Dear colleagues,

thank you for the revision of the manuscript. I still have some minor issues before I can accept the manuscript for final publication.

1) I tried to get the data from the website http://www.ec.gc.ca/glaces-ice/ to reproduce the plots. It seems the data are distributed over two tables and it is not clear for me how to merge both into a consistent time series. Moreover, only for Alert there are data for the period 2003 until present. Thus, it was not possible for me to reproduce your results. Could you please add the data used for the analysis to a supplement? It would be good if you could also include the supporting air temperature from reanalysis. Please note the data source in Fig. 7.

Howell et al.
The stations we used (Cambridge Bay, Resolute, Eureka and Alert) have indeed remained unchanged over the long term record and merging them is very easy because the date of the measurement is given. In order to add value to our paper we have included a merged Excel spreadsheet for both ice and snow thickness. We have also included a spreadsheet for the temperature data which is observed by Environment Canada and not reanalysis.

For this observed data we have added in the data description that the data is available in "supplementary material" We have changed the caption in Figure 7 to reflect the data source as well as in the Data Description.

Data Description Changes:
"The other source of observed data used in this study was Environment Canada's monthly mean air temperature records at Alert, Eureka, Resolute, and Cambridge Bay (see supplementary material) for which a complete description is provided by Vincent et al. [2012]."

"Values of maximum or end-of-winter ice thickness and corresponding snow depth during the ice growth season were extracted from the weekly ice and snow thickness data at the selected sites (see supplementary material)."

New Figure 7 Caption:
Figure 7. Time series observed mean air temperature by Environment Canada during winter (DFJ), spring, (MAM), summer (JJA) and autumn (SON) at the Cambridge Bay, Resolute, Eureka and Alert.

2) Section "Models" includes the methods of statistical analysis mixed with the climate/ocean models. Please distinguish between the trend analysis and the description of climate/ocean models and change the structure. The "methods" are now under the section "Data description". You may add another section "Methods" to describe the statistical test for significance in more detail. I have not fully understood your procedure for the test of significance and would like to learn more about the noise model. The reference to the two papers is not enough to reproduce your results.

Howell et al.
We modified the Data Description Removing lines 109-116 replacing them with the following:
"We obtain the multi-model mean of trends and their statistical significance at each grid point by creating the distribution of trends through a Monte-Carlo simulation. We use a t-distribution for the interannual variability and build a noise model to account for internal variability as in Swart et al. [2014] and Laliberté et al. [2016]. We obtain the multi-model mean of Pearson correlations and their statistical significance by first performing a Fisher transform and then applying the same method as for the trends. The inverse Fisher transform is applied after obtaining the multi-model mean and its significance. See the appendix for a complete description of the method."

We then added an Appendix describing our statistical approach as follows:

"The Monte-Carlo simulation used to combine trends and Pearson correlations is applied at each grid point independently. Models that have a land mask at a grid point are discarded before starting the procedure.

A noise model is created to ensure that internal variability is comparable for models with different ensemble sizes, following Swart et al. [2014] and Laliberté et al. [2016]. To generate the noise model, we discard models that have fewer than two realizations. From the remaining models, we pick one and then one of its realizations. We then record to the noise model the difference of this realization's trend from the mean trend of the model's realizations, multiplied by $(n/(n-1))^{1/2}$, with n being the number of realizations, to account for the fact that some models have such a small number of realizations that it cannot completely account for the internal variability. We repeat this procedure 1000 times and compute the variance σ_n of the noise model.

We then pick a model from which we select 1000 realizations, allowing repetitions. For each one of these realizations, we select a random value from its trend t-distribution. If the inter-realization trend variance σ_m is smaller than the variance of the noise model σ_n, we then draw a random value from the noise model, multiply it by $(1-\sigma_m/\sigma_n)^{1/2}$ and add it to the random value from the trend t-distribution.

We repeat this procedure with the remaining models. We then average the 1000 values across models, creating a distribution for the multi-model mean trend with 1000 values. The mean of this distribution gives our multi-model mean and its two-sided p-value is given by twice its survival function or cumulative distribution function at 0, whichever is smallest.

The Pearson correlations are analyzed in the same way except that a Fisher transform (obtained by the hyperbolic arctangent of the correlation) is applied first and random values are drawn from a normal distribution (instead of the t-distribution) with variance 1/(T-3), with T the number of years used for the correlation. The multi-model mean Pearson correlation is then given by the inverse Fisher transform (obtained by the hyperbolic tangent of the mean) of the distribution mean."

3) COV is usually used for covariance, at least I do so. Please use the more common CV for coefficient of variation.
Howell et al.
Changed.

4) Table 2) PIOMASS -> PIOMAS
Howell et al.
Changed.

Additional Note:
We have included a new version of Figure 14. The original text referred to this version of the Figure but we included an older version by mistake in the original submission.