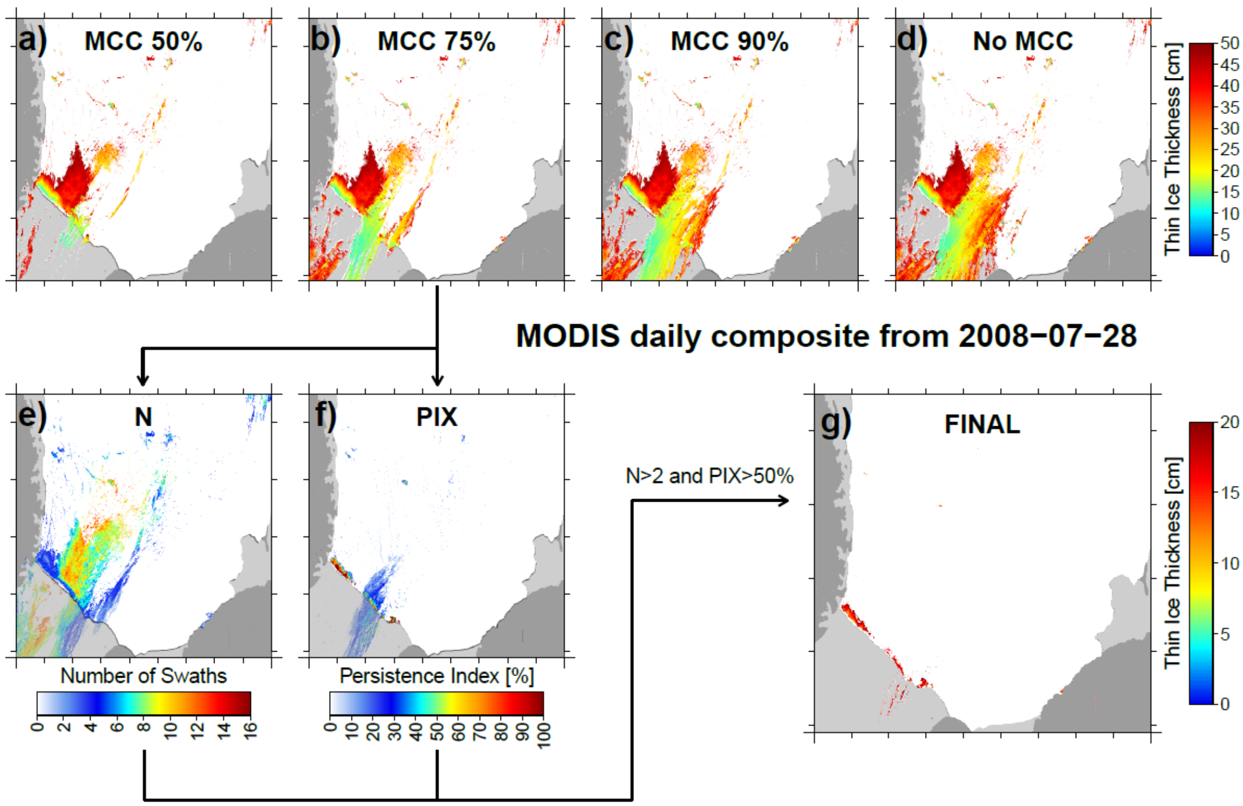
We would like to thank Stefan Kern for his in-depth review and the time and effort he put into this. We appreciate the feedback and would like to go through it point by point and highlight our changes accordingly.

General comments:

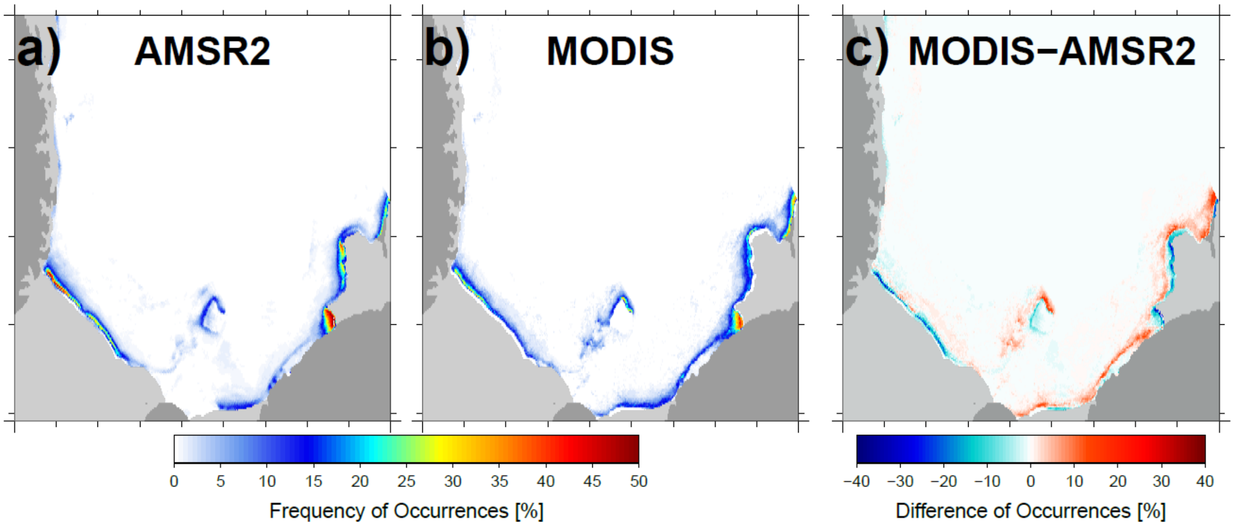
1) The paper shows many results which integrate over the entire time series produced. Since this - to my knowledge - is the first publication where the approach in its current version has been used to derive polynya area I would find it useful to see a) some specific examples (maps) of the polynya area detected for good (clear sky), bad (close to the worst case), and (perhaps) the bulk of the examples with a mix of clear sky and cloudy conditions b) some effort on inter-comparison with independent data on the scale of the daily maps, i.e. an inter-comparison with i) independent optical imagery such as Landsat and ii) SAR imagery. Even if these two points were issues in the other papers referenced in the context of explaining the method I would find the above-mentioned indeed useful - mainly because the method has been developed further.

We appreciate the comment of the reviewer and would like to comment on this:

a) We agree, that a case study of examples is so far lacking in the manuscript. As complete clear-sky conditions are very rare in a large area as the Weddell Sea, we decided to combine the reviewers' suggestion for a case study with his question for details on the persistence index and the lack of an example for the ERAI cloud cover usage later on. A high-resolution version of the Figure shown below along with some additional comments was added to the manuscript in the methods section ("2.3.2 Identification of cloud-cover affected data").



b) The spatial coverage of LANDSAT and its infrequent coverage of the area let us discard the use of LANDSAT for an inter-comparison. Also we did not have SAR data readily available to us, which excluded this suggestion as well. Both types of intercomparison would justify a paper by themselves. In order to prepare a best possible inter-comparison, we settled for the use of ASI AMSR2 3.125km grid data distributed by the University of Bremen (which the reviewer also mentions later). We projected the AMSR data set and a down-scaled MODIS data set onto a common equirectangular reference grid with roughly 3.5 times 3.5 km grid size (similar to what we did with our MODIS data in the first place, just with coarser resolution). We then compare the frequency of open-water/thin-ice/low sea-ice concentration occurrences with each other. While there are certain flaws in this comparison as for e.g. different definitions for POLA (thickness based vs. sea-ice concentration based) and reducing MODIS resolution, we believe it was the best possible way to add a comparison to the manuscript.



We added two more subsections to the Data & Methods section to explain our proceeding:

"For comparison, we use daily sea-ice concentration data derived from the Advanced Microwave Scanning Radiometer 2 (AMSR2) provided by the University of Bremen (Beitsch2014, Spreen2008). The ARTIST Sea Ice (ASI) algorithm estimates sea-ice concentration from polarization differences in brightness temperatures obtained from the AMSR2 89GHz channel. Since 1 August 2012, ASI sea-ice concentrations are available with a spatial resolution of 3.125km x 3.125km."

"A threshold of 70% was used for the definition of a polynya pixel from AMSR2. The comparison of frequency of polynya pixel occurrence from AMSR2 (Fig.3a) and MODIS (thin-ice thicknesses equal or below 0.2m, Fig.3b) for the years 2013 and 2014 reveals an overall good agreement. Both data sets were projected onto a common equirectangular reference grid, similar to the one used in the general data processing procedure, but with a coarser average resolution of approximately 3.5km x 3.5km.

The difference in frequencies is shown in Figure 3c. The majority of pixels lie in a ±10 % range, and there are positive as well as negative differences. Along the coastline of Filchner Ice Shelf and Coats Land, MODIS shows overall higher frequencies, whereas especially close to the ice-shelf edge of Ronne Ice Shelf and Brunt Ice Shelf AMSR2 shows higher values of polynya occurrences. However, effects of a different land/ice-shelf mask in AMSR2 data compared to MODIS are visible due to a gap in-between ice shelf and thin-ice areas (e.g. at Brunt Ice Shelf). As we will discuss later in the study, there are difficulties in discriminating between fast ice, adjacent icebergs, ice shelves and thin ice in passive-microwave data (Tamura2008). Additionally, this effect of a per-pixel shift between both data sets may partly result from the applied nearest-neighbor forward projection, when transferring the data to the common reference grid. The small underestimation by MODIS near the Ronne Ice Shelf may also be caused by persistent cloud cover in that area particularly during the presence of a polynya."

2) A minor, rather editorial general comment is the usage of "seasonal". I connect seasonal usually with something that varies over the course of the year (–> seasons), i.e. from summer over fall into

winter to spring. Here you talk about winter-time polynya area (April to September). I guess, you could avoid confusion if you choose "winter" or "winter-time" instead of "seasonal" throughout the paper.

We changed all instances of “seasonal” to either “winter” or “winter-time” throughout the manuscript.

Specific Comments:

P3960 L25/26: You might add "during winter" to this sentence

We changed that.

P3961 L3: "ice production occur predominantly under thin-ice areas ..." While that is true, an even larger amount of ice production might occur in the open water areas and/or in the areas where the ice is predominantly composed of frazil and grease ice.

The reviewer is correct in pointing this out and we changed this sentence to read:

“high rates of ice production occur predominately under thin-ice/frazil-ice areas as well as in open-water areas during austral winter.”

L7/8: This statement is true as well - however, what would one need to measure in-situ to obtain the ice production. Is this technically feasible?

While our intention was the use of measurements of the energy balance components, e.g. based from a ship maneuvered into the polynya, we admit that this was not clear from the given phrasing of the sentence. We decided to remove the sentence to avoid any misunderstandings, as it is not essential for understanding the scope of the paper at this point.

L13: "polynya dynamics" I am wondering what you understand under polynya dynamics? Variability in area? Variability in ice production? Variability in ice type? Variability in formation?

In the context of this paper and from parameters and sea-ice properties that we are able to derive from our MODIS data set, inter-annual changes in area/ice production, frequency distribution of polynya sizes and thin-ice classes, inter annual variations in polynya days as well as the corresponding energy –balance components are summarized by the term polynya dynamics. The term “dynamics” can be understood in its general meaning in describing processes and characteristics of changes. We do not think that we have to give a specific definition of “polynya dynamics”.

L21: I personally would add to the Nihashi and Ohshima citation that they provide a combined product of polynya area and fast ice area.

We changed that.

P3962 L3: Now, at the end of the introduction I have a few questions: i) Why did you pick the Southern Weddell Sea? Is it because polynyas are particular persistent? Is it because the AABW formation is particularly high? Is it because cloud cover is particularly low? Is it because there is a particularly high number of MODIS swaths covering the area of interest? In short: I would like to learn more about the motivation. ii) What is your opinion about why IR data have not been used yet for polynya monitoring in the Weddell Sea to the extend you are presenting it here? iii) Your definition of "Southern" Weddell Sea is a bit vague. Is there any specific motivation for the selection as it happened to be?

The Southern Weddell Sea was picked for several reasons. The main reason is that the Southern Weddell Sea polynyas are of particular importance to the formation of High Salinity Shelf Water (HSSW), which again contributes to the formation of AABW. Additionally, we can find relatively large polynyas of frequent occurrence along the complete coastal area that are to our knowledge not fully represented in other studies. This might be attributed to the lack of capability to resolve especially smaller features with established AMSR-E or SSM/I algorithms (with better algorithms and resolution in AMSR-2 this might change in the future).

A general problem with IR data not only in the Weddell Sea but also in other regions in Antarctica (and probably everywhere else in Polar Regions) is the problem of dealing with clouds and accounting for cloud-induced data gaps in a best-possible way. We consider this the primary reason for the lack of previous IR studies in the area.

The Southern Weddell Sea extent as defined in this study was chosen to i) cover all important sub regions that were studied by the majority of other studies, ii) was limited in its northern extent due a far retreating MIZ in Summer by potentially overestimating polynya area from open-water areas, which cannot be considered polynyas by definition. In order to reduce this effect, instead of delaying the start of our investigation period to May each year, we limited the spatial extent to maximize the temporal extent.

We also added the following paragraph for clarification:

"While the Weddell Sea features polynya activity along almost its complete coastal area, the majority of studies focus only on selected subregions. The investigation area was limited spatially to the Southern Weddell Sea, since this area is considered to be a key region for the formation of High Salinity Shelf Water (HSSW), and observational reference data of polynya dynamics are needed for sea-ice/ocean models (Haid et al. 2015)."

L7-17: Consider the reader might be not aware of the MOD/MYD29 product. Wouldn't it, in the case, make sense to give more details? To my opinion not clear are: 1) Is this a gridded product? You write that the data have 1 km x 1 km spatial resolution at nadir which poses the question: And what is the resolution off nadir? And if this is different from the nadir resolution how do you cope with the different grid resolutions in your approach? 2) Two different satellites contribute to this data set. 3) You write the data product was corrected already for the cloud influence. What can you say about the reliability of the used cloud mask that far south and during the dark season of austral winter. It is known that MODIS cloud masks are - not ideal - over cold surfaces like snow and ice in the high latitudes. You might want to comment on this – particularly because the cloud mask plays an important role in this paper later on. 4) As this is a swath product and we are in the high southern latitudes it might be interesting to know how many overpasses per day cover your region of interest (during day and during night). 5) It is said that the overall IST accuracy is 1-3 K. First of all, what is the impact of this accuracy on your results. Secondly, the IST is computed from the IR temperature measurement and an infrared emissivity. Do you know whether a constant emissivity is used regardless of surface (ice) type in the MOD/MYD29 product? If so, given the variation the infrared emissivity can have with a) the ice type while the ice is young and b) the surface incidence angle, what would you expect in terms of the contribution this has on IST accuracy and subsequently your results?

We agree with the reviewer that there is room for clarification in the manuscript:

1) The MOD29 product is gridded but without an applied projection and based on referenced lat/lon values. Due to MODIS acquiring its data from a constant-rotation scan mirror with constant focal length optics in a whiskbroom scanner, the spatial coverage of a pixel gets deteriorated towards the edge of a swath. Hence, the spatial resolution decreases off nadir with an increasing factor for along and across track direction. For the outer margin pixels of a MODIS swath, Fraser et al. (2009) state these factors to be 2.01 and 4.93, respectively.

2) This is correct and we added a clarification to the manuscript (see below).

3) We added a remark to the manuscript. Please see below.

4) Based on all swaths, including both day and night pixels and only discarding swaths with only daytime pixels, a total number of 140056 partially covering the complete Weddell Sea were obtained. However, only 78696 swaths were found to contain nighttime pixels in the Southern Weddell Sea. Hence, an average of 33 swaths is available per day. We added a clarification to the manuscript (see below).

5) In their latest study, Hall et al. (2015) stated that a constant emissivity is used for snow/ice areas in the MOD/MYD29 product. However, they did not state any numbers.

For clarification we rephrased the whole paragraph to read:

"In this study we make use of the MODIS Sea Ice product (MOD/MYD29, Riggs2006, Hall2004), which is a swath-based product that is derived from both MODIS sensors on-board the NASA Aqua and Terra satellites and distributed by the US~National Snow and Ice Data Centre (NSIDC). Our work is based on the ice-surface temperature data set (IST), which is one part of the sea-ice product. The provided data has

a~spatial resolution of 1km x 1km at nadir and each swath covers an area of 1354km (across track) x 2030km (along track). The spatial resolution decreases off nadir to about 2km x 5km at the swath edge.

The overall IST accuracy of the MOD/MYD29 sea-ice product under ideal (i.e. clear-sky) conditions is 1--3K (Riggs2006,Hall2004) and derived based on a constant emissivity for snow/ice (Hall2015) The MOD/MYD29 data product was already corrected for cloud cover using the MODIS cloud mask. However, the performance of the MODIS cloud mask is drastically reduced during polar nighttime due to the lack of visible channels and the low thermal contrast between clouds and the underlying snow/ice (Liu2014,Frey2008,Liu2004). We used this data for the austral winter period from 1~April to 30~September for the years from 2002 to 2014 that comprise a~total of 78,696 MODIS swaths which cover the Southern Weddell Sea area. This equals an average of 33 single swaths per day."

L19-22: I have a couple of question to the ERAI data. I guess it would make sense to include the answers to these also into the manuscript. 1) What is the temporal resolution of the ERAI data? 2) What was the criterium in terms of the time difference between MODIS swath and ERAI data to use the latter for the approach? 3) What is the (average) minimum time difference between MOD/MYD29 and ERAI data? 4) You write data the ERAI data are interpolated such that the spatially and temporally fit the MODIS data. Good. But what is the temporal and spatial resolution of the MODIS data (see above)?

We agree with the reviewer that there is also room for clarification on the ERA-I in the manuscript:

1) As mentioned in the manuscript, the temporal resolution of the ERA-I data is six hours (00, 06, 12, 18 UTC). We added the additional information on the UTC times to the manuscript.

2)/4) The ERA-I data is used regardless of time difference as it is necessary to derive ice thickness via the energy-balance and sea-ice model. While the coarse resolution of ERA-I, both temporal and spatial, is not perfect for the combination with satellite data, there is to our knowledge, no better data set available that yields the necessary meteorological parameters. In order to reduce the effect of the coarse temporal resolution of the ERA-I data, we interpolate it to 15-minute steps. As the MODIS data is acquired in steps of 5-minute satellite travel time, each swaths has a clear "closest in time" interpolated ERA-Interim field.

3) While we are not exactly sure what the reviewer is aiming at, based on the interpolated ERA-I data sets, the time difference between a swath and an ERA-I data point is either five or zero minutes. In comparison to the original ERA-I data points, the temporal difference can be a maximum of three hours.

P3963 L2-8: I don't understand the sentence :"which was found to show the in general best agreement with MODIS satellite data and the MODIS cloud mask during daytime." Which "MODIS satellite data" are meant here? The statement made here is based on daytime data ... and it seems that this is confirmed by the given reference ... but how about the agreement beteen ERAI medium-level cloud data and MODIS cloud mask during night time? Can we expect a similar agreement and if so, why?

Given the fact that the MODIS cloud mask might not be that ideal (see my comment above) one could have followed a different approach and not use the ERAI cloud information which fits best with the MODIS cloud mask (because the latter might be not correct) but perhaps consider all ERAI cloud data? Please comment on this.

We removed “MODIS satellite data” from this sentence. Concerning the daytime quote, we quoted the Liu and Key paper incorrectly, as they also did their analysis for Arctic nighttime as well. In both cases for the year 2013, the ERA-I medium cloud cover showed the best agreement with cloud estimates from MODIS as well as CALIPSO data. We changed that error in the manuscript and thank the reviewer for pointing this out to us. So despite day or nighttime, the ERA-I medium cloud cover appears to be a reliable additional procedure to account for clouds. The complete ERA-I cloud cover data set is not a good choice due to the dominating and higher than expected low-cloud-cover fraction as shown by Zygmuntowska et al. (2012). Due to comments of the first reviewer, we already introduced this remark in another paragraph concerning the use of long-wave downward radiation in ERA-I.

L10-22: 1) Please define what you understand under "thin" ice. 2) In Line 11 you write that you restrict the analysis to nighttime. How critical is this for the chosen time period lasting from April to September at the latitude considered in this study? Couldn’t one even take into daytime data as well for months May to July? 3) Is there a threshold IST above which a retrieval does not make sense anymore because of, e.g., a too large uncertainty? 4) For which young and thin ice types does this approach work?

1) Based on the reviewers’ question, we added a thin-ice thickness limit to the paragraph:

“The derived thin-ice thickness (TIT) of up to 0.5 m is calculated by using a surface energy balance model (Adams et al., 2013).”

2) The limitation to nighttime (defined as “no solar radiation”) is based on a per pixel check of the solar incidence angle. Daytime pixels are then masked out and are not used for analysis. In general, this only has a negligible effect on the overall coverage with swaths, given the high average number of swaths covering the investigation area (33on average, see above). While there are a few daytime swaths in the mentioned months, their number is rather small. However, the use of daytime swaths would introduce short-wave radiation in to the model which would make the use of albedo products necessary as mentioned in the manuscript. Therefore, we prefer to focus on nighttime only. Of course, during polar winter “nighttime” is 24 hours per day.

3) As long as IST does not exceed the freezing point of sea ice, there are no limitations.

4) The algorithm does not discriminate between ice types and assumes that a linear temperature profile can be used for the calculation of heat conduction in the ice.

L26: Unless statistically proven I suggest to replace "significantly" with "substantially" or "considerably".

We changed that.

P3964, L8/9: Which "common grid" is used and what is the grid resolution? You write not interpolation is applied. So the ERAI and IST data have gaps?

In accordance with the first reviewers' suggestion we changed that paragraph to read:

"First, all MODIS swaths were projected onto a common equirectangular grid with an average spatial resolution of 2km x 2km using a nearest-neighbor approach. The spatial resolution of this type of grid decreases slightly with decreasing latitude and vice versa. No interpolation between projected MODIS pixels was applied (i.e. data gaps are possible). Subsequently, ERA-Interim data, which exhibit a much coarser spatial (0.75°) and temporal (six hourly) resolution than the MODIS data, were linearly interpolated to spatially and temporally fit the MODIS data on the common reference grid."

ERA-I data does not have any data gaps. However, due to the use of the nearest-neighbor approach to match MODIS pixels to the best fitting lat/lon position on the reference grid, there is the potential for data gaps. However, these data gaps will be filled and accounted for by the SFR approach in a similar way as are the cloud-cover induced data gaps.

L12: "cloud covered data was identified" ... here the reader asks himself/herself: Why? Is this because the cloud mask is not good enough? Or what will the announced "correction" be made for? It is not clear here where the trip goes. It might have been good to already in the introduction say something about this, about the gaps in the data and that you intend to fill them. This would increase readibility of the paper I guess.

We agree with the reviewer and changed to last paragraph in the introduction to read:

"In this study, we present long-term results from coastal polynyas in the Southern Weddell Sea that were derived from Moderate-Resolution Imaging Spectroradiometer (MODIS) thermal-infrared imagery. Remote sensing of sea ice using thermal-infrared data yields the opportunity to monitor thin-ice thicknesses and distribution on a regular basis (Adams2013,Willmes2010, Yu1996). While the MODIS cloud mask (MOD35) shows in general good results for polar daytime (e.g., Frey2008, Liu2004), its performance decreases during polar nighttime (e.g., Holz2008,Frey2008,Riggs2006,Hall2004). This results from the lack of visible channels, as well as the low thermal contrast between clouds and the underlying snow/ice (Liu2014,Frey2008, Liu2004). We therefore use additional criteria to identify cloud-covered areas and use a new approach for gap filling with respect to thin ice thickness (Paul2015a). While the basic thin-ice retrieval was used before in Arctic regions (e.g. Preusser2015, Adams2013, Willmes2011, Willmes2010), we present now the first continuous and cloud-cover corrected time-series of polynya

dynamics during the austral winter period (April to September) and for the complete coastal area in the Southern Weddell Sea for years from 2002 to 2014."

L19/20: You write "This data is also aggregated into daily composites" Why "also"? I haven't read about any aggregation into daily composites yet. Yes, further up (L10) you write that you compute TIT as a mean daily composite. But also here it is not clear how this is done. I assume there are several swaths in the MOD/MYD29 IST product per day so that you could have 1,2,3, ... x swaths within one pixel. Do you average these? Or do you always use the latest swath per day?

We acknowledge the lack of clarity in this paragraph and thank the reviewer for pointing this out to us. We rephrased the paragraph in 2.3.1 to read:

"Based on these adjusted swath-based data sets, TIT is calculated pixel-wise and daily TIT composites are calculated based on the median thin-ice thickness and corresponding ice-surface temperature of all available swaths per pixel. The resulting composites comprise TIT and IST data together with the daily swath-based median energy-balance components of each thin-ice pixel. In the next step, cloud-covered data was identified and flagged as will be described in the following subsections."

Additionally, we rephrased the mentioned example in 2.3.2 to read:

"The binary cloud-cover information (i.e. depending on whether a pixel is cloud influenced or not) of both the MODIS cloud mask as well as the ERA-Interim medium-level cloud-cover fractions is pixel-wise aggregated into daily composites and separated into four different cloud-cover dependent classes."

P3965, L1-3: Here it would be good to know how many swaths could cover one pixel and whether this number is latitude dependend - or in other words does this number across your area of interest?

We added the following sentence to clarify this for the potential reader:

"On average, about five swaths cover each pixel in a daily composite. This number of swaths can vary mainly due to cloud cover with up to 20 different swaths covering a single pixel or region."

L8: "swaths per pixel" I suggest to add "per day" L8-10: Will this persistence index be explained later? Will we see "typical" values of it in a table and/or a map?

We added the reviewers' suggestion to the text. We also agree with the reviewers' implication that the manuscript in its current state lacks additional information and details on the use of the persistence index. In addition to the changes made for the first general comment we therefore added a new paragraph to give insight to the reader.

"In the two-step procedure, we first apply the SFR approach to all cloud-free pixels (i.e. pixels in the ccs class and mcp class where the majority of pixels show clear-sky conditions) that also feature a PIX value greater than 0.5 (i.e. pixels with a thin-ice thickness <0.2m present in more than 50% of the swaths covering that pixel) in the 7-day interval and are covered by at least three swaths. Pixels that do not match these two criteria are considered of lesser quality and are discarded from the SFR approach.

Based on Equation 4, the binary information (i.e. thin-ice or no thin ice) of the six days surrounding the day of interest is matched with their corresponding weights and the polynya probability estimated. A probability value above 0.34 is classified as polynya area in a resulting binary image.

Subsequently, cloud-contaminated pixels with a probability above 0.34 are assigned a~pixel-wise weighted average ice-thickness and ice-surface temperature value based on the six days surrounding the initial day of interest (doi) (TITadd/ISTadd, Fig. 2). Weights are applied in the same configuration as for the SFR approach itself. Additionally, pixels categorized as lesser quality that feature a probability above 0.34 can be up-valued and are assigned their original thin-ice thickness and ice-surface temperature values."

L12-27 to P3966 L2: I understand that you here try to very briefly describe what the SFR and the PE is about and how these are used together with the new thing done in this paper. I have the feeling that this comes a bit short here. At least I have some open questions. 1) In L12/13 you write about "low quality data and that these are accounted for" ... In which sense low quality? What is the problem with the data? Then you are accounting for these low quality data ... Why? Do you want to remove them? Do you want to replace them? Here, the answer again would be to take the reader by the hand already in the introduction, telling what will be done step by step and why. L18: What are the "drawbacks" mentioned here? L19: "no coverage above 50%" ... Coverage of what? Clouds? L22: Instead of "three before and three after" etc. you could perhaps write: " ... information of a 7-day interval centered around the day of interest is weighed directly ..." L24: What is a "daily median composite"? For what parameter this composite is computed? L26ff I suggest to move the "The PE ... as thin ice" further up to where things are said about the PE

In accordance with the first reviewers' suggestion we substituted "low-quality data" with "cloud-contaminated data" and hope that this move clarifies some of the issues in this paragraph also for the potential reader.

Additionally, we rephrased the whole paragraph in order to increase the readability in accordance with the reviewers' comments to read:

"Employing the above mentioned two procedures of cloud-cover dependent classification and persistence index calculation, we are able to identify cloud-contaminated MODIS data in each daily TIT composite. These are then complemented by a~two-step procedure utilizing a combination of spatial feature reconstruction (SFR, Paul2015a) and proportional extrapolation (PE, Preusser2015).

In the SFR approach, the information of a 7-day interval ($doi_\textrm{-3...-1}$ and $doi_\textrm{1...3}$) centered around the day of interest (doi) is weighed directly proportional to its temporal proximity to the initial day of interest. This yields a~probability of thin-ice occurrence for the day of interest based on the surrounding six days. (Eq. \ref{eq:sfr}).

Information about polynya area is on average significantly correlated within at least three days and $>$90\,\unit{\%} per-pixel gaps are shorter than four days (Paul2015a). We use the set of weights (w_3 = 0.02, w_2 = 0.16, w_1 = 0.32) and probability threshold (th = 0.34), which featured the highest spatial correlation in the analysis of \citet{Paul2015a}. A~detailed description and analysis of the SFR approach and its setup is given in \citet{Paul2015a}.

The PE approach on the other hand assigns thin ice to cloud-covered areas in the same proportion as it is detected in the cloud-free area. For example, if a~region is 80% cloud free and 50% of the cloud-free area features a~thin-ice signal, then 50% of the cloud-covered region is considered as thin ice.

In the two-step procedure, we first apply the SFR approach to all cloud-free pixels (i.e. pixels in the ccs class and mcp class where the majority of pixels show clear-sky conditions) that also feature a PIX value greater than 0.5 (i.e. pixels with a thin-ice thickness $\leq$ 0.2m present in more than 50% of the swaths covering that pixel) in the 7-day interval and are covered by at least three swaths. Pixels that do not match these two criteria are considered of minor quality and are discarded from the SFR approach.

Based on Equation \ref{eq:sfr}, the binary information (i.e. thin-ice or no thin ice) of the six days surrounding the day of interest is matched with their corresponding weights and the polynya probability estimated. A probability value above 0.34 is classified as polynya area in a resulting binary image.

Subsequently, cloud-contaminated pixels with a probability above 0.34 are assigned a~pixel-wise weighted average ice-thickness andice-surface temperature value based on the six days surrounding the initial day of interest (doi) (TIT$_{\text{add}}/$IST$_{\text{add}}$, Fig.~\ref{fig02}). Weights are applied in the same configuration as for the SFR approach itself. Additionally, pixels categorized as minor quality previously, but feature a probability above the threshold of 0.34 in the SFR output are "up-valued" from minor to high quality and are assigned their original thin-ice thickness and ice-surface temperature values.

The remaining coverage gaps that could not be corrected for by this approach, e.g. due to temporal gaps longer than three consecutive days, are filled by the proportional extrapolation (PE) scheme (Preusser2015).

In case that after the application of the SFR approach more than 50% of the investigated sub region is cloud-contaminated, daily estimates of polynya area and ice production will be interpolated between neighboring days with sufficient (i.e. above 50%) cloud-free coverage.”

P3966, L5: "surrounding days" ... are we still talking about the 7-day interval?

Please refer to the changes made above.

L7: How is the POLA actually derived then? Is a TIT threshold used? Is a IST threshold used? Is the non-used parameter used to check the "skill" of using the other parameter? In other words is IST used to cross-check whether the TIT-based POLA makes sense? How accurate is the POLA? What is the minimum (average) change in POLA which can be reliably determined?

Based on a comment made by the first reviewer, we changed the first paragraph to read:

"From our cloud-cover corrected daily thin-ice thickness composites, we then derive daily polynya area (POLA, defined as area with open water and thin-ice between 0.0m and 0.2m thickness) as well as the accumulated winter-time ice-production (IP) from heat loss for each POLA pixel (e.g. Tamura et al., 2011, Willmes et al., 2011)."

No cross-validation is conducted between TIT and IST as this type of validation would be potentially affected by i) the stated uncertainty induced by a diurnal cycle in a daily IST composition and ii) the fact that there is no fixed temperature range which can be associated with thin ice aside from the freezing point of sea water for open-water areas. The temperature associated with thin ice can vary depending on the surrounding air temperature that controls the heat flux through the ice and into the atmosphere from a thin-ice cover.

Without available cross-validation data from at best independent approaches it is hard to correctly assess the accuracy. The smallest detectable average change would be the average spatial coverage of a pixel in the common reference grid of 2x2km.

L10: Am I right assuming that delta t in Equation (1) is 24 h because you base the computation of the ice volume production on a daily polynya area? Am I right assuming that you are not accounting for sub-daily variations in ice production?

The reviewer is correct with his assumptions. We hope to also clarify this for the potential reader with the following statement added to the manuscript:

"All estimates of volume ice production are integrated over the period of 24 hours, assuming constant ice-thickness and ice-surface temperature. Hence, we are not able to account for sub-daily variations with this procedure."

L15-17: These TIT distributions and frequency distributions ... are these computed for each polynya separately or only for the entire Southern Weddell Sea as a whole?

These parameters are calculated based on POLA pixels for each of the six sub regions. We slightly altered the phrasing of the sentence to hopefully clarify this concern for other potential readers:

“Furthermore, thin-ice thickness distributions of daily POLA as well as frequency distributions of thin-ice occurrence are calculated for each subregion. The results are then put in context with other recent remote sensing and model studies.”

L20: "MODIS coverage" ... of what?

We exchanged *"MODIS coverage"* with *“The spatial coverage of cloud-free MODIS data”* to clarify our intention.

P3967, L16 "RAW" See comment at Table 1.

We changed that.

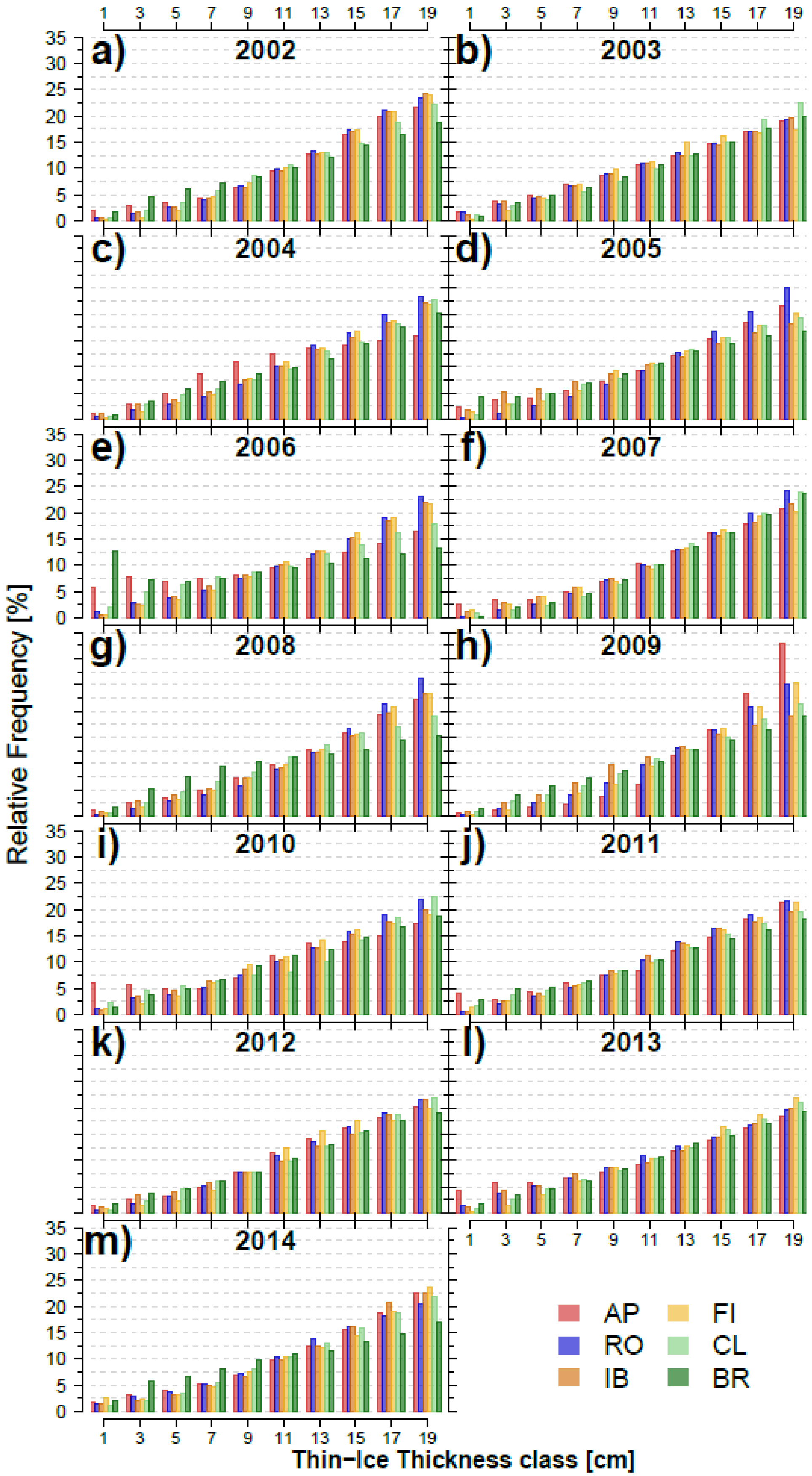
L17: Coverage of what?

We added *“…whenever no coverage with cloud-free information…”* to the sentence.

P3967/3968, L26 to L3: I am not too happy with the very global interpretation of Figure 3. I cannot read from this that "does not show strong regional nor seasonal differences". You are averaging over such a large number of pixels and over 13 years that I doubt that such a global statement can be given here. A particular persistence of one polynya in terms of TIT might outrule a high variability in another polynya. My suggetion would be to add a sentence that more detailed results (like Figure 8) are shown later in the paper. My suggestion would further be to take a look into your dataset and figure out the different types of polynyas in terms of persistence and in terms of their typical opening / closing scenarios. Why? Because a polynya which is steadily kept open by persistent katabatic winds is supposed to contain a large fraction of frazil and grease ice, maybe even small pancakes until at its leeward side one enters the more consolidated ice where the so-called frazil ice collection depth (a term from numerical modeling of polynyas) indicates that here typical ice thickness could be 0.1m already on average. In contrast, a polynya without persistent katabatic winds will have the above-mentioned situation paired with periods where the polynya simply freezes over with nilas (under calm conditions). These two different processes cause a different typical ice thickness distribution within a polynya and I am wondering how your approach is reflecting this. A second thing coming into my mind is that the variation in the standard deviation for the thickness classes shown could perhaps be caused by the variation of the numbers of values falling into these thickness classes.

Based on the reviewers’ question, we included the following Figure to the response letter in order to highlight the inter-annual and regional thickness distributions per year. With very few outliers (e.g. 2006 for BR), there is not much new insight gained from a split up version of the given plot and we still feel confident in our statement made in the manuscript.

While we welcome the suggested idea, we are not sure if our method really holds the potential to resolve these differences. Furthermore, we expect the additional work without several independent studies to compare our results with, which are out-of-the scope of this manuscript. We would rather focus on the comparison of annual and multiyear results where we have comparison studies at hand. However, a more detailed analysis of case studies is a great idea for a future publication based on the presented method. However, we do agree with the reviewer that starting the results section with this rather global statement was a poor choice, which is why we swapped the part about thickness distribution with the frequency distribution in the manuscript.



a) 2002
b) 2003
c) 2004
d) 2005
e) 2006
f) 2007
g) 2008
h) 2009
i) 2010
j) 2011
k) 2012
l) 2013
m) 2014
Relative Frequency [%]
Thin−Ice Thickness class [cm]
AP
RO
IB
FI
CL
BR

P3968, L4-13: I agree with the statement made here, that only considering the RO is likely to lead to an underestimation of the ice production and associated potential water mass modification associated with the polynyas considered. However, you could perhaps question here that it is not clear whether the contribution of these other polynyas is that relevant and that you will indeed show this later in the paper. You might also take a look at Kern, 2009, where in Figure 1b you can see that this author did indeed as well look into the contribution of different polynyas and not just the RO. For that paper the author needed to focus on a few polynyas / polynya regions from which it was thought that these are key for the paper. As far as I know this was work carried out in a German National Funding project and perhaps you could contact the author for the final report of that project. In Figure 2 of that paper you also find information about the average maximum number of polynya days for the regions which were selected in the Weddell Sea. From there you could see that indeed the region which is termed Halley in that paper which includes your BR contribution, for example, is a region with a high average number of polynya days (during winter) (Fig. 2 a) but that the persistence is quite low (Fig. 2b).
Finally - and this goes back to my general comment 1, I would have loved to see more than just that full-period winter-time average map of thin ice thickness distribution given in Fig. 4. This figure does not give information about the inter-annual variability. It does in particular leaves the question open whether areas showing 40% thin ice thickness occurrence have these because the ice is thin there anyways or because there have been 2 winters out of the 13 winters where there was no ice at all for a long time of the winter season ... (ok, this is not the case, but perhaps you see my motivation to ask this question). As we see in Figure 5 (b, d and f) there is indeed some interannual variation in the POLA. How about you prepare a set of maps where you show the thin ice thickness distribution for those years where POLA is particularly large and particularly low in a respective polynya. You could focus on regions shown in Figure 5 b), d) and f) and provide a panel with 6 maps, three for maximum POLA and three for minimum POLA in the respective regions. I guess this would be extremely informative also in the context how the thin ice distribution looks in the other polynya regions during years where a maximum or minimum was reached in the selected region.

We agree with the reviewer that the current Figure does not properly display inter-annual change which is clearly present in the data from later analysis. We therefore exchanged the current plot with a high-quality version of the presented quick-look Figure below. Instead of limiting us to the three major regions, which would have been an overall improvement as well, we feel like a complete analysis agrees better with our overall scope for the manuscript to analyze the complete coastal area.

We changed the corresponding paragraph about the thin-ice frequency of occurrence distribution to read:

"Almost the complete coastal area in the South and East of the investigated Southern Weddell Sea features a~recurrent thin-ice signal for the years from 2002 to 2014 (around 30\,\unit{\%} and above, Fig.~\ref{fig04}). The overall high recurrence of thin ice is a~very important finding when considering that the primary focus of many studies lies solely on the Ronne Ice Shelf (RO) region when investigating the Weddell Sea (e.g. \citealp{Nihashi2015}). This neglects the importance of e.g. the Brunt Ice Shelf (BR) region to the overall ice production and ocean-atmosphere heat exchange. It also underestimates the inter-annually highly variable contribution of the area around the grounded iceberg A23A (IB), Filchner Ice Shelf (FI) and Coats Land (CL) to the bottom-water formation due to salt release during ice formation.

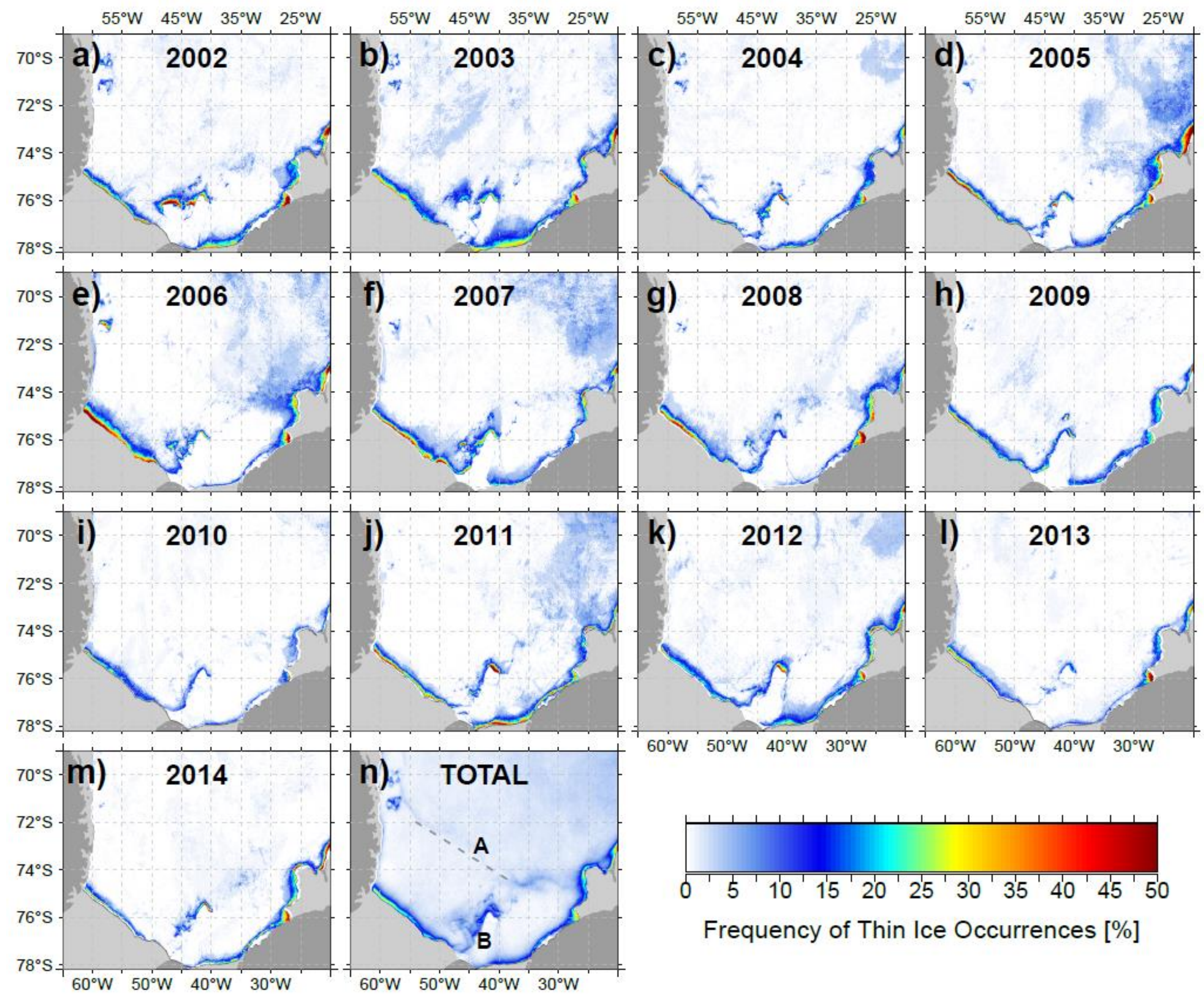
The relative frequency of thin-ice occurrences in the Antarctic Peninsula (AP) region is spatially focused around smaller grounded icebergs and rather low compared to the other sub regions. The inter-annual contribution also decreases during our investigation period in the years 2007 to 2010 (Figs.~\ref{fig04}f-\ref{fig04}i), when the group of icebergs detach from the ground.

The very light blue areas in the overall thin-ice frequency distribution (Fig.~\ref{fig04}n) found in the northeast of the investigation area correspond to a~low sea-ice concentration area due to the far south lying marginal ice zone (MIZ) in April in the years 2005, 2006 and to a smaller extent in 2007 near the Brunt Ice Shelf region (at around 75{\degree}\,S, taken from AMSR-E observations, not shown). This is also reflected in our thin-ice frequency distributions of said years (Figs.~\ref{fig04}d-\ref{fig04}f).

There is a~sharp separation between a~zone with present activity in the North and North-East to a~zone with almost no activity in the South and South-West closer to the Ronne and Filchner Ice Shelves in the total frequency distribution (Fig.~\ref{fig04}n, A). This line of separation (Fig.~\ref{fig04}n, A) coincides very well with the course of the continental slope (Fig.~\ref{fig01}, \citealp{Arndt2013}). However, no conclusive explanation can be given.

\citet{Markus1996} studied the effect of the grounded iceberg A23A on the ice production in front of the Filchner Ice Shelf and found a~drastic average increase in sea-ice concentration (i.e. a~decrease in the amount of thin-ice area and hence ice production) during the freezing period. While the recurring formation of a~fast-ice bridge is also visible in our results (a~low frequency area between the coast and the grounded iceberg A23A, Fig.~\ref{fig04}n, B), we still find high polynya activity in front of Coats Land (CL) and the Filchner Ice Shelf (FI). Especially the area around the grounded iceberg A23A (IB) and the westward side of the ice bridge show high thin-ice occurrences throughout all years (Figs.~\ref{fig04}a-\ref{fig04}m). It is also possible to see differences in extent and persistence of this ice bridge from our data set via the absence of thin-ice occurrences.

In comparison to results found by \citet{Nihashi2015}, our frequency estimates for the Ronne Ice Shelf (RO) region between 30 and 40\,\unit{\%} are in good agreement."



L10: I would add "heat" between "ocean-atmosphere" and "exchange"

We added “heat”.

L16: I would refer to Fig. 4 at the end of the first sentence of this paragraph.

We changed that.

L19: "Those two years" ... here you try to explain whether the marginal ice zone located quite south could have had an impact and you could not come to a solution. Would it perhaps help to redo Figure 4 without taking April 2005 and April 2006 into account?

We corrected that issue in the changes made above.

L23-25: Could it be that the line you are talking about here coincides with the border between the region where the Weddell Gyre transports sea ice from the northeast into the southern central Weddell Sea and the region where sea ice from the southern Weddell Sea polynyas is advected north? Here could lie kind of a shear zone where you once in a while might encounter increased lead formation. Maybe taking a look into ice drift fields could help here.

We thank the reviewer for the additional suggestion and consider this a valuable remark. Indeed, the position of this line corresponds to the shelf break, which separates warmer Atlantic water from colder shelf water. However, we would prefer to focus on a pure description of the mentioned observation without including speculation that could not be proven within the scope of this manuscript.

P3969, L16-25: I have some comments / remarks here. 1) You mention correctly that the time-series is quite short and that a temporal analysis with regard to a trend might not be appropriate. But then you come up with the 10-year trend and talk that this is significant etc. ... well, if you would have chosen 2006 as the start year of your time series analysis I bet the trend would have been even steeper. What is the relevance here to try to come up with different trend periods. Are you aiming to connected these to changes in ocean and/or atmospheric circulation? Could it be that 2006 & 2007 are simply positive excursion and 2013/2014 negative excursions from an otherwise "no trend" temporal development? 2) For the 10-year trend in POLA you give a value of 347.60km^2 / year. I am wondering in this context how accurate your POLA estimates are. I haven't found any information about this yet. Independent of that it might look better to not use any decimals here but write 348 km^2 / year. 3) L23: The POLA and ice production numbers given here do not fit what is written in the figures

1)/2) We removed the mentioned section of the paragraph.

3) As stated in the manuscript the numbers presented here represent the change over the whole 13 years in contrast to the figures which state an annual trend. We agree that this might be misleading the reader and changed all numbers in the text to annual trends to agree with the numbers presented in the figures.

P3970, L25: Here you switch to a comparison with model results. On P3972, L5, you switch back to satellite data inter-comparison. Wouldn't it make sense to stay with satellite data inter-comparisons first and then consider the inter-comparison with the model? In the entire results / inter-comparison section I suggest to either speak of "our results", "our estimates" etc. OR of "MODIS results", "MODIS estimates" etc.

We followed the reviewers' suggestion in using "our" results/estimates. However, we did not move the model part. As the model comparison is also a rather big part of this manuscript we feel it is necessary to go into this data set in depth based on inter-annual values before we move to the inter-comparison with global values based on satellite as well as again the model results. As we do in general not have inter-annual estimates from other studies aside from the model study, we moved that part to the end to finish up the POLA/IP inter-comparison with the available global values.

I have a few questions to FESOM: 1) How is the POLA defined / found in FESOM? How is it computed? 2) How is ice production computed in FESOM? 3) How does FESOM treat different ice types occurring in a polynya? 4) You POLA and TIT product is a composite how 1 km to x km MODIS information and 0.75 degree interpolated to same resolution ERAI data. How about FESOM? Particularly interesting information could be the spatial resolution, the computation time step, the time with which variables like ice production are output, and the atmospheric forcing. The latter is mentioned later but it might be good to have this information upfront.

Most answers to the reviewers' questions can be found in Haid and Timmermann (2013) and Haid et al. (2015).

1) Model nodes are defined as polynya area when either the sea-ice concentration is below 70% or the ice thickness is below 20 cm.

2) From the fully coupled ocean/sea-ice model and the calculation of fluxes and energy-balance components, the authors calculated ice production from the release of latent heat. However, they did not neglect the contribution of the ocean heat flux instead of compensating all heat loos to the atmosphere by release of latent heat.

3) FESOM simulates snow and sea ice thickness, which are both assumed to be evenly distributed over the ice-covered part of each area unit of the finite elements.

4) The spatial resolution of the used NCEP data is given by the authors with 1.875° and was spatially and temporally interpolated to fit with the model data (3-5km resolution near the Ronne ice shelf), with model time steps of three minutes. All variables are output as daily mean values.

We added the following paragraph to the manuscript to clarify this given information for the potential reader:

"Haid and Timmermann (2013) used NCEP reanalysis data (Kalnay1996) with a spatial resolution of 1.875° as their atmospheric forcing and defined areas with either a sea-ice concentration below 70% or an ice thickness below 0.2m as POLA. Haid and Timmermann (2013) calculate sea-ice production from the release of latent heat but also take the oceanic heat flux into account. The model resolution is 3-5km near the Ronne Ice Shelf."

P3971, L6: "is" or "could be"?

We changed "is" to "could be".

L11/12: You write "Topographic effects on the wind such as katabatic winds and barrier winds influence a broader region due to the smoothed topography in NCEP." Is this true? I would question, whether with a smoothed topography katabatic winds are resolved or even present from the model physics in NCEP at all. Maybe we simply see a distortion of the surface flow due to some topography but this is basically a mix of geostrophic and thermal influences. In this context: What do you think is the influence of the katabatic and barrier winds for the other polynyas, i.e. not the AP? Maybe you can state this difference in the paper to clarify what is presumably the dominant atmospheric forcing of the polynyas considered.

Clearly NCEP data include topographic effects such as katabatic winds, but the smoothing leads to errors in the field of the Antarctic coastal regions (see Haid et al. 2015 for a direct comparison of NCEP with high-resolution atmospheric modelling). Therefore the original statement is correct.

L18: I guess it is 2003 only here.

The reviewer is correct and we changed the text accordingly.

L16-23: If I look at the average values and their standard deviations then I would feel confident to make the statement that for RO the values almost agree within one standard deviation. Since you are discussing POLA as obtained from model and MODIS for AP quite a bit, I feel provoked to ask the following questions: Why, when I look at Figure 9, is the modeled POLA always < MODIS POLA for RO except for 2003? What is / was so different in 2003 compared to the other years? Why are modeled POLA always < MODIS POLA in general (except 2003, RO) for RO and BR? Why is in 2006-2008 the modeled POLA that much below the MODIS POLA? Could this be explained by the way how in FESOM polynya area is defined?

Unfortunately, Haid and Timmermann (2013) do not give a statement on what was going on in the year 2003, which yielded the highest average POLA value in their study. However, from their data there are at four more years with relatively high average annual POLA values in the nineties (not shown in our study) and 2002.

P3971, L24 to P3972, L4: For RO the average modeled and observed ice production agrees within one standard deviation. This is cool. But what happened in 2002 and 2003? Why is the modeled ice production in these two winters so much off the MODIS based ice production? And why is this not the case for BR? How about regions IB and FI? You write that model estimates of the ice production are presumably smaller than the observations because oceanic heat fluxes are neglected in the computation of the ice production in your MODIS data based method. That is true with regarding that

a positive oceanic heat flux reduces ice production per area. A positive oceanic heat flux could, however, also have an impact on the size of the polynya and hence on POLA; it would increase POLA. Now the question is - since I don't know from the paper - how is POLA defined? If by the TIT then neglecting oceanic heat fluxes in your method does not only reduce the ice production per area but at the same time could cause a smaller POLA. You could write this double effect in the paper

The increased ice production also correlates with the increased average POLA found in the paper of Haid and Timmermann (2013). So it is not a one-time thing but just not represented fully in our comparison. From the mentioned study of Tastula et al. (2013) NCEP was found to show warm bias of 2 K over ERA-I, which in turn reduces ice production. One might speculate, if these biases found in the average for a particular winter are always present, i.e. colder and windier extremes are present more frequently in the NCEP data which causes these *"anomalies"*. These changes might be of a higher impact in the coldest region near the Ronne Ice Shelf compared to the Brunt region. Nothing can be said about IB and FI as they are not discussed in the model study. However, Haid et al. (2015) show a model intercomparison between NCEP, a global mesoscale analysis and two high-resolution mesoscale models and a cold bias up to 10K is found for NCEP near the Ronne ice shelf (average of a winter season). This study shows that the differences in temperature and wind vary strongly in space, and that they are most pronounced for the Filchner-Ronne ice shelf region. The oceanic heat flux is found to be around 50 W/m² in this region, which is small compared to the total atmospheric heat flux of 400-500 W/m².

We thank the reviewer for his suggestion on the topic and we added the following paragraphs to also state the made comments to the potential reader:

"A positive oceanic heat flux reduces sea-ice production and at the same time potentially increases polynya area due to an increased amount of thinner ice. Hence, neglecting the oceanic heat flux increases ice production and potentially reduces polynya area in our satellite observations. However, Haid and Timmermann (2013) show that the oceanic heat flux is only of the order of 10% of the total atmospheric heat flux."

"One can only speculate about the reasons for the increased average annual POLA and IP in the years 2002 and 2003 as there is no statement in the study of Haid and Timmermann (2013). Potentially, a mix of below-average air temperatures in combination with above average wind speeds caused this increase in ice production with at the same time increased polynya size/persistence."

P3972, L5: I suggest to add "based on satellite observations" after "different studies"

We added *"based on satellite observations as well as models"* as the stated problem is not limited to satellite data but also applies to the presented model study as this is a major problem also discussed in the manuscript.

L13-18: I am not sure whether the authors are mixing two things here which one could perhaps write separately. 1) One thing is the difference in the spatial resolution between using MODIS IR data and passive microwave (PMW) data. The fraction of mixed pixels with an influence of the different surface types potentially encountered in an Antarctic coastal polynya to the fraction of "clear" polynya pixels is for sure much larger for PMW data than for IR data. Depending on how POLA is derived from PMW data I would assume that both, an over- and under-estimation of the actual polynya area and/or an IR-based estimate of the POLA is possible. If mixed pixels are counted as POLA then PMW over-estiamtes POLA, if mixed pixels are excluded from POLA, then PMW under-estimates POLA. 2) The second thing is that indeed radiometrically, fast ice, ice bergs and thin ice can exhibit similar emissivities and hence brightness temperatures so that fast ice / icebergs could be interpreted as thin ice and vice versa. I am kind of buying the argument that "fairly narrow coastal polynyas" cannot be observed by PMW ... however, the PSSM applied to SSM/I data (Kern, 2009) may resolve coastal polynyas reliably down to 10 km width of the thin ice area; the same method applied to AMSR-E or AMSR2 data has the potential to resolve smaller structures; finally, AMSR2 sea ice concentration maps as - e.g. - provided by University of Hamburg has a grid resolution of 3.125 km ... maybe - and this is what I am hoping for - you could add to your analysis how accurate your method allows us to delineate a polynya, what is the minimum size a polynya needs to have so that its area is derived reliably and what is the minimum detectable change in polynya area? This way the reader gets a quantitative assessment of the advantages of IR-based polynya monitoring over PMW based polynya monitoring.

With our now previously stated average grid size of 2x2km our IR data set has an overall higher accuracy and resolution compared to the presented studies based on PMW. Sadly, to date, there is no study available to our knowledge that utilizes the improved capabilities of AMSR2.

We rephrased the mentioned section to read:

“The effect of mixed pixels causes passive-microwave approaches to potentially misinterpret polynya signals. Due to the generally rather coarse resolution, algorithms based on passive-microwave sensor data can only resolve relatively large polynyas while fairly narrow coastal polynyas, like those in the Weddell Sea, may remain partly undetected (Tamura2007).”

L19-24: I am not sure how relevant this paragraph is here, in the satellite section. This might be better in the part where you do the inter-comparison with the model data?

We prefer to keep the section associated to the Table instead of dividing it into satellite vs model. As we finish the POLA comparison in this paragraph we feel it rounds up the POLA comparison to keep the model comparison here as well before starting with the IP comparison.

L25 ff: I agree that the study of Kern, 2009, does not include useful information for your study in terms of a proper inter-comparison. However, as stated above, the paper contains just a subset of results. There is even a daily polynya area (on request even a sub-daily) data set available from

http://icdc.zmaw.de which could be used for inter-comparison purposes for a future study. Maybe - if you intend to keep Kern (2009) in Table 2 - a thing worth mentioning could be that the way Kern (2009) defined POLA is completely different from the way Tamura et al., Nihashi and Ohshima, and presumably you defined POLA. Hence it would perhaps be not too surprising that your results agree quite nicely with these latter two studies while they don't agree with Kern (2009).

Based on the additional information in the final report provided by Stefan Kern we were able to provide a better comparison and changed the paragraph to read:

"Kern2009 derived polynya areas from SSM/I data for an investigation period from April to September for the years 1992--2008. While our results in Table~\ref{tab:pa}} only present four-year averages compared to the much longer averaging period in Kern2009, results for the Brunt Ice Shelf (BR) area are in good agreement. For the Ronne Ice Shelf (RO) area, our results are lower than the estimates by Kern2009. However, Kern2009 summarizes our subregions of RO and FI as well as parts of IB, CL and AP, which makes comparisons in this region very difficult compared to the BR region."

P3973, L11: I suggest to write "average cumulative winter-time ice production" instead of just "average"

We changed that.

L13: I suggest to write: "are about three times larger than our results of the ice production" instead of using "(by almost triple)"

We changed that.

L15-22: I have difficulties to follow your argument here. If I recall correctly, Nihashi and Ohshima are combining 37 GHz and 89 GHz information from AMSR-E and hence their product is based on a combined grid resolution of 12.5 km (37GHz) and 6.25 km (89 GHz). Their method allows in particular a better definition of the POLA as long as TIT is below 10 cm. Above that their method relies on 37 GHz data. So one can expect - following your argumentation above and assuming that usage of finer spatial resolution decreased the fraction of mixed pixels - that POLA based on Nihashi and Ohshima approach agrees better with your approach while POLA derived from the other mentioned groups (Drucker et al., Tamura et al.) is larger. Can I assume - as it is not entirely clear - that the cited Tamura et al. (2011) also used AMSR-E data at 12.5 km grid resolution? In this case the statements make sense - provided that the same data and methodologies are used to compute sea ice thickness. Is this true?

The cited Tamura paper by Nihashi used SSM/I 85-GHz and 37-GHz data on a 12.5 km grid. The point to make here was the apparent decrease of IP estimates with increasing resolution to give an additional explanation for the discrepancy between high-resolution MODIS estimates and coarse resolution (only the 36 GHz channel was used by Drucker Group) PMW data. However, the combined use of 36 GHz and 89GHz seems to result in a better agreement.

L25: Good! What is the accuracy of your method?

Please refer to above changes.

L3974, L4ff: I would say that the additions suggested to the remainder of the paper would justify that the subsection 3.4 could find perhaps place in a future paper. I don't see this connected too much to the remainder of the paper and find it "round" enough without this section. If you decide to keep this then I would encourage you to ask the comments I have for this section.

While we appreciate the suggestion of the reviewer, we feel this section is also an important part of what our approach can deliver and with a recent at hand to compare our results to, we would like to keep this part in the manuscript.

P3976, L5/6: "Especially ..." As you have not done any estimations into this direction I suggest to delete this sentence.

We changed that.

P3980, Table 1: I suggest to not introduce another abbreviation (RAW) but explain that in the table caption and use the following lines in the table: "uncorrected" "only PE" "SFR and PE"

We changed that.

P3981, Table 2: You might want to delete the number derived by Kern, 2009. It seems not appropriate because of the much different area used.

Please refer to the changes made above.

P3984, Figure 2: I would encourage to make this figure bigger. The year of Preußer et al. is 2014 in the figure but 2015 in the text. The figure contains "IST". Is this the IST from the MOD/MYD29 product? It reads here as if your are producing this parameter

We changed that. The reviewers' assumption is correct as it is indeed the MOD29 IST. However, at the indicated stage, the output IST is the IST corresponding to the median TIT value for a daily composite in our reference grid. So yes, this parameter is actually "produced" as indicated in the Figure and Caption.

P3986, Figure 4: I suggest to delete "thickness" in the legend caption. In the print-out version of this figure it is relatively difficult to see color (and hence value) variation above about 15%.

We changed that and altered the color bar to allow for a better interpretation in the new plot (see above).

P3987/P3988, Figure 5 & 6: What are the units of the trends?

Trend units are stated in the corresponding Figure caption. However, we also introduced them now directly in the plot for the readers convenience.

P3989, Figure 7: Caption says that estimates below 250 km^2 are not shown. Is this the "natural" lower boundary of detectable polynya area?

No but simply chosen for readability issues.

Typos / Edits P3967, L24: "leads a decrease" –> "leads to a decrease" P3969, L22: "IP.However" –> "IP. However" P3971, L1: "Due different" –> "Due to the different

We thank the reviewer for pointing these typos out to us. We changed them all.