Reviewer 2

Summary

In this work the authors demonstrate the synthesis of hydrographic and sea ice concentration data into a 16-km horizontal resolution Arctic and North Atlantic coupled sea ice-ocean model. The reduction of an uncertainty-weighted model-data difference cost function was achieved by iteratively optimizing a set of adjustments to a set of atmospheric and initial condition control variables using gradient information provided by the adjoint of the numerical sea ice-ocean model. The final multiyear state estimate was constructed by optimizing each single year between 2000 and 2008 in succession - the final optimized state of year X is defines the initial state for year X+1. The authors demonstrate improvements of the model's reproduction of the data. The largest reduction in terms of percentage is found with sea ice concentration and SST with lower relative cost reduction for other data, including T and S profiles, SSH, and mean dynamic topography. The largest sea ice concentration cost reductions in terms of RMS are found during summer months. Discrepancies between simulated and observed sea ice extent are found to increase in some months even when discrepancies in simulated and observed total sea ice area decrease. After synthesizing ocean and sea ice data, little impact is seen in ocean volume, heat, and freshwater fluxes through Fram Strait and Davis Strait.

We thank the Reviewer for the thoughtful evaluation of our manuscript. In the following Reviewer's comments are in italic, our answers are in usual font, text from the manuscript is in the quotation marks and the new text is in blue color.

Specific Comments

1) With respect to the title, assimilation is not "into a Coupled Ocean-Sea Ice Adjoint Model". The assimilation is "into a Coupled Ocean-Sea Ice Model using its adjoint".

We thank the Reviewer for this suggestion and changed the title accordingly.

2) Abstract: Better to provide the actual spatial resolution of the satellite sea ice concentration data that is assimilated rather than refer to it as 'high resolution'.

We decided to remove the reference to the resolution altogether because the sea ice data are assimilated, as the other data, on the model grid. The sentence now reads as follows:

"Satellite sea ice concentrations (SIC), together with several ocean parameters, are assimilated into a regional Arctic coupled ocean-sea ice model covering the period 2000-2008 using the adjoint method."

3) Page 1, Line 6: 'values of sea ice extent become underestimated' doesn't define a metric. Is the metric the sum of model minus data or weighted model minus data difference or the RMS of model minus data or something else?

We have tried to make the statement more precise, now it reads as follows:

"During summer months, values of sea ice extent (SIE) integrated over the model domain become underestimated compared to observations,…"

4) Page 1, Line 6-7: Characterizing a state estimate of a system as complex as the Arctic Ocean requires that one analyzes a suite of metrics. The author's statement that one the sea ice extent metric is "not suitable to characterize the quality of the sea ice simulation" is odd and out of place. To whom is this statement aimed? This seems to be a straw man argument.

The statement is more related to the practice of characterising the quality of a sea ice simulation (not the complete Arctic system) in the model by only considering one metric, namely the integrated Northern Hemisphere September sea ice extent. This is still common in many publications and authors are also guilty of this sin. However, we agree with the Reviewer in that the abstract is not the right place for such a statement and now the sentence reads as follows:

"During summer months, values of sea ice extent (SIE) integrated over the model domain become underestimated compared to observations, however the root-mean-square difference of mean SIE to the data is reduced in nearly all months and years."

5) Page 1, Lines 10-11: The atmospheric control variable adjustments that one finds during any optimization are intimately related to the magnitudes of the prior uncertainties of the individual terms of the first-guess atmospheric state. The author's statement that biases in sea ice are reduced 'mainly due to corrections to the surface atmosphere temperature' is difficult to interpret because the reader does not know the magnitude of prior uncertainties used during e optimization. Are you referring to the sum of the squared normalized adjustments? How is surface atmosphere temperature identified as the main control variable correction since atmospheric forcing units are arbitrary?

As suggested by the Reviewer, we have re-evaluated our analysis of the corrections to control variables and decided to remove the statement about the 2-m air temperature contribution to the improvement of the model state.

6) Page 2, Line 8. The authors may consider using the term 'state estimation' to describe the model-data synthesis methodology used in this study instead of the term 'data assimilation'. An uninformed reader may think that the work conducted here referring to sequential data assimilation, a technique that has been applied to sea ice data for decades. The adjoint method used in this work is rather special and yields quite a different product (namely a physically-consistent ocean and sea ice state). Below I post an excerpt from Wunsch and Heimbach, 2007 in which they argue for their choice of the term 'state estimation' when describing the application of the adjoint method to combine data with a model (emphasis mine): "In physical oceanography, the problem of combining observations with numerical models differs in a number of significant ways from its practice in the atmospheric sciences. It is these differences that lead us to use the terminology "state estimation" to distinguish the oceanographers' problems and methods from those employed under the label "data assimilation" in numerical weather prediction. "Data assimilation" is an apt term, and were it not for its prior use in the meteorological forecast community, it would be the terminology of choice. But meteorologists, faced with the goal of daily weather forecasting, have developed sophisticated techniques directed at their own particular problems, along with an opaque terminology not easily penetrable by outsiders. Because much of oceanography has goals distinct from forecasting, the direct application of meteorological methods is often not appropriate."

The term “adjoint data assimilation” is used in the community and even the usage of the term “state estimation” exist for applications with the Kalman filter. Since there is no agreement in the community, the reader has to anyway carefully read the methods section of the paper in order to understand how exactly data were used to improve the model. We agree with the Reviewer that the term “state estimation” is probably better to put the reader on the right path. However, “state estimation” inherited the flavour of trying to estimate a static climatological state, as it was attempted in the first applications of the adjoint method during the WOCE era, therefore we find it to be a less appropriate term; but again, current usage of both terms show that the nomenclature is not well defined and although we could live with the term “state estimation”, we don’t find it really better.

7) Page 4 Line 10-11: List the control variables.

The list of all control variables can be found in the text below. We did not use the longwave radiation as a control variable for final simulations. We adjusted the text accordingly.

8) P4 Line 20: Describe why the atmospheric control variable frequency was changed to daily.

We incorrectly used information from a different experiment setup, so the frequency of updates is actually once per three days, which is still higher than 10 days used by Köhl (2015), who has chosen 10 days due to computational (memory) limitations. We added the requested information and now the sentence reads as follows:

“In contrast to Köhl (2015), additional control variables are optimized and the frequency of the updates is enhanced to once per 3 days in order to reflect shorter time scales of sea ice variability.”

9) P4 Line 23: As atmospheric adjustments are an important control parameter in this work, the authors should (a) explicitly state how they were derived as Kohl (2015): "For the atmospheric state, errors are calculated as before from the [standard deviation] of the NCEP fields." And (b) show maps of their magnitudes in the main text or in supplemental materials. Also, because they are so important, more discussion about your choice of standard deviation of NCEP fields is appropriate. The standard deviation of Arctic near-surface atmosphere temperatures is considerable given the large seasonal cycle. In much earlier versions of ECCO/GECCO the use of atmospheric state standard deviations could be justified because in mid-latitudes and the tropics they partially captured "random" variations due to synoptic variability. At high latitudes the standard deviation for near-surface atmosphere temperature and shortwave radiation is mostly due to the seasonal cycle.

Although there is a large seasonal cycle the difference between the STD with and without the seasonal cycle is actually not that large for most of the globe; but it is true that in the Arctic region the error is with values around 12-30˚C overestimated by a factor of 2. The STD is in both cases relatively homogeneous, such that a figure would not provide valuable information.

We have added the following text to the manuscript:

”For the atmospheric control variables, uncertainties are specified as the maximum of the STD of the NCEP fields and the errors for the mean components of air temperature, humidity, precipitation, downward shortwave radiation and wind were specified as 1°C, 0.001 kg/kg,

1.5×10^{-8} mm/s, 20 W/m^2 and 2 m/s, respectively. For the downward shortwave radiation both mean and time varying parts were set to 20 W/m^2."

10) Page 4, Line 24: Why are the sea ice data assigned a constant 50% error? Satellite SIC products have errors that are far smaller than that everywhere except in the MIZ and in summer when meltponds are present.

We added an explanation in the text:

"We verified the sensitivity of our results by using space-time varying uncertainty estimates as they became available, as well as different values of a constant error. Results of the sea ice assimilation with variable uncertainties were very similar to the ones with constant error value of 50%."

11) Page 4, Line 27: To clarify, each year after the first uses initial conditions that are identical to the final state of the previous year, correct?

Yes, this is correct. We feel that the sentence: "After the first year assimilation, we move to the next year using the final state of the previous year's successful iteration as initial conditions.", describes this sufficiently.

12) Page 4, Lines 3-4: Some SST products have nonzero values beneath sea ice. Is that the case in the RSS dataset?

The RSS data have a "sea ice" flag that, in practice, means missing value. We didn't take the SST data if the "sea ice" flag was set.

13) Table 1: I understand that the PHC climatology had large biases relative to modern Arctic T and S because it was derived with observations mainly from the 1970's and 1980's and before the recent shifts in Arctic heat and freshwater (McPhee et al, 2009). Can you comment on how the simultaneous use of the PHC climatology alongside contemporary data may have affected the T and S cost reduction?

We used the PHC climatology only for model initialisation and it was not used in the data assimilation. We have removed it from Table 1. Since the model was started from PHC and in situ data is sparse, the model cannot be corrected very much away from the first guess, a point made in the text.

As mentioned below in the paper, the changes of the deep ocean state are quite small due to both the predominance of sea ice and sea surface temperature observations over interior hydrographic observations and the short assimilation periods (yearly chunks). Using a more recent Arctic Ocean state as initial conditions would certainly be beneficial due to an initially smaller cost in T and S. However, judging from the experience we gained during this exercise, we believe that there would still be hardly any significant cost reduction of T and S beyond surface layers in the Arctic Ocean, mainly again because of our experiment design and a much larger amount of sea ice data compared to hydrography data.

14) Page 4, Paragraph 1: Cost function reduction percentages are important but obviously they are dependent on how close to the data you were when you began your simulations. The first-guess solution of Fenty et al., (2015) could have been further from the data than your *first-guess solution. While both may end up in the same state, their reduction percentage would be higher. The most important information is how well one's final state estimate fits the data. Much less important is the magnitude of the improvement relative to one's (somewhat arbitrary) starting point.*

We agree that just stating the percentage reduction is problematic, but it is nevertheless important information about the performance of the assimilation. We cannot really assume that we found a minimum, and therefore success is usually evaluated by the amount of reduction. It would not be a too bad assumption that both controls perform about equally well. We have added the caveat of this comparison in the text:

"In 2004 the cost reduction of sea ice area was about 30%, less than that reported by Fenty (2015) (49%), which may partly be explained by differences in the first guess solution."

15) Page 4, before line 31: It may be useful to mention how many iterations were conducted before the 1% threshold was achieved. In Figure 3 I see "iteration 3" as the final iteration for 2005 and 2007. That strikes me as unusual. If your cost was dominated by SIC and SST data and the adjoint method quickly reduced the misfits of those data, then I can see how you hit the 1% total cost reduction threshold quickly. However, it is possible that if those two datasets were ignored, the adjoint machinery could have continued to substantially reduce misfits in other datasets. Can you comment on that?

The Reviewer is correct. The cost is dominated by SIC and SST. These data easily respond to the surface controls. We added this explanation:

"The cost is dominated by SIC and SST data, which easily respond to the surface controls, and the adjoint method quickly reduced the misfits of those data, so that the number of iterations was usually less than five."

16) Page 5, Line 15: There may be a missing figure. I cannot match up Figure 2 to the description offered here. Fig 2 is % cost reduction in different years vs. data.

Thank you very much for spotting this. The first paragraph of the section "Sea ice concentration changes" was from a previous draft version. We did not intend to include it in the manuscript. It is now removed.

17) Page 5, Line 24: Good to additionally mention why most models overestimate sea ice in the Greenland Sea with a reference.

This statement was referring more to results of climate models (see for example Figure 9.23 in IPCC AR5 Chapter 9). It is not correct to transfer this result to regional ocean-sea ice models because much of the bias in climate models result from biases in the atmosphere. So we have removed this part of the sentence. Now the end of the sentence reads as follows:

"Most noticeable is the decrease in the SIC along the east coast of Greenland after data assimilation"

18) Page 5, Line 35: This is probably because in these extreme months the location of the sea ice edge is relatively stable compared to spring and fall months when the ice pack contracting and expanding.

We thank reviewer for this explanation, which we added in the text with a slight modification. Now the text reads as follows:

"Interesting to note, values of RMSE in March and September are quite similar, despite the large differences in ice cover in the two months. One of the possible reasons is that the location of the ice edge in those extreme months is relatively stable compared to spring and fall when the ice pack is contracting and expanding."

19) Section 4: This entire discussion must be rewritten. Atmospheric control variable adjustments seem to be compared by their relative magnitudes but their relative magnitudes are not meaningful because these physical variables have different, arbitrary, units. By all means show the magnitude of the adjustments but to make a meaningful comparison one should first normalize them by their prior uncertainties. a. This includes Figure 8, which should be updated to show all control variable adjustments normalized by their uncertainties. Also include longwave radiation.

Neither dimensional, nor normalized values provide the impact of the changes per se. The prior errors of controls have nothing to do with the impact or even the anticipated impact but describe only our knowledge about them. Moreover, our choice of STD does not make a difference because there is no reason why one STD of perturbation should have a similar impact across all parameters.

Nevertheless, in the optimization the parameters enter normalized, and corrections are generated according to the normalized sensitivities and the approximation of the Hessian matrix. Since we have only a few iterations completed, the Hessian stays not very far away from its initial value, which is the identity. Therefore, in this special case the normalized corrections will still more or less reflect normalized sensitivities. Since the impact is the product of the sensitivity and the corrections, normalized corrections will provide a reasonable measure of the relative importance of the parameters.

As suggested by the Reviewer we have added the normalized corrections to Fig. 8 and reworked the section. The additional text reads as follows:

"Dimensional values of the corrections do not directly provide information about the relative importance of changes in the controls for bringing the model into consistency with observations. However, due to the relatively small number of iterations, we can use values of the corrections normalized by uncertainties as a reasonable measure of the relative importance of changes in control parameters. Spatial distributions and monthly means of absolute values of normalized corrections for the year 2005 are shown in Fig. \ref{fig:8}.

Wind corrections seem to play integrally a larger role, with a maximum in May. This agrees well with results of \citep{Kauker2009}, who used an adjoint sensitivity analysis to determine the relative contribution of different atmospheric and ocean fields to the September 2007 sea ice minimum and found that the May-June wind conditions are one of the main factors in setting up extremely low sea ice conditions in Summer 2007. The maximum contribution of air temperature corrections occurs in June and it is about a factor of five smaller than the contribution of the wind corrections. However, using free drift in the adjoint biases the sensitivities towards larger sensitivities of sea ice to wind changes. Since measuring the impact by the normalized corrections relies on the assumption of correct sensitivities, the results may be also biased to too large an impact by the wind.

Given the absence of proper sea ice dynamics in the adjoint model (only free drift is used) and lack of many important processes in the forward model (such as tides or waves), the question remains to what extent corrections to control variables reflect deficiencies in the forcing fields or a compensation to the sea ice model or sea ice data deficiencies, particularly since in the Arctic the NCEP reanalysis seems to perform well near the surface \citep{Jakobson2012}."

Showing only normalized adjustments maybe tells a lot to people involved in the adjoint community, but for most people it's just easier to look at absolute values that have some physical meaning.

The long wave radiation was not a control variable in our final simulations, so we can't show it. We have changed the text in the method description accordingly.

20) Page 8, Line 20-22. The "probably realistic" spatial distribution of the Kwok Arctic sea ice thickness field deserves a reference. Are the 0.7 m errors spatially correlated or uncorrelated?

The 0.7 m is a mean error; the individual values would vary of course, depending on the sea ice thickness. Reference to Kwok et al. (2008) was added.

21) Page 9, Line 22-24: Neither the length of the simulation nor the number of T/S profiles is a fundamental impediment to magnitude of model-data misfit reduction. An iteration 0 state with T and S close to the data as measured by the prior uncertainty could be responsible. Maybe averaged normalized costs should be added to Figure 2 for each cost category for iteration 0 and the final iteration.

The reason why we think the number of data and time period of assimilation matter is that the data information has to be able to reach the sensitivities to the controls in the adjoint model. All data from year 2 and later is excluded from modifying the initial condition due to the separation into 1 year windows. The time window of one year, on the other hand, is too short for deep data signals to be able to reach the surface. Sparse data is in general a problem because, due to the lack of covariance information, sparse data is likely to produce unphysically small-scale corrections, which are likely to be not beneficial for the simulation of the dense SST and SIC data that determine most of the cost.

22) I may be incorrect but it seems that no Arctic Ocean T and S pro- files were used in this work. I do not see Arctic Ocean data in the Ingleby and Huddleston report and the NISE database doesn't show data north of the Norwegian Sea. Given that the assimilation period overlaps with the existence of ice-tethered profilers, why were ice-tethered profile data not included (http://www.whoi.edu/page.do?pid=20781)? As for the CTD data in the Arctic, both the ICES database (http://www.ices.dk/marine-data/data-portals/Pages/ocean.aspx) and the World Ocean Database v3 (https://www.nodc.noaa.gov/OC5/WOD13/) have data for the time period considered in this work. There may be perfectly fine reasons for excluding these data but the reasons should be offered.

At the time we have started our assimilation efforts (year 2012), the combination of the EN3 (which includes a good amount of Arctic Ocean T and S profiles) and NISE dataset was the best available option in terms of data coverage and technical efforts were required to interpolate observations to the model grid. Later we decide to stick with this choice for consistency.

We now added the following:

“The collection of hydrographic observational data in the Arctic Ocean used in the present work is not comprehensive and does not include, for example, ice-tethered profile data. In the present pilot study we decided to stick to two well-structured data sets available at the time we have started our efforts.“

Technical Corrections

1. Page 1, Line 5: change ’become’ to ’are’ as in ’values of sea ice extent are underestimated’

Corrected.

2. Page 1, Line 5: first comma to semicolon. Or split this long sentence into two before ’however’

We changed it to semicolon.

3. Page 1, Line 14: strike ’to date’

Corrected.

4. Page 1, Line 16: reference?

We now cite Overland and Wang (2013).

5. Page 1, Line 17: strike comma before ’is therefore of utmost importance’

Corrected.

6. Page 1, Line 24: strike ’if not possible’

Corrected.

7. Page 2, line 2, strike comma before ’the community’. Strike ’heavily’.

Corrected.

8. Your doi for Detlef’s 2016 paper is incorrect. It should be DOI: 10.1146/annurev-marine-122414-034113

We double checked and could not find a difference between the DOI that you have provided and what appears in the paper. Can you please specify what exactly is wrong in the DOI?

9. Page 2, Line 19: strike "usually in general"

Corrected.

10. Page 5, Line 12: strike "are going to"

Corrected.

11. Page 5, Line 28: replace "very good" with "improved"

Corrected.

12. Page 5, Line 29: strike "thus"

Corrected.

13. Page 5, Line 31-32: This sentence deserves a rewrite for clarity. As mentioned above, relative percentage sea ice cost reductions are also a function of the (unknown) first guess states.

The year in Fenty et al. (2015) is different as well as the first guess, so we remove the sentence completely.

14. Add 'bears' before 'a good resemblance'

Corrected.

15. Page 5, Line 24: For clarity consider saying 'since a perfect total sea ice area evolution...' and the following sentence is redundant.

We removed the redundant sentence and modified the sentence in question to comply with Reviewer 1 request as follows:

"Both metrics suffer from the inability to guarantee that improvements in this metric also lead to an overall improved match in the spatial sea ice coverage, since a perfect total SIC or SIE evolution may still correspond to considerable differences to the data in their regional distribution."

16. Page 8, Line 20-22. Strike "except for the" and simply say that "Sea ice thickness are not provided by Kwok for the Barents and Kara Seas and the Canadian Archipelago because ..." with a reference.

We deleted part of the sentence after "except for the" since it does not make sense to discuss uncertainty or realism of the data in the regions where they are not present.

17. Page 8, Line 26: change "variables" to "variables'"

We guess the Reviewer meant "to variable's". Corrected.

18. Page 9, Lines 1-2: Why is it hard to provide quantitative estimates? You could plot time series of the uncertainty-weighted squared model-data misfit (normalized cost) before and after the assimilation.

The Reviewer is right. Quantitative metrics are not hard to provide in general; we think the visual comparison of spatial distribution is more instructive than just a few numbers. We removed the respective sentence.

19. Plotting model minus data or model minus data squared in Fig 5 might simplify comparison.

Although your suggestion allows for an easier evaluation of the improvement, we decided to continue showing absolute values since we believe it is easier for most readers to interpret. Adding separate panels with differences would just duplicate the information and make the figure unnecessarily verbose.

20. Section 4: Fonts on the time series of Fig 8 are also small and difficult to read. One subplot is cut off. After normalizing the summed control variable adjustments they could all be shown in the together in the same plot.

We now made the fonts of the time series in Fig. 8 larger.