

Interactive comment on “Climate change and Northern Hemisphere lake and river ice phenology” by Andrew M. W. Newton and Donal Mullan

Anonymous Referee #2

Received and published: 19 October 2020

General comments: The authors have collected northern hemisphere ice phenology data from several large databases and climate data from an online portal. They have split the ice phenology data into three large geographic regions: North America, Fenoscandia, and Russia and within each region they have examined the overall ice phenology trends and also site (lake or river) specific ice phenology trends over four time periods. Finally, the authors have used a correlation analysis approach to examine potential relationships between several climate drivers (air temperature, precipitation, wind speed) and ice phenology metrics for the three large regions. The manuscript contains at least two novel components: its geographic scope is large and is split into three reasonably large regions (North America, Fenoscandia, Russia), and the analysis con-

C1

tains an assessment of potential changes in ice-phenology for both lakes and rivers. The general statistical approach of using Mann Kendall trend tests with Sen's slopes to evaluate both the significance and slopes of monotonic trends has been commonly applied to similar datasets. The use of correlation analysis to assess potential drivers is also common, but there is a strong potential for false positives and negatives due to the large number of tests. Further, the authors appear to have used mean climate conditions to within each study region for the correlation analyses. However, given the large size of the study regions and large inter-regional variation of trends (e.g. to the east and west of the Laurentian Great Lakes) such an approach may not be appropriate to assess the linkages between ice phenology trends and climate drivers. There is also some uncertainty of how the regional averages were calculated and if these changed over time as the number and location of sites within the regions changed. For example, there were no data included for Canada during the last time period (only US sites)...did the area used to calculate the 'mean' climatic conditions change? The use of the three time periods is helpful but the last time period was unequal in terms of length (the first two periods were of 30 year durations, while the last was only 15 years). I suspect the choice of years may have been logistical, but the authors should evaluate and discuss the implications of this decision. Third, as part of the online comments, a reviewer suggested the authors evaluate the implications of oscillatory dynamics that may affect the slopes of their relationships, particularly in the shorter 30 or 15 year time periods. This is quite important and the authors should consider this. In terms of methods, the authors included too much detail in some areas (i.e. mathematics supporting Mann Kendall are not needed here), but not enough information in others (see specific comments). While I found the manuscript and analysis interesting and useful, my largest criticisms relate to its length, lack of clear objectives and hypotheses, attempts to assess changes at both large regional and site specific scales, and limited attempts to compare their results with other large scale analyses in the literature. At 50 pages, 17 figures and 7 tables, the manuscript is far too long and unfocussed. I suggest the authors remove all site-specific analyses and focus the manuscript on the

C2

broader regional trends and drivers and novel aspects (e.g. lakes vs. rivers). If the large regions require some additional partitioning due to interregional variability that is fine, but the analysis and discussion of individual sites is not particularly useful for the reader and takes away from the more important regional assessment. From there, the authors should more clearly state their objectives and hypotheses, and focus the discussion on not only their results, but also improve the discussion of the implications of their findings and how they compare and contrast with other large-scale analyses (i.e. what is new and novel). The authors would also need to defend their correlation approach to identifying potential drivers, acknowledging the high likelihood of false positives and negatives, potential challenges with using regionally averaged values over such large geographic areas where there is both high intra-region variability and where the geographic distribution of sites change, and the use of climate indices over large geographic areas. I have provided some more detailed comments below.

Specific points – The introduction could use a clear set of objectives and hypotheses. This would greatly assist with improving the conciseness of the manuscript, which is quite long and unfocused. – P9 L133+ The authors use climate variables (mean monthly temperature, precipitation) from KNMI Climate Explorer. The data were ‘downscaled’ (averaged?) over 3 geographic regions (Fennoscandia, North America, Russia) and then used to assess relationships with spatially averaged ice-freeze up or break up dates. Is spatially averaging over such a large geographic area a good idea? There is evidence that there are large differences in the magnitude and even direction of ice phenology trends within the North American dataset. Would the strength of the analysis not be affected by geographic scale with weaker correlations expected as the study area increased? – Should section 3 (L. 153) not read ‘3. Results’? – With 60 pages of text (including 17 figs and 9 tables), this manuscript is quite long and highly unlikely to fit into the space requirements of the journal. The manuscript is also largely unfocused, moving from broad discussions of phenological changes among regions to significant text devoted to individual sites. Much of the site-specific material is extraneous and distracts from the broad patterns. The objectives and hypotheses

C3

of the study are not clearly articulated and the closest we get is on page 3 where the authors state ‘The paper explores the hemispheric spatiotemporal trends in ice phenology...Observed changes are then compared with climate records...’ I would encourage the authors to devise more clear objectives and testable hypotheses. Further, I would encourage the authors to focus the current manuscript on a concise description of the broad phenological changes among the regions without delving into specific site responses. These are unnecessarily distracting. Such individual sites responses could either go into supplemental material or separate manuscripts. – In relation to methods, the Mann Kendall trend tests with Sen’s slopes is an appropriate statistical test for the monotonic trends, however the methods section does not indicate at what level the trends were considered significant. The correlation analysis, examining the potential drivers, is more suspect for a few reasons. First, there are a large number of correlations being tested (>800) and therefore a large potential for false positives or negatives. Second, at least two of the regions (NAM and RUS) have large geographic extents and the authors acknowledge in other locations in the manuscript that there are large regional differences in phenological trends (e.g. east and west of the Laurentian Great Lakes). Thus the use of a single averaged climate (temperature, precipitation, wind, etc.) value for a study region would not be expected to correlate well with the phenological trends. While considerably more work, it would be more useful to examine the relationships between downscaled climate drivers and phenological responses at individual sites (or smaller regions). This would also avoid the changing locations of sites between study periods. – While I greatly appreciate the extent of analysis undertaken, the presentation of results is at times confusing and potentially misleading. There are many figures (e.g. Figure 5, 7, 8, 11, 13, 14, 15; Tables 2, 3 [first panel]) where trends are reported. However, the reader is unable to distinguish if these individual site trends are significant and, if so, at what level? If they are not significant, then they should not be reported as so. The legend should also indicate what the white boxes represent (e.g. insufficient data, no trend, something else?). – Table 1. Assel et al. 2003. What is meant by ‘maximum fraction of lake surface

C4

ice coverage'? Should this be maximum extent of lake ice coverage'? – Table 1, while useful, is quite large and is probably best suited to supplemental material. It also misses many references (Magnuson, Benson, Sharma, etc. etc.). – Surprised Benson et al. 2013 was not included in Table 1 given it covers such a large number of lakes – P6 L86. The indication of rivers in brackets. Are these 88 sites on three rivers, or 85 lake sites and 3 river sites? – P6 L88 Change to are clustered around the Laurentian Great Lakes. – P7 L99 Change Julian days to Ordinal days? – Note: Mann Kendall examines unimodal trends only – P8 L103+ More information is required on the statistical approach. At what level were trends considered significant? The first I see of this is on page 16 L180 where an alpha value of 0.1 was considered significant. – P8 L108+ The Mann Kendall test is widely used for time series analysis and the mathematical details are widely available. No need to detail here. – Figure 2. The 3 shades of grey are not easily distinguishable. – P10. L154-166. This section reads more like introductory material and methods than results. – P10. Lake size or elevation can have large effects on freeze up or break up dates. Perhaps this (e.g. few lakes, highly variable lake size) might account for the lack of latitudinal trend. But note that in the description they indicated a fairly tight geographical extent in Russia (51.5-52 N, 104.5-105E). – P10. Comments on changes in European changes (L174+) are qualitative... not assessed using a statistical approach. These are not easily distinguishable in the figure. – P10. L171-172. With the exception of 3 sites on the eastern portion of the map, the sites in Russia (Figure 1a) appear largely to be of similar latitude, which precludes a latitudinal assessment. Thus, the case that 'This is the case for ...but not Russia' is not technically correct. Rather, a thorough assessment cannot be made due to limited data. – Figure 2. Why are the grids at different scales between the 3 figure panels? – Figure 3. The figure itself is good, but the caption requires editing. E.g. 'Trends in ice breakup, ice freeze-up (is 'freezeup' a word?), and the duration of the open water period for lakes across North America, Europe (is this not referred to in other locations such as Figure 1 as Fennoscandian? Be consistent), and Russia. – P12. L196. 'When all sites are considered there is a clear increase...'

C5

I don't believe a statistical test was used to make this assertion. While it may appear obvious by the graphic, such visual assessments require a statistical assessment. – P12. L201. 'For Russian sites it is clear...' Again, there is no statistical test on which to base these claims. These are qualitative assessments only. – P13 L203. Again a reference to the 'large area' of the Russian sites. – P15 L60. Note there are many lakes in Canada being monitored for ice-phenology, they are just not recorded in the database or captured in this manuscript. – What was the criteria for whether a site could be included or not. Was it 90% of records for the 15, 30, or 75 year periods? For the 75 year trend test, were only lakes that had 90% of records included for the whole period included? – P14. Table 2. Please include some additional details in the table caption: The statistical test used, the significance value of the trend. Within the table n values and confidence intervals would be helpful. There is value in the table, but it is hampered by the large differences in study areas (e.g. latitudinal extent) and differences in the number of sites between time periods (esp. in Russia and North America). This can influence the qualitative assessments provided in the text, for example in North America between the middle and last time periods where a large number of northern sites dropped off. Further, it would be useful for Figure 1 (panels b-c) could have lake and river sites distinguished independently. This would help to see if lake and river sites were distributed homogeneously or if they are clustered at certain latitudes. – P15 L261. Awkward sentence. Reword. – P16. L212 'The data show...'. Larger than what? Where is the statistical test? This is also true for the general pattern discussed for Europe and North America. The wording 'do not appear to change significantly...', was this an eyeball assessment? – P16. L279. While there is some value pointing out that there are exceptions to the general trend, this is obvious in Figure 3. I think this figure can be referred to rather than describing specific lake or river systems that behave differently. – P 16. The use of standard deviation to assess interannual variability should be introduced in the methods. Why use numbers (14) and letters (eight). – P17. L301. The reporting of mean values should be accompanied by estimates of variance, the degrees of freedom, and the significance of

C6

the trends. – P17. L303+. ‘Only four North American sites experience later breakup during...’ The challenge here is that I can’t tell if the authors are assessing significant or significant + non-significant trends. Was the example of Frame Lake a significant or non-significant change? The slope is reported (which appears quite close to zero), but no p value or measures of variance. – P17. L309+ Given the large scope of the manuscript, it seems ill advised to drill down to the extent here for individual lakes or rivers. A blow out box may be more appropriate for dealing with specific case studies if there is a strong need to provide an example. The removal of these details will make it much easier for the reader to follow the broader relevance of the study and more space for the authors to discuss the relevance of their analysis compared to similar studies. – P49. L826+. I’m not sure I follow the logic here. How do summer wind speeds alter ice break up in the spring (either the preceding or following spring)? This is likely a false positive and one would expect some false positives and negatives given the large number of tests. Second, the link between summer windspeed and effects on turbulence are only relevant once air temperatures drop below 0 deg C. Summer windspeeds could be important by deepening thermoclines and higher heat storage, leading to delayed ice out dates. Again, potential for false positive here. Further, I am not surprised by the poor relationships over North America and Russia given the large geographic area and changes in the distribution of sites. As noted by the authors in other locations (e.g. p18 L325), there can be large regional differences in climatic patterns within the study area that can lead to regional differences in ice phenology trends (e.g. NAM sites). While more labour intensive a better approach would be to examine the linkages between climate and ice phenology at the individual sites, and then bringing these relationships into a more global analysis (i.e. does ice phenology respond similarly to climate across sites with regional differences in climate causing regional differences in ice phenology, or are other factors (e.g. lake or watershed morphometry, lake vs. river, etc. critical). – Figure 4. Caption should indicate that trends were tested using Mann Kendall and Sens slopes. After looking at Figure 3 from Duguay et al. 2006 (which I think is largely the same data), most of the ice freeze up values

C7

(panel e in this manuscript) would appear to be non-significant. This makes me wonder whether all trends are reported regardless of significance. The authors need to clearly state in the figure caption whether the trends being reported are significant and what level of significance. – Table 3 is confusing. The authors must indicate what statistics were used (Mann Kendall with Sen’s slopes?). What do the numbers represent in the panels represent (Sen’s slopes?). What are their units (days/decade?). Why are significance values included only in the last three panels/time periods but not the first panel? Does the dash represent ‘no data available’? Why not simply use the first panel data? Note, that another interpretation of the data from this table is that few lakes or rivers are displaying significant trends. For example in the 1931-90 period only 4/15 sites show significant changes at the $p=0.05$ level, or 7/15 at the $p=0.10$ level. Overall the table is interesting, but these site specific tables should probably be moved to supplemental material. – Table 3 shows that from the period 1991-2005 the majority of N America Lakes included in the analysis had either no significant trend, or had later ice on dates (significant positive numbers). There were no significant negative numbers. Yet, when I examine Figure 4, I see a large number of earlier freeze up dates reported (particularly to the east of the Laurentian Great Lakes). Am I missing something? – Figure 5 caption is poorly written. What does ‘Comparison of how sites in North America with an open water season calculation’ mean? Perhaps ‘Patterns of ice freeze-up and ice breakup...’ What level of significance? What statistical test? Were non-significant trends reported as trends? Do white boxes represent no significant trend or no data? – Figure 17. This figure is important as it represents the only example of where the drivers of the phenological trends were being tested. Although the methods are not entirely clear, the analysis seems to have a few issues. I am assuming that the authors assessed the trends of the ‘downscaled’ annual mean or median values of the freeze-up breakup dates over the full study period (1931-2005) with each study region (FEN, NAM, RUS). How was the air temperature data treated? Was it the mean value of the whole study area? This presents some problems because in both North America and Russia, the study area is quite large and the number and

C8

geography of sites changed dramatically over the study period. For example, no Canadian sites were available during the most recent period. In this case, I would think that this would have a large effect on the relationship between air temperature and ice phenology. For Russia, there was a wide spread of sites (25-27 sites) longitudinally during the 1961-1990 period, but much fewer sites during the periods before (5-6 sites) and after (1 site). A better approach would be to examine the downscaled air temperature values for each site specifically. This would also remove potential problems associated with elevation. Further, a large number of correlations are being tested (e.g. >400 in panel b), which leads to the potential for false positives. A minor point, but presumably the bolding pattern refers to the level of significance of the trends. This would be important to point out in the caption. Perhaps I have misunderstood a few aspects of these analyses and in general more clarification of the statistical approach is needed.

Interactive comment on The Cryosphere Discuss., <https://doi.org/10.5194/tc-2020-172>, 2020.