Dear Editor,

Hereby we submit a revised version of the manuscript entitled "The benefit of using sea ice concentration satellite data products with uncertainty estimates in summer sea ice data assimilation "for publication in TC.

We substantially revised the text in order to address the reviewer's concerns. Please find our detailed responses to the reviewers' comments below as well as a description of how the manuscript has been improved.

With best regards,

Qinghua Yang
On behalf of the co-authors

Response to reviewer 1

We appreciate very much the constructive and helpful comments from the reviewer. Addressing the revisions recommended by the Reviewer #1 (Author's Response follow "AR:" in text).

We also have re-organized the texts to make the storyline of this MS more constructive:
1) We do data assimilation experiments in summer,
2) Using the provided uncertainties for sea ice concentration in summer does improve the sea ice concentration forecast,
3) No improvement (and sometimes worse) ice thicknesses,
4) We link the sub-optimal thicknesses on two little spread of the model fields after assimilation,
5) We link this two little spread on the uncertainties provided with the observations,
6) This leads us to the mismatch between the radiometric and physical concentrations in summer, which is something the community (both modelling and satellite) is only recently recognizing.

From here two paths:
7a) for this short paper, LSEIK-3 was a pragmatic solution;
7b) for future research, we need to develop better DA methodologies.

Authors perform sea ice concentration assimilation for 3-month period (June-August) in 2010 (April). They use two observational datasets, and different uncertainties associated with this data. The paper explores performance of the model in terms of sea ice concentration and thickness after data assimilation with different uncertainties.

##General comments
The topic of the paper is very interesting. It is one of the first studies that explores effects of spatially and temporarily variable uncertainties of sea ice concentrations on its assimilation in to the model. However the paper leaves an impression of being written in hurry, with a lot room for improvement. It certainly has to be expanded to make results more conclusive, especially in terms of sea ice thickness analysis. I do not recommend publishing the paper in the present form in "The Cryosphere".

##Specific comments
Description of the data, that is used for assimilation have to be very clear. Now it is hard to understand where exactly the data came from. Use of the selected time period (summer 2010) have

to be also justified, especially considering attempts to perform sea ice thickness analysis. There are some satellite data on sea ice thickness in recent years, which can serve for comparison. Not including September in the analysis, the month with maximum melting, also have to be justified. Evaluation of the sea ice concentration simulated by the model is based on comparison with NSIDC dataset that can hardly serve as an independent data source. Moreover it is not shown how NSIDC data compares with OSISAF and SICCI and if being closer to NSIDC data is actually mean being closer to reality.

I find discussion on the sea ice thickness comparison very weak. It is based only on two point stations, and can't serve as a basis for very broad conclusions presented by the authors. All discussions about thickness should be ether excluded, or better expanded to compare with more representative data.

AR:
- *We have corrected the texts to describe the sea ice concentration data clearly.*
- *At present, there are some available satellite based sea ice thickness data, e.g., Cryosat-2 and SMOS. However, both data sets are only available in the cold season, there is no useful satellite based ice thickness data in summer, and so the validation of sea ice thickness forecasts are much more difficult than the validation of sea ice concentration. In the revision, we further calculated the mean in-situ ULS ice thickness using two state-of-the-art satellite based sea ice concentrations (SICCI and NSIDC), so the comparisons are more conclusive than in the previous version.*

- *We acknowledge that September could have been included in the analysis. In this particular year, however, there was open water in the interior pack near the North Pole as early as Jul12, so that we can assume that there are melting conditions everywhere in the Arctic Ocean in August. Hence, including September does not add new or different melting situations.*

- *As suggested by the other reviewer, it is not necessary to compare LSEIK-1 and LSEIK-2 because the different sensors and different resolutions between SICCI and OSI-401-a. So we decided to focus on the LSEIK SICCI assimilation series, and removed the original LSEIK-1 experiment, which assimilated the OSISAF OSI-401-a data set.*

In the revised MS, we compared the assimilation results with both the assimilated SICCI (Fig. 3a) and the non-assimilated NSIDC (Fig. 3b). We report only the RMSE for grid location where the satellite products reports and ice concentration lower than 0.35. See our response below to the specific comment of P2550 Line 10-12.

We also add Table 1 to better show this comparison.

Table 1. RMSE of the four forecasting experiments from mean ice thickness calculated by the ULS moorings BGEP_2010A, BGEP_2010D and the satellite ice concentration observations. The two values refer to the calculation using two different data sets SICCI-NSIDC.

		BGEP_2010A	BGEP_2010D
1	*MITgcm*	*0.86-0.89 m*	*0.93-0.97 m*
2	*LSEIK-1*	*0.43-0.46 m*	*0.55-0.59 m*
3	*LSEIK-2*	*0.61-0.64 m*	*0.51-0.55 m*
4	*LSEIK-3*	*0.43-0.46 m*	*0.59-0.62 m*

Detailed comments
Abstract
Abstract reads rather strange – you begin with description of the results, skipping the setup of the experiments (data assimilation with constant and varying uncertainties). So the "how" section of your abstract is incomplete.

AR: We have re-written the abstract:
Recently, the European Space Agency Sea Ice Climate Change Initiative (ESA SICCI) released ice concentration data complete with error estimates that depend on space and time. These data are used to in data assimilation experiments that aim at improving summer ice concentration and thickness forecasts in Arctic. The data assimilation system uses the MIT general circulation model (MITgcm) and a local Singular Evolutive Interpolated Kalman (LSEIK) filter. The effect of using sea ice concentration satellite data products with appropriate uncertainty estimates is assessed by three different experiments: in one experiment the SICCI concentration data is used with constant uncertainties; in two further experiments the same SICCI data are included along with their provided uncertainties; they differ only in imposing different minimum uncertainties. Using the observation uncertainties that are provided with the data improves the ensemble mean state of ice concentration compared to using constant data errors, but ice thickness is not affected in a systematic way. Further investigating this lack of impact on the sea ice thicknesses leads us to a fundamental mismatch between the satellite-based radiometric concentration and the modelled physical ice concentration in summer: the passive microwave sensors used for deriving the vast majority of the sea ice concentration satellite-based observations, cannot distinguish ocean water (in leads) from melt water (in ponds). New data assimilation methodologies that fully account or mitigate this mismatch must be designed for successful assimilation of sea ice concentration satellite data in summer melt conditions. In our study, thickness forecasts can be slightly improved by adopting the pragmatic solution of raising the minimum observation uncertainty, to inflate the data error and ensemble spread.

P2544
7-9: I don't think this information belongs to the abstract.
AR: Removed.

Introduction
21-22: Using IPCC report as a reference for such a statement is a bad practice. You should at least point the reader to the chapter in the report, or even better just cite individual researches that support your statement.
AR: We updated the reference to the exact chapter in the IPCC report: Vaughan, D.G., J.C. Comiso, I. Allison, J. Carrasco, G. Kaser, R. Kwok, P. Mote, T. Murray, F. Paul, J. Ren, E. Rignot, O. Solomina, K. Steffen and T. Zhang, 2013: Observations: Cryosphere. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

P2545
13-29: Here you discuss OSISAF and SICCI datasets that have temporal coverage of 1978-2009 and 1992-2008. On the next page you state, that you are going to use this datasets to study summer 2010 sea ice concentration. This sounds a bit strange. In the next section, you mention that it is actually OSI-401-a and SICCI AMSR-E. Please make it very clear what you use exactly. If you still want to discuss sea ice re-analysis products in the introduction, then you have to connect them to the data you are actually using.
AR: Corrected. Currently, the OSISAF OSI-409 and SICCI are the only two data sets that provide uncertainty estimates, so we gave a brief review here. Following the other reviewer's comments, we removed the original LSEIK-1 experiment that assimilates the OSI-401-a data set (See AR to General Comment 4)
Further, we have rephrased the text to make the description of these data sets clearer.

Forecasting experiment design
P 2546

22: Define SEIK. Before you define LSEIK with the same references. Is it the same thing?

AR: SEIK and LSEIK are different. We corrected the texts: "..and using the same ensemble-based Singular Evolutive Interpolated Kalman (SEIK) filter (Pham et al., 1998; Pham, 2001) in its local form (LSEIK, Nerger et al., 2006)."

P2547

5: Why this time period is chosen?

AR: This period was chosen as it was very particular. The open water was first found in the interior pack ice near the North Pole as early as July (http://nsidc.org/arcticseaicenews/2010/07/). We have rephrased the text to clarify this point

7-9: You use EOF information, but how s exactly initial conditions for your ensembles were generated? Please clarify.

AR: We have extended the text to "The leading Empirical Orthogonal Functions (EOFs) of the considered model variability are transformed by second-order exact sampling to generate the initial ensemble of ice concentration and thickness."

23: Here you use LSEIK again. Is there a difference between SEIK and LSEIK, and if there is, please explain it.

AR: Now we have explained their differences in the Introduction.

P2548

4-9: I strongly doubt that this product can be considered to be independent. The SSM/I and SSMIS are deliberately made quite comparable, so that the satellite measurements record started in 1978 can be continued. So SSMIS is improved version of SSM/I but it is in no way it can be considered as producing results "independent" of SSM/I.

AR: This is correct, the two datasets were not enough independent. We now assimilate the SICCI product from AMSR-E (using a blending of the Bootstrap Frequency Mode, and Bristol algorithms) and compare with NSIDC data (NSIDC; Cavalieri and others, 2012; http://nsidc.org/data/docs/daac/nsidc0051_gsfc_seaice.gd.html). For summer 2010, it uses SSMIS from DMSP-F17 and the NASA-Team algorithm. We also add the sentence: We note that both the SICCI and NSIDC products are computed from channel combinations of relatively similar passive microwave instruments and that they cannot be regarded as strictly independent. Using a different instrument and a different algorithms is nevertheless often the best we can use for passive microwave sea ice concentration data."

Results
P2549

18: You don't mention how you handle missing data around the North Pole in NSIDC during the comparison. Was this region excluded, or you assume some constant concentration?

AR: During the comparison, the North Pole area were excluded. We corrected the texts.

18: Your model and NSIDC data have different resolutions and different grids, so I assume for comparison you have to interpolate sea ice concentrations at some common grid. Details of this interpolation should be provided. Such interpolation can lead to quite significant local changes in sea ice concentration, so these effects must be considered in your comparison.

AR: We did the comparison after interpolating the observation to the model grids. This could have some influences on the comparison, but it is very difficult to quantify them. We added one more sentence to remind the reader that the possible effects of the interpolation were not discussed in the comparison.

P2550

2-5: The sentence is hard to follow, consider rephrasing.

AR: As the original LSEIK-1 experiment which assimilated the OSISAF OSI-401-a data was

removed, so we deleted this sentence.

10-12: I don't think that you can make such a statement. What you show is that LSEIK-3 is close to NSIDC data, but you did not show, that being close to NSIDC data mean being more realistic. It might be quite opposite. NSIDC data have a number of problems, especially in summer. What you doing here are comparing model after assimilation of more advanced sea ice products against presumably less accurate product.

You at least have to show how NSIDC sea ice concentration compares with OSISAF and SICCI in terms of RMSE.

AR: The reviewer is correct. In the new Fig. 3, we compared the assimilation results with both the assimilated SICCI (Fig. 3a) and the non-assimilated NSIDC (Fig. 3b). We report only the RMSE for grid location where the satellite products reports and ice concentration lower than 0.35. The texts below are added in the MS:

These are thus mostly location along the ice edge. Fig. 3 thus mostly assesses how the data assimilation experiments constrain the envelope of Arctic sea ice, not the interior (cyan color on Fig. 1). The reason for choosing this range is that all sea ice concentration products from passive microwave instruments have challenges with high concentration values in the summer (Ivanova et al. 2015). In such a case, documenting that the assimilated state is closer to the NSIDC product is not very conclusive, since NSIDC and SICCI products are probably likewise challenged at high concentration values. Looking away from the ice concentration values and focusing on the outskirt of the sea ice cover make the conclusions somewhat more robust as the influence of melt-ponds is reduced, and the approaches over open water are different in both products (Weather Filters in NSIDC and explicit correction for atmosphere perturbations for SICCI).
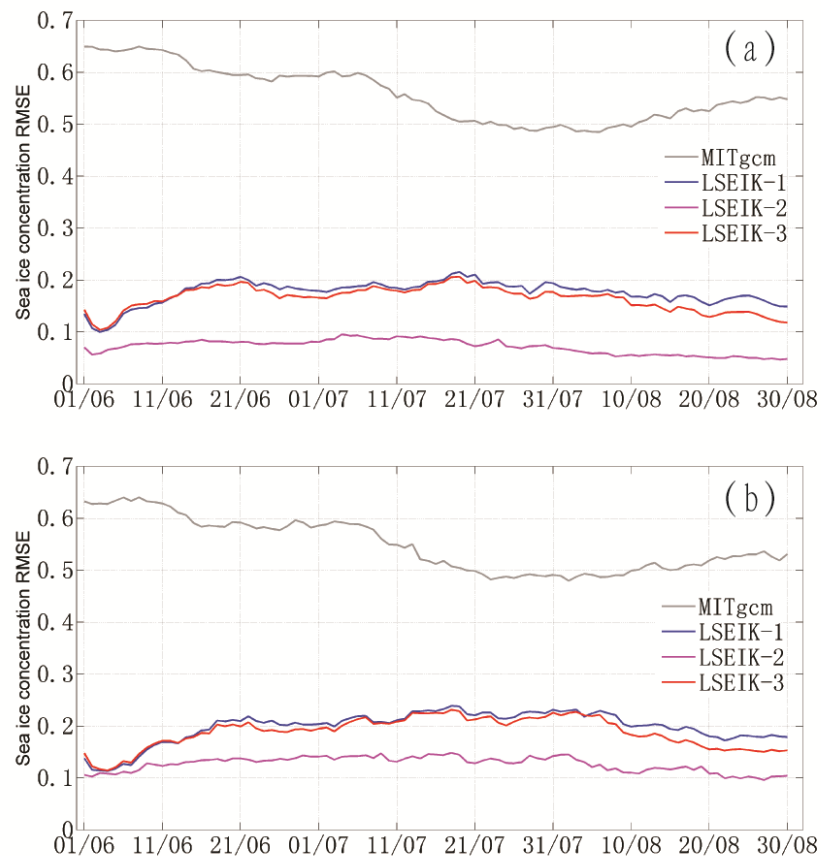

Figure 3.Temporal evolution of RMSE differences between sea ice concentration forecasts and the SICCI (a) and NSIDC (b) ice concentration data. The RMSE of the MITgcm free-run, LSEIK-1, LSEIK-2 and LSEIK-3 24-h forecasts are shown as gray, blue, magenta and red solid lines, respectively.

13-29: It is hard to estimate performance of the model using only two observational points. This analysis can't serve as a ground for your statement in the abstract that "SICCI concentrations outperforms the assimilation of OSISAF data in both concentration and thickness forecasts", it is just simply too local. In my opinion this comparison should be excluded from the study, or considerably expanded by adding analysis of spatial thickness distribution

AR: We agree that the two data points cannot lead to general statements. We have toned down our conclusions (also in the abstract) and now say there cannot be any definite conclusions. We would still like to keep this comparison.

21: You use abbreviation DA here for the first time. Define it.
AR: Corrected. We change "DA" to "data assimilation".

Discussion
P2552
16: You use abbreviation SD here for the first time. Define it.
AR: We corrected to standard deviations (STDs).

Conclusions
P2553
14-15: This is just contradicts your statement at P2551 L13:14.
AR: In the revision, as the original LSEIK-1 which assimilated OSISAF OSI-401-a was already removed, so we deleted this sentence.

Figures
When plotting maps of the Arctic Ocean most of the time the 0th meridian is used as a central longitude. What is the advantage to use 45th meridian in this case?
AR: The 45th meridian was used in our previous studies, and we think this format would be easier for us and the readers to follow the different works.

Response to reviewer 2

We appreciate very much the constructive and helpful comments from the reviewer. Addressing the revisions recommended by the Reviewer #2 (Author's Response follow "AR:" in text).

We also have re-organized the texts to make the storyline of this MS more constructive:
1) We do data assimilation experiments in summer,
2) Using the provided uncertainties for sea ice concentration in summer does improve the sea ice concentration forecast,
3) No improvement (and sometimes worse) ice thicknesses,
4) We link the sub-optimal thicknesses on two little spread of the model fields after assimilation,
5) We link this two little spread on the uncertainties provided with the observations,
6) This leads us to the mismatch between the radiometric and physical concentrations in summer, which is something the community (both modelling and satellite) is only recently recognizing.

From here two paths:
7a) for this short paper, LSEIK-3 was a pragmatic solution;
7b) for future research, we need to develop better DA methodologies.

This paper uses the MITgcm sea ice model to forecast the Arctic sea ice cover during summer (June to August) in terms of the evolution of its concentration and its thickness. For this purpose the MITgcm is assimilated using LSEIK with two different sea ice concentration data sets. These do have uncertainty estimates. Different realizations of uncertainties are tested: two constant onces and to varying ones. The influence of using these for forecasting Arctic Ocean is investigated for concentration and thickness. The paper is an important contribution to current knowledge and the paper is a good example for the usage of sea ice products WITH uncertainty estimates. Most of the paper is clearly written and well to understand already. A number of things and questions deserve more attention in my eyes, though, because of which I recommend to carry out some major revisions before acceptance of the manuscript for "The Cryosphere".

I detail my general comments in the following paragraphs. These will be followed by a number of other detailed comments before I will close the review with some hints towards typos etc.

General comments:
While the description of the methodology is fine - in my eyes - as far as it concerns the model and LSEIK some important questions and motivations remain open for the observational data sets.
a) I have difficulties to understand why the authors compare a coarse-resolution (25 km) but newer sensor sea ice concentration (SIC) data set (AMSR-E SICCI) with a finer resolution (10 km) but older sensor SIC data set (SSM/I OSISAF). AMSR-E offers finer spatial resolution than SSM/I and I guess the producers of the AMSR-E SICCI data set had a good reason for keeping the grid resolution of this data set similar to the SSM/I SICCI data set. On the other hand I doubt that the 10 km grid resolution offered by SSM/I OSISAF is a "real" resolution because footprint size and sampling of SSM/I data usually allows for 25 km grid resolution if using the lower frequency (19 and 37 GHz) channels. I recommend the authors to motivate their choice a bit better and to also discuss whether the different grid resolution and different actual resolution of the different products might have had an influence on the results.
AR: The reviewer is right, the different sensors and different resolutions between AMSR-E SICCI and SSM/I OSISAF made this comparison not necessary. So we focus on the LSEIK SICCI assimilation series, and have removed the original LSEIK-1 which assimilated the OSISAF OSI-401-a data set.
The influence of different resolution and different actual resolution is also very interesting, however, as this study focuses only on the effects of using the provided uncertainties, we would like to investigate this in the future work.

b) I am wondering why the authors did not also use the uncertainty information provided by the OSI-SAF SIC data set. Perhaps the uncertainty retrieval is the same for OSISAF and SICCI and therefore it is sufficient to look at SICCI only?

AR: In the previous version, we mentioned two OSISAF datasets: the reprocessed OSISAF OSI-409 and the near real-time OSI-401-a datasets. The OSISAF OSI-409 dataset has uncertainty estimates, but at the time of writing this MS, it only covered from 1978 to 2009 (this data has been extended very recently to June 2015). For the study period of summer 2010, the OSISAF ice concentration data that has been assimilated was the OSI-401a dataset which has no uncertainty estimates. In the revision, we removed experiment LSEIK-1 which assimilated OSISAF OSI-401-a (See General comment a). Also as the reviewer mentioned, the uncertainty retrieval is similar for OSISAF and SICCI, so it is not necessary to further assimilate OSISAF-409 data in this study.

c) The authors use an NSIDC SIC product - presumably based on the NT2 algorithm - to evaluate their model and assimilation results. While this is a fair approach to use the reader might miss some information about the why this product was chosen, what would have been the alternatives, whether it is important to have alternatives (at all), and what are the potential difficulties with the NSIDC product used. Currently the authors are risking that they rate the quality of their results with regard to SIC to a SIC data set which uncertainty and which bias are unknown.

AR: To address both reviewers' concerns, in the new Fig. 3, we compared the assimilation results with both the assimilated SICCI (Fig. 3a) and the non-assimilated NSIDC (Fig. 3b). We report only the RMSE for grid location where the satellite products reports and ice concentration lower than 0.35. The texts below are added in the MS:
These are thus mostly location along the ice edge. Fig. 3 thus mostly assesses how the data assimilation experiments constrain the envelope of Arctic sea ice, not the interior (cyan color on Fig. 1). The reason for choosing this range is that all sea ice concentration products from passive microwave instruments have challenges with high concentration values in the summer (Ivanova et al. 2015). In such a case, documenting that the assimilated state is closer to the NSIDC product is not very conclusive, since NSIDC and SICCI products are probably likewise challenged at high concentration values. Looking away from the ice concentration values and focusing on the outskirt of the sea ice cover make the conclusions somewhat more robust as the influence of melt-ponds is reduced, and the approaches over open water are different in both products (Weather Filters in NSIDC and explicit correction for atmosphere perturbations for SICCI).
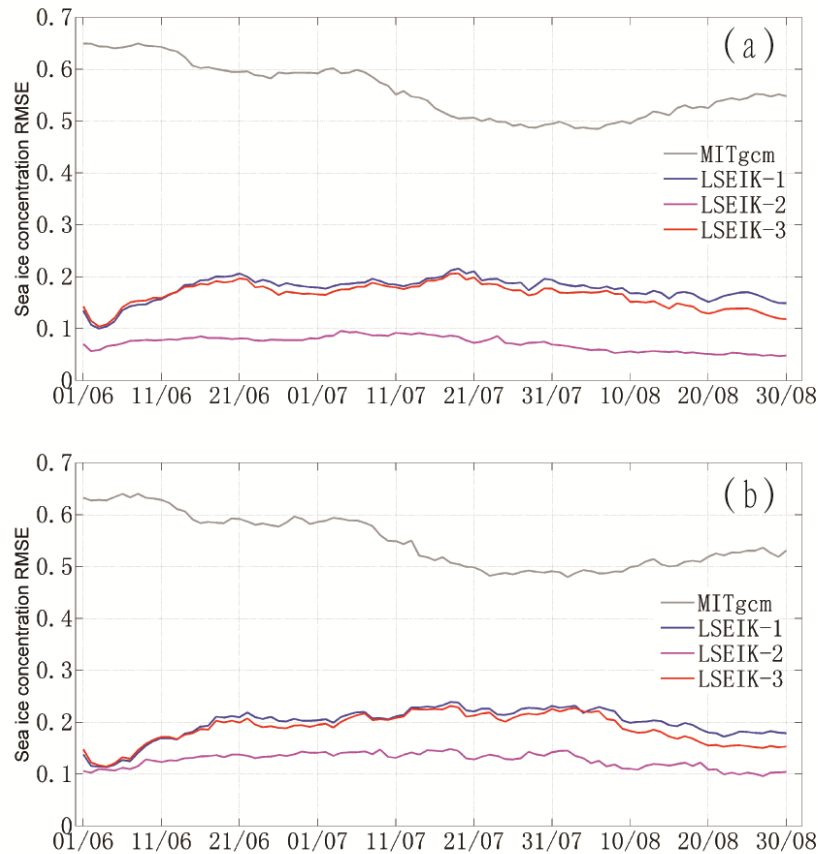
Figure 3.Temporal evolution of RMSE differences between sea ice concentration forecasts and the SICCI (a) and NSIDC (b) ice concentration data. The RMSE of the MITgcm free-run, LSEIK-1, LSEIK-2 and LSEIK-3 24-h forecasts are shown as gray, blue, magenta and red solid lines, respectively.

d) The authors use BGEP ULS data to get a view of the sea ice thickness (SIT). I am missing two things in their investigation of that data set. First I would have liked to see more discussion about the large different in the spatial representativity of the SIT data from ULS compared to the model. Secondly, the authors used the SICCI SIC to convert the SIT into a sea ice volume to more easily inter-compare it with the model data. Why did they use SICCI? Why didn't they use OSI-SAF and how the results looked like with OSI-SAF (and its finer grid resolution)? Wouldn't it be more reasonable - from the point of view of that you are evaluating the impact SICCI SIC has in the model - to multiply the BGEP SIT data with the NT2 data against which you also reference the SIC model results? Currently, one might argue that sea ice volume as computed from BGEP data and model output in terms of SIT are not independent because both use SICCI SIC.

AR: Currently, although there are some available satellite-based sea ice thickness data set, e.g., Cryosat-2 and SMOS, but they are only valid in the cold season. Thus, there is no such reliable data in summer and the validation of sea ice thickness forecasts are much more difficult than the sea ice concentration. We have added some text in the Introduction part to explain this: Currently, satellite-based observations of ice thickness are a challenge (Kwok and Sulsky, 2010; Kern et al. 2015), and there are very few reliable summer sea-ice thickness products available. Instead of remote-sensing data we compare ....

As the SICCI dataset is the state-of-the-art sea ice concentration product, so in the previous version, we used SICCI SIC to calculate the mean thickness. To make this comparison more convincing. In this revision, we also show the calculated mean ice thickness based on the non-assimilated NSIDC SIC. We add Table 1 to better show this comparison.

In the following I abbreviate page with P and line with L

Detailed comments: P2544, L22-26: I am wondering whether only the economic opportunities are driving this research. I would have thought that maybe seasonal weather forecast, climate model and also ordinary weather model data might be influenced by the changes we witness. Maybe the authors could be a bit broader here and motivate their study also from the science point of view. In this context I am wondering about the "risks" you mention (which are these?) and in particular about how these are "managed"?

AR: Corrected. The motivations were extended to also include the importance of Arctic sea ice reduction to the climate system and weather research.

P2545, L1: The authors wrote that data assimilation (DA) is important for accurate sea ice prediction because it is important to have a realistic initial state. Is this really the case? I would have thought that model assimilation with observations data is rather a tool to "push" or "keep" the model close(r) to reality ... i.e. not the initial state is important (this is the case with any model, right?) but the potential to continuously "supervise" where the model goes.

AR: We have corrected the texts: Sea ice data assimilation (DA) plays a pivotal role in sea ice forecasting, as it can provide realistic initial model states, and continuously constrains the model state closer to reality.

P2545 L9-10: I am wondering whether the authors also looked aside their own work a bit and maybe found other literature. how about Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., Kwok, R., 2011. Uncertainty in modeled Arctic sea ice volume. J. Geophys. Res. 116, C00D06?
AR: Corrected. We now also account for Schweiger et al. (2011) and refer to this study.

Also, in L10, the authors write "efforts". Which efforts are meant here? Are you referring to the previous studies mentioned above? Or other studies?
AR: "efforts" were corrected to "the studies" mentioned above.

P2545 L20: Here one could add a line that at the time of writing these two, SICCI and OSISAF, are the only two algorithms or products which come with a physically based sea ice retrieval uncertainty information - instead of an estimate of the spatio-temporal variation of the SIC within a certain grid area and time window which is a measure of the variation of the SIC due to actual SIC changes and due to articifial SIC changes implied due to algorithms' deficiencies to work under certain weather and/or surface property conditions.
AR: Corrected.

P2546, L5: I would find it helpful to find a half-sentence saying that the motivation for using the LSEIK is given in the following section.
AR: We now give the motivation in the Introduction. The SEIK filter algorithm is selected to assimilate the sea ice concentration because it is computationally efficient when applied to nonlinear models (Nerger et al., 2005), and its local form (LSEIK) has already been successfully applied in the sea ice concentration data assimilation (Yang et al., 2015a).

P2546, L5 and L22: One time the authors write LSEIK, the other time they write SEIK. If there is a difference between those and if the authors wish to highlight this then it does not become clear from the paper currently. If this is the same and/or if the authors mean LSEIK all the time, then they might want to change the name accordingly.
AR: Yes, the SEIK is different to LSEIK. We have corrected this and now clearly describe their relationship in the introduction:"..and using the same ensemble-based Singular Evolutive Interpolated Kalman (SEIK) filter (Pham et al., 1998; Pham, 2001) in its local form (LSEIK, Nerger et al., 2006).

P2547, L19: "grid points" ... I would recommend to write "model grid cells".
AR: Corrected.

P2547, L27: Can the authors please check whether the SICCI data set is avaliable on polar-

stereographic grid? I doubt so. Also see my general comment a).

AR: The "polar-stereographic grid" was corrected as "polar-centered EASE2 grid".

P2548, L1: "revised algorithm merging method" ... this remains unclear as long as the reader does not know that two (or more?) algorithms are combined in a hybrid approach.
AR: We removed this part.

P2548, L3: As the authors write "total standard error" they might want to also explain how this total error is composed. This is in a way also needed later in the paper in the discussion.
AR: We add more description on the data uncertainties in the last paragraph of Section 2: The SICCI total uncertainties are indeed the sum of two components, one characterizing the algorithm uncertainties, and the other measuring the uncertainties due to representatives of 25 km daily averages, geo-location and instrument foot-print mismatch (Lavergne and Rinne, 2014). The second component to the total uncertainties is only pronounced in areas of gradients in the sea ice concentration observations –typically at the ice edge-, and amount for the inability of such coarse resolution satellite observations to accurately locate sea ice edge. Should the SICCI sea ice concentrations be assimilated in models with significantly better spatial resolution, the enlarged uncertainties allow the model to freely locate its ice edge within the 25x25 km grid cells showing intermediate ice concentration values in the data.

P2548 L7: "grid spacing" ... which grid is used here? EASE2?
AR: This part has been deleted.

P2548, L7-9: See my general comment c). Also the statement that the NSIDC SIC is independent from the SICCI and OSI-SAF data sets is not entirely true because SSM/IS is the successor of SSM/I and hence share the same channels and same viewing geometry.
AR: Yes, as SSMIS is also a passive microwave sensor, so NSIDC is not completely independent measure. To address the concerns from both Reviewer 1 and Reviewer 2, we now provide the comparison results with the state-of-the-art SICCI and NSIDC data sets. We also add the sentence: We note that both the SICCI and NSIDC products are computed from channel combinations of relatively similar passive microwave instruments and that they cannot be regarded as strictly independent. Using a different instrument and a different algorithms is nevertheless often the best we can use for passive microwave sea ice concentration data."

P2548, L13-15: I am wondering whether the relatively old paper of Melling et al. is a proper reference for the BGEP moorings. How about in addition: Krishfield, R. and Proshutinsky, A.: BGOS ULS Data Processing Procedure Report, http://www.whoi.edu/fileserver.do?id=85684pt=2p=100409, Woods Hole Oceanographic Institute, 2006. This is at least a more recent reference.
AR: Corrected.

Also, I am wondering whether the paper by Nguyen et al. is suited as a reference about how to convert ULS draft to thickness. I guess, when the authors look into that paper more carefully they will find a proper reference which can be used for this.
Finally: Did you use the conversion factor of 1.1 regardless of ice type? If so, you might want to discuss this in the discussion section.
AR: We corrected the reference to Rothrock et al. (2008).
As we have no information of ice types the effects of ice type on the conversion were simply ignored in this study. We added one more sentence to note this.

P2548, L25: What do the authors intend to say with this last sentence? How do the authors define "open water"? When I look at the SIC maps provided later in the paper then I see SIC as low as 50
AR: This sentence was used to show that this summer 2010 is a particular period.
The open water was found with MODIS visible-band image, please see the news report here: http://nsidc.org/arcticseaicenews/2010/07/. This report is also referred in the MS.

P2549, L2: What do the authors mean by one standard deviation? Does the model treat the constant uncertainty estimate of 0.25 as one standard deviation?

AR: We corrected the statement to: e.g., the observation errors are assumed to be Gaussian distributed with STD 0.25.

P2549, L4: What do the authors mean by "representation error"? Does this refer to the smearing uncertainty which is given separately in the SICCI SIC product as well?

AR: We refer to a representativeness error due to the used projection of the observation to the model space (Yang et al., 2014). It is a different value from the smearing uncertainty provided in the SICCI SIC product. We have clarified this in Section 2.

P2549, L10-13: See my general comment b). In addition to that what happened in these to cases with grid cells where the total error was below the given thresholds. Were these not used for the model assimilation or were these used but with the uncertainty set to the respective value?

AR: These data were also used for the assimilation but with the uncertainty set to the respective value.

P2549, L14-16: The SICCI SIC product has flags where the SIC stems from interpolation and/or has been set to 0 or flagged invalid for various reasons. How about flagged grid cells not belonging to the north pole disc? Did you use this data as well? Did you also choose an uncertainty of 0.3 here? As you show later the total error in the area around the north pole disc is quite small, usually below 0.05. How does the assimilation cope with the jump in uncertainty from < 0.05 to 0.3? Another question regarding the SICCI SIC data set is whether the authors used the "regular" SIC product or whether they also tried to use the off-range SIC?

AR: 1) The flagged grid cells not belonging to the north pole disc were also used with an uncertainty of 0.30. Now we have clarified this in the MS. This is a simple approach, and should be considered more carefully in the future work.
2) As the concentration data belonging to the north pole were interpolated values, so we assigned them a relatively larger uncertainty. This jump was only related with the uncertainty, e.g., we can see a larger ensemble spread near the north pole, but for the SIC, there was no jump near the north pole area.
3) We only used the "regular" SIC in this MS.

P2549, L20-23: This reads as if large retrieval errors (only) can be expected at SIC < 0.05. This is a quite small value and I am wondering how this value is justified. I can only speak from the remote sensing side and would say that elevated uncertainties start to pop up at considerably higher SIC than just 0.05. Also it appears to me that SIC = 0.05 is within the (total) retrieval uncertainty at this SIC range?!

AR: Following Lisæter et al. (2003), we used the value of 0.05. As the reviewer said, it is also not perfect. In the new Figure 3, we report only the RMSE for grid location where the satellite products reports and ice concentration lower than 0.35. See our AR to General Comment c).

P2550, L2: "... from, on average, 0.24 ... 0.11, respectively. ..."
AR: Corrected.

P2550, L11: "time and space dependent" ... how about writing "the full range of"?
AR: Corrected.

P2550, L12: Here the discussion of results in Figure 2 seems to be done. What explains the maximum RMSE at the end of June followed by a net decrease in RMSE into August for all but LSEIK-3?

AR: In the current revision, we show the RMSE evolution with respect to the SICCI and NSIDC data sets. We add one sentence in the MS: The RMSE temporal evolutions are associated with the number of available data points that can be used for comparison or with surface forcing.

P2550, L16: See comment d)
AR: Currently, although there are some available satellite based sea ice thickness data sets in the

cold season, e.g., Cryosat-2 and SMOS. However, there is no such data in summer. Thus, the validation of sea ice thickness forecasts is much more difficult than the sea ice concentration. As the SICCI dataset is the state-of-the-art sea ice concentration product, so in the previous version, we used SICCI SIC to calculate the mean thickness. To make this comparison more convincing, now we also show the calculated mean ice thickness using the NSIDC data (see Section 3, Figure 4 and Table 1). And the results show that the differences between ULS-SICCI and ULS-NSIDC are small.

P2551, L5: At the end of this paragraph I ask myself: Why? Why does NOT using the full range in SIC uncertainty seemingly result in better results with regard to SIT? Where are the uncertainties of the range 0.01 - 0.1 located? Can we use the location of these grid cells as a potential explanation?
AR: In Section 2, a new paragraph was added in Section 2, which describes the uncertainty maps from Figure 2: The original observational data uncertainties of ice concentrations that are provided with the SICCI data set and used in LSEIK-2 and LSEIK-3 are displayed in Fig. 2. In Fig 2, we show the provided observation uncertainties on 1 June, 16 June, 1 July, 16 July, 1 August and 16 August 2010. The uncertainties are about 0.05 over packed ice and open water, but larger uncertainties up to and beyond 0.3 are present at the ice edge, and region of intermediate ice concentration values.

The reasons of the ice thickness performance are discussed in the following Section 4:
The ensemble-represented standard deviations (STDs) of sea ice concentration for LSEIK-2 turn out to be relatively small. For example, on 30 August 2010, most of the STDs in the Arctic central area and the sea ice edge area are less than 0.01 and 0.03, respectively (Fig. 5c). This means that all members are very close to the ensemble mean and the data assimilation will have only little effect. LSEIK-3 has a similar spread distribution pattern of higher STDs in the sea ice edge area and lower STDs in the concentrated central ice area but overall higher STDs than LSEIK-2. Together with the fact that LSEIK-2 does not fit the thickness observations as well as LSEIK-3, this suggests that the ensemble forecast spread for sea ice concentration is too low and cannot reflect the uncertainty. As only observations of sea ice concentration are assimilated, sea ice thickness is influenced indirectly during the data assimilation through the point-wise covariance between the ice concentration and thickness, thus through a linear update. Here, the very small sea ice concentration variance leads to a very small sea ice thickness spread (Fig. 6b). This probably explains why the LSEIK-2 system is not very effective at improving the sea ice thickness estimates while LSEIK-3 does somewhat better. The increased spread in the sea ice concentration allows the system to better represent the uncertainties and leads to a larger ice thickness spread (Fig. 6c). The sea ice thickness forecasts are improved accordingly.

P2552, L1: The authors write that "the satellite based concentration estimates are known to underestimate the sea ice cover". I would add that you talk about microwave radiometry here. I would add references here which underline your statements written? For which algorithms is this statement valid? Is this a general phenomenon for ALL algorithms or are there better or worse algorithms to compute SIC? Maybe the SICCI reports can help or the paper by Rösel et al., in IGARSS 2012.
AR: The texts was corrected to "the microwave radiometry based...". Now Ivanova et al. (2015) is referred to.

Also, one important comment at the end of this paragraph ending on P2552, L2: The uncertainties provided with the SIC data are a physically based retrieval uncertainty and NOT an estimate of the / a potential bias. A potential over- or under-estimation of SICCI SIC during summer is hence NOT reflected in the uncertainty estimate given.

AR: Following the reviewer's comment, we revised the texts here:
The relative enhanced skill of sea ice thickness forecasts by LSEIK-3 with respect to LSEIK-2, does thus point to a possible issue with assimilating the summer SICCI ice concentration with the provided uncertainties. At first sight, the data uncertainties in summer sea ice pack seem to be too low (Fig. 2). For example, on 16 July 2010 when surface ice melting prevails and the microwave radiometry based ice concentration estimates are known to underestimate the physical sea ice cover (Ivanova et al. 2015), the provided uncertainties at the sea ice pack area are still lower than 0.06 with few regions exhibiting values around 0.1 (Fig. 2d).

In fact, Ivanova et al., (2015, section 5.3 "Melt ponds") report that AMSR-E and SSM/I, like all other passive microwave sensors, cannot distinguish ocean water (in leads) from melt water (in ponds) because of the very shallow penetration depths of the microwave signal in water. Therefore, these radiometric sea ice concentrations are closer to one minus the open water fraction (ponds and leads), than to the physical sea ice concentration in our models. This mismatch between the observed and modelled ice concentration (radiometric vs. physical) does not exist in winter when there is no surface melting. But in summer melt conditions, the observed ice concentration includes an unknown area of pond water. The provided uncertainties are not larger since the radiometric concentration is not more uncertain. This mismatch results in a systematic difference between the two quantities (the physical concentration is larger than the radiometric concentration) that cannot be fully mitigated by enlarged standard deviations of a Gaussian uncertainty model in Ivanova et al. (2015). The influence of melt-ponds on the accuracy of the SICCI dataset is documented in Lavergne and Rinne (2014, section 2.2.1.1 "summer melt-ponding").

P2552, L13-15: I agree to what you write here. However, these results have to be seen against what is used to convert draft –> thickness –> volume. The authors should always keep in mind what their reference data are and how well these are known. So, compared to NSIDC NT2 SIC, LSEIK-3 is better than LSEIK-4. Compared to BGEP SIT LSEIK-4 is a bit better than LSEIK-3.

AR: Comparing the satellite based SICCI and NSIDC sea ice concentrations, the original LSEIK-3 ensemble mean concentration was better than LSEIK-4, but the LSEIK-3 ensemble spread of sea ice concentration was too low in LSEIK-3, and this affects the performance for ice thickness. This is further discussed in the third paragraph of Section 4: This means that all members are very close to the ensemble mean and the data assimilation will have only little effect. LSEIK-3 has a similar spread distribution pattern of higher STDs in the sea ice edge area and lower STDs in the concentrated central ice area but overall higher STDs than LSEIK-2. Together with the fact that LSEIK-2 does not fit the thickness observations as well as LSEIK-3, this suggests that the ensemble forecast spread for sea ice concentration is too low and cannot reflect the uncertainty. As only observations of sea ice concentration are assimilated, sea ice thickness is influenced indirectly during the data assimilation through the point-wise covariance between the ice concentration and thickness. Here, the very small sea ice concentration variance leads to a very small sea ice thickness spread (Fig. 6b). This may explain why the LSEIK-2 system is not very effective in improving the sea ice thickness estimates.

P2553, L3: I would not use the term "under-estimation" here because as I wrote before the SICCI SIC retrieval uncertainty does not include potential biases and is not meant to do so.

AR: The reviewer is indeed correct here. In the revised manuscript, the section was re-written and now includes: (...) in summer melt conditions, the observed ice concentration includes an unknown area of pond water. The provided uncertainties are not larger since the radiometric concentration is not more uncertain. This mismatch results in a systematic difference between the two quantities (the physical concentration is larger than the radiometric concentration) that cannot be fully mitigated by enlarged standard deviations of a Gaussian uncertainty model."

P2553, L7: Having read the discussion I see the following 4 points missing: 1) The largest part of the SICCI SIC total error comprises smearing uncertainty outside areas of compact sea ice. this could / should be discussed. 2) See my comment d) 3) ULS SIT and its uncertainty are not that clear to me. Also see my comment d). How would the SIT curve from ULS whould be like?

1) AR: A new paragraph was added in Section 2, which describes the uncertainty maps from Figure 2 and particularly the reasons for enhanced uncertainties along the ice edge: The uncertainties are about 0.05 over packed ice and open water, but larger uncertainties up to and beyond 0.3 are present at the ice edge, and region of intermediate ice concentration values. The SICCI total uncertainties are indeed the sum of two components, one characterizing the algorithm uncertainties, and the other measuring the uncertainties due to representatives of 25 km daily averages, geo-location and instrument foot-print mismatch (Lavergne and Rinne, 2014). The second component to the total uncertainties is only pronounced in areas of gradients in the sea ice concentration observations – typically at the ice edge-, and amount for the inability of such coarse resolution satellite observations to accurately locate sea ice edge. Should the SICCI sea ice concentrations be

assimilated in models with significantly better spatial resolution, the enlarged uncertainties allow the model to freely locate its ice edge within the 25x25 km grid cells showing intermediate ice concentration values in the data.

2) AR: Currently, although there are some available satellite based sea ice thickness data in the cold season, e.g., Cryosat-2 and SMOS, but there is no such reliable data in summer. Thus, the validation of sea ice thickness forecasts are much more difficult than the sea ice concentration. We have added a clarification in the Part of "Forecasting experiment design": Currently, satellite-based observations of ice thickness are a challenge (Kwok and Sulsky, 2010; Kern et al., 2015), and there are very few reliable summer sea-ice thickness products available. Instead of remote-sensing data we compare .... Also we extended the discussion with: "However, as there are still no available satellite based sea ice thickness data in summer, in this study, the ice thickness validation are only based on two local ULS based observations".

3) As the SICCI dataset is the state-of-the-art sea ice concentration product, so in the previous version, we used SICCI SIC to calculate the mean thickness. To make this comparison more convincing, in this revision, we also show the calculated mean ice thickness based on NSIDC SIC. We also added some texts in the discussion: Also because we calculate the mean ice thickness using the local SICCI or NSIDC sea ice concentration data which is not real and certainly has potential bias, this further introduces uncertainties to the thickness calculations.

4) The original ULS SIT curves are shown below (in black dotted lines). As the numerical model carries mean thickness (volume over area) as a variable, the observed thickness is multiplied by SICCI or NSIDC local ice concentration to arrive at the observed ULS-SICCI or ULS-NSIDC mean thicknesses shown in the Figure below.
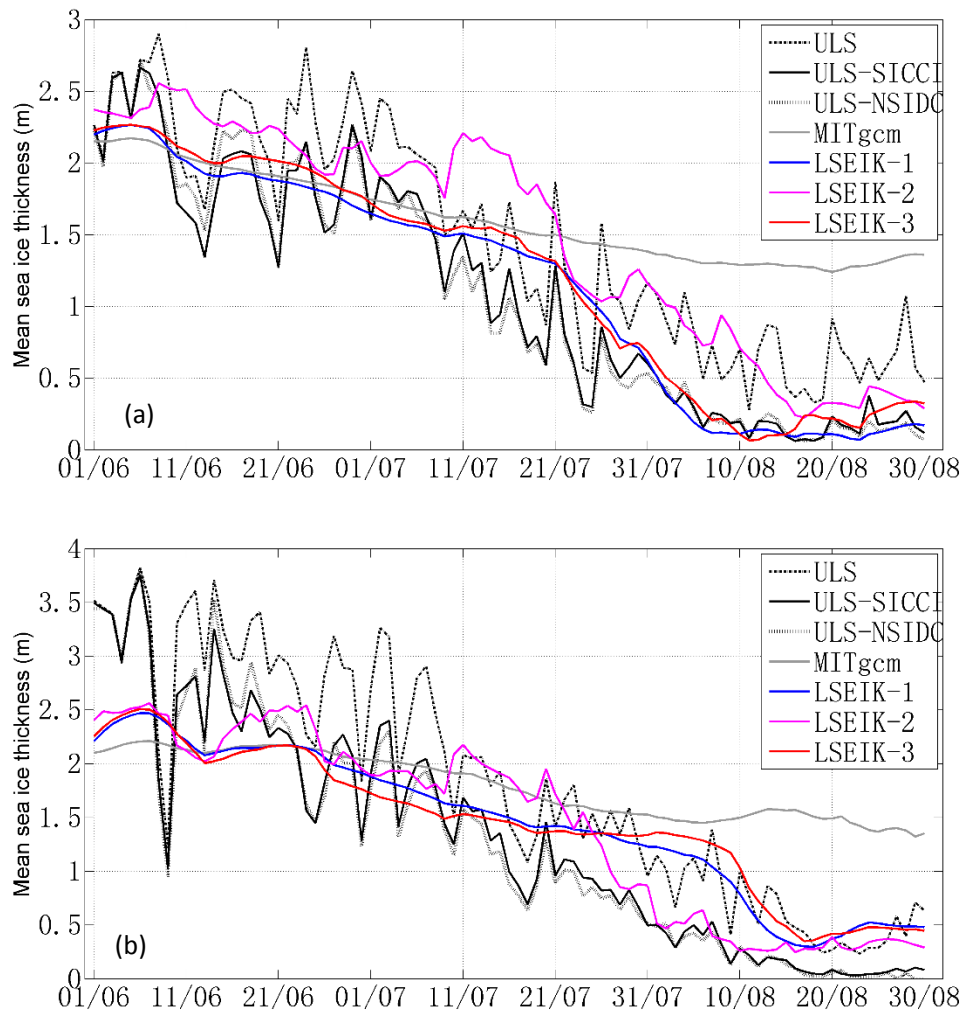
Figure xxx. Evolution of mean sea ice thickness (m) at (a) BGEP_2009A and (b) BGEP_2009D Beaufort Sea from 1 June to 30 August 2010. The black dotted lines show the original ULS thickness, the black solid and dashed lines show the obtained mean ice thickness using SICCI and NSIDC sea ice concentrations, respectively. The MITgcm free-run, LSEIK-1, LSEIK-2 and LSEIK-3 24 h ice thickness forecasts are shown as blue, magenta and red solid lines, respectively.

P2553, L10-12: Please rewrite sentence "While the ... distributors."
AR: Corrected.

P2553, L13: "better estimates" ... implies how accurate the reference is. It is important to tell what is the reference.
AR: Corrected.

P2553, L19: To me it is not clear what is meant by "mismatch"
AR: Now we clarify this in Section 4 of this MS : In fact, Ivanova et al., (2015, section 5.3 Melt ponds) report that AMSR-E and SSM/I, like all other passive microwave sensors, cannot distinguish ocean water (in leads) from melt water (in ponds) because of the very shallow penetration depths of the microwave signal in water. Therefore, these radiometric sea ice concentrations are closer to one minus the open water fraction (ponds and leads), than to the physical sea ice concentration in our models. This mismatch between the observed and modelled ice concentration (radiometric vs. physical) does not exist in winter when there is no surface melting. But in summer melt conditions, the observed ice concentration includes an unknown area of pond water. The provided uncertainties are not larger since the radiometric concentration is not more uncertain. This mismatch results in

a systematic difference between the two quantities (the physical concentration is larger than the radiometric concentration) that cannot be fully mitigated by enlarged standard deviations of a Gaussian uncertainty model in Ivanova et al. (2015). The influence of melt-ponds on the accuracy of the SICCI dataset is documented in Lavergne and Rinne (2014, section 2.2.1.1 "summer melt-ponding").

P2553, L22: Again the under-etimation of uncertainties used here.
AR: This was removed in the revised version, see our answer to your earlier comment on P2553, L3.

P2553, L24: How do you know? What is the basis for the knowledge? And further: What makes you so sure that the mismatch does not occur during winter?
Did you check it out by yourself?
In general: What would you expect using different assimilation processes?
AR: Now we clarify this in Section 4: This mismatch between the observed and modelled ice concentration (radiometric vs. physical) does not exist in winter when there is no surface melting. But in summer melt conditions, the observed ice concentration includes an unknown area of pond water. The provided uncertainties are not larger since the radiometric concentration is not more uncertain. This mismatch results in a systematic difference between the two quantities (the physical concentration is larger than the radiometric concentration) that cannot be fully mitigated by enlarged standard deviations of a Gaussian uncertainty model in Ivanova et al. (2015). The influence of melt-ponds on the accuracy of the SICCI dataset is documented in Lavergne and Rinne (2014, section 2.2.1.1 "summer melt-ponding").

This mismatch between the measured and modelled quantities calls for adopting more advanced data assimilation methodologies, e.g. embedding a matching relation in form of an observation operator, that would necessarily include modelled melt pond fraction, for successful assimilation of sea ice concentration satellite observations (from passive microwave instruments). Given the scope of this study and the comparisons with the in-situ BGEP ice thickness, the solution implemented in LSEIK-3, that is to enlarge the observation uncertainties using a minimum value of 0.10, is a pragmatic but effective approach.

Typos, etc.:
P2544, L12: "... uses improved ..."
AR: Corrected.

P2544, L12: "atmosphere weather" ... maybe the authors wish to be a bit more specific here? Or they simply write "weather"?
AR: Corrected.

P2548, L1: "tunes" –> "computes"
AR: Corrected.

P2548, L7: "... NSIDC ice concentration ..."
AR: Corrected.

P2548, L11: "Experiment Program" –> "Exploration Project"
AR: Corrected.

P2548, L21: "... are some differences between ..."
AR: Corrected.

P2548, L22: "... both data sets show ..."
AR: Corrected.

P2548, L23: "heavy" is maybe not the correct term here. Do you mean thick pack ice or closed pack ice or something alike?

AR: Thick pack ice. Corrected.

P2553, L 15: "... uses improved algorithms ..."
AR: Corrected.

P2553, L26: "...melting. Fully resolving ..."
References: I did not find Losa et al., 2012, and Nerger et al, 2006, cited in the text somewhere.
AR: Corrected.

Figure 1: I guess the small symbols deserve an a bit larger font, and thicker lines and also some more possibilities to act.
AR: Corrected.