Response to: Interactive comment on The Cryosphere Discuss., doi:10.5194/tc-2016-108, 2016.

We thank the reviewer for the useful comments and recommendations, which have helped clarify the manuscript. Specifics of changes in response to the review are outlined below and are shown in track changes in the manuscript.

We had missed the supplemental material in Eisenman et al. (2014) that examined the inconsistencies between different versions of the Bootstrap and NASA Team algorithms for the Arctic. This is an important omission, and we have included references to the paper and the supplemental material in the revised manuscript. Of the fields considered in the supplement, ice extent findings are analogous to the open water area measure we are using. We note that we use the 95% and 99% confidence intervals, which are well within the 68% error bars suggested by, for example Comiso and Nishio (2008) and 90% confidence interval used by the IPCC AR5.

Comiso, J. C. and Nishio, F. 2008. Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research*, 113, C02S07, doi:10.1029/2007JC004257.
Eisenman, I., W.N. Meier and J.R. Norris. 2014a. A spurious jump in the satellite record: has Antarctic sea ice expansion been overestimated? *The Cryosphere*, 8, 1289-1296, doi:10.5194/tc-8-1289-2014.
Eisenman, I., W.N. Meier and J.R. Norris. 2014b. Supplement of: A spurious jump in the satellite record: has Antarctic sea ice expansion been overestimated? *The Cryosphere*, 8, 1289-1296, doi:10.5194/tc-8-1289-2014-supplement.

We apologize for the brevity of the description of the exploratory phase of the analysis. We elaborate on the process here and have added this to the text. There are several different ways to test for structural changes in a time series. We used several and checked using an additional one on the basis of this review. As noted in the text, in an exploratory mode, we split the record into sub-periods and tested the differences in the means. In addition, we examined the data by looking for the years that displayed the largest change in one moving three year average to the next. The two largest changes that were already not included in a previously identified six year period (two three-year moving average periods) using this measure were 1988-1989 and 2006-2007. The next stand-alone change had a much smaller change and the magnitude of the change was relatively similar to the next few candidates. While there is no particular justification for choosing the top 2, 3 or N values, the first two stood out as noticeably different in magnitude than the ones following and suggested reasonably long periods. We also ran a statistical test of the quality of various models of the time series. For example, we tested two structural shifts using a model that compared the statistical significance of a model that defined three different means and no trend to a model that defined a single trend. The former model was significant and the trend was not.

In order to increase confidence in the signal, we have now applied the method of Rodionov et al. (2004) to the NASA Team (NSDIC) record, which yielded the same two breakpoints. While it also suggested a third breakpoint around 2001 (which is

also suggested in the graph in Figure 3), using that breakpoint would create a rather small structural break. In addition, a breakpoint at around 2001 was not as strongly supported by the three year rolling average test. We therefore limited our analysis to two breakpoints. We have also added this new analysis to the paper.

Rodionov, S.N. 2004. A sequential algorithm for testing climate regime shifts. Geophysical Research Letters, 31, L09204, doi:10.1029/2004GL019448.

As noted in the manuscript, however, the short record of the satellite derived time series is a limitation. In addition, as raised by Eisenman et al. (2014) and the reviewer, there is potential for processing artifacts. As noted on page 7, line 31, we do not rule out the possibility that these shifts may be associated with artifacts in the satellite record. As the reviewer observes, it is difficult to obtain a truly independent record of the Arctic sea ice record, which is why we often use the term "quasi-independent". That said, the Barrow-Prudhoe Bay navigable days time series is a good candidate for independence, and yielded the same behavior. Nevertheless, we agree that it is not possible to completely rule out the conclusion that these shifts are *not* driven by physical processes. We feel that we have taken this as far as we can in the data realm and are conducting ongoing analysis using a model.

Finally, if as we suggest the structural shifts *are* physical in nature, there is no reason to expect that they should be periodic. If these shifts represent, for example, tipping points in ice thickness, then there is no inherent oscillation in the system that is likely to generate this behavior (although of course that is an hypothesis that could be tested in a climate model.)

Minor revisions

2, 24-25: Note that the CDR simply combines the NASA Team and Bootstrap estimates, so the different products are not completely independent.

The reviewer is correct in that there is a merged product 1979-2014 and a CDR 1988-2014. It may be semantics, these products in our view are less of a merging of Team and Bootstrap and more of choosing the least bad one, at least in September. But indeed, none of these products are independent and we have added a clarification to that effect.

2, 27: The "merged" product uses data that underwent manual quality control over its entire record, 1978-2014, not just the period prior to 1988.

We have corrected this error.

2, 30-31: It is true that PM products do not detect ice well at low concentrations, but the 15% concentration threshold is also selected based on the relatively low spatial resolution of the sensors, which makes any ice edge somewhat "fuzzy". I believe the 15% threshold was originally defined because it was found to best match the "true" ice edge during validation studies.

We have added this criterion.

3, 10: I think HadISST2 uses OSI-SAF concentrations since 1979, at least according to Titchner et al., 2014.

Titchner and Rayner (2014) confirm that HadISST2 uses OSI-SAF (Ocean and Sea Ice Satellite Applications Facility) passive microwave retrievals, which is product using SMMR, SSM/I, SSMIS, and EMSR (1972-1979). HadISST2 does *not* use EMSR. Interestingly, the preferred method for processing the data is a combination of the NASA Bootstrap method and the Bristol algorithm (Smith 1996, Hanna and Bamber 2001). The Bristol algorithm applies to Antarctic sea ice, meaning the NH product is effectively Bootstrap.

Hanna, E., and J. Bamber (2001), Derivation and optimization of a new Antarctic sea-ice record, Int. J. Remote Sens., 22(1), 113–139, doi:10.1080/014311601750038884.

Smith, D. M. (1996), Extraction of winter total sea ice concentration in the Greenland and Barents Seas from SSM/I data, Int. J. Remote Sens., 17, 2625–2646.

Titchner, H. A., and N. A. Rayner (2014), The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, J. Geophys. Res. Atmos., 119, 2864–2889, doi:10.1002/2013JD020316.

3, 13: It's worth noting that Radarsat data started in 1995. This represents a substantial change in NIC's chart. Before then, they used a lot of PM data; after, it was much less. Also it would be good to include a journal reference for the NIC charts, e.g., Dedrick et al.

We have added this observation and included the reference.

4, 4: Not sure what you mean by "rasterized"? The source data are already on a grid. Maybe rephrase this sentence.

The original ice data is on a grid but then the tracking scheme is Lagrangian and hence no longer grid based. As a result, the derived ice age must then be re-gridded.

5, 23: The HadISST does use NIC charts, but only as a bias correction.

Titchner and Rayner (2014, see reference above) on page 2869 state that "The NIC chart record appears stable relative to OSI SAF after this time; therefore, we chose to use the NIC chart data from 1995 onward as the representation of the "true state" against which to adjust the relative biases in the other data sources used." For this reason, we feel that the characterization "only as bias correction" underplays the role of the NIC charts. We clarified this role on the manuscript.

5, 25: I think the resolution likely accounts for a lot of the differences between the 100 km HadISST and the 25 km PM records.

We agree, and have added a comment to this effect.

7, 15: It would seem more relevant to use the same period as the PM record, i.e., 1979-1988 instead of 1953-1988, etc. Would that potentially affect the analysis?

This would not affect the analysis except near the start and end of the records, which is why we note that significance is lower in the case of the passive microwave data records for the late 1980's breakpoint. But the breakpoint is still detected, using the multiple means described above.

7, 25: Adding the Eisenman reference (above) somewhere here would be useful

The Eisenman et al. (2014) references have been added to the discussion in several locations.

7, 26: What is meant by "errors in the data stream in the first months of the SSMI instrument"? Do you mean the first SSMI, launched in 1987, on DMSP F8? I'm not aware of such errors. The Harvey reference mentions recent errors in 2016, but these don't seem relevant to analysis at hand.

Yes, this is correct (the first SSM/I, launched in 1987, on DMSP F8). We have this information in an email from J. Stroeve on May 2nd 2016, in which she documented the small error, and she estimated it would not impact the results by more than 15% (and hence unlikely to be the cause the structural shifts we detected.) We have removed the other (Harvey) reference, as the reviewer is correct – it is discussing a more general issue.

7, 33: The authors say Week 37 is used for ice age. Is this consistent past the sea ice "birthday"

According to the documentation for the sea ice age product, the sea ice age "birthday" is defined by the September minimums; that is, it does not increment in age until a minimum is passed. This typically occurs in September, according to the documentation, although not, as the reviewer notes, on a particular week. Our figure 5 is based on the data in the week 37 file; we have not applied interpolation or extrapolation in order to maintain consistency with the manner in which we have calculated average ice areas.

8, 4: The age algorithm doesn't so much "favor" the oldest ice as it effectively records only the oldest ice in a given cell.

We have changed this phrasing.

Figure 2: The caption says the units are % open water, but the y-axis is 0-0.8. I assume the y-axis is actually sea ice fraction (unit-less, 0-1 range)?

We have corrected the caption.

Figure 3: Same comment as for Figure 2. Both figures would also benefit from a bigger font for the axis numbers and legend.

The caption does use fraction, not %. We will supply larger fonts in the publication quality figures.

References: Just a style comment and this may be a formatting standards issue, but it would be helpful to indent the Reference list. It's very hard to pick out individual references.

We have conformed the Cryosphere format template which did not include indentation of references.