The Cryosphere Discuss., 9, C3035–C3060, 2016
www.the-cryosphere-discuss.net/9/C3035/2016/

The Cryosphere
Discussions
Open Access

Interactive comment on "Seasonal sea ice predictions for the Arctic based on assimilation of remotely sensed observations" *by* F. Kauker et al.

F. Kauker et al.

frank@oasys-research.com

Received and published: 23 February 2016

Response to Reviewer comments on the manuscript "Seasonal sea ice predictions for the Arctic based on assimilation of remotely sensed observations"

23 February 2016

0 General Response

We thank the editor and the reviewers for their careful inspection of the manuscript and their valuable comments. Their main point is that the complex material is quite condensed and its presentation needs to be improved to be attractive for the wide audience of TC. Indeed we find this very challenging and made an effort in the revised version to meet: "The primary recommendation of this review is that the authors put themselves in the position of someone trying to replicate their experiments and edit their paper accordingly."

In particular we have now

- extended the introduction to better motivate our objective and approach with the steps required, including the crucial role of the calibration (model tuning);

- extended the conclusions section and also added a number of general recommendations/take home messages derived from the results;
- extended the description of the data assimilation approach (section 2.1), including definitions of the jargon;
- extended the description of the numerical model (section 2.3), in particular the processes affecting the sea ice simulation.

In the following we address comments by the anonymous referees (quoted in *italics*) point-by-point. Where we quote from the revised manuscript changes are marked in **bold faced letters**.

1 Comments by Anonymous Referee #1

General: This paper describes a tuning exercise and a set of sea ice data assimilation experiments using the NAOSIM model, designed to improve the performance of this system in producing seasonal sea ice forecasts for the Arctic. General comments: This study's merit lies in the fact that it has performed a multivariate sea ice assimilation (concentration, snow and thickness), which is a new area. However, the authors should make more effort to describe the papers unique contribution and importance in the preamble. The study is quite technical and the authors make little effort to make the material accessible to a wide audience and it is not clear to me what the studies key findings/recommendations to other forecasting centres are. For example, lots of DA jargon is used, which is not explained. I found that the Conclusions section was more a summary of what has been done than a digested message to take away from the paper. Due to the technical nature of the study, I think it is particularly important that the conclusions are clear, since many readers will be interested in the papers message, but not full details. I suggest that in the authors add a paragraph to the conclusions section

discuss recommendations about the usefulness of this method for other forecasting centres.

We have modified the abstract such as to better stress our unique contribution:

"The recent thinning and shrinking of the Arctic sea ice cover has increased the interest in seasonal sea ice forecasts. Typical tools for such forecasts are numerical models of the coupled ocean sea ice system such as the North Atlantic/Arctic Ocean Sea Ice Model (NAOSIM). The model uses as input the initial state of the system and the atmospheric boundary condition over the forecasting period. This study **combines, for the first time, remotely sensed observations of four variables of the ocean sea ice system in a data assimilation system. The four data streams are the Alfred Wegener Institute's CryoSat-2 ice thickness product, the University of Bremen's snow depth product, and the OSI SAF ice concentration and sea surface temperature products. The assimilation system, built around NAOSIM, uses a variational approach with a two-month assimilation window, in which all observations act simultaneously as constraints on the initial model state.** We investigate the skill of predictions of the summer ice conditions issued in **May** for three different years. Straightforward assimilation of the above combination of data streams results in slight improvements over some regions (especially in the Beaufort Sea) but degrades the over-all fit to independent observations. A considerable enhancement of forecast skill is demonstrated for a bias correction scheme for the CryoSat-2 ice thickness product that uses a spatially varying scaling factor."

My understanding is that this is more a model development paper, rather than a scientific one. I therefore suggest the authors change the title to: "Developments to a seasonal sea ice prediction system using remotely sensed observations", or similar and frame the rest of the paper like this. Then it is clearer that it is mainly a development paper.

A lot of information is presented in the paper, but it is not always clear why certain *aspects of the analysis are important, or why certain graphs have been included. It is therefore quite impenetrable. I suggest the authors carefully edit the paper to stitch the different parts of the analysis together as part of a coherent story.*

We have addressed most of the general comments through the substantial revisions as explained in section 0 of this response.

We also have changed the title to: Towards seasonal sea ice predictions for the Arctic based on assimilation of remotely sensed observations.

Specific suggestions:

1. *The paper uses a lot of technical jargon, such as "Data stream" and "Operational Products". I suggest these are changed, or at least defined, to make the paper more accessible to a wider audience.*

   The introduction of the data assimilation method is now much more detailed. We also added a definition of data stream where the term is first mentioned.

   "The present article describes the **development** of an assimilation and prediction system of the Arctic sea ice conditions. **Ideally, such a system will combine with a numerical model with observational data of various types, for example in terms of variables (e.g. sea ice or ocean), scale of representativeness (e.g. point or two dimensional area), or observational approach (e.g. in-situ or satellite-derived). We term each of these types a data stream, and for the use in a combined assimilation/prediction system, an obvious requirement is their availability close to near real time. In this study, we use four data streams which fulfill this requirement,** namely, the OSI SAF sea ice concentration and sea surface temperature products, a snow depth product provided by the University of Bremen, and the CryoSat-2 ice thickness product derived at the Alfred Wegener Institute (AWI). **All above data streams are available from**

   2012. Also there is only one single two month period per year (March and April) for which the CryoSat-2 product is currently available (and note that the followup version covers October to April/May). We thus restrict our study to assimilation of the above four data streams in the spring of each of the three years **from 2012 to 2014** and to prediction of the ice conditions in the following summer."

   And also for "operational products":

   "For this purpose, the data streams have to be available operationally or have to become operational in the near future**, i.e. the product time series need to be continuously extended until the recent past (near real time).**"

2. *P5523:L21-22: See also Day et al. (2014) I think the introduction would benefit from some discussion of why we think sea ice should be predictable on seasonal and longer timescales. I suggest at least mentioning predictability studies such as Koenigk and Mikolajewicz (2009) or Holland et al. (2010), which provide this justification. A review of the state of the art in sea ice prediction is also presented by Guemas et al. (2014).*

   We added an introductory sentence on predictability to the relevant paragraph of the introduction:

   "**The potential for sea-ice predictions has been addressed by a set of studies (e.g. Kauker et al. (2009); Koenigk and Mikolajewicz (2009); Holland et al. (2010); Day et al. (2014)).** Predictions by a dynamical model depend on the state of the system at the beginning of the simulation period (initial state). Previous studies have highlighted the role of the initial ice thickness distribution Kauker et al. (2009); Holland and Stroeve (2011); Lindsay et al. (2012); Chevallier and Salas-Mélia (2012) for the forecast quality. Systematic use of observational information in a data assimilation system can help to derive an improved estimate of the initial state (Lindsay et al., 2012; Chevallier et al., 2013; Yang et al., 2014; Massonnet et al., 2015)."

3. *P5531: It is not clear what the purpose of the comparison with PIOMAS and TOPAZ is. I don't really think it adds too much to the paper. I suggest this is deleted.*

   As now elaborated in the introduction, the experimental section and the conclusions we think this preparatory calibration step is essential to correct biases. The comparison with PIOMAS and TOPAZ is used to evaluate the result of the calibration effort.

4. *P5532:L1-5: Why is NCEP reanalysis chosen as the driving data? There is evidence that ERAInterim performs much better in the Arctic (Lindsay et al., 2014).*

   We are using NCEP-CFSR and not NCEP for the most recent time period. The statement that Lindsay et al. (2014) have shown that ERAInterim performs much better is a bit too general: ERAInterim performs certainly better with respect to precipitation but performs worse with respect to the surface temperature where ERA-interim shows a strong warm bias over sea ice in summer (see Jacobsen et al. 2012, grl, VOL. 39, L10802, doi:10.1029/2012GL05159.) which prevents us from using ERAInterim.

5. *P5532;L6-11: I suggest the authors reiterate what they expect to learn from the different experiments here.*

   We have extended this introductory paragraph to section 3 and also added a table:

   "In this section three different sets of experiments are described **(see Table 1)**: The first set performs a straightforward initialisation through simultaneous assimilation of all four data streams described in section 2.2.2. **This set of experiments will exhibit a low forecast skill for summer ice concentration, pointing at remaining inconsistencies (despite our thorough calibration of the model). A second set of experiments is used to investigate the mechanisms for such inconsistencies: We do this by inferring the evolution of the ice thickness**

distribution when it is constrained only by the other three data steams but not by the CryoSat-2 observations. Based on these "reconstructed" ice thickness fields we will derive a bias correction scheme for CryoSat-2 ice thickness, which is then applied in the final set of experiments."

assimilation experiment	constraints					
	prior information	daily ice concentration	monthly ice thickness	daily SST	daily snow depth	
heightstraightforward	yes	Mar-Apr	Mar-Apr	Mar-Apr	Mar-Apr	height
heightreconstruction	yes	Jul-Sep	-	Mar-Sep	Mar-Sep	
heightbias-corrected height	yes	Mar-Apr	Mar-Apr	Mar-Apr	Mar-Apr	

Table 1. List of sets of experiments performed and the constraints used. Each set of experiments consists of three experiments, one each for the years 2012-2014.

6. *P5533:L20-30: It is really not very clear to me how the lines in figure 7 were calculated, or how to interpret them. Was the DA scheme used to update the state vector all through the simulations, or was a control simulation run without data assimilation? This needs to be much more explicit. It is very important that this explanation is improved as it is impossible to assess the validity of the paper, when the method is not clear.*

   The caption of the Figure is extended to read:

   "Prior **(black)** and posterior **(red)** misfit **(as defined in Eq.$\tilde{(}$3))** per data stream and month for a) the CryoSat-2 ice thickness (data are currently only available for March, April and November; not scaled, see text), b) the OSI SAF ice concentration, c) the snow depth (UB), and d) the OSI SAF SST. **The three black (red) lines are representing respectively the three years 2012-2014 and are marked by different symbols (circle: 2012, triangle: 2013, square: 2014)**."

The extended description of the DA system (section 2.1) should also help readers to understand the Figure.

7. *Figs 6, 8 and 10. Suggest adding dates to the plots to make it easier for the reader.*

   Dates are added to the plots.

8. *Fig 11: I suggest the more information is included in the caption of this figure.*

   The caption gives more information now and reads:

   “The **three-year mean (2012-2014) of the ratio of the reconstructed and the CryoSat-2 ice thickness** for a) March and b) April.”

2 Comments by Anonymous Referee #2

Review overview: The paper needs substantial editing and better communication. There may be some useful work here, but it is difficult to tell. If readers do not understand the work, it will go to waste.

This paper investigates the utility of data assimilation strategies within the NOASIM ice/ocean model via experimental procedures.

Several model parameters are calibrated using observations form the period 1990-2008. Experiments are then performed for 2012-2014 where data is assimilated during March-April and the model is evaluated in September. CryoSat-2 ice thickness, ice concentration (OSI SAF), SST's (OSI SAF) and snow depth (Uni. Bremen) are assimilated at times. Using a "straightforward" assimilation strategy in March-April still produces biased results in September.

"Reconstruction" runs are then performed, which aim to find March-April ice thicknesses that give good results in September. The ratio of CryoSat-2 ice thicknesses and those found in the "reconstruction" runs is then used as a form of "bias correction" in a final set of assimilation experiments. The final runs give better September results than the original ("straightforward") case, however, the March ice thickness used in this case seems to be unrealistically thick, with vast areas being +3.5m thick.

The overall concept of calibration, then assimilation and analysis is quite good. Unfortunately, it is difficult to learn much from this paper. The description does not effectively communicate exactly what was done, nor is there an analysis of why the assimilation failed to make expected improvements. The primary recommendation of this review is that the authors put themselves in the position of someone trying to replicate their experiments and edit their paper accordingly.

As mentioned above in section 0 of our response, we made an effort to put ourselves “in the position of someone trying to replicate our experiments” and worked on the communication through substantial extension of several sections. We think that this better communication already answers many of the issues/questions raised in the reviewers specific comments below and will partly give only brief response to those.

My impression after reading the paper is as follows: since the "straightforward assimilation" of realistic data did not produce a good result and the "bias correction" assimilation uses ice that is too thick, the NAOSIM model suffers from structural, parameter or input errors. Data assimilation can be used to highlight these issues (which is perhaps the most useful contribution made by this paper), but it should not be used as a crutch for trying to correct such systematic problems - that goes against the theory and underlying assumptions of optimal data assimilation.

It seems that a final conclusion could be that either: 1) the CryoSat-2 ice thickness data is incorrect thus adds no useful information or 2) the NAOSIM model has issues that need to be rectified before it can make a reliable forecast... Case (2) seems more *likely, but one cannot say based on the information in this paper and it is up to the authors to demonstrate either case.*

We now state clearly (for example in the conclusions) that the use of a bias-correction scheme is only the second-best choice (after tracing back the origin of inconsistencies and removing them):

"**We took the bad forecast skill as an indication of remaining biases in the system. Guided by biases in the autumn ice thickness simulated by our and the two above-mentioned sequential assimilation systems we suspected inconsistencies between the CryoSat-2 data streams and the rest of our assimilation/forecasting system. Until we fully understand and will be able to remove the origin(s) of this bias, we need to devise a strategy for enhancing the forecast skill through correction of this bias. Our procedure is based on a second set of experiments with a longer assimilation window (March to September). We deliberately omitted the CryoSat-2 data stream in the system and constructed an initial state on March 1 that is consistent with the rest of the system. From this initial state we simulated the ice thickness distribution in March and April. Since this set of experiments made use of the summer ice conditions, we called these simulated ice thickness fields 'reconstructed'. As the ratio of this 'reconstructed' and the CryoSat-2 ice thickness fields for March and April is very similar for all three years, our bias correction scheme uses the three-year mean of this ratio field as a point-wise multiplier for the CryoSat-2 product.**"

In the forecasting business one is, however, always confronted with a situation in which imperfect models and data are to be combined, and (until a bias can really be eliminated) the standard procedure is to correct for it. The justification is always an increase in the forecast skill. There are numerous examples of empirical bias correction schemes, in particular from NWP. Some of them are even adaptive, in the sense that parameters of the bias correction scheme are estimated within the analysis system.

We also state now that the variational assimilation system is an ideal tool to detect inconsistencies. In fact these inconsistencies are typically combinations of imperfections in the data streams, the prior, the model (formulation and parameters), and the boundary conditions, and it is a substantial effort to pinpoint their origin.

"**The main recommendation from this study to the ice forecasting community is to try and eliminate potential biases between the model (including the boundary conditions) and the observational data streams by a thorough calibration of the model and examination of the calibration results with independent information. Such a procedure can only be a first step, as it cannot eliminate all sources of bias. Careful inspection of assimilation results for inconsistencies is thus essential but requires a tedious analysis. The variational data assimilation approach in the form used here is a powerful basis for detection of inconsistencies and ultimately their removal. As an intermediate step bias correction schemes such as the one we constructed for assimilation of the CryoSat-2 ice thickness product are helpful for enhancing the forecast skill through compensation of model-data inconsistencies. This appears to apply as well to two established sequential assimilation systems (PIOMAS2.1 and TOPAZ4) as they underestimate the autumn ice thickness in a similar manner as our system.**"

Specific:

1. *Section 2.1 - How does the variational code deal with potential data inconsistencies? What is the error assigned to the model in the C(Xo) matrix? Results in Figure 5 would suggest that the model error is set to be very high as very little of the total cost function is from this term (Figure 7).*

   We extended various parts of the manuscript to highlight how the variational system handles inconsistencies, for example in the introduction:

   "The assimilation system is built around the regional North Atlantic/Arctic Ocean

Sea Ice Model (NAOSIM, Gerdes et al. (2003), Kauker et al. (2003)). **We apply a variational assimilation approach that determines an initial state (of sea-ice and ocean fields) on March 1. This means through variation of the model state on March 1 we search a trajectory that provides the best match to the four data streams over a subsequent two-month interval (assimilation window). The seasonal forecast is then performed by a simulation from that initial state along that trajectory into the future. Moving on a model trajectory means that we simulate a temporal sequence of states that fulfills the dynamical constraints imposed by the equations governing the model, including the fundamental conservation laws of mass, momentum, and energy. In summary, all observational information in the two-month observation window is used to constrain the initial state on March 1. All subsequent changes to model variables (including the forecast) are consequences of the changes in the initial state.**

It is evident that this approach has the potential to reveal inconsistencies between the model and our four data streams that otherwise would remain undetected. Such an inconsistency could be, for example, a huge ice thickness in a grid cell with little ice concentration. It could yield to some combination of a bad match of the observations, an unrealistic initial state, and an unrealistic forecast. Exactly this type of inconsistencies were detected in initial preparatory experiments. Hence, a first part of this study attempts to minimise biases by adapting some of the process parameters in the formulation of our model to better match observations over the 19 year period from 1990-2008. To evaluate the success of this calibration exercise we compare the calibrated model with two independent models that are typically used as reference."

C(Xo) is the inverse weight for the prior term (Eq. 1). The inverse weight for the model term is $C(d_{mod})$ (Eq. 2). We stress again that we do not use a sequential

approach here. The prior term has indeed little weight (Fig 5), which gives the observations the freedom to act on the control vector, i.e. the initial state. We think this is o.k.

2. *One suggestion would be to make a table with four columns: Assimilated data, State/control variables impacted by the data, Observational error, Model error. This would give the reader a better picture of what was used. Maybe even consider schematic diagrams to explain the assimilation experiments.*

   We hope that the substantial revision explained in section 0 clarified that in our approach all assimilated data streams impact all components of the control vector. The (space and time variant) observational uncertainty is described with each data stream (section 2.2.2), and the uncertainty from model error is set to zero (see section 2.1) as we assume it is dominated by the observational uncertainty (Eq. 2).

   But in general a table is a good idea to give an overview (see Table 1 in this response, i.e. Table 2 in the manuscript).

3. *The M operator needs to be described. Does M treat all control variables independently or are there other matrix elements which allow information from one variable to propagate to others? In either case, how does the M operator impact the final result? For example, when ice thickness is NOT assimilated, but concentration is assimilated, what happens to thickness when concentration is increased or decreased? If new ice is added, what is its thickness? If concentration goes down does thickness stay the same?*

   Here we also hope that the substantial extension described in the above general response (section 0) clarified this question, in particular the more detailed description of the method:

   "**A coupled ocean sea ice model computes a sequence of states of the ocean sea ice system (trajectory) from an initial state. By varying the initial**

state we can control the trajectory along which the model evolves. Thus we denote the vector composed of all initial model fields as control vector, x. The task of a variational assimilation is to minimise the difference of the model simulation to the available set of observations, d. This observation vector is typically a subset of all variables that can be simulated with the model at any point in time (within a given assimilation window) and space (within the model domain). To formalise the assimilation methodology it is useful to consider the model as a mathematical mapping (function) $M(x)$ from the control vector to the observation vector, more precisely to its equivalent simulated with the model. The assimilation system seeks a control vector, x_{post} such that $M(x_{post})$ achieves the best possible match to the observation vector d. At the same time we want to use any extra information we already have on x, the so-called prior information x_0. As the information on the model, the observations, and the prior is only approximate the theory is most conveniently formulated in terms of probability density functions (PDFs) (Tarantola, 2005). For computational convenience one typically assumes Gaussian distributions of the prior and the observations and can then show that x_{post} minimises the following cost function:

$$J(\mathbf{x}) = \frac{1}{2}\left[(M(\mathbf{x}) - \mathbf{d})^T \mathbf{C}(d)^{-1}(M(\mathbf{x}) - \mathbf{d}) + (\mathbf{x} - \mathbf{x_0})^T \mathbf{C}(x_0)^{-1}(\mathbf{x} - \mathbf{x_0})\right] \quad (1)$$

where $\mathbf{C}(x_0)$ denotes the uncertainty covariance matrix of the prior and the superscript T is the transposed."

So, in brief: M maps the state on March 1 onto all observations in the two-month assimilation window.

4. *Section 2.3 - The sea ice portion of the model is the part receiving assimilation so it would be good to give some information about it. How many ice layers are*

there? Are there any thickness categories per grid box? Is snow a bulk layer? How does the M(x) function map the model states to the observed quantities?

We have also elaborated on the model description, which should clarify this set of questions (on the role of M see the response to the previous comment).

5. *Only parameters related to dynamics are changed in the calibration; what was the reason for excluding the thermodynamic parameters? For example, albedo? The result of the calibration is that thickness is better, but extent is not - are you not just pushing the ice around at this point rather than removing it via a thermodynamic process?*

We don't look at the extent but the concentration. The misfit in winter gets worse but in summer (where our focus lies) gets slightly better (see Table 1 in the manuscript). We also elaborated on the effect of the calibration, for example, in the conclusions:

"**To limit the computational effort, we restricted the calibration to a sub-set of parameters that, in initial sensitivity studies, showed high impact on the ice conditions. The calibrated model showed a horizontal ice thickness distribution much closer to the ICESat-JPL observations.** A positive bias in the Beaufort Sea ..."

6. *"As this strong deviation is absent when forced with NCEP reanalysis this can be attributed to deficits in the CFSR surface forcing" - this goes against the vast majority of investigations which indicate that the NCEP forcing is rather poor. As one recent example, see Lindsay et al., 2014 which show that CFSR is superior to NCEP. My guess would be that the authors interpretation is not correct. Was the model was tuned for NCEP and only limited correction could be made for the CFSR forcing? In any case better justification or explanation needs to be given for such a statement.*

OK. Sentence removed.

7. *Section 3: "the price we have to pay for more reliable ice margin is misfit to the CryoSat-2 ice thickness" - this statement suggests to me that there are some fundamental problems with either the model, the CryoSat-2 data or the assimilation scheme. Ideally the rest of the paper would go on to investigate where the problem lies rather than trying to find a factor to compensate for this problem. It seems that the authors have side-stepped the real problem.*

   Partly this is correct but we have included the comparison with PIOMAS2.1 and TOPAZ to illustrate that not only NAOSIM has problems but most likely PIOMAS2.1 and TOPAZ4 as well, as both model do not outperform NAOSIM with respect to independent observations (i.e. not assimilated by PIOMAS 2.1 and TOPAZ4).

   We added a paragraph on the reviewer's suggestion to the conclusions:

   "**The main recommendation from this study to the ice forecasting community is to try and eliminate potential biases between the model (including the boundary conditions) and the observational data streams by a thorough calibration of the model and examination of the calibration results with independent information. Such a procedure can only be a first step, as it cannot eliminate all sources of bias. Careful inspection of assimilation results for inconsistencies is thus essential but requires a tedious analysis. The variational data assimilation approach in the form used here is a powerful basis for detection of inconsistencies and ultimately their removal. As an intermediate step bias correction schemes such as the one we constructed for assimilation of the CryoSat-2 ice thickness product are helpful for enhancing the forecast skill through compensation of model-data inconsistencies.**"

8. *"prize" should be "price"*

   Corrected.

9. *For the "straightforward" case, how often is an assimilation step performed? You are using daily and monthly average data - how are these items assimilated in the same system? My guess at the moment is that it is once a day for March 1 - April 30, but with a special step once a month where thickness is assimilated (and somehow the monthly averaging process is accounted for)? - but I could not tell from the paper. Where does the extra weight of 180 get applied? From the text, I cannot tell how this experiment is performed. It might be a good experiment for other model groups to try, but we can't tell.*

   We hope our elaborated description of the method has answered most of the questions. In brief: We use a single two month assimilation window and compute equivalents (including temporal averaging) to all observations.

10. *For the Reconstruction runs: If ice concentration and thickness are not assimilated in March, how/why are the concentrations in these simulations so different from the control experiment? Where did these new concentrations and thicknesses (as suggested in Figure 9) come from? Are these initial thicknesses simply guesses? A better explanation is required.*

    We hope our elaborated description of the method has made clear that we are not operating in sequential mode but are shifting model trajectories through variation of the initial state on March 1.

11. *If the model was well calibrated (as described in Section 2), why does it need further bias correction? The purpose of calibration is to remove biases (low frequency error) so that data assimilation can account the more random and higher frequency error (e.g. initial conditions error). This point should be reconciled either with reference to the model or the observations. A situation requiring double bias correction often points to a case of overfitting.*

    We are not risking overfitting because the observations for the calibration and for assimilation are different.

We elaborated on the need for bias correction for example in the conclusions:

"**In essence the calibrated model matched the observational data within their (partly large) uncertainty ranges. Still it was difficult to assess whether the result of the calibration exercise is sufficient for our objective (bias elimination). What we could do, however, was to compare with output of two established (sequential) assimilation systems, which also confirmed the success of the calibration. We note that this calibration is, however, limited by the reliability of the available data streams. For example, given the difference in measurement approach (laser vs. radar altimetry), it is not clear, how consistent the ICESat-JPL ice thickness product (used for calibration) and the CryoSat-2 product (to be used for the assimilation/forecasting) are. As both products don't overlap in time a direct comparison is not possible.**"

12. *As noted before, ice thickness and concentration are related variables - it seems unlikely to have a concentration of 20% and 3m thick for a grid box. Updating one of these variables without consideration of the other can lead to physical conflicts and/or unlikely situations that would not happen in an "open loop" simulation. How is this accounted for?*

    The fact that we are moving on model trajectories saves us from this type of problem that can occur in sequential assimilation schemes.

13. *Figures 7 & 9: To many readers the CostFunction might be an abstract quantity, particularly when it is not scaled for ice thickness, thus the Y-axis has no real meaning. These values might be better understood as percentage changes from the control? However, what readers would be most interested in (and comprehend the easiest) is if you were to transform the cost function back to quantities such as mean increment in ice thickness and snow depth etc.*

    We think the misfit (i.e. the cost function) is a good metric to quantify the model-data difference in a single number per data stream and month, because it ac-

counts for uncertainty in the data and the model (as defined by the data term in Eq. 1 and the newly introduced Eq. 3). We assume by increment the reviewer means the difference between posterior and prior runs for each quantity and month. We could plot this, of course, but think it would be a bad measure of the change. For example when shifting ice around the mean difference could be zero but the misfit could change significantly. Note also that for the ice thickness the spatial structures of posterior and prior fields are already shown in Figure 6.

3 Comments by Anonymous Referee #3

General Comments: The assimilation of ice thickness data and its potential impact on the predictability of seasonal sea ice is an interesting topic and one that deserves publication in The Cryosphere.

I was excited to read about this work when first I saw this study but sadly after finishing the paper I was left feeling disappointed.

Unfortunately, in its current guise, this article does not advance the scientific understanding of this topic because the paper is not easy to follow and not well motivated. In particular I find that the article has nowhere near enough detail to allow the reader to understand what was done and why but at the same time feels a bit dry and technical to read! This article will therefore require considerable revision before it can be accepted for publication.

One of the reasons I find this difficult to read is that this work appears to be documenting two different pieces of work: a) an Arctic model tuning exercise and b) implementation of ice thickness data assimilation. The result is that neither of these pieces of work are described in adequate detail.

There is also not very much discussion of the results and the Conclusions section *is more a summary of the tasks performed rather than anything else. Therefore it is not obvious what this study is trying to tell the scientific community about such an interesting and high profile issue.*

I also find that the figure captions are not descriptive enough. They should include more details about the plot (i.e. what the individual coloured lines etc. mean).

As mentioned in section 0 we extended the introduction, the description of the assimilation approach and the conclusions significantly. In particular we better motivated the need for the model calibration. In the response to this reviewer's specific comments we try to avoid repeating quotations from the manuscript made in the response to the previous two reviewer's comments. All figure captions have been thoroughly inspected and extended where necessary.

Specific Comments:

1. *There is little or no mention about the underlying sea ice model being used in these studies. I would like to know what processes are included (or not) and what the performance of the model is. In particular there is no discussion of model biases which really need to be understood before assimilating data. The data assimilation system used is simply referred to as multi-variate or variational but there is little discussion of the mechanics involved. Questions that are left unanswered include (but are certainly not limited to): 1. is 3D-Var or 4D-Var being used? 2. how are the ice concentration, ice thickness and snow thickness contributions balanced? 3. what length-scales are used? 4. what ice/snow thickness is used when ice is added? 5. is the sea ice model single or multi-category (and if the latter then how does the DA deal with this)?*

   The expression 'multi-variate' is not mentioned at all in the manuscript. In the revised manuscript, points 1 to 4 have been taken care of in the introduction and the description of the assimilation system (section 2.1). We have elaborated on

   the model description, in particular for the sea ice model (point 5: section 2.3):

   "NAOSIMDAS is constructed around the North Atlantic/Arctic Ocean Sea Ice Model (NAOSIM) (Kauker et al., 2003). The **ocean** model is derived from version 2 of the Modular Ocean Model (MOM-2) of the Geophysical Fluid Dynamics Laboratory (GFDL). The version of NAOSIM used here has a horizontal grid spacing of 0.5° on a rotated spherical grid. The rotation maps the $30°W$ meridian onto the equator and the North Pole onto $0°E$. In the vertical it resolves 20 levels, their spacing increasing with depth. The ocean model is coupled to a sea ice model with viscous-plastic rheology (Hibler, 1979). **The thermodynamics are formulated as a zero-layer model following Semtner (1976), and its parameters (e.g. albedos) are set in accordance to the AOMIP protocol (Nguyen et al., 2011). Freezing and melting are calculated by solving the energy budget equation for a single ice layer with a snow layer and an ocean mixed layer according to Parkinson et al. (1979). In contrast to the original formulation the energy flux through the ice is calculated by a PDF for the distribution of ice thickness based on EM-bird measurements (Castro-Morales et al., 2014). The sea ice model's prognostic variables are ice thickness, ice concentration, and snow depth. Ice drift is calculated diagnostically from the momentum balance. All quantities are mean quantities over a grid box. When atmospheric temperatures are below the freezing point, precipitation is added to the snow mass. The snow layer is advected jointly with the ice layer. The surface heat flux is calculated using prescribed atmospheric data and sea surface temperature predicted by the ocean model. The sea ice model is formulated on the ocean model grid and uses the same time step. The models are coupled following the procedure devised by Hibler and Bryan (1987).** At the open boundary ..."

2. *I am confused as to what the motivation is behind the PIOMAS Topaz comparisons. This section should be better explained or removed. I find it odd that there*

is a lot of motivation for near-real-time, "operational" observations when there is no mention anywhere that this system is used for operational forecasting or that it produces any products. Perhaps more information could be included about this?

We elaborated on the necessity for the comparison with the other two models. The data assimilation system is used, among other applications, for the Sea Ice Outlook, and we included a reference to Kauker et al. (2010). Future forecasts will also be published on meereisportal.de run by AWI.

3. *The "reconstruction" and "bias correction" techniques need to be better motivated and described. I find myself wondering what their purpose is here – save for to account for biases in the underlying model (which are not discussed). The fact that the reconstruction shows that unrealistically high March ice thickness is required to get a good September forecasts suggests to me that the model is melting too much ice.*

   The model is melting too much ice but this holds as well for the two assimilation systems used as a reference. It is not clear, however, which component of the prediction system is responsible for the excess melting: Is it the model (model formulation or process parameter values) or the surface forcing. Unfortunately this is almost impossible to answer, because we do not have in-situ observations of the surface forcing in the region were the largest March and April thickness is needed by the system. Indeed, the other two models show similar biases (in October/November in the ice thickness) although the model formulations are different.

4. *Furthermore it seems to me that the "bias correction" is derived using information about the future ice concentration. How would this work in a real forecasting situation?*

   There are two points to mention in this context. First, the necessity of atmospheric boundary conditions, on which we comment in the conclusions:

"To focus on the effect of constraining the initial state of the ocean sea ice system, we assumed to have perfect seasonal atmospheric forecasts providing perfect surface boundary conditions (for a use of this system in operational mode, uncertainty in boundary conditions is handled through an ensemble approach, **see e.g. Kauker et al. (2010)**)."

Second, the availability of the bias correction scheme, on which we also commented in the conclusions:

"We note that **this does not provide a completely independent assessment, because our prediction target, namely the summer ice conditions of 2012 to 2014, was used in the construction of the bias correction scheme. Future assimilation/forecast experiments for years beyond the period from 2012 to 2014, will, however, be completely independent.**"

References

K. Castro-Morales, F. Kauker, M. Losch, S. Hendricks, K. Riemann-Campe, and R. Gerdes. Sensitivity of simulated arctic sea ice to realistic ice thickness distributions and snow parameterizations. *Journal of Geophysical Research: Oceans*, 119(1):559–571, 2014. ISSN 2169-9291. doi: 10.1002/2013JC009342. URL http://dx.doi.org/10.1002/2013JC009342.

M. Chevallier and D. Salas-Mélia. The role of sea ice thickness distribution in the arctic sea ice potential predictability: A diagnostic approach with a coupled gcm. *Journal of Climate*, 25 (8):3025–3038, 2012. doi: http://dx.doi.org/10.1175/JCLI-D-11-00209.1.

M. Chevallier, D. Salas y Melia, A. Voldoire, M. Déqué, and G. Garric. Seasonal forecasts of the pan-arctic sea ice extent using a gcm-based seasonal. *Journal of Climate*, 26(16): 6092–6104, 2013. doi: 10.1175/JCLI-D-12-00612.1.

J. Day, E. Hawkins, and S. Rietsche. Will arctic sea ice thickness initialization improve seasonal forecast skill? *Geophys. Res. Let.*, 2014. doi: 10.1002/2014GL061694.

R. Gerdes, M. J. Karcher, F. Kauker, and U. Schauer. Causes and development of repeated arctic ocean warming events. *Geophysical Research Letters*, 30(19):n/a–n/a, 2003. ISSN 1944-8007. doi: 10.1029/2003GL018080. URL http://dx.doi.org/10.1029/2003GL018080. 1980.

W. Hibler. A dynamic thermodynamic sea ice model. *Journal Geophysical Research*, 9:815–846, 1979.

W. I. Hibler and K. Bryan. A diagnostic ice-ocean model. *Journal Physical Oceanography*, 17: 987–1015, 1987.

M. M. Holland and J. Stroeve. Changing seasonal sea ice predictor relationships in a changing arctic climate. *Geophysical Research Letters*, 38:L18501, 2011. doi: 10.1029/2011GL049303.

M. M. Holland, D. A. Bailey, and S. Vavrus. Inherent sea ice predictability in the rapidly changing arctic environment of the community climate system model, version 3. *Clim. Dyn.*, 36(7): 1239–1253, 2010.

F. Kauker, R. Gerdes, M. Karcher, C. Köberle, and J. Lieser. Variability of Arctic and North Atlantic sea ice: A combined analysis of model results and observations from 1978 to 2001. *Journal of Geophysical Research Oceans*, 108(C6):3182, 2003. doi: 10.1029/2002JC001573.

F. Kauker, T. Kaminski, M. Karcher, R. Giering, R. Gerdes, and M. Voßbeck. Adjoint analysis of the 2007 all time arctic sea-ice minimum. *J. Geophys. R.*, page L03707, 2009. doi: 10.1029/2008GL036323. URL http://www.agu.org/pubs/crossref/2009/2008GL036323.shtml.

F. Kauker, R. Gerdes, M. Karcher, T. Kaminski, R. Giering, and M. Voßbeck. June 2010 Sea Ice Outlook - AWI/FastOpt/OASys. Sea Ice Outlook web page, June 2010. URL http://www.arcus.org/search/seaiceoutlook/index.php.

T. Koenigk and U. Mikolajewicz. Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Clim. Dyn.*, 32(6):783–798, 2009. doi: 10.1007/s00382-008-0419-1.

R. Lindsay, C. Haas, S. Hendricks, P. Hunkeler, N. Kurtz, J. Paden, B. Panzer, J. Sonntag, J. Yungel, and J. Zhang. Seasonal forecasts of arctic sea ice initialized with observations of ice thickness. *Geophys. Res. Lett.*, 39:L21502, 2012. doi: 10.1029/2012GL053576.

F. Massonnet, T. Fichefet, and H. Goosse. Prospects for improved seasonal arctic sea ice predictions from multivariate data assimilation. *Ocean Modelling*, 88:16–25, 2015.

A. T. Nguyen, D. Menemenlis, and R. Kwok. Arctic ice ocean simulation with optimized model parameters: Approach and assessment. *J. Geophys. Res.*, 116(C04025), 2011. doi: 10.1029/2010JC006573.

C. Parkinson, , and W. Washington. A large-scale numerical model of sea ice. *Journal Geophysical Research*, 84:311–337, 1979.

A. Semtner. A model for the thermodynamic growth of sea ice in numerical investigastions of climate. *Journal pf Physical Oceanography*, 6:379–389, 1976.

A. Tarantola. *Inverse Problem Theory and methods for model parameter estimation*. SIAM, Philadelphia, 2005.

Q. Yang, S. Losa, M. Losch, X. Tian-Kunze, L. Nerger, J. Liu, L. Kaleschke, and Z. Zhang. Assimilating smos sea ice thickness into a coupled ice-ocean model using a local seik filter. *Journal of Geophysical Research: Oceans*, 119(10):6680–6692, 2014. ISSN 2169-9291. doi: 10.1002/2014JC009963. URL http://dx.doi.org/10.1002/2014JC009963.