	0.00	2.70	0.02	0.34	-0.14	0.66

Last date	-5.11	5.10	0.00	2.41	0.25	0.34	-0.66	0.66
Number	-6.10	6.39	0.00	3.43	0.04	0.34	-1.12	0.66
Duration	-8.70	8.12	0.00	4.08	0.16	0.34	-0.63	0.66

* All results were calculated with IBM SPSS Statistics v20.

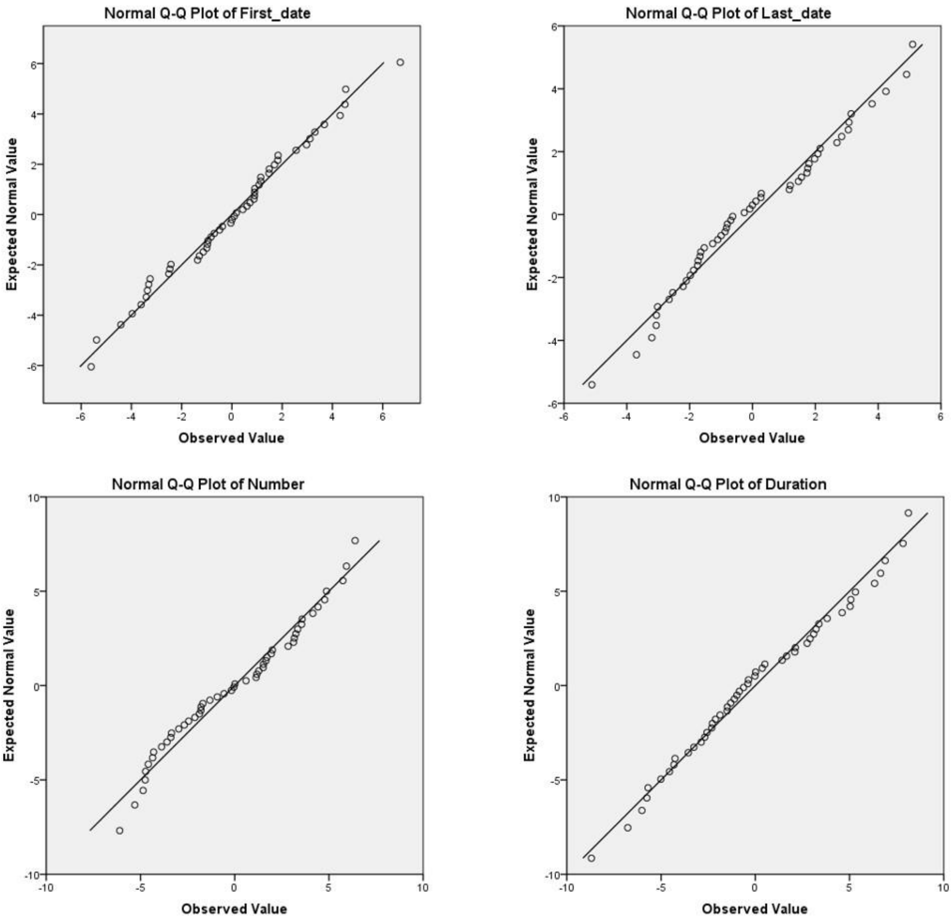

Figure 1. Q-Q plots of residual for each variables.

* All subplots were created with IBM SPSS Statistics v20.

7) **P3791, L14. How was this number (300 days) selected? If the missing days occur more frequently during winter than in summer, it will bias the statistics. How was this issue addressed?**

   Response: Before we selected the threshold, we also used years with 365 daily records to calculate the indicators. Through comparison, the results used 300 days as threshold have errors of +/- 2 days. In this case, we believe that the 300-day threshold and will have very limited effects on results. However, it will give us more stations to cover more spatial areas.

8) **P3791, L20-21. It took me a while to understand what is shown in Fig. 2a. Please remind the reader that you are showing the anomalies in this graph, both in the figure caption and in the texts.**

   Response: We have made changes that they are anomalies, and modified this sentence as following.

   “Overall, departure of FD showed a significantly increase across China by nearly 5 days, or a trend of 0.10±0.03 day yr−1, for the period 1956–2006 (Fig. 2a).”

9) **P3791, L21. How is the range (+/- 0.03) defined?**

   Response: It was estimated by a width of 1-standard variation in linear regression. In other words, any estimated coefficient is also a random variable and has a given statistical distribution. So we gave a range of estimated coefficient.

10) **P3791, L23. How is the coefficient of multiple determination defined?**

   Response: We described the long-term change through two aspects. One of them was estimated linear trend, and another was total change over the period. For example, here, the estimated linear trend is -0.58 day/yr, since the early 1990s; the ending of study period is 2006, so the total change is estimated as 0.58*15~=9 days.

11) **P3791, L26. How is the p value defined? Do the data meet the assumptions for the calculation of p value (e.g. normal distribution)?**

   Response: P value is for F statistic of the hypotheses test that the corresponding coefficient is equal to zero or not. For example, the p-value of the F-statistic for ‘slope’ is less than 0.05, so this term is statistically significant at the 5% significance level.

   To examine the normalization, we have shown in **Response 6)**. Here is only a Q-Q plot. If the distribution of residual is normal, the plot will be close to linear. This figure shows the precondition should be met.

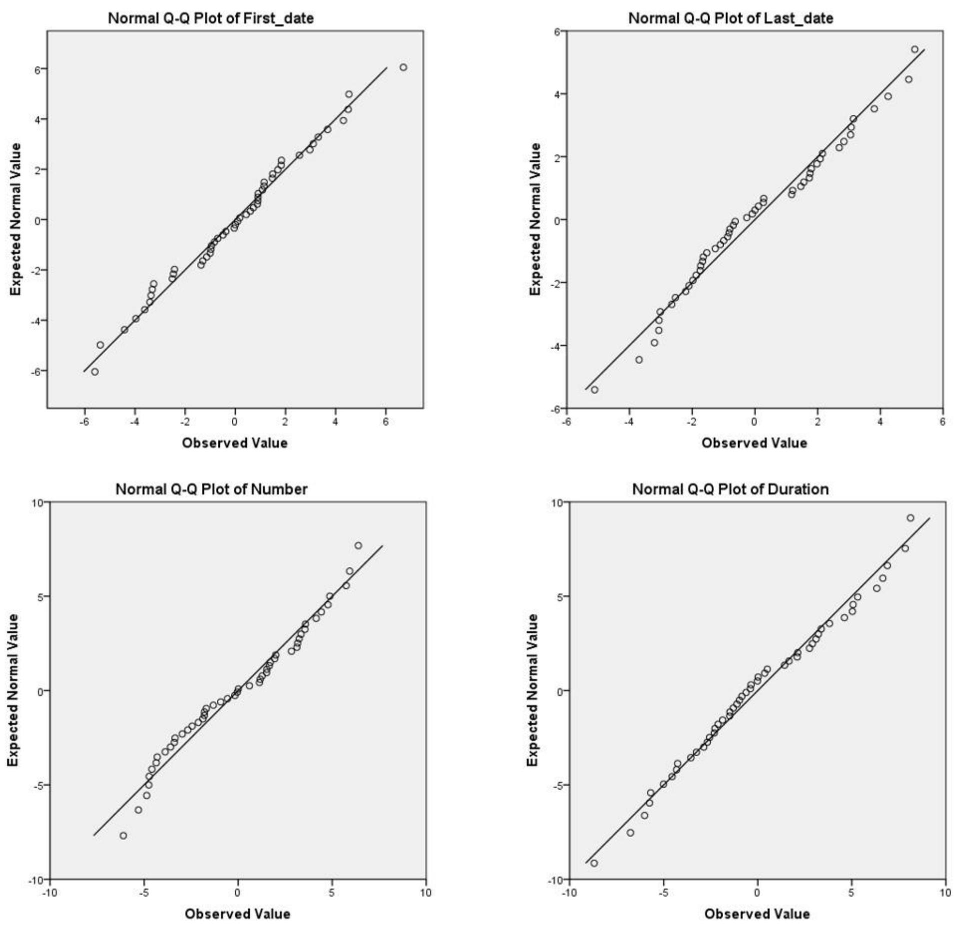

Figure 1. Q-Q plots of residual for each variables.

* All subplots were created with IBM SPSS Statistics v20.

12) P3792, L1-2. A "major" increase in the early 1970's is not visible in Fig. 2a. The data seem to be more or less steady until after 1990.

Response: Before the early 1970s (or exactly 1969-1970), there was a short cold period. From Figures 2-5, we can find this characteristic. Hence, we deem that an increase occurred since the early 1970s. The original adjective 'major' may mislead readers. We have corrected it as following.

"An increase started from the early 1970s when a short cold period was ended."

13) P3792, L6. This sentence contradicts with the statement in L1-2 (see above).

Response: We have corrected it as following.

"An increase started from the early 1970s when a short cold period was ended."

14) P3792, L19-20. Is this statistically significant?

Response: It was insignificant thus we did not show any estimated trend. We have modified this sentence to make it clear:

"Prior to the early 1990s, LD occurred slightly and insignificantly earlier."

15) P3792, L26. How is "west China" defined?

Response: West and east China is defined according to 110E, i.e., west of 110E in China is defined as "west China". It also was referred in this article (P3792, Line 10) as well as in caption Figure 6 (P3808).

16) P3792, L28. What does "dramatically" exactly mean?

Response: We mean "significantly".

17) P3794, L6. "NF varied significantly". This needs to be explained a bit more carefully, as there was a period with very little change, followed by a period of visible change. What is the statistical significance of NF during the two periods?

Response: Here we showed stations with $p<0.05$. We believe they varied significantly. Of cause, some of stations may also have one or more break points. Using 1991 as unified break point, we also created a map to show stations with significant trends (~130 stations, not shown in the manuscript).

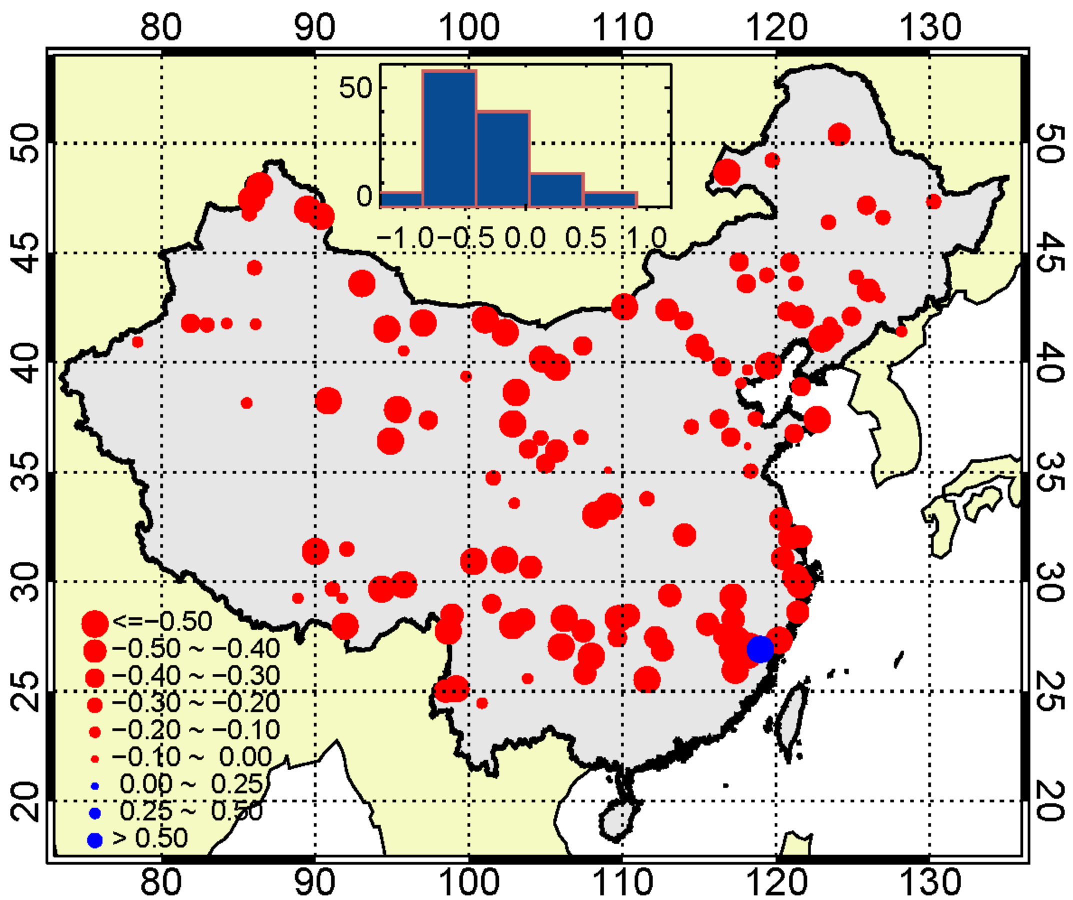

Figure 1. Changes in NF from 1956 through 2006 across China; center-top is histogram of changes in NF.

18) P3794, L11-12. Please indicate Qinghai-Tibet Plateau and Yangtze River on a map.

Response: We have added the boundary of Qinghai-Tibetan Plateau and also Yangtze River on Figure 1.

19) P3794, L15-16. The issue of urbanization needs to be investigated much more carefully. Is the urbanization restricted to the lower reaches of the Yangtze? I imagine urbanization has been occurring in other parts of China. How many of the monitoring stations are located in rural areas? Should the analysis be conducted separately for urban and rural stations?

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.

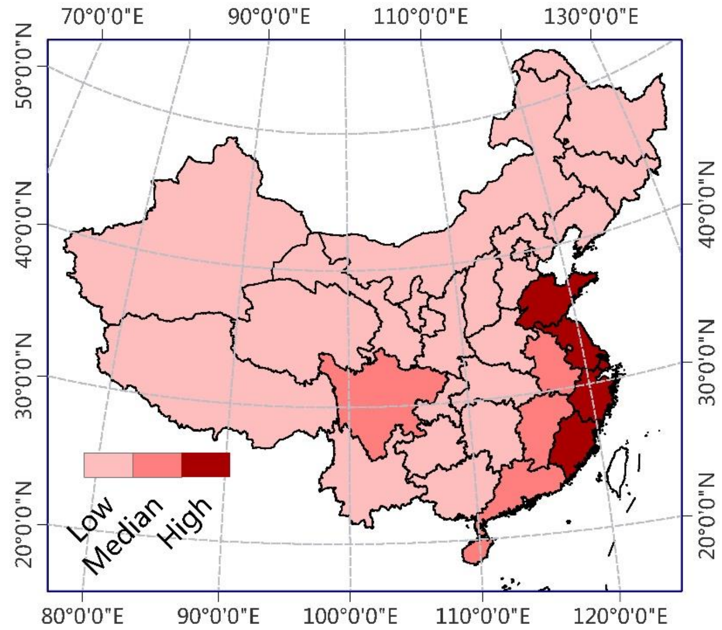

Figure 8. Rates of urban expansion from 1990s through 2010s. (Reclassified from Wang et al. (2012))

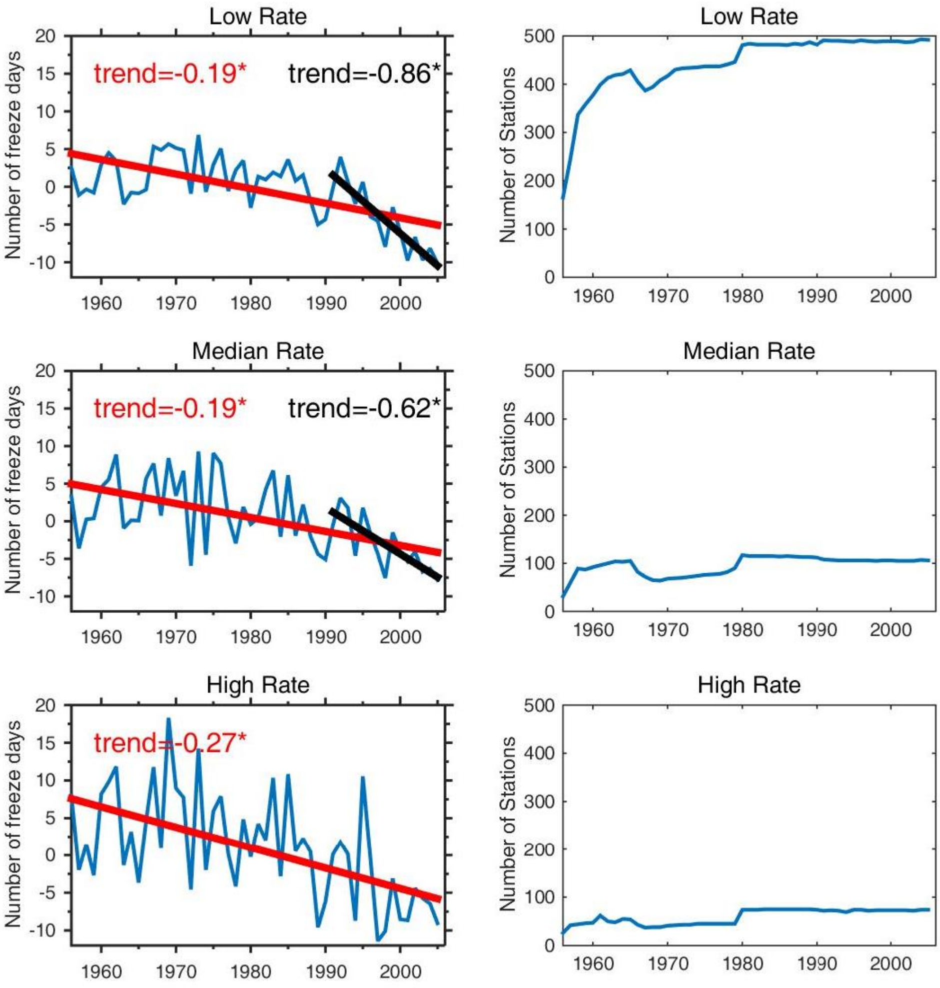

Figure 9. Regional changes of NF in regions with different urbanization rates (left). Black lines and red lines depict respectively the linear regression for the period after 1990 and the period since 1956. Symbol '*' indicates a statistically significant at 95% confidential level. Rights are number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

20) **P3794, L16. Another related issue is the relocation of stations and changes in measurement methods and depth. If any station has been affected by relocation or other changes, it should be excluded from the data analysis. Also, did all stations have continuous data from 1956 to 2006 (P3790, L5)? If new stations have been added over time, how does it affect the statistical analysis? Can there be a bias in the regional distribution of new stations? These are very important issues, which should be discussed thoroughly.**

Response: Most stations were built in the early 1950s. A known large scale of stations adjustment (about three provinces) was taken in 2007. Some new stations was built since 21st century, however, we defined the baseline period as July, 1971-June, 2001. In this way, we excluded many newer stations with less than 30 years of records. We have also done thorough data quality control to make sure consistency at annual scale through plotted and screened each individual time series.

21) **Figure 1. This map appears to show the stations that are not likely affected by soil frost at all. For the purpose of this paper, it will be better to include only those stations that are subjected to soil freeze-thaw. Also, this map shows Taiwan and southern islands that are not relevant to this paper at all. They should be removed.**

Response: We agree with the reviewer's comments. In order to keep the completeness of meteorological station distribution, we have them all stations in the map. However, we added information for stations that have no frost.

22) **Figure 2. In the caption for (a), what are "composite variation" and "low-pass filter"? Please explain these in the texts. What are the units of the values shown in these figures? Please indicate. This map appears to have a smaller number of data points than in Figure 1. Why?**

Response: The composite variation is combined by all valid station's time series. Low-pass filter is a method generally used in a long-term climatic series analysis; it can smooth high-frequency noises according to a cut-off frequency. Its mathematical foundation is fast Fourier transform (FFT). More detail numerical algorithm can be found in many references (for example, Brault, J. W. and White, O. R., 1971, The analysis and restoration of astronomical data via the fast Fourier transform, Astron. & Astrophys., 13, pp. 169-189.)

Figure 2 showed only the stations with significant trends at 95% confidence level. Due to the first date may be influenced by some extreme weather; its long-term trend was not strongly significant. Thus not all stations changes significantly over our study period, and that is why a small number of stations was shown in Figure 2.

23) The paper presents the data in the form of anomalies, but the reader is given no information on the actual number of frost days, etc. It will be useful to include a map or histograms showing the distribution of duration of frost, or number of frost days.

Response: The climatology, mainly focused on spatial characteristics, of these indicators was analyzed in another paper, which is under reviewing for publishing.