Dear Dr. Dupont,

Thank you very much for your helpful comments. We did the revision according to your comments. The reply details are given below.

1-page 1, line 17: "freshwater" here sounds awkward. Why not just "precipitation"?
reply: Changed to “precipitation” (p1, l17)

2-page 2, line 14: "[...] the changes are further accelerated by processes of Arctic amplification" does not tell anything. Please elaborate or drop.
reply: Dropped.

3-page 4, line 6: "As the first baroclinic Rossby radius is very small in the Arctic Ocean (Nurser and Bacon, 2014) [...]" Please amend. "very small" is not very telling but I assume the authors mean <5km. Then this statement is only true in the shallow parts of the Arctic and around the GIN seas. It also seemingly contradicts the authors goal to nearly resolve the first Rossby radius in the deep parts of the Arctic where it is about 10km or more with at least 2 points.
reply: the range of Rossby radius is added (p4, l3)

4-page 4, line 17: I am not sure what the author meant by "practically optimal": "Almost optimal" or "practical (useful) and optimal"?
reply: Changed to „In practice, however, optimal ...“ (p4, l14)

5-page 7, line 9: "looses" -> "loses"
reply: changed (p7, l22)

6-It would probably be telling if the authors could map an instantaneous field for high- lighting the model capacity to (nearly?) resolve mesoscale activity where resolved.
Reply: We do not consider this model setup as well eddy-resolving with which one can focus on mesoscale processes.

7-Fig 5: please show exact boundaries for domain averaging
reply: Definition is added in the text (p8 l15): „The two basins are defined as the Arctic region where the ocean bottom is deeper than 500 m, and separated by the Lomonosov Ridge.“

8-page 12, line 4: "Different from" sounds awkward. "Contrary to"?
reply: Changed. (p12, l22)

9-Fig10: Given the success of the CAA run to reproduce the same FW pathways as HIGH, I am curious to understand if the CAA run reproduces HIGH in other aspects: profiles, AW layer, SSH... It may be that the eddy-resolving resolution in the deep ocean is not necessary after all, only a realistic throughflow of the CAA (the eddy parametrization seemingly providing sufficient physics for the rest)!
Reply: We had mentioned this aspect in the second last paragraph of the paper. In the revision we added one sentence, „We also found that only better resolving the CAA channels (in the simulation where only the CAA is resolved with 4.5 km) did not significantly impact the representation of the AW layer.“ This additional sensitivity experiment shows the impact of CAA on FW spatial distribution, but not on AW circulation. Further extended studies, if required, can only be done in separate work considering the current paper length. (p33, l5)

10-Fig.12 FWC anomaly relative to which period?
Reply: Relative to the mean of the plotted period. Added to the figure caption.

11-page 14, line 8: definition of FWC from manuscript: "defined as the amount of pure FW that could be taken out of the upper ocean so that the ocean salinity is changed to 34.8[...]". Just for clarity could you provide the exact depth that defines the upper ocean in your calculation of FWC?
Reply: We clarified it now: „In the calculation of the modelled FW content presented below, the integration is taken from ocean surface to the depth where salinity is equal to the reference salinity“ (p14, l8)

12-page 22, line 17: "2D FW content". why 2D here? FWC is assumed implicitly to be a vertical integral. "maps" maybe?
Reply: changed to „vertically integrated“ (p22, l13)

13-Fig 14, maybe a little outside the scope, but given the pattern of thick ice, I suspect that the ice velocity are too slow. Have the authors compared their sea-ice velocity against buoys or derived-satellite products?
Reply: We checked the simulated sea ice velocity. On the contrary, the simulated sea ice drift is higher than observed from satellites. We consulted colleagues working on observations, and they suggested that the observation products might underestimate the drift was well. Further efforts are required in understanding the behavior of sea ice velocity, both in the model and in the observation.

14-page 26, last paragraph. Can the authors comment on the spurious diffusion on LOW. What are the value of the explicit horizontal diffusion in both simulations? For that matter, it would be nice to have background vertical diffusion value as well...
Reply: We did not quantify the spurious numerical mixing coefficients. However, we performed sensitivity simulations, and found that changing the background vertical diffusivity on the order of 1e-6 m^2/s can significantly change the model results. This indicates that numerical mixing, even small, can still significantly impact model results. Lateral mixing coefficients were mentioned in the model setup section. The importance of vertical mixing coefficients was mentioned in line 10 on page 29. The value of background vertical mixing coefficient is added (line 13 on page 6).

15-page 28, line 28: "obtains" sounds ill-chosen in this context. "displays" instead
reply: Changed to „has“

16-page 30, line 25: "Practically" is again a bit ambiguous. "For practical reasons" maybe?
reply: Changed to „in practice“ (p31, l7)

17-page 30, line 30: "Besides, maintaining high resolution measurements of ocean transports is of great importance for model development too." switch to observation-related subject a bit brutal to the reader. Maybe elaborate a bit?
reply: The sentence is removed.

Best regards
the authors

Dear reviewer,

Thank you very much for your helpful comments. We did the revision according to your suggestions and the detailed reply is listed below.

Specific comments

1. Focus.
As stated, the research is very timely and I am pleased to see this being covered and particularly in a global context. I think the basis of the research idea here is excellent. As far as I can read into the execution and analysis, this too appears excellent. Saying this, I found the write up difficult to follow. It is very long and feels like it wants to act as: (a) a review of physical processes in the Arctic, (b) introduce a multi-resolution global model with a high resolution Arctic, whilst also (c) evaluating the multi-scale model. I understand it is a complex system, and there is some need to discuss expected physics, but as a GMD model evaluation paper, I think it would help the reader significantly if focus is directed at (c). Details of (b) to be included as required for reproducibility (since the model 1.4 is introduced in past papers) and (a) discussed for context of the analysis of model output. Related to (a), some paragraphs go into a detailed description of physical processes in the ocean, only to end with a general suggestion that more work is required. It is good to include a description of the physics, but only if this is relevant to specific analysis of model output, or in direct relation to the basis of this work – with increased resolution. Cleaning up these and the more general statements that appear but seem unnecessary and do not add to the paper – would help focus the paper, reduce verbosity and make it more accessible. Some examples in the introduction:

–

"Numerical modeling can be used to understand the dynamics of the ocean and predict its future changes."

Reply: removed.

–

"Large model biases in the upper Arctic Ocean are another common issue in many ocean general circulation models." which models? and links to why? Is this a resolution issue? These statements need to be specific about models and approaches when this paper is exactly about a different modelling approach that could solve these issues.

Reply: citation is added to this sentence. (p2 l25) The sentences after it explains details.

–

"Model simulation results can be sensitive to model configuration, including the choice of numerical and physical schemes, parameters and grid resolution."

surely model simulation results *are* directly dependent on model configuration?

Reply: sentence removed.

–

"As computational resources grow with time, the modeling community tends to use higher and higher model resolution. Certainly there is a need in the modeling community to evaluate high resolution models with respect to the common model issues identified in previous model studies."

Again this is very verbose, is not specific and does not add anything significant. Which "common model issues" Which "previous model studies"

Reply: We make this sentence more exact: "Certainly there is a need in the modeling community to evaluate high resolution models with respect to the common model issues identified in previous model studies as mentioned above." (p2, l8)

–

"These studies provided background knowledge on the Arctic Ocean representation in those models and identified their common issues." Agaig a weak summary of what has already been said previously. This repetition with no specifics is not helpful.

Reply: "These studies provided background knowledge on the Arctic Ocean representation in those models and identified their common issues." Therefore, "In this paper we will mainly focus on the key diagnostics used in these studies for evaluating our simulations." Because of the experience gained in these studies, we now know what we should focus on in this study. (p4, l25)

General sentences like these should be removed.
Other examples of superfluous content:

– Page 5, line 12:
"It helps to preserve monotonicity and eliminate overshoots. When compared to a second order scheme without flux limiter and an implicit second order scheme in idealized 2-D test cases, at coarse resolution this FCT scheme tends to slightly reduce local maxima even for a smooth field, but at high resolution it well represents sharp fronts and shows least dispersion errors (Wang, 2007)." This is a model detail under (b), but it is not clear how this is relevant to the focus of the paper. Please either elaborate on how this is relevant or remove.
Reply: removed.

– This is again repeated on page 31, line 32, and does not appear necessary to include. Here a link is at least suggested to the work, "which can explain the obtained improvement of the AW layer in the high resolution simulation" but is only loose conjecture. This is something that could be tested here in this modelling framework. Please either test and include or remove.
Reply: It was shown that the AW layer can be better represented (the thickness in particular) with reduced numerical mixing by Holloway et al. (2007), and it was shown that numerical mixing can be reduced with increasing resolution by Wang (2007) in FESOM idealized experiments. It is then logical to link the improved AW layer with 4.5km resolution to reduced numerical mixing in the model. Here we decided to keep it.

– The following paragraph (beginning "The diapycnal mixing is..") could also be potentially removed. These are useful details of the model setup, but standard for FESOM1.4, and all detailed in the model description paper Wang et al. (2014). If these choices are different to others (in CORE-II, for example) and pertinent to the success of the model here, please elaborate (as per the later on SSS restoring), otherwise remove.
Reply: This paragraph provides some of the key information required by modellers. Even though they are the standard settings, we always repeat them when describing the model setups in papers, as readers want to know key parameters directly in the paper they are reading rather than looking for other papers. We certainly did not mention those secondarily important settings in the paper.

–
"Besides identifying the impact of horizontal resolution on the Arctic Ocean main circulation, we also discussed scientific questions and model issues that need to be explored in future work, and some of the illustrated model issues are common in many other ocean-sea ice models. Overall, increasing model resolution does considerably improve the performance of the Arctic Ocean simulation, while further efforts are necessary to solve remaining issues that are not linked to applied model resolution, and to develop/improve parameterizations that are still required even with best resolution affordable now."
This lengthy paragraph adds very little except a very general comment that it has been shown that increased resolution helps. Please remove. In general, please ensure all discussion of physical processes directly links to quantitative analysis of model performance and output. Detailed descriptions of physical process in the Arctic that end with loose general statements that these could be important / need to be considered when modelling, do not add significantly to the paper and its aim to evaluate the 3 configurations. The increased length only makes the paper more inaccessible.

Reply: This paragraph is the last one of the whole paper, and we summarized the paper and provided the outlook. In terms of the outlook, we emphasize that many other aspects should be investigated besides increasing model resolution, which is worth addressing at the end. We are inclined to keep this paragraph.

2. Structure

This paper is part of a concerted effort to develop FESOM, improve estimation in the Arctic and this is the subject of many recent related papers. (The lead author is lead author on 3 cited works in 2016 alone). Given this, the paper would greatly benefit from a focused "positioning" section. Before jumping straight into"Model setup" I suggest a section briefly reviewing the effort to date, how each of the relevant papers fits in and solves/identifies lines of research to arrive at the work here. Some of this can be collected from snippets spread throughout the paper – including for example parts of the introduction, and a large part of "5.3 Unstructured-mesh modeling" (which I found odd positioned at the end of the paper). On the introduction, the beginning 3 paragraphs are good and give a nice motivation for the work. The fourth paragraph jumps straight into the aim of the paper to investigate resolution – which at that point is not motivated – but then jumps back into motivation discussing the narrow straits. The fifth paragraph then jumps back out to talk about the overall modelling framework in general development. I suggest the fourth paragraph discusses the narrow straits, to then motivate and justify why an increase in resolution could help. There is some good justification given in Section 5.3 (e.g. "An adequate representation of the CAA throughflow is found to be very important. With an unstructured-mesh model like FESOM, one can locally increase model resolution to accurately resolve the narrow channels and faithfully simulate the FW export (Wekerle et al., 2013) ..., and more therein). There is also a review of higher resolution modelling efforts there that is appropriate for the motivation in the introduction, and to give context to this approach. This then naturally leads into a short paragraph outlining the aims here – bringing in the 4.5km high resolution, potentially addressing issues prevalent in CORE-II. Please make this concise. For example, the following, all in the latter part of the introduction, largely say the same:
– "In this paper we will evaluate a high resolution Arctic Ocean simulation and elucidate the sensitivity of model results to resolution."
– "As one of the first steps towards designing such a system, in this paper we evaluate the simulated Arctic Ocean by FESOM on a global mesh with 4.5 km resolution inside the Arctic Ocean".
– "We will compare the 4.5 km model results with those from this coarse setup to understand the impace of model resolution." (note this typo – I think should be "impact") Then I suggest the "positioning" section to make it clear what has been achieved with this approach, related studies (e.g. the 4.5km sea-ice modelling study"Werkerle et al. (2013) The Canadian Arctic Archipelago throughflow in a multiresolution global model: Model assessment and the driving mechanism of interannual variability" and how this studies fits / differs / is an important step in this continuing development.

Reply:

We removed "Model simulation results can be sensitive to model configuration, including the choice of numerical and physical schemes, parameters and grid resolution. We will evaluate a high resolution Arctic Ocean simulation and llucidate the sensitivity of model results to horizontal resolution" from the fourth paragraph, and start directly with the motivation of using high model resolution. This indeed makes this part better. (p2, l32)

The previous work (especially the COREII intercomparison papers) is already mentioned in the third paragraph where we describe the common issues in current Arctic models. In purpose, we tried to avoid lengthy description of FESOM's own applications in the introduction, as there is less intention to review FESOM in this work. Readers can better focus on points of "model resolution" and "Arctic Ocean" without being stopped by paragraphs about unstructured-mesh modelling. Any way, we try to have the short section of Section 5.3, for those who might want to know more about

FESOM and unstructured-mesh modelling in general. We think this is better considering our motivation of the paper, the length of the paper and the readability.

3. Figure 1
This is a good figure and helpful to include for reference. It is also useful to locate the vertical section in figure 8. I suggest the “curves” are better identified as “arrows”, since they are orientated markings. In hardcopy, it was difficult to make out the blue arrows and blue text “Transpolar drift” on top of the scalar blue background. I suggest different colours are used to make a clear distinction between the arrows and text vs. the background. Possibly the background scalar depth field could be coloured in green? The arrows outlined in white?
Reply: Figure 1 is revised.

4. Figure projections
Figures 2(a) and (b) are compared – and it is good to discuss resolution of the Rossby radius in this way – but they are presented in different projections, which makes a comparison by eye difficult. Moreover, Figure 1 is used for orientation – again this is helpful – but it is oriented differently to 2(b) and the subsequent 4, 6 7, 10, 11, 14 and 15. Please make the orientation consistent, such that Figure 1 can be used to easily identify regions when later output figures are discussed.
Reply: The projection of Fig. 1 is changed to the same direction (North Atlantic is at the bottom) as in other figures. We try to keep the previous projection of Fig. 2a because it nicely shows model resolution in a much larger area, and this will not cause any difficulty in understanding the paper content to readers.

5. Definition of HIGH and LOW
The introduction refers to Figure 2 on page 4. Figure 4, on page 5, refers to HIGH, but this is not defined until later in the text on page 6. Please fix so the reader knows what HIGH refers to before it is used in text.
Reply: The caption of Fig. 2 is modified as “The horizontal grid size of a mesh with 4.5 km in the Arctic Ocean (referred to as mesh HIGH in this paper)”

6. Eddy resolving considerations
Figure 2 and the connected text make an analysis of where the fixed mesh can resolve mesoscale eddy activity. This quantitative analysis, including the spatial dependence presented in 2(b), is then significantly undermined by the Figure 2 caption comment: “Note that effective model resolution usually is coarser than the grid size due to numerical dissipation.” Please change this into a quantitative argument. As it stands the reader really is at a loose end in knowing where the model under consideration is mesoscale eddy resolving. This point is part of the discussion, and the reason for figure 2(b) – and link to / motivation for the increased resolution – so it is important it is completed. Is there a quantitative study of FESOM1.4 (or the underlying discretisation/numerical dissipation present) that can be used to infer the effective model resolution, and where in fact, in this configuration being evaluated here, the model is mesoscale eddy resolving?
Reply: Two grid cells per Rossby radius is the boundary of resolution where models may start to resolve mesoscale eddies. We write “Judged by comparing the Rossby radius and grid size (Fig. 2b) and inspecting the simulation result, mesh HIGH is not well eddy resolving in the Arctic Ocean, while it permits eddies in the Eurasian and Canadian Basins.” (p6, l6) to be more clear. So far we did not carry out quantitative analysis of effective resolution. By inspecting the model results, we believe that the model permits mesoscale eddies, but does not well resolve them in the Arctic Ocean at 4.5 km resolution. Higher resolution simulations are planned for the future work.

7. Reproducibility and Zenodo archive
The authors have provided a DOI link to a Zenodo archive.
I was able to access the link at https://doi.org/10.5281/zenodo.831484

and download the archive fesom1.4.tar.bz2
After attempting various methods, unfortunately I could not access the contents of the archive.
For example, trying:
tar tvjf fesom1.4.tar.bz2
tar: This does not look like a tar archive
tar: Skipping to next header
tar: Exiting with failure status due to previous errors
Please take a look at the archive. It appears the format is not as expected.
Please provide instructions on how to access the data. Alternatively, and a better solution in my opinion, I suggest uploading files as they are, and not compressed or contained in an archive (Zenodo can handle large single files 50GB) – to avoid problems in opening/uncompressing files.
Reply: Bugs in the uploaded files are corrected, and the new DOI is provided in the code and data availability section.

8. Mesh set-up and reproducibility
A large focus of the paper is on the three mesh configurations, LOW, HIGH and high only in the CAA. Sometimes the labels "24 km" and "4.5km" are used for LOW and HIGH in the figures. It could be helpful to keep this consistently to the defined labels LOW and HIGH. For the third mesh configuration with high resolution only in the CAA, various labels are used. "CAA HIGH" appears in figure 10 but is not defined. The graphic uses "CAA 4.5 km". It would be helpful to stick to one label for all throughout. Please define a clear label for this third case, and ideally together with the definitions of the LOW and HIGH cases. It would also be helpful to make it clear in this same place early on that there are 3 different mesh configurations considered here.
Reply: We now define the 3 setups in the model description section. In terms of naming, we use LOW, HIGH and HIGH-CAA in the text including figure captions. (p6 l8) In the title of figure panels and figure legends, we try to write the resolution explicitly, in purpose, as many readers will only look at figures first; they will get the model resolution quickly without spending time on looking into the text.

The description of how the resolution varies and in particular how it varies between the three is not described sufficiently. This makes it difficult to reproduce or even compare results. Another flexible mesh model may wish to use 4.5km resolution in the Arctic in the same way to compare and contrast. It is not clear how this would be achieved without significant ambiguity over how resolution is varied, bar the broad, general description "On the second mesh (HIGH) the horizontal resolution is further increased to 4.5 km inside the Arctic Ocean (defined by the Arctic gateways of Bering Strait, CAA, Fram Strait and Barents Sea Opening, Fig. 2a)" and the image in figure 2a (which would be difficult to extract this data from and also includes only half the globe). This paper is evaluating the sensitivity of output to mesh resolution. It is important this is characterised fully. In what way does the resolution vary from 4.5km to 24km in mesh HIGH? The 4.5km region is defined as the Arctic Ocean, closed with reference to the gateways. Yet this description appears incomplete, since in figure 2a it appears higher resolution is applied in the Baltic Sea also? On reproducibility, can the bounds of the 4.5km region used in this case be defined by a mathematical function, in terms of longitude and latitude for this purpose? If not, can shapefile definitions be provided? Other multi-resolution models may not use a triangular mesh, and require accurate description of regions for generation. Please provide sufficient information such that the 3 three mesh configurations can be regenerated – in storage that is persistent and can be reliably depended on, ideally with a DOI. The description of the third configuration is even less well-defined: "It has resolution similar to LOW outside CAA, but the same resolution as in HIGH inside CAA." and leaves the reader questioning why is it not the same resolution as LOW outside the CAA? How is the CAA region defined? How does the resolution vary as you move to outside of the CAA? This is an important study that others will want to compare to. It is important the 3 configurations can be regenerated. Also, ideally there will be sufficient information that others would be able to re-run the same FESOM simulations. The authors advocate a model intercomparison using high resolution cases. It seems this would be facilitated with sharing as much as possible of the model configurations and model output. Why not make the 3 mesh configurations available on a persistent resource such as Zenodo? It would also benefit a model intercomparison if key outputs from FESOM1.4 here were also made available in this way – for others to contrast and compare.

Reply: We agree that storing all the 3 meshes and making them available to others is helpful. And actually all the concerns in this comment can be solved if we do it. Now they are put to the same place as the code with the same DOI. In the paper we stick to what we have and the information is enough for readers who do not want to carry out exactly the same configuration. As the generation of meshes is very complicated with much manual work in between, readers need to download the mesh files and get into the details if they are interested in it for some reason. People who are using different models can generate their own meshes following our mesh data.

9. Spin up and simulation time frames

The spin up part of model runs is repeatedly referred to throughout, but it is not clear how long this is and what forcing data is used. Section 4.2 mentions it takes nearly 30 years for salinity to reach a quasi-equilibrium state. Is this enough for the whole model? In the section on freshwater 5.2, 20 to 30 years is mentioned for a specific sensitivity run exploring the effect of SSS restoring. Please clarify. Page 6, line 29 implies the 3 configurations are run from 1950 to 2009 (with some spin up assumed beforehand, relaxing to a set climatology?). This is supported by figures such as 3, 8 and 9 with time axes over this period.

Reply: As mentioned in the model setup section, we start the model simulations from steady ocean, and run them using forcing from 1950 to 2009. The first 20 years are considered as spin up. This is the spinup time for the Arctic basin, but the global ocean needs much longer time to reach equilibrium. We focus on the 60 years period and analyze both the spinup phase and the mean state of the last 20-30 model years. Discussing the spinup phase is very important because the drift taking place in this period, especially the initial surface salinity drift, is the largest (as shown in Wang et al, 2016b). Longer simulations with high resolution will be carried out in future projects.

Figure 4 shows an average from 1970-1999, figure 8 from 1980-1999, figure 6 compares the 70s to 90s, figure 7 is not clear on range considered. Figures 10 and 11 considers different periods 1993-2002, 2003-2007 and 1996-2009, and figure 14 and 15 other ranges. In some cases these choices are justified, but others not. Please make sure it is clear why particular time ranges have been chosen. For example, why in figure 8 miss the contribution from 1970-1979 which is included in figure 4? Indeed why not an average of the entire simulated record over the CORE-II time range?

Reply: The choice of analyzed periods is mainly due to two reasons: the availability of observations and published model results; the events/phenomena that should be addressed. The former is related to Figures 4, 6, 10, 11, 14, the latter involves figures 8e,f and 15. The first 20 years are spinup and should be excluded when computing mean and variability, but are addressed when discussing the model spinup behavior. In the figure captions of Figs. 4, 6, 8e,f, 10, 11, 14 we add the reasons for the periods used. In Fig. 7 we show the standard deviation, for which we do not have observations, so we just take the last 30 years for analysis, leaving out the first 30 years; And the main conclusion about the impact of model resolution based on Fig. 7 does not change if we take 10 years more or less. In Fig. 15 we take passive tracers from year 2000, a year after long release of the tracers; but we did not take results from the very last years because the observed warming of Atlantic Water was not well simulated, which might be due to uncertainty in atmospheric forcing used (addressed on page 28, line 12), a topic for future work.

For a paper on model evaluation and more so when different spatial resolutions (and by CFL, time stepping) are involved, one would expect an analysis of model performance with respect to time. This is touched on with a loose reference to 7 days throughput per day. Please include details of time steps used in all three configurations and simulation time per time step on each mesh. If

possible include a scaling plot. It would also be interesting to compare simulation time per time step for each of the model components (e.g. ocean, sea-ice) on each mesh configuration. This information is very important for comparison studies, to guide choices in the development of other models and for users in model choice – possibly even more so here given FESOM2.0 is being developed with a different dynamical core and discretisation choice. On "The deficiency indicates a clear requirement for eddy resolving resolution in the Fram Strait region in order to faithfully simulate the amount and property of AW that enters the Arctic basins through the Fram Strait. In ocean climate simulations, however, it is hardly possible to afford 1 km model resolution in the near future." (page 28, line 3): it would be helpful to be more quantitative here – how much computational effort do the authors believe a 1km version of the model would require? Analysis in Holt et al. (GMD, 2017) implies it will be possible to run global models more routinely down to coastal scales of 1.5km in the next 10 years.
Reply:
For climate scale applications, we do not expect modellers can afford 1.5 km in the next 10 years. It will only be practical for process studies and short prediction purposes. That is why we emphasize the importance of improving parameterizations, and Holt et al. (2017) did so too.
We modified the paragraph and added a new table (Table 2) to summarize the computational performance. Thanks for this comment; we noticed an error in the performance estimation (the throughput of HIGH should be about 8, not 7) when we were revising this part.
For us, the model throughput is the most important aspect to consider. We cite a recent paper showing the scaling property of FESOM1.4. We limit our discussion to the comparison of the three grids shown in Table 2, and do not intend to describe model details and do not extend the discussion too much to avoid going beyond the main scope of the paper. (pages 30-31)

The following "Accordingly, efforts on parameterizations are required to improve the simulation of AW circulation in Fram Strait." does not add much. Please add clarification on what parameterisations are required or remove.
Reply: changed to "Accordingly, further effort on parameterizing mesoscale eddy effects is required to improve the simulation of AW circulation in the Fram Strait." (p28, l1)

Page 30, line 21:
"However, if the finest grid size is used in narrow straits in FESOM, the model time step and the overall model throughput could be constrained by this grid size." What else is it likely to be constrained by? It seems very likely rather than "could". A breakdown of costs of model components (e.g. ocean, sea-ice, ...), including the cost impact of moving to 4.5km in the Arctic region or just to the CAA – suggested above – would help give a quantitative answer here. As it stands the reader has no feeling for what the next step might be. What is the computational impact of HIGH vs HIGH CAA? – an important aspect of evaluating these model configurations.
Reply: The sentence is changed to "However, if the finest grid size is just used in narrow straits, the model time step and the overall model throughput can be constrained by this grid size (the Courant–Friedrichs–Lewy (CFL) constraint)." In terms of computational performance, see our reply above. (p30,l20)

10. Atlantic Water core temperature prediction
Examining Figure 4(a), it appears the LOW model over-estimates the Atlantic Water core temperature (AWCT) compared to PHC climatology. Moving to HIGH this over-estimation increases significantly. Notably warmer waters (of over 1 degree above the PHC climatology) appear to propagate through the Fram Strait into the Arctic Ocean region. In this regard, AWCT temperature *distribution* appears to worsen in the move to using high resolution. Please comment.
Reply: This is the feature when the model has higher but not eddy-resolving resolution, shown by Fieg et al. 2010. We have discussed this in the first paragraph of 5.1.b.

How do instantaneous distributions appear? A significant error is seen in HIGH's LFW transport anomaly across the Fram Strait between 1950-1970 (not apparent in LOW) – is this linked? (although Figure 4 contains model results averaged 1970-1999).
Reply: The first 20 years are the spinup phase, when the two setups behave very differently. The variability afterwards becomes more similar. So we do not expect a close link between the two phenomena.

On discussing heat content, with both LOW and HIGH integrated content higher, further analysis is dismissed with the following:"Due to inaccuracy in diagnosing heat budget terms (e.g., caused by interpolation) and ignoring heat diffusion in model output, the mismatch between the ocean heat content changing rate and Arctic net heat flux can have the same order of magnitude as this value." (page 26). It is not clear what is meant here. Please reword and improve the explanation. It does not seem satisfactory that the increased heat content cannot be explained. "Our model results show that the magnitude of the AW temperature is not lower in LOW" (page 26) – is this shown in Figure 5? Is this the maximum or depth integrated magnitude of AW temperature?
Reply: The first few sentences of this paragraph are removed to avoid misleading. In our practice we found that we cannot close the heat budget for the Arctic ocean domain: summing up all heat fluxes, we get the net heat flux into the Arctic Ocean; there is misfit between this net heat flux and the total heat content change in the Arctic Ocean. The model conserves tracers, so we believe the misfit is mainly due to interpolation errors when we calculate heat transports during offline analysis of the model results, although neglecting diffusion of temperature across gateways in the analysis can also contribute to the misfit. The misfit is small, compared to individual components of heat source/sink fluxes (ocean heat fluxes through different gateways and different components of atmospheric heat fluxes), but it is as large as the heat flux source required to explain the difference of heat content between the two simulations. Therefore, we analyze passive tracers to help us to understand the difference of AW water masses. The paragraph is modified to make it clearer.

11. Over-estimation of liquid freshwater content
Both LOW and HIGH over-estimate the liquid freshwater (FW) content. This is noted on page 18, line 11. What are the reasons for this over-estimation? Is it expected this can be further reduced by further increasing resolution?
Reply: We do not expect that the reason is model resolution. Discussion on this was given in section 5.2.a.

"The variety of FW content distributions simulated in different ocean models shown by Wang et al. (2016b) presumably can be partly attributed to different model representations of the CAA region." – this is could be true, but this does not add significantly. Can it be backed up by other studies? What are the different representations? How do they relate to the HIGH CAA configuration here? Do other representations – that are most closely related to the one here – show similar changes/improvements/indications? Some of this is discussed much later on in 5.3 and 6. It appears a key consideration here given focus on resolution and representation and might be good to have its own section?
Reply: We used the HIGH-CAA experiment to show that the representation of CAA region can influence the FW content spatial pattern (in section 4.2). It is based on this finding that we speculate that the variety of FW content spatial pattern shown by Wang et al. 2016b can be partly attributed to the CAA representation in coarse models (in section 4.3). In that study, it was shown that climate ocean models have very different width and resolution in the channels, which can certainly produce very uncertain ocean transports. The finding that better resolving CAA transport can impact also the Arctic basin FW distribution and the fact that currently CAA is not resolved in ocean climate models are the major information that we want to address. Each model development group has the task: Inspect the ocean transports through main ocean gateways.

12. Passive tracer implementation

Figure 11 includes the comments “Note that the passive tracers were set to zero south of Fram and Davis Straits. The passive tracers are averaged over the upper 100 m.” These are details better included inline in the text. This raises additional questions. For the tracer
φ, the above implies:
$\frac{\partial}{\partial t}\int_V \varphi \, dV = 0$ (1)
So tracer total is not conserved? Please elaborate on why this is done and its implications.
What is the reason for averaging the tracer over the top 100m? Is this type of analysis done elsewhere, under the same conditions?
Reply: We did not manually change passive tracers after they enter the Arctic Ocean in the model simulations. We just did not plot them in the figure for the paper. The caption is changed to avoid misleading. As we are discussing the surface water (the FW), we show the passive tracers in the upper ocean.

13. FW variability
Referring to these sentences in section 4.4:
–
“Compared to the period of 1980-2000, the mean Arctic FW content averaged over the 2000s has increased by about 4500 km3 based on observations (Polyakov et al., 2013b; Haine et al., 2015), while the increase is only about 1700 km3 in our two simulations (Table 1).”
–
“On average the 13 CORE-II models analyzed in Wang et al. (2016b) underestimated the observed upward trend also by half.” Observations show mean Arctic FW content has increased nearly 3 times that seen in the model simulations presented. Why is this? The paper highlights that on average CORE-II models underestimate by a half. These models have a significantly lower resolution than the HIGH case considered here, yet the above implies this higher-resolution simulation performs more poorly than the average. Why is this?
Reply: the first two paragraphs discussed two “different diagnostics”. The first is about the “difference of FW content” between two chosen periods which have been considered by Haine et al., 2015. This diagnostic was not analyzed in Wang et al. 2016b. The second diagnostic is the “linear trend of FW content” computed from 1996 to 2009, which we compare to the estimate based on observations (Haine et al, 2015 and Polyakov et al. 2013b) and the analysis of Wang et al. 2016b. The linear trend is slightly higher and closer to the observational estimate in HIGH than in LOW as represented in the second paragraph. Anyway, the difference between HIGH and LOW in terms of the trend of FW content is not very significant. This implies that other model parameters may play large roles for this diagnostic, and we tried to address this point in general including the very last paragraph of the paper (see also the reply to another comment about the very last paragraph above).

The paragraph on temporal variation (page 22, line 7) ends with “Simulation HIGH consistently obtains positive changes in the Eurasian Basin, but with a larger magnitude. It has negative values north of Greenland, which is not present in the observation.”
which appear unexplained. Can you suggest why negative changes are seen north of Greenland? and the positive change in the Eurasian Basin?
Reply: The FW content in the Eurasian Basin increases between the two periods in the model, consistent to the observation. The background of FW increase in the Arctic Ocean was given in the introduction subsection (4.1). For the change of FW content in the small region north of Greenland, the model shows local decrease. Too many reasons can cause the difference of the model from the observations, including the pathway of liquid FW (of different sources) and the local sea ice condition, and uncertainties due to very sparse observations as well. But such speculation is too general and will not give more information. Based on the fact that the model largely captures the spatial pattern of FW increase, and we just pointed out the place of the clear difference from observations without much general speculation. We add “Further efforts are required to understand the reason.” (p22, l9)

On “The interannual variability of FW transport through the Arctic gateways shows large similarity between the two simulations (Fig. 13a-c).” It does not seem correct to characterise this as a large similarity. There is some agreement in the anomaly trend, but also large deviations. Fig. 13(a) shows HIGH has a very significant deviation in the first two decades 1950-1970 of the six decades considered. The magnitude of the deviation is of the order of the max change seen, so arguably significant. Why the deviation? Is it linked to the excessively warmer waters see entering through the Fram Strait in the model output?
Reply: As mentioned in one of the above comments, the first 20 years are spinup phase, when the FW transports behave differently in the two simulations especially for the Fram Strait. Afterwards, the correlation is high (Table 1). The period 1980-2009 is considered in the Table. To be clear, in the text we change to “The interannual variability of FW transport through the Arctic gateways shows large similarity between the two simulations *after the spin-up phase* (Fig. 13a-c). The correlation coefficients between the FW transports from the two simulations are similar at the Davis and Fram Straits (0.75 and 0.78, respectively *for the period of 1980--2009*, Table 1). (p23, l1)

14. Outcomes
“Instead, we often modify the geometry of the CAA channels to allow adequate CAA throughflow. Such model adjustment, however, is not trivial as shown by the large model spread in CAA FW transports among the ocean climate models analyzed in Wang et al. (2016b).” (Page 30, line 27). This can be considered yet another parameterisation and the same approaches made to analyse its impact. “When developing global climate models, the modeling groups certainly need more efforts to better adjust the CAA representation. Besides, maintaining high resolution measurements of ocean transports is of great importance for model development too.” (Page 30, line 29). This does not add anything significant. Can the authors suggest a solution to CAA representation/parameterisation?
Reply: CAA representation means how to adjust the geometry and grids in the CAA channels. In the work of Wang et al. 2016b, it was found that model development groups often treat the narrow straits without examining the results, and people often forget who did the mesh and how it was done. Therefore, our purpose here is to repeat the importance of adequate treatment of the straits and warn model developers to take care. We cannot suggest solutions beyond this, which can be model dependent.

“Most of the models analyzed in past CORE-II model intercomparison studies have relatively coarse resolution. For developing our unstructured-mesh model system with regional focus, it would be helpful to communicate experience with the large structured-mesh model community, for example, in future high resolution ocean climate model intercomparison projects.” (Page 31, line 1). This could be better worded. Do you mean “Most of the test cases/configurations in past CORE-II...”? – and you are suggesting a suite of higher resolution intercomparison test cases?
Rely: we modify the last sentence to “ .... (for example, through the future CMIP projects where increasing model resolution will be pursued (Haarsma et al. 2016)).” (p31 ,l 11)

15. Other small points
(a) Last line of the abstract sounds disjoint, given the paper concerns increased resolution. Maybe something like: “Along with increased resolution, we additionally discuss other issues that could benefit from development to help increase accuracy in the region, including the improvement of parameterizations, for example.”
Reply: changed to “... appear not to be very sensitive to the increase in resolution employed here. By highlighting the issues that are independent on model resolution, we address that other efforts including the improvement of parameterizations are still required.” (the last sentence of the abstract)

(b) Lars Smedsrud et al. have recently published (January 2017 – you cite as “Smedsrud et al., 2017”) on Fram Strait sea ice export – can this help update the 11+ year old figure and error in Table 1? “Fram Strait sea ice export variability and September Arctic sea ice extent over the last 80 years” The Cryosphere, doi:10.5194/tc-11-65-2017. There appears to be some large deviations in Fram Strait fluxes here, comparing FESOM model output and observations.
Reply: Uncertainties in observations come mainly from ice thickness estimate. The Smedsrud paper does not intend to solve this issue, but rather focuses on “*area* flux”.

(c) Page 24, line 21: Part of a general request for more quantitative statements: “Note that much higher model resolution is required in order to simulate sea ice leads with realistic width.” What resolution is required? Is it therefore expected that increase in sea-ice model resolution which see no advantage until this is reached?
Reply: changed to “... because they are typically narrower than 1 km in reality (Tschudi et al., 2002).” At least when such width can be properly resolved, the related physical processes can be directly simulated. Otherwise parameterizations are required. (p24,l21)

Technical corrections

1. Use of the definite article “the” appears to have been skipped in multiple places throughout the paper. e.g. Page 10, line 6: “The maximum temperature in Eurasian Basin in simulation” , Page 11, line 8: “correct circulation direction in Canadian Basin”.
Reply: we corrected these place and tried to proofread the text.

2. Page 1, line 14:
“including improving parameterizations” , better: “including the improvement of parameterizations”.
Reply: changed. (p1, l14)

3. Page 1, line 22: “deep water formation regions”
better worded as “regions where deep water is formed”?
Reply: changed (p1, l22)

4. Page 2, line 16:
“lower latitudes ocean and climate”, change to “lower latitude ocean and climate”.
Reply: changed (p2, l16)

5. Page 2, line 16:
“societal” is better here than “social”.
Reply: changed (p2, l16)

6. Page 2, line 16:
“that remains under debate” better.
Reply: changed (p2, l16)

7. Please check capitalisation in references, e.g.
– Wekerle et al. (2013) should contain “Canadian Arctic Archipelago”.
– Schauer et al (2002) “Amundsen” and “Makarov”.
– “Arctic” , “Fram” and “Barents” in Maslowski et al. (2004).
– …
Reply: changed

8. Page 25, line 3:
“(e.g., Smedsrud” comma not needed – please check throughout.

Reply: this depends on the journal requirement. We leave this for the production editor to decide and change.

9. Figure 14(f) is not labelled “(f)”. Also, the coastline is not identified like the others (a)–(e).
Reply: changed

10. Figure 2 caption:
“Values larger than one indicate mesoscale eddy resolving.” the meaning is understood, but please make into a sentence.
Reply: changed to “With resolution finer than two grid cells per Rossby radius models may start to resolve mesoscale eddies....”

11. Page 5, line 3:
“It work with” , change to “It works with”.
Reply: changed (p4, l42)

12. Page 5, line 15:
“shows *the* least”.
Reply: these sentences are removed now.

13. Page 6, line 21:
“one over the ocean column”
– is this one over the ocean depth? water column thickness?
Reply: changed to “in the whole ocean column”. (p6 ,l1)

14. Pages 4 and 6: Please avoid repetition of the
“Time frame of 7 model years per day”.
Reply: the second-time phrase is removed (page 6)

15. Min and max values are missing in many colour bars – e.g. Figures 2b, 3a, 3c, 7a, 7b, 8a-e, 10, ... – This is very helpful for future studies comparing to this work. Please ensure all min and maxes are labelled.
Reply: Revised.

16. Some axis do not have end values – e.g. time axes in Figure 9. Is this data until 2009? Please make it clear where the axis ends (mark with year, or make it a decade to 2010). Same with Fig 3, 12, 13, …
Reply: the shown integration period is added in the figure captions.

17. Use of the definite article “the” appears inconsistent – e.g.
“towards the Fram Strait” on page 7 (twice), but “toward Fram Strait” on page 8.
Reply: missing “the” is added.

18. Spelling of “elucidate” page 2, line 34.
Reply: sentence removed.

19.
“Dupont, F., and others (2015)”
– please elucidate on the “others”.
Reply: corrected

20.

“CAA” unnecessarily redefined in figure 10.
Reply: For readers who only read figures, we prefer to explain abbreviation in the figure captions.

21. Possibly move the reference to Holloway et al. (2007) on page 11 up to the introduction of topostrophy – so readers know where to find its definition.
Reply: the reference is added after “topostrophy”. (p11, l1)

22. Page 11, line 10:
“Indeed, the Arctic Ocean hydrography obtained on mesh LOW was found to be one of the well simulated when comparing the state-of-the-art ocean climate models (Ilicak et al., 2016)”
“one of the well simulated” does not read well and the sentence is verbose. Last part better rephrased as “when compared to the suite of state-of-the-art climate models in Ilicak et al. (2016)”?
Reply: changed to “Indeed, the Arctic Ocean hydrography obtained on mesh LOW is well simulated when compared to the suite of state-of-the-art ocean climate models analyzed in Ilicak et al. (2016).” (p11,l5)

23. Page 12, line 10:
“between the two models” or “between models”.
Reply: changed

24. Page 18. line 2: Please refer directly to the section number rather than this loose reference.
Reply: “Section 4.2” is added. (p20, l6)

25. Page 24, line 11:
“in *the* 2000s”.
Reply: added (p24, l6)

26. Page 30, line 4:
“without necessity” , please reword – “without *the* necessity”?
Reply: added (p30, l3)

27. Page 33, line 1: Space in URL.
Reply: space removed (p33, l16)

Best regards
the authors

A 4.5 km resolution Arctic Ocean simulation with the global multi-resolution model FESOM 1.4

Qiang Wang[1], Claudia Wekerle[1], Sergey Danilov[1,2], Xuezhu Wang[3,1], and Thomas Jung[1,4]

[1]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research (AWI), Bremerhaven, Germany
[2]Jacobs University Bremen, Bremen, Germany
[3]Hohai University, Nanjing, China
[4]University of Bremen, Bremen, Germany

Correspondence to: Qiang Wang (Qiang.Wang@awi.de)

Abstract. In the framework of developing a global modeling system which can facilitate modeling studies on Arctic Ocean and high-mid latitude linkage, we evaluate the Arctic Ocean simulated by the multi-resolution ocean sea-ice model FESOM. To explore the value of using high horizontal resolution for Arctic Ocean modeling, we use two global meshes differing in the horizontal resolution only in the Arctic Ocean (24 km vs. 4.5 km). The high resolution significantly improves the model's representation of the Arctic Ocean. The most pronounced improvement is in the Arctic intermediate layer, in terms of both Atlantic Water (AW) mean state and variability. The deepening and thickening bias of the AW layer, a common issue found in coarse resolution simulations, is significantly alleviated by using higher resolution. The topographic steering of the AW is stronger and the seasonal and interannual temperature variability along the ocean bottom topography is enhanced in the high resolution simulation. The high resolution also improves the ocean surface circulation, mainly through a better representation of the narrow straits in the Canadian Arctic Archipelago (CAA). The representation of CAA throughflow not only influences the release of water masses through the other gateways, but also the circulation pathways inside the Arctic Ocean. However, the mean state and variability of Arctic freshwater content and the variability of freshwater transport through the Arctic gateways appear not to be very sensitive to the increase in resolution employed here. By highlighting the issues that are independent on model resolution, we address that other efforts including the improvement of parameterizations are still required.

1 Introduction

The Arctic Ocean is the smallest among the world oceans, but it is a very important component of the global climate system due to its geographical location. The atmosphere transports moisture to northern high latitudes and supplies precipitation to the land and ocean. By receiving freshwater through river discharge and direct precipitation the Arctic Ocean is thus a large freshwater reservoir (*Serreze et al.*, 2006; *Dickson et al.*, 2007; *Rudels*, 2015; *Carmack et al.*, 2016). The inflow through Bering Strait is also considered as an Arctic freshwater source because the salinity of Pacific Water is lower than the mean Arctic salinity (*Roach et al.*, 1995; *Woodgate and Aagaard*, 2005). The Arctic Ocean feeds the North Atlantic with its excess freshwater through Fram and Davis Straits (Fig. 1). The released freshwater passes by the regions where deep water is formed, which could have significant impacts on the large scale ocean circulation (*Aagaard et al.*, 1985; *Goosse et al.*, 1997; *Hakkinen*, 1999; *Holland et al.*, 2001; *Wadley and Bigg*, 2002; *Jungclaus et al.*, 2005; *Arzel et al.*, 2008; *Jahn and Holland*, 2013). The liquid freshwater stored in the upper Arctic Ocean results in a strong stratification and helps to form a permanent halocline. This limits the upward heat flux from the underlying warm water and allows for a persistence of sea ice cover (*Rudels et al.*, 1996). The latter plays a crucial role for the climate by constraining air-sea heat, momentum and constituents exchange. The Arctic Ocean is also fed by warm and saline Atlantic Water, which circulates mainly cyclonically under the cold halocline and provides a possible heat source of Arctic sea ice basal melting (*Polyakov et al.*, 2010, 2013a). The intermediate Arctic waters leave the Arctic basins through Fram Strait, the only deep Arctic gateway, supplying part of the dense waters that feed the Atlantic overturning circulation (*Rudels and Friedrich*, 2000; *Karcher et al.*, 2011).

The Arctic air temperature increased more strongly than the global mean temperature in the recent decades (the "Arctic Amplification", e.g., *Serreze and Barry*, 2011). In the meantime the Arctic Ocean is undergoing unprecedented changes, with a freshening trend in the surface layer (*Proshutinsky et al.*, 2009; *McPhee et al.*, 2009; *Rabe et al.*, 2011; *Polyakov et al.*, 2013b; *Haine et al.*, 2015), warming events (*Dmitrenko et al.*, 2008; *Polyakov et al.*, 2012, 2013b), and significant sea ice decline (*Kwok et al.*, 2009; *Comiso*, 2012; *Cavalieri and Parkinson*, 2012; *Stroeve et al.*, 2012; *Laxon et al.*, 2013). These changes are accompanied not only by a shift in ocean circulation regimes and physical conditions, but also by substantial changes in biogeochemical processes (*Arrigo and van Dijken*, 2015; *Tremblay et al.*, 2015). The on-going and future Arctic changes could have a large influence on lower latitude ocean and climate with potential societal impact, although it is a subject that remains under debate (e.g., *Vihma*, 2014; *Wallace et al.*, 2014).

Despite a lot of success in modeling studies of the Arctic Ocean, the state-of-the-art ocean general circulation models still show non-negligible model biases, as illustrated by different model intercomparison studies (*Jahn et al.*, 2012; *Johnson et al.*, 2012; *Aksenov et al.*, 2016; *Wang et al.*, 2016a, b; *Ilicak et al.*, 2016). For example, in the earlier Arctic Ocean Intercomparison Project (AOMIP, *Proshutinsky and Kowalik*, 2007; *Proshutinsky et al.*, 2011), it was identified that too thick Atlantic Water layers in the Arctic Ocean were simulated in the models, very possibly due to spurious numerical mixing (*Holloway et al.*, 2007; *Karcher et al.*, 2007). In the recent Coordinated Ocean-ice Reference Experiments, phase II (CORE-II, *Griffies et al.*, 2012) project, it was found that this issue still remains one decade later (*Ilicak et al.*, 2016). Large model biases in the upper Arctic Ocean are another common issue in many ocean general circulation models as shown by *Wang et al.* (2016a, b). They found that the mean state of the liquid freshwater and sea ice simulated in the CORE-II models, including their storage and the Arctic-Subarctic fluxes, shows a very pronounced spread among models, although the temporal variability is more consistently represented. They also showed that all the CORE-II models experienced a dramatic increase in their Arctic liquid freshwater content during the first few decades of model simulations, and the spread of simulated liquid freshwater transport through Fram and Davis Straits amounts to as much as 50% of the mean transport values. The large model uncertainty identified in previous studies calls for further model development efforts in the community.

There are many narrow straits in the world's oceans, which play important roles in connecting different ocean basins but are difficult to resolve with resolution typically used in large scale ocean models. The Arctic Ocean is enclosed by continents and connected to lower latitude oceans via narrow straits. Especially, the three main Canadian Arctic Archipelago (CAA) channels have widths of about 10 km, 30 km and 50 km at their narrowest locations (*Melling*, 2000). It was shown that the

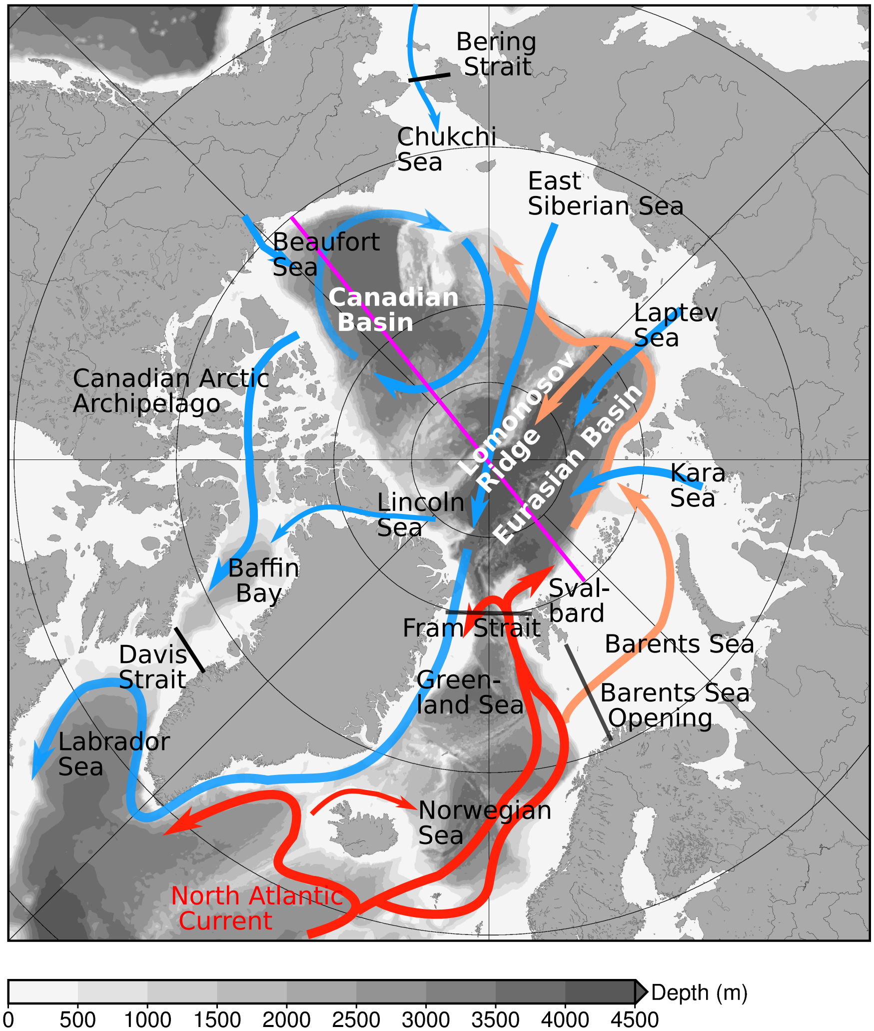

Figure 1. Schematic of main ocean circulations in the pan-Arctic Ocean. The freshwater circulation is shown with light blue arrows, and the Atlantic Water (AW) circulation is shown with red/orange arrows. The gray patch in the background shows the ocean bottom bathymetry. The black lines indicate the four Arctic gateways. The magenta line crossing the North Pole indicates the location of the transect shown in Fig. 8e,f.

ocean fluxes through these narrow channels can be reasonably resolved when using very high horizontal resolution in model simulations (about 4 km, *Wekerle et al.*, 2013). High resolution is also required to resolve small scale dynamics which could have impact on larger scale circulation and water mass properties. As the first baroclinic Rossby radius is small (about 1 to 15 km) in the Arctic Ocean (*Nurser and Bacon*, 2014), mesoscale-eddy resolving is difficult to achieve for long simulations even in regional Arctic Ocean models. For process studies, however, simulations with 1–2 km resolution have been used to resolve mesoscale dynamics and ocean circulation in Fram Strait (*Kawasaki and Hasumi*, 2015; *Hattermann et al.*, 2016; *Wekerle et al.*, 2017b). As computational resources grow with time, the modeling community tends to use higher and higher model resolution. Certainly there is a need in the modeling community to evaluate high resolution models with respect to the common model issues identified in previous model studies as mentioned above.

In the framework of our own model development, we aim to develop a coupled model system that can facilitate to carry out climate research with focus on the Arctic Ocean and Arctic lower-latitude linkage. We use the global Finite Element Sea ice Ocean Model (FESOM, *Wang et al.*, 2014) as its ocean/sea ice component. This model employs unstructured meshes and allows for variable resolution without traditional nesting. With it we are able to allocate finer resolution in the northern high latitudes than in many other parts of the global ocean. In practice, however, optimal ocean resolution in the Arctic Ocean needs to be decided. The finally chosen resolution should help to adequately simulate the key ocean dynamics with confined model biases. At the same time the model system should not be too costly as it will be used in long climate simulations. As one of the first steps towards designing such a system, in this paper we evaluate the simulated Arctic Ocean by FESOM on a global mesh with 4.5 km resolution inside the Arctic Ocean (see Fig. 2). This resolution is higher than typical resolutions used in current climate models (one fourth to one degree) while we still obtain a reasonably high model throughput of about 8 model years per day.

For our purpose we carried out ocean simulations driven by prescribed CORE-II atmospheric forcing. A coarse resolution setup of FESOM (with 24 km in the Arctic Ocean) has been used in previous CORE-II studies. We will compare the 4.5 km model results with those from this coarse setup to understand the impact of model resolution. The most climate-relevant metrics of the Arctic Ocean, that is, the Atlantic Water property, the Arctic freshwater budget and sea ice state were used to evaluate the state-of-the-art ocean climate models in the CORE-II Arctic studies (*Ilicak et al.*, 2016; *Wang et al.*, 2016a, b). These studies provided background knowledge on the Arctic Ocean representation in those models and identified their common issues. In this paper we will mainly focus on the key diagnostics used in these studies for evaluating our simulations.

The model setups will be described in Section 2. The results about Atlantic Water and freshwater budget of the Arctic Ocean will be presented in Sections 3 and 4, respectively, followed by discussions (Section 5) and summary (Section 6).

2 Model setup

The latest version of FESOM (*Wang et al.*, 2014) is used in this study. The ocean dynamical core stems from the early study of *Danilov et al.* (2004) and *Wang et al.* (2008). It works with unstructured triangular meshes, so variable grid resolution can be conveniently applied without the necessity of using traditional nesting. It is coupled to a dynamic-thermodynamic sea ice model

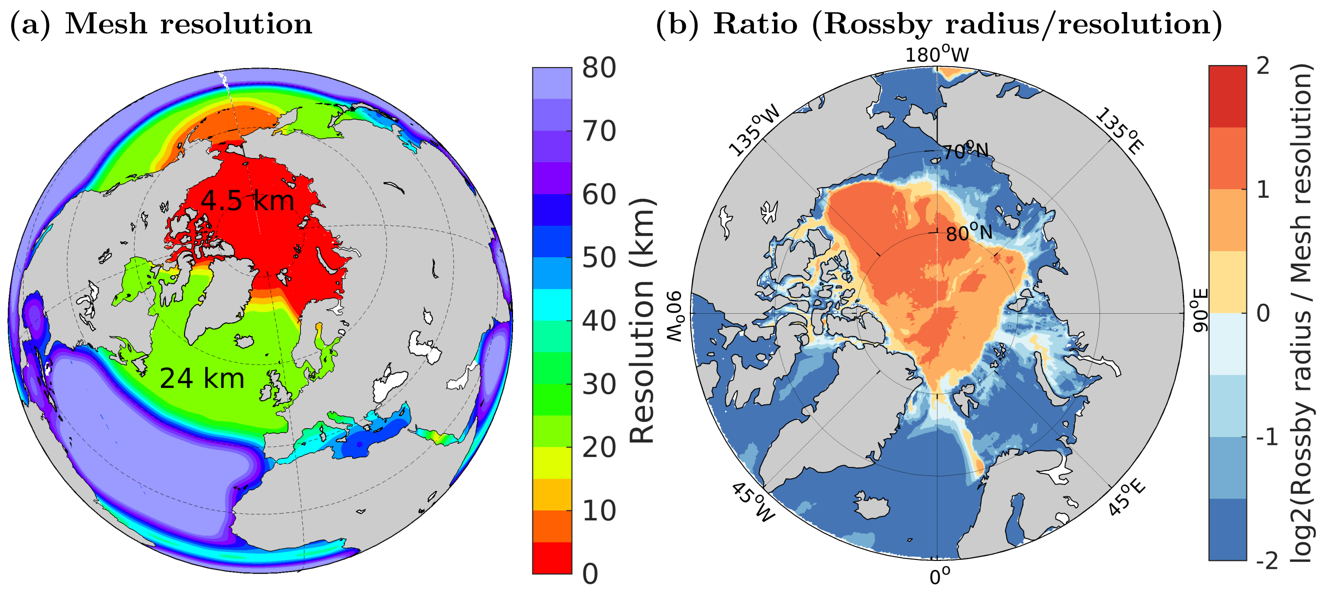

Figure 2. (a) The horizontal grid size of a mesh with 4.5 km in the Arctic Ocean (referred to as mesh HIGH in this paper). (b) Ratio between the first baroclinic Rossby radius and grid size shown with the log2 scale for mesh HIGH. With resolution finer than two grid cells per Rossby radius models may start to resolve mesoscale eddies depending on numerical mixing in the model. The Rossby radius is calculated for each season and the local minimum is used for (b).

(*Timmermann et al.*, 2009; *Danilov et al.*, 2015), which is based on the *Parkinson and Washington* (1979) thermodynamics and uses an updated version of the elastic-viscous-plastic (EVP, *Hunke and Dukowicz*, 1997) rheology. The sea ice model is discretized on the same surface mesh as the ocean model by using an unstructured-mesh method too.

A blend of two bottom topography data sets is used. North of 69°N the 2 km resolution version of the International Bathymetric Chart of the Arctic Oceans (IBCAO, *Jakobsson et al.*, 2008) is used, while south of 64°N the 1 min resolution version of the General Bathymetric Chart of the Oceans (GEBCO) is used. Between 64°N and 69°N the topography is taken as a linear combination of the two data sets. An explicit second-order flux-corrected-transport (FCT) scheme (*Löhner et al.*, 1987) is employed in the tracer equations. It helps to preserve monotonicity and eliminate overshoots.

The diapycnal mixing is parameterized with the k-profile scheme proposed by *Large et al.* (1994). In case of static instability the vertical mixing coefficients are increased as a parametrization for unresolved vertical overturning processes. We apply biharmonic friction with a *Smagorinsky* (1963) viscosity, which is flow-dependent. The *Redi* (1982) isoneutral diffusion with small slope approximation and the *Gent and McWilliams* (1990, GM) parameterization in a skew diffusion form (*Griffies*, 1998) are used. A reference value is determined for neutral diffusivity and GM thickness diffusivity at each surface grid location by considering the local horizontal resolution. It is then scaled by the squared buoyancy frequency to obtain 3D diffusivity (*Wang et al.*, 2014).

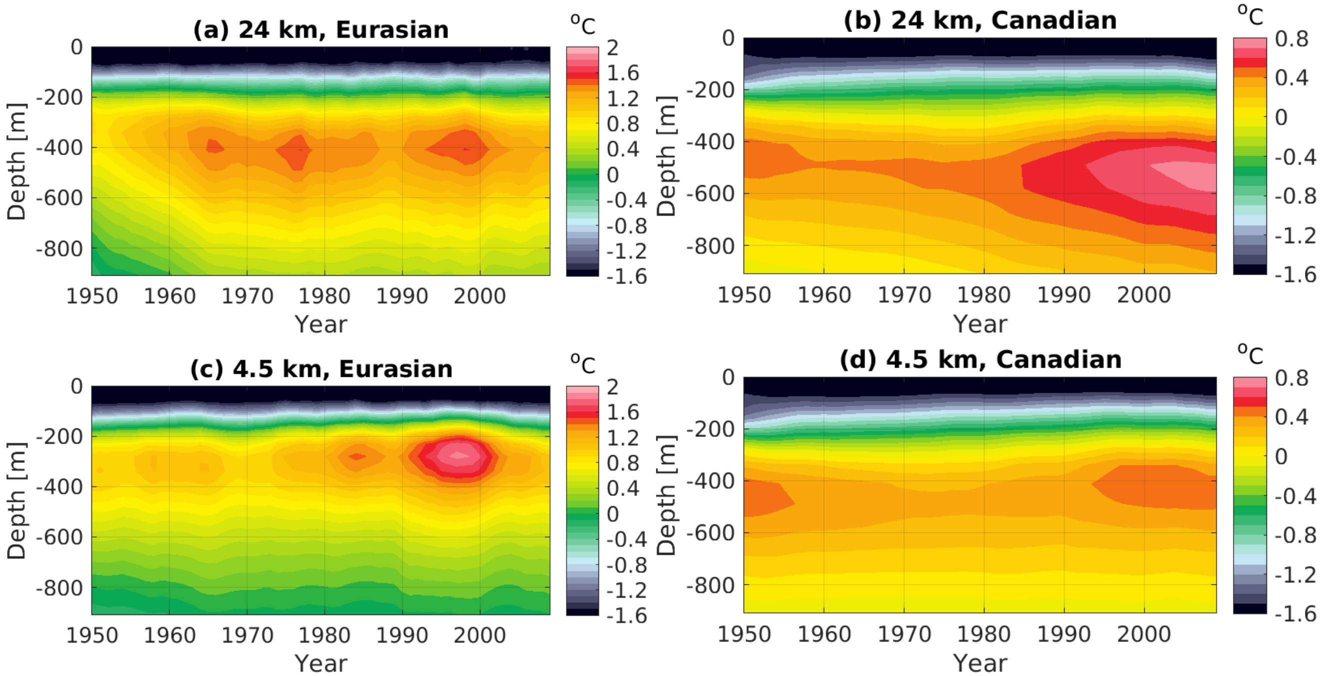

Figure 3. Hovmöller diagram of mean potential temperature for the (a) Eurasian Basin and (b) Canadian Basin obtained in the simulation LOW. (c),(d) are the same as (a),(b) but for simulation HIGH. The whole integration period 1950–2009 is shown.

Two global meshes are compared in this study. The first one (LOW) has 1 degree nominal horizontal resolution in most part of the world's ocean. In the equatorial band the resolution is doubled, and north of 45^{o}N the resolution is set to about 24 km. On the second mesh (HIGH) the horizontal resolution is further increased to 4.5 km inside the Arctic Ocean (defined by the Arctic gateways of Bering Strait, CAA, Fram Strait and Barents Sea Opening, Fig. 2a). Mesh LOW has been used in the CORE-II model intercomparison studies, and mesh HIGH has been used in a recent study on Arctic sea ice leads (*Wang et al.*, 2016c). Judged by comparing the Rossby radius and grid size (Fig. 2b) and inspecting the simulation result, mesh HIGH is not well eddy resolving in the Arctic Ocean, while it permits eddies in the Eurasian and Canadian Basins. To explain the impact of model resolution on the spatial distribution of FW in the Arctic basins, we carried out one sensitivity experiment on an additional mesh (HIGH-CAA). It has 24 km resolution in the Arctic Ocean, except inside the CAA straits where the resolution is increased to 4.5 km. This mesh has been used in the CAA throughflow study by *Wekerle et al.* (2013).

The reference value of the neutral and GM thickness diffusivity is $50\,\mathrm{m}^2/\mathrm{s}$ at 24 km resolution, and $4\,\mathrm{m}^2/\mathrm{s}$ at 4.5 km. In the vertical 47 z-levels are used with resolution of 10 m in the top 100 m and gradually decreased downwards. The background vertical diffusivity is set to $10^{-5}\,\mathrm{m}^2/\mathrm{s}$. Three passive tracers are used to illustrate the pathways of different inflow water masses from Fram Strait, Barents Sea Opening (BSO) and Bering Strait. Initially the concentration of these tracers is set to zero. During the simulations they are restored to one in the whole ocean column within these straits. Adding these passive tracers reduces the model throughput by about 20%.

As discussed by *Griffies et al.* (2009), ocean climate models without a coupled active atmospheric model lack many of the feedbacks present in a fully coupled system, which necessitates restoring of model sea surface salinity (SSS) to observed climatological SSS in global ocean models. In addition, SSS restoring helps to avoid unbounded local salinity trends that can occur in response to inaccuracies in, for example, precipitation. The strength of SSS restoring (defined by a piston velocity) in our simulations is $50\,\mathrm{m}$ over $300\,\mathrm{days}$, a value used in many CORE-II models (*Danabasoglu et al.*, 2014). The impact of SSS restoring will be discussed in the discussion section.

The model is forced by the CORE-II interannual atmospheric data set (*Large and Yeager*, 2009) from 1950 to 2009. The ocean is initialized with temperature and salinity fields from the Polar Science Center Hydrographic Climatology v.3 (PHC3, *Steele et al.*, 2001) and starts from a steady state, and sea ice is initialized with climatological fields obtained from a previous simulation. Interannual monthly mean river runoff is taken from the data provided by *Dai et al.* (2009), and in the model the river water is spread over a range of 300 km near river mouths to count for unresolved processes (*Wang et al.*, 2014).

3 Atlantic Water in the Arctic Ocean

At the beginning of this and the next sections we will briefly introduce the present understanding of the Arctic Ocean dynamics and changes, and the major issues to be discussed. Then the model results will be presented.

3.1 Background

A schematic of Atlantic Water (AW) circulation in the pan-Arctic Ocean is shown in Fig. 1. Saline and warm AW enters the Nordic Seas via the northern limb of the North Atlantic Current through the Greenland-Scotland Ridge, and continues northwards in the Nordic Seas in two branches of the Norwegian Atlantic Current (NwAC, *Orvik and Niiler*, 2002). When approaching the BSO, the eastern branch bifurcates with one branch entering the shallow Barents Sea and the other flowing towards the Fram Strait. The AW that enters the Barents Sea loses most of its heat (*Skagseth et al.*, 2008; *Smedsrud et al.*, 2013), and these modified waters flow into the intermediate layer of the Arctic Ocean via St. Anna Trough, or contribute to the halocline (*Karcher and Oberhuber*, 2002; *Dmitrenko et al.*, 2011, 2015; *Aksenov et al.*, 2011). The western branch of the NwAC and the remainder of the eastern NwAC branch continue towards the Fram Strait, and form the West Spitsbergen Current (WSC). At Fram Strait, a fraction of AW carried in the WSC recirculates and flows southwards in the East Greenland Current. The remaining part of the WSC enters the Arctic Ocean at depth, carrying the heat of the AW (*Rudels and Friedrich*, 2000; *Schauer et al.*, 2008; *Beszczynska-Moeller et al.*, 2012).

The AW below the halocline circulates mainly cyclonically along the continental slope and mid-ocean ridges as topographically steered boundary currents (*Rudels et al.*, 1994; *Karcher et al.*, 2003). The warmer Fram Strait branch and colder BSO branch converge north of the Kara Sea (*Schauer et al.*, 2002; *Karcher and Oberhuber*, 2002; *Maslowski et al.*, 2004) and continue eastward along the Eurasian slope. After passing the Laptev Sea slope, the boundary current bifurcates into one branch following the Lomonosov Ridge and another following the continental slope (*Woodgate et al.*, 2001). The former brings the AW toward the Fram Strait, while the latter continues into the Canadian Basin. Interannual changes in AW temperature can propagate into the Arctic Ocean via the Fram Strait inflow, leading to temperature variability along the AW boundary current (*Gerdes et al.*, 2003). Pronounced warming events in the Arctic AW layer have been observed in recent decades (*Polyakov et al.*, 2012, 2013b). These recent unprecedented warming implies that the Arctic deep basins are undergoing significant changes.

In previous model intercomparison studies with focus on Arctic AW (*Holloway et al.*, 2007; *Karcher et al.*, 2007), it was found that one outstanding issue in most ocean models is the unrealistic deepening and thickening of the AW layer. Numerical mixing associated with the advection operator was suggested to be the major cause (*Holloway et al.*, 2007). The recent CORE-II study indicates that the state-of-the-art ocean general circulation models which are currently used in climate studies still suffer from the deepening of the AW layer (*Ilicak et al.*, 2016). In the following we will explore whether and to what extent this problem can be alleviated by increasing horizontal resolution. Besides the mean state of the AW, we will investigate the model representation of decadal warming and variability on seasonal and interannual time scales.

3.2 Spin-up of the AW in Arctic basins

The annual mean temperature horizontally averaged in Eurasian and Canadian basins is plotted as a function of time and depth in Fig. 3. The two basins are defined as the Arctic region where the ocean bottom is deeper than 500 m, and separated by the Lomonosov Ridge. The basin mean temperature shows a very different temporal evolution in the two simulations. In the Eurasian Basin the warm AW layer thickens with time during the first 15 model years in the low resolution simulation (LOW), while the layer thickness remains quasi-steady (up to interannual variability) in the high resolution simulation (HIGH). After initial spin-up the depth of temperature maxima is located at about 400 m in LOW, while in HIGH it remains at about 300 m, the observed depth suggested by the hydrographic climatology. In the Canadian Basin the thickening and deepening of the AW layer is also very obvious in LOW. In this simulation the core of the AW layer deepens by about 100 m, changing from about 450 m to 550 m during the 60 model years. The model drift in the AW layer occurring during model spin-up is irreversible afterwards. The longer simulation presented in the CORE-II model intercomparison work indicates that the depth of the AW layer temperature maxima in the Canadian Basin continues deepening and stays at around 600 m depth after 300 model years (*Ilicak et al.*, 2016). In HIGH no thickening and deepening trend is found in the Canadian Basin. In both simulations the Eurasian Basin is featured with decadal warming events, and the Canadian Basin shows more pronounced warming in recent years. Besides the mean state, the two simulations are also different in their representation of variability and decadal changes, which will be assessed below.

3.3 Mean state of AW

To assess the spatial distribution of the AW in the Arctic Ocean, we show the Atlantic Water core temperature (AWCT) derived from 30 years mean model results in Fig. 4a. The AWCT is defined as the maximum temperature over the depth at each location. The typical spatial pattern of AW is shown by the climatology. The WSC brings warm AW into the Fram Strait, with a fraction recirculating southwards and the remaining part entering the Arctic Ocean. The latter passes the northern slope of Svalbard

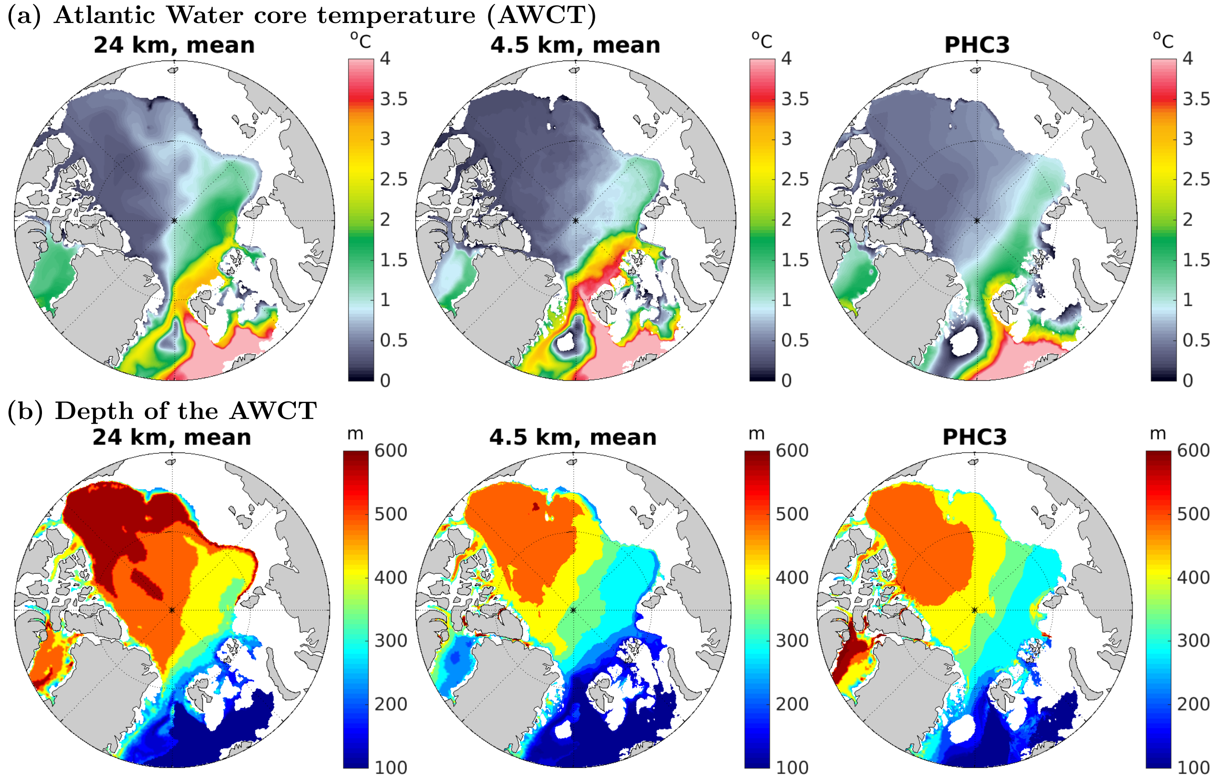

Figure 4. (a) Atlantic Water core temperature (AWCT) for (from left to right) simulations LOW and HIGH and the PHC climatology. (b) The same as (a) but for the depth of AWCT. Shelf regions (<200 m) are not shown. The model results are averaged from 1970 to 1999 for the purpose to compare to the PHC climatology.

and flows along the continental slope eastward in the Eurasian Basin. There is a strong contrast in temperature between the Eurasian and Canadian Basins, separated by the Lomonosov Ridge. The cold Barents Sea branch of AW enters the basin at the St. Anna Trough and circulates cyclonically as boundary current over the continental slope. Although both simulations can capture these main features, the warm AW is more confined in the Eurasian Basin in simulation HIGH than LOW. The AW boundary current starting from the St. Anna Trough towards the Lomonosov Ridge is much narrower in simulation HIGH, while it is horizontally more spread in LOW. The observed AWCT is located above 300 m depth in most part of the Eurasian Basin, and deepens towards the Beaufort Sea (Fig. 4b). Simulation HIGH largely reproduces the spatial change of the AWCT depth, and the depth in both the Eurasian and Canadian Basins is well represented. In simulation LOW the AWCT is deeper than the observation in most of the Arctic regions, and the contrast between the Eurasian and Canadian Basins is not as obvious as in the observation and simulation HIGH.

Simulation LOW obtains a vertically extended AW layer on basin scales, as shown by the mean temperature profiles in the two basins (Fig. 5a). In this simulation the depth of temperature maxima deepens by about 100 m in both Arctic basins, with the

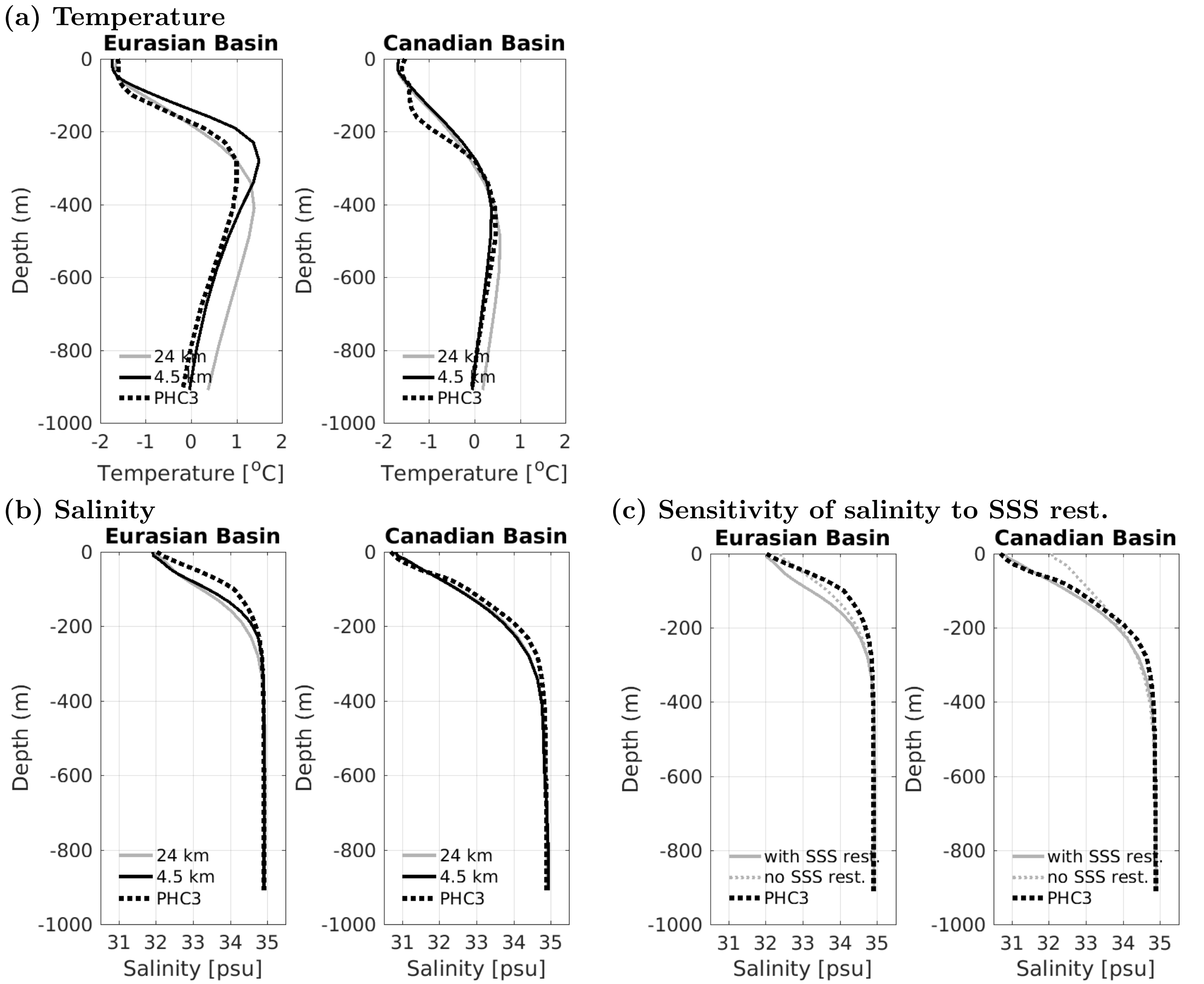

Figure 5. (a) Mean temperature profiles in the Eurasian and Canadian Basins in the low and high resolution simulations. (b) The same as (a) but for salinity. (c) The same as (b) but for comparing two low resolution simulations with and without sea surface salinity (SSS) restoring. Model results are averaged from 1980 to 1999.

vertical extent of the warm layer reaching much deeper depth. The maximum temperature in the Eurasian Basin in simulation HIGH is about 0.4^{o}C higher than that in the PHC3 data, but the observed depth of the temperature maxima, at about 300 m, is captured by this simulation. In the Canadian Basin the temperature of the Pacific Winter Water (located between the Pacific Summer Water and about 200 m depth) is overestimated in both simulations, implying too strong vertical diffusion, which mixes the cold water with warmer AW below. This feature is obviously not linked to model horizontal resolution and will be discussed in Section 5.

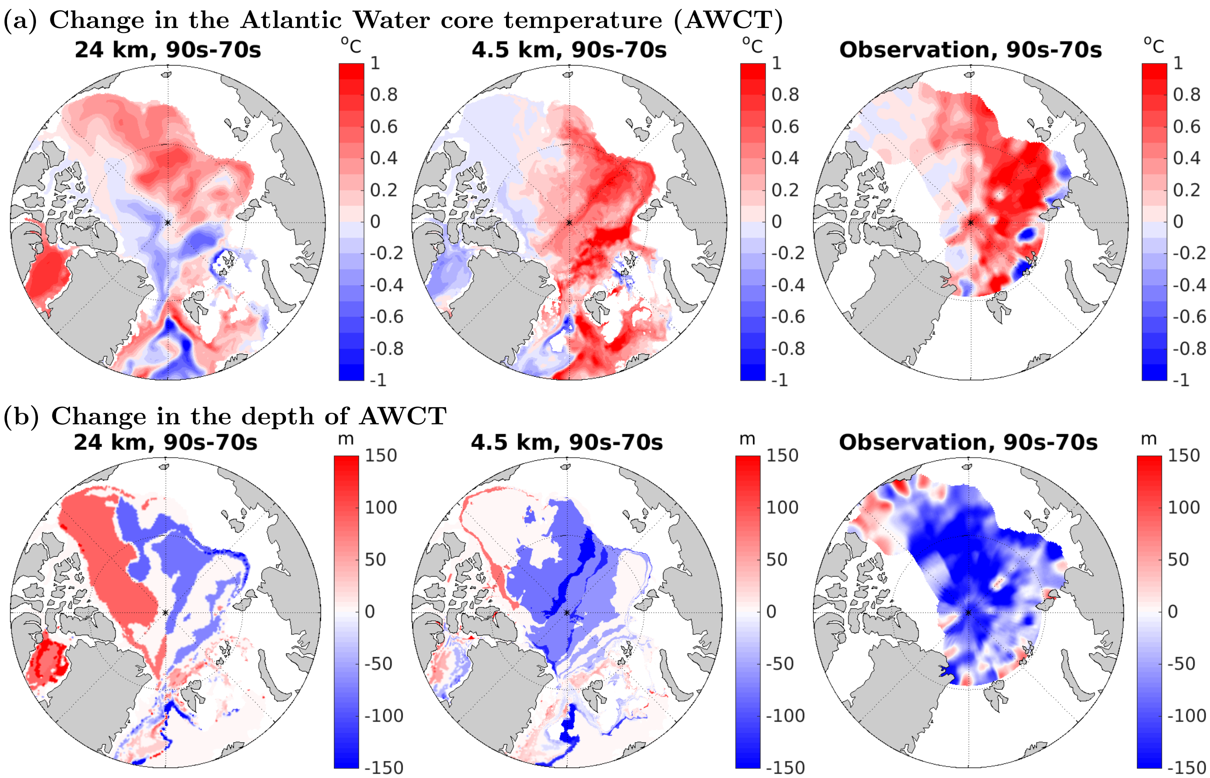

Figure 6. (a) Difference of the Atlantic Water core temperature (AWCT) between the 1990s mean and 1970s mean (the former minus the latter) in simulations LOW and HIGH and in observations (from left to right). (b) The same as (a) but for the depth of the AWCT. For the model results the shelf regions (<200 m) are not shown. The observations of the two periods compared are derived from *Polyakov et al.* (2012).

The AW circulation pattern was examined by comparing the topostrophy (*Holloway et al.*, 2007) in the two simulations. Cyclonic circulation dominates the AW layer boundary currents in both ocean basins similarly in the two simulations (not shown). The fact that simulation LOW also has the correct circulation direction in the Canadian Basin is very possibly just because its resolution is already fine compared to those with problems reported in previous studies (e.g., *Holloway et al.*, 2007). Indeed, the Arctic Ocean hydrography obtained on mesh LOW is well simulated when compared to the suite of state-of-the-art ocean climate models analyzed in *Ilicak et al.* (2016).

3.4 Variability of AW

Although the Arctic Ocean is at the far end of the North Atlantic Current northern limb, strong warming events have been observed in the Arctic AW layer at the end of the 20th century and beginning of the 21st century (*Gerdes et al.*, 2003; *Karcher et al.*, 2003; *Polyakov et al.*, 2012). Despite very limited observations in the remote Arctic deep basins, averaged over decadal time scales the warming events are outstanding and the compiled datasets can be used to assess the model representation of the AW warming (*Polyakov et al.*, 2012). In Fig. 6a the warming in the 1990s relative to the 1970s in the two model simulations and observation is shown. The observation indicates a basinwise warming by about $1°\mathrm{C}$ in the Eurasian Basin, which propagates into the Canadian Basin crossing the Lomonosov Ridge along the continental slope. Simulation HIGH similarly shows a basinwise warming in the Eurasian Basin, and slightly weaker penetration of the warming signal into the Canadian Basin compared to the observation. In this simulation the boundary current along the continental slope and Lomonosov Ridge shows stronger warming than the basin interior, which is not seen in the observation. This could be partly due to the sparseness of hydrography observations. Simulation LOW also obtains a warming signal in the 1990s, but mainly in the eastern Eurasian Basin and over a large part of the Canadian Basin. As shown by the time series of the basin mean temperature in Fig. 3, there is a strong warming and deepening trend in the Canadian Basin throughout simulation LOW. This model drift can explain part of the warming in LOW shown in Fig.6a .

The depth of the AWCT became shallower in the 1990s compared to the 1970s in the observation (Fig. 6b). Simulation HIGH shows a consistent pattern in the change of AWCT depth, while the magnitude is about half of the observed. In simulation LOW the depth becomes shallower in a small region in the sector of the East Siberian and Chukchi Seas, but it becomes deeper by about 100 m north of the CAA. The latter can be attributed to the deepening trend of the AW in the model as shown in Fig. 3.

In the following we will focus on the resolution dependency of the interannual variability of AWCT in the two simulations. We use the standard deviation (std) of annual mean AWCT as an indicator of the interannual variability. As shown in Fig. 7, the interannual variability is stronger in the Eurasian Basin and weakens along the AW advection pathway in both simulations. In simulation HIGH, the std is more than $0.4°\mathrm{C}$ in front of the Eurasian continental slope and along the Lomonosov Ridge toward the North Pole. The highest std is found in the western Eurasian Basin where the boundary of the inflowing AW on the interior side changes its location most significantly. In simulation LOW, the std is in the range of $0.2-0.3°\mathrm{C}$ along the path of the AW circulation. Contrary to HIGH, there is no clear indication of stronger interannual variability along the topographically steered boundary current in LOW. The interannual variability is advected to a larger area in the Canadian Basin in LOW, which is consistent to the larger horizontal spreading of AW (Fig. 4a). Using a different model with resolution of 10 km, *Lique and Steele* (2012) showed that the std of AWCT is in a range of about $0.1-0.4°\mathrm{C}$ in the Eurasian Basin, similar to that in our simulations. However, the spatial pattern of the std is very different from any of our simulations. In their simulation the highest interannual variability is found starting from the Laptev Sea coast toward the Lomonosov Ridge directly crossing the Eurasian Basin (figure 13 of *Lique and Steele*, 2012). In this respect the difference in the AW interannual variability induced by different resolutions, although significant, is less pronounced than the difference between two different models.

The magnitude of mean seasonal cycle of the AWCT is also compared in Fig. 7. Both simulations show that the seasonal variability is advected from Fram Strait into the Arctic interior along the AW boundary current, and then the variability is re-energized at the St. Anna Trough by the BSO branch. In simulation HIGH the magnitude of the AWCT seasonal cycle is nearly $0.5°\mathrm{C}$ in the boundary current downstream the St. Anna Trough, and decreases to about $0.2°\mathrm{C}$ over the Laptev Sea continental slope. In simulation LOW the magnitude of the seasonal variability along the continental slope is about half of that in HIGH. When the boundary current bifurcates, with one branch circulating northward along the Lomonosov Ridge and

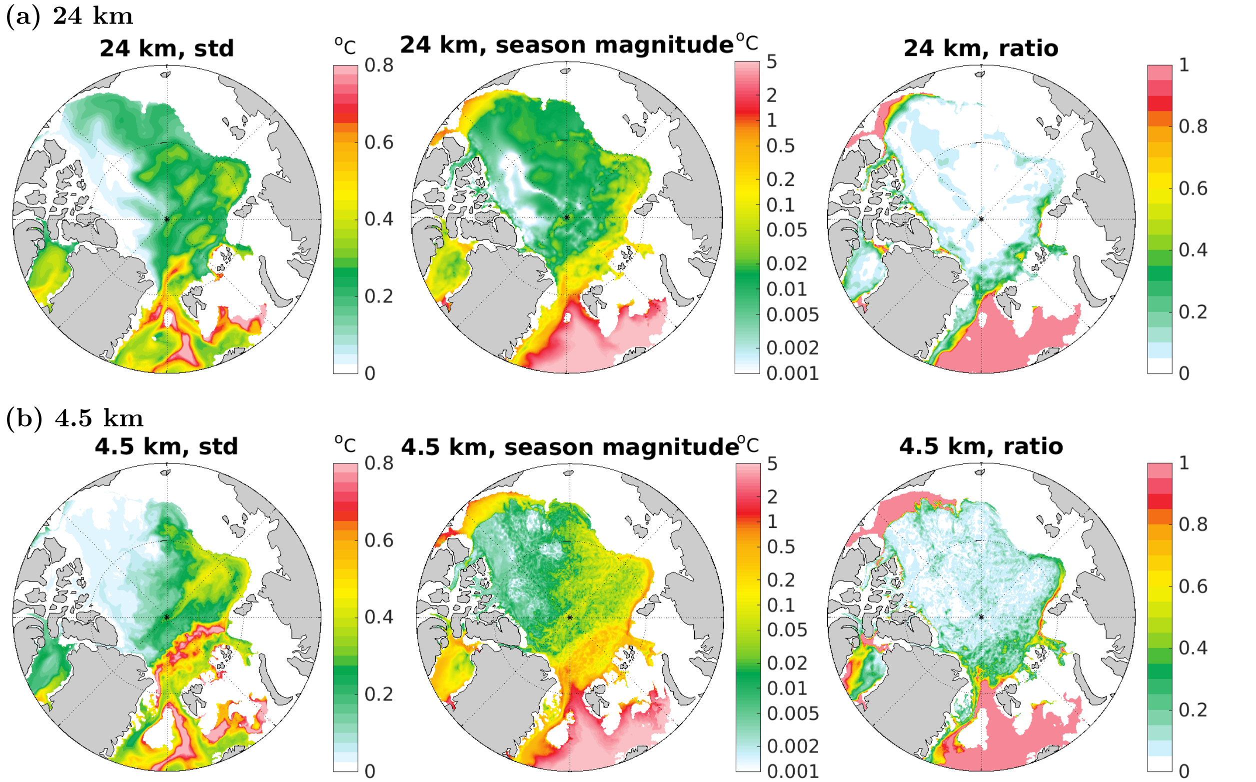

Figure 7. (left) The standard deviation (std) of the annual mean Atlantic Water core temperature (AWCT), (middle) the magnitude of the AWCT mean seasonal variability, and (right) the ratio between the seasonal magnitude and the std of the annual mean, in (a) simulation LOW and (b) HIGH for the period of 1980-2009. Note that nonlinear color scales are used in the plots of seasonal variability.

another penetrating into the Canadian Basin, the seasonal variability also propagates further along these branches. However, the magnitude becomes smaller with distance, which is less than 0.05°C in HIGH and even lower in LOW. When compared to the strength of interannual variability, in both simulations the seasonal variability is negligible except in the region north of Svalbard and within the narrow boundary current. The most significant seasonal variability is found within the narrow boundary current between the St. Anna Trough and the Laptev Sea continental slope in simulation HIGH, where the ratio between the magnitude of the AWCT seasonal cycle and the std of the annual mean AWCT is about 0.8. Therefore, the BSO branch supplies a large part of the seasonal variability shown in this slope region. The spatial pattern of the AWCT seasonal variability in the Arctic Ocean in our simulations is similar to that derived from a different model at 10 km resolution shown by *Lique and Steele* (2012). However, the strength of the AWCT seasonal variability in their model simulation is similar to that obtained in our simulation LOW and lower than in HIGH.

4 Salinity and freshwater budget

4.1 Background

A schematic of freshwater (FW) circulation in the pan-Arctic Ocean is shown in Fig. 1. The Arctic Ocean receives a large amount of FW from river runoff, net precipitation, and Pacific Water through Bering Strait (*Serreze et al.*, 2006; *Dickson et al.*, 2007; *Haine et al.*, 2015; *Carmack et al.*, 2016). Liquid FW is stored in the Arctic Ocean with a very non-uniform spatial distribution. The Canadian Basin is characterized by the largest FW content (Defined as the amount of pure FW that could be taken out of the upper ocean so that the ocean salinity is changed to 34.8, the Arctic Ocean reference salinity (see *Aagaard and Carmack*, 1989). In the calculation of the modelled FW content presented below, the integration is taken from ocean surface to the depth where salinity is equal to the reference salinity). Especially in the Beaufort Gyre the FW amounts to about 20 m, whereas it is about 5-10 m in the Eurasian Basin (e.g., *Rabe et al.*, 2011). The anticyclonic Beaufort Gyre is driven by the Beaufort Sea High in atmospheric pressure, which changes the FW content in Beaufort Gyre and the FW distribution between the ocean basins by modulating convergence/divergence of Ekman transport (e.g., *Proshutinsky et al.*, 2002, 2009; *Giles et al.*, 2012). Wind variability over continental shelves can locally induce more significant changes in FW content than the variability from river fluxes, and the variation in large scale atmospheric circulation (Arctic Oscillation) can modify the pathway of river runoff, thus changing the FW distribution between the Arctic basins (*Dmitrenko et al.*, 2008; *Morison et al.*, 2012).

Both liquid FW and sea ice are drained by the Transpolar Drift and released through the Fram Strait. The liquid FW exported through the Fram Strait is slightly larger than sea ice export (*Serreze et al.*, 2006), but the difference has increased during the last decade (*Haine et al.*, 2015). Arctic liquid FW is also released to lower latitudes through the CAA and then the Davis Strait, with an amount similar to that released through the Fram Strait (*Serreze et al.*, 2006; *Curry et al.*, 2014). Sea ice export through the Davis Strait is much less than that from the Fram Strait. The possible climate relevance of the FW cycle in the Arctic Ocean and FW release to the North Atlantic is one of the main reasons for continued research on the Arctic FW budget in both the observation and modeling communities (*Carmack et al.*, 2016).

The liquid FW stored in the Arctic Ocean has been increasing starting from the mid-1990s as shown by observations (*Proshutinsky et al.*, 2009; *McPhee et al.*, 2009; *Giles et al.*, 2012; *Polyakov et al.*, 2013b; *Rabe et al.*, 2014), while sea ice has a persistent declining trend in thickness and volume (*Kwok et al.*, 2009; *Laxon et al.*, 2013). The liquid FW export through the Davis Strait was observed to be lower in the 2000s than in the 1990s (*Curry et al.*, 2014), while the Fram Strait liquid FW export has slightly increased in the 2000s compared to the climatological value (*Haine et al.*, 2015). In recent CORE-II model studies using a suite of global ocean-sea ice models, it was shown that the recent increase in Arctic liquid FW content is caused by both sea ice melting and reduction of total liquid FW export, with the former being more significant in most of the models (*Wang et al.*, 2016b). However, current observations, especially those of liquid FW budget, are still too sparse for the purpose of quantitative verification of the finding based on models.

In the CORE-II model intercomparison project, it was found that the simulated mean state of Arctic FW (FW content and its spatial distribution, and FW transport through Arctic gateways) has significantly large model spreads, even though the same atmospheric forcing was used (*Wang et al.*, 2016a, b). Interannual variability of FW export via the Fram Strait is the least

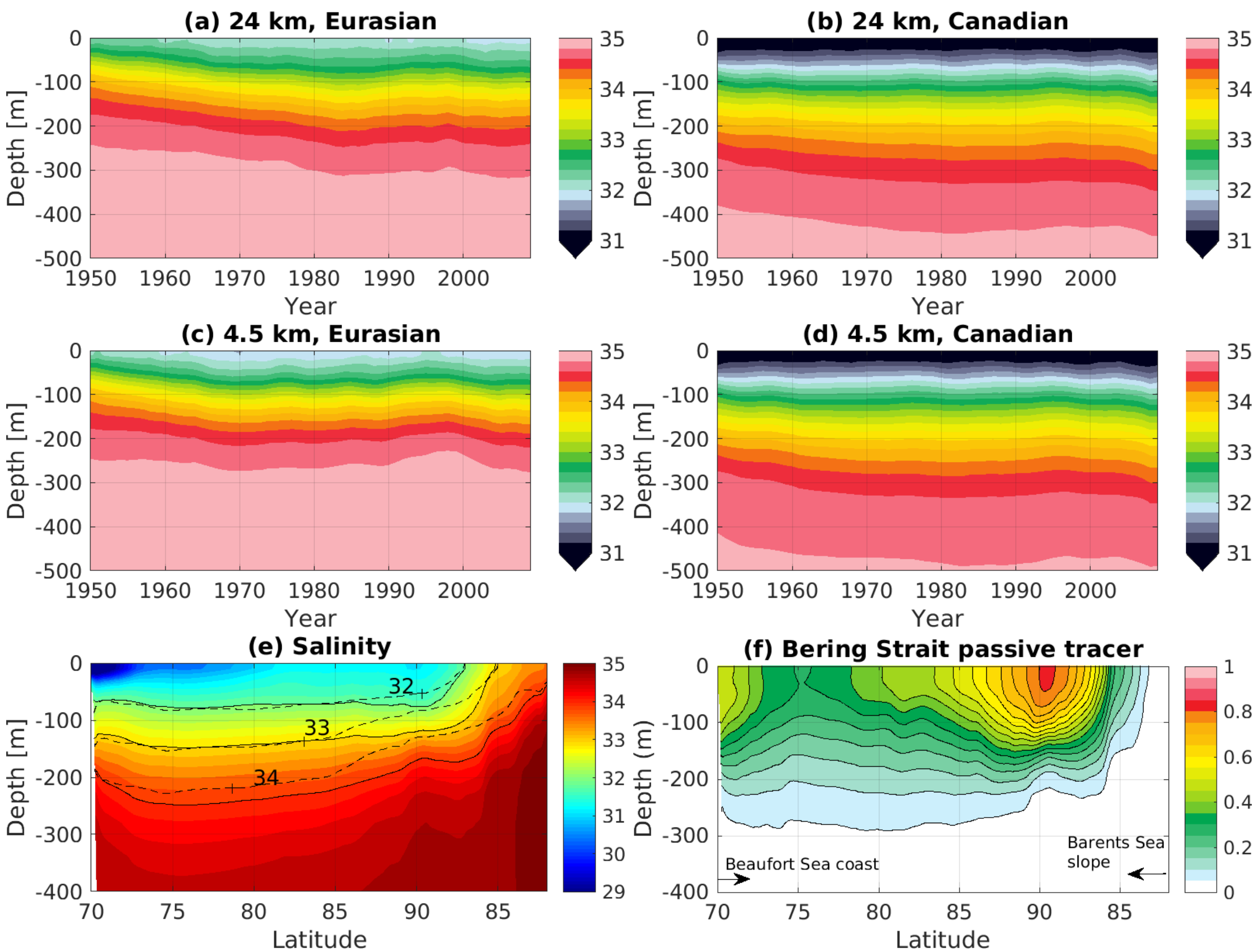

Figure 8. Hovmöller diagram of mean salinity for (a) the Eurasian Basin and (b) the Canadian Basin obtained in simulation LOW. (c),(d) are the same as (a),(b) but for simulation HIGH. (e) Salinity at the transect along the $140^{\circ}\mathrm{W}/40^{\circ}\mathrm{E}$ longitude averaged over the 1980-1999 period in simulation HIGH, shown by the color patch and solid contours. The dashed contours indicate the salinity at the beginning of the model simulation. This comparison between the initial state and the mean state when salinity does not show a strong trend is intended to explain the salinity drift during spinup. (f) The same as (e) but for the passive tracer released in Bering Strait. The location of the transect is indicated by the magenta line in Fig. 1. The whole simulation period 1950–2009 is shown in a-d.

consistently simulated among different Arctic gateways, in both AOMIP and CORE-II models (*Jahn et al.*, 2012; *Wang et al.*, 2016b). Besides, all the CORE-II models show a dramatic increase in the simulated Arctic liquid FW content during the model spin-up phase; afterwards, the FW content stays at the overestimated level (unless overestimated AW salt inflow causes it to drop in one particular model, *Wang et al.*, 2016b). There are indications in some studies that higher model resolution might improve the pathway and spatial distribution of liquid FW (*Koldunov et al.*, 2014; *Aksenov et al.*, 2016). In the following we will compare the Arctic FW budget simulated with FESOM using two different horizontal resolutions. The focus will be on the impact on model spin-up, mean state, and interannual to decadal variability of the FW budget.

4.2 Spin-up of salinity and freshwater content

The annual mean salinity horizontally averaged over the Eurasian and Canadian Basins is plotted as a function of time and depth in Fig. 8a-d. In both basins the salinity decreases with time, and it takes nearly 30 years for salinity to spin up to a quasi-equilibrium state in both basins. The two simulations show very similar results, except that the salinity drift in the Eurasian Basin takes place in a relatively shorter period (about 20 years) in the high resolution (HIGH) than in the low resolution simulation (LOW). In the Eurasian Basin the salinity drift takes place mainly in the upper 200 m, while in the Canadian Basin mainly between 100 and 400 m depth. The different behavior implies that processes associated with the salinity drifts are different in the two basins.

The freshening of the Eurasian Basin in HIGH is illustrated in a transect along the $140^{o}\mathrm{W}/40^{o}\mathrm{E}$ longitude line in Fig. 8e. Compared to the mean salinity in the first model year, the salinity becomes considerably lower near the Lomonosov Ridge (located near the North Pole in this transect) after the model spin-up phase. The location of strong freshening coincides with the pathway of the Pacific Water from the Bering Strait (Fig. 8f), which is carried by the Transpolar Drift together with FW from the Eurasian riverine. Therefore, the freshening of the Eurasian Basin could be linked to model representation of the upper ocean circulation pathway and the spatial distribution of FW from the Bering Strait and river runoff. In simulation LOW we obtain similar results, so changing model resolution does not influence the occurrence of this salinity drift.

The salinity drift is manifested in the time series of Arctic Ocean FW content (Fig. 9). In both simulations the total Arctic liquid FW content increases nearly linearly in the first 20 years. The increase takes place mainly in the two basins, with a similar magnitude. In the Canadian Basin the FW contents are almost identical in the two simulations for all the time, while the FW content in the Eurasian Basin is about 20% higher in simulation LOW after 30 model years. To explain the latter we carried out one sensitivity experiment on mesh HIGH-CAA. Its resolution in the Arctic Ocean is the same as LOW (24 km), but it has 4.5 km resolution inside the CAA straits. The spatial patterns of mean FW content (in m) from the three simulations are shown in Fig. 10a-c. HIGH-CAA shows a pattern very similar to simulation HIGH, characterized by a large FW storage in the Beaufort Gyre and a decrease of FW content from the Canadian Basin towards the Eurasian Basin as expected from observations (Fig. 10d). In the simulation LOW, more FW takes the release route through the Fram Strait, because the CAA straits are poorly resolved with the coarse resolution and the CAA outflow is restricted (see more details in the section of mean state). This increases the FW content in the western Eurasian Basin. Therefore, it is mainly resolving the narrow straits

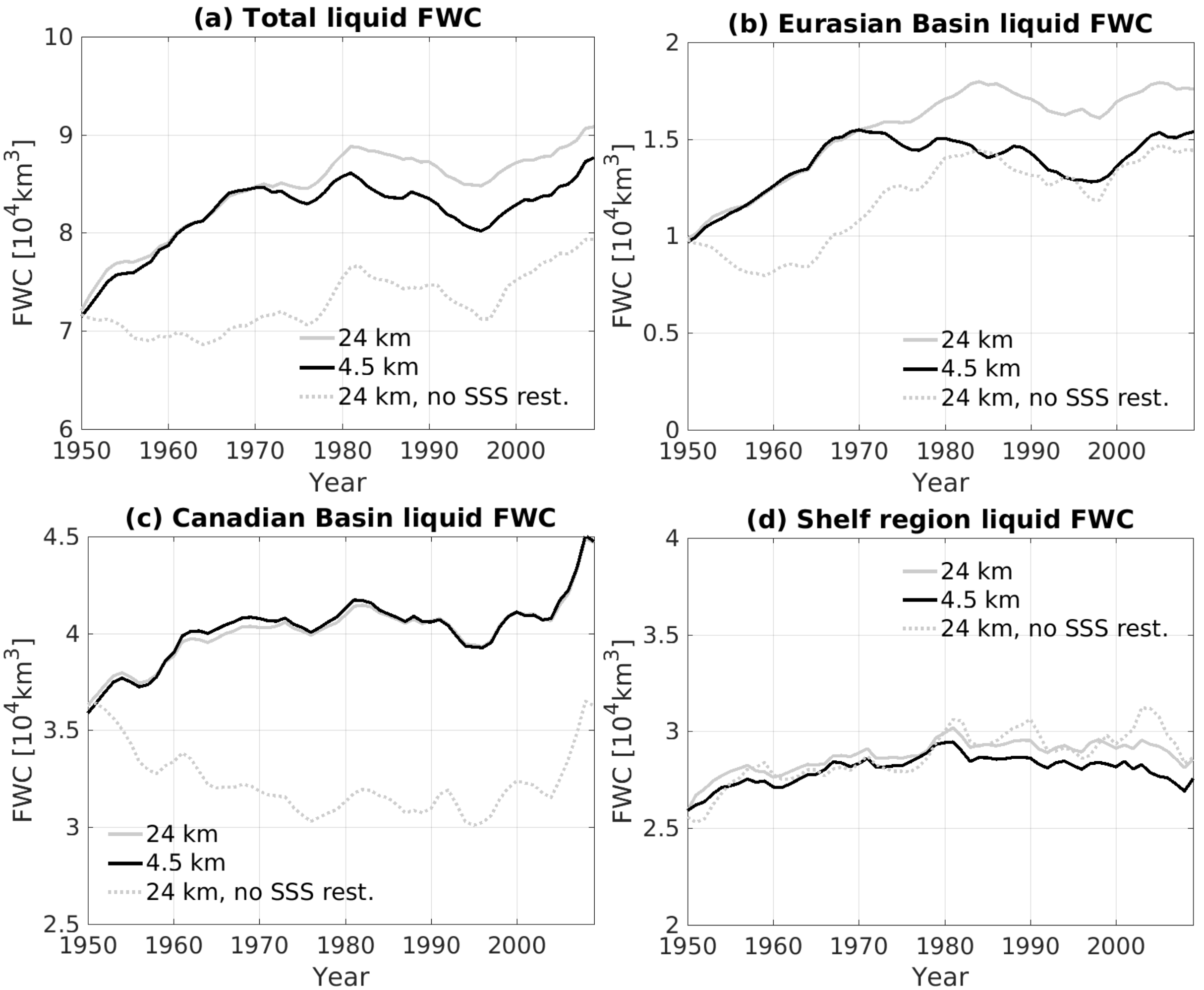

Figure 9. (a) Time series of annual mean total Arctic liquid freshwater (FW) content. The liquid FW content in the Eurasian Basin, Canadian Basin and shelf regions are shown in (b)(c)(d), respectively. The FW content is calculated using a reference salinity of 34.8. The whole integration period 1950–2009 is shown.

Table 1. Arctic Ocean liquid and solid freshwater (FW) budget relative to a reference salinity of 34.8, and the net ocean volume transport through Arctic gateways. The FW budget terms are shown for the periods 1980-2000 and 2000-2009 separately. The correlation coefficients for fluxes obtained from the two simulations (LOW and HIGH) are shown for the period 1980-2009 in the last column, and all correlations are significant at the 95% level. Liquid FW contents for the 2000-2009 period are shown with the **changes** relative to the 1980-2000 period. FW fluxes are shown in $\mathrm{km}^3/\mathrm{year}$, FW contents are in $10^4\,\mathrm{km}^3$, and ocean volume transports are in Sv. Positive fluxes indicate sources for the Arctic Ocean.

	1980-2000			**after 2000**			Model
	Observation	LOW	HIGH	Observation	LOW	HIGH	correlation
Liquid freshwater							
Fram Strait	-2660 ± 528^a	-2306	-2115	-2800 ± 420^b	-1979	-1861	0.78
Davis Strait	-3200 ± 320^a	-2263	-2887	-2900 ± 190^b	-2199	-2722	0.75
Bering Strait	$2400? \pm 300^a$	2029	2170	2500 ± 100^b	1932	2079	0.98
BSO	-90 ± 94^a	-591	-441	-90 ± 90^b	-779	-664	0.90
Arctic FW content	6.92 [c]	8.69	8.19	$(+0.45)^{b,d}$	(+0.17)	(+0.17)	
Solid freshwater							
Fram Strait	-2300 ± 340^a	-2369	-2488	-1900 ± 280^b	-2065	-2154	0.95
Davis Strait	$-160 \pm ?^a$	-416	-427	-320 ± 45^b	-320	-342	0.98
NH FW content	1.8[e]	2.28	2.21	1.44[e]	1.84	1.81	

	1980-2009			Model
	Observation	LOW	HIGH	correlation
Ocean volume flux				
Fram Strait	-2 ± 2.7^f	-2.18	-1.84	0.88
Davis Strait	-3.2 ± 1.2 to -1.6 ± 0.2^g	-1.03	-1.69	0.90
Bering Strait	$0.8 \pm 0.2^{h,i}$	0.87	0.95	0.99
BSO	2.0 to $2.3^{j,k,l}$	2.36	2.52	0.93

[a] *Serreze et al.* (2006), [b] *Haine et al.* (2015), [c] computed from PHC3 (*Steele et al.*, 2001), [d] *Polyakov et al.* (2013b), [e] based on the PIOMAS Arctic sea ice volume reanalysis (*Schweiger et al.*, 2011) by assuming sea ice density of $910\,\mathrm{kg/m^3}$ and salinity of $4\,\mathrm{psu}$, [f] *Schauer et al.* (2008), [g] *Curry et al.* (2014), [h] *Roach et al.* (1995), [i] *Woodgate and Aagaard* (2005), [j] *Smedsrud et al.* (2010), [k] *Skagseth et al.* (2008), [l] *Smedsrud et al.* (2013).

in simulation HIGH that leads to the difference of FW content spatial distribution from simulation LOW, rather than the high resolution inside the Arctic Ocean.

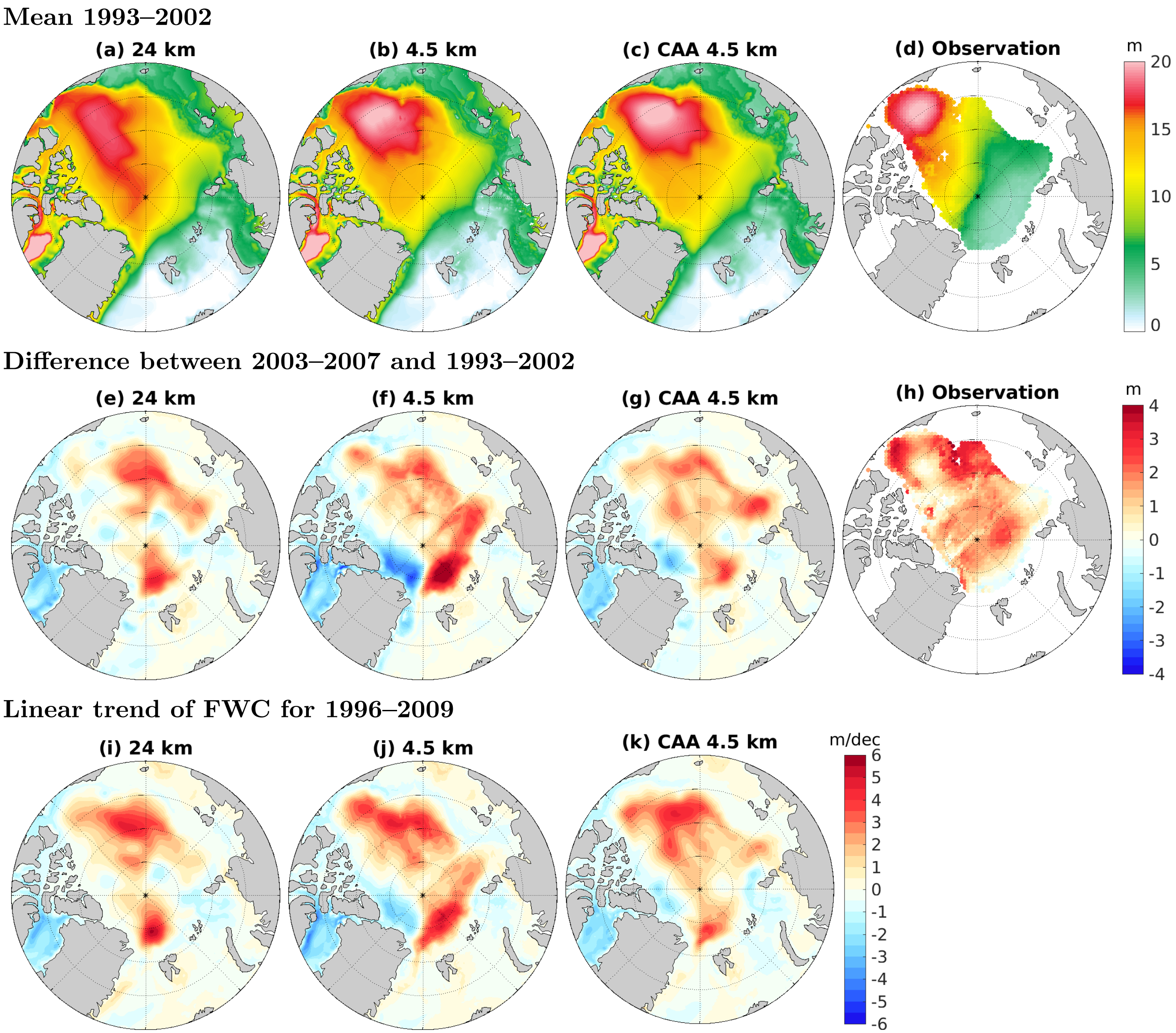

Figure 10. Mean liquid freshwater (FW) content (in m) for the period 1993–2002 for (a) LOW, (b) HIGH, (c) HIGH-CAA, and (d) observation of *Rabe et al.* (2011). Difference in liquid FW content between the periods 2003–2007 and 1993–2002 for (e) LOW, (f) HIGH, (g) HIGH-CAA, and (h) observation. These two periods are chosen to be the same as those used by *Wang et al.* (2016b) for the convenience of direct comparison. Linear trend of FW content (m/decade) for the period of increasing FW content (1996–2009, see Fig. 9a) for (i) LOW, (j) HIGH and (k) HIGH-CAA. The FW content is calculated using a reference salinity of 34.8. The sensitivity experiment HIGH-CAA is introduced to isolate the impact of resolution in the Canadian Arctic Archipelago (CAA) from that in the Arctic interior. It has 4.5 km resolution (the resolution in HIGH) only inside the CAA straits and 24 km resolution (the resolution in LOW) in other parts of the Arctic Ocean.

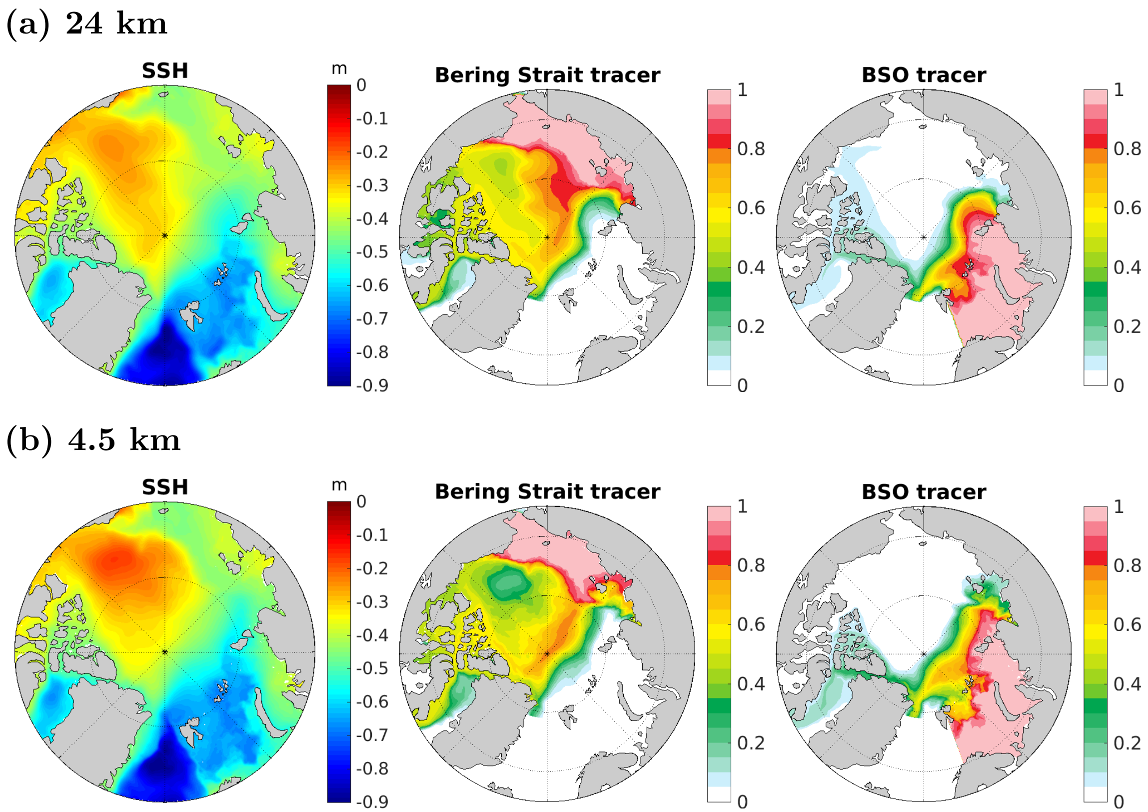

Figure 11. (a) Mean sea surface height (left), Bering Strait passive tracer (middle), and Barents Sea Opening (BSO) passive tracer (right) in simulation LOW for the period of 1993–2002; for this period the mean state of FW content is shown in Fig. 10a,b,c. (b) The same as (a) but in simulation HIGH. Note that the passive tracers are set to zero south of the Fram and Davis Straits in the plots. The passive tracers are averaged over the upper 100 m.

4.3 Mean state of liquid freshwater

As a consequence of salinity drift during the model spin-up, the basin mean salinity shows biases in the halocline in both Arctic basins (Fig. 5b). The biases are largest at the mid-depth of the halocline, as the salinity is restored to the climatology at the ocean surface, and below the halocline the salinity is determined by that of the AW. As the Eurasian Basin bias in simulation LOW is larger than in simulation HIGH, the overestimation of Arctic FW content is more significant in LOW (26% compared to 18%, Table 1). As mentioned above (Section 4.2), the spatial distribution of liquid FW content is better reproduced in HIGH than in LOW (Fig. 10a-d), albeit with overestimation in both simulations, because the high resolution more faithfully represents the narrow channels in CAA. The variety of FW content distributions simulated in different ocean models shown by *Wang et al.* (2016b) presumably can be partly attributed to different model representations of the CAA region.

The spatial pattern of FW content is manifested in the simulated sea surface height (SSH, see Fig. 11), since the steric height is dominated by the halosteric component in the Arctic Ocean (e.g., *Griffies et al.*, 2014). In simulation HIGH the SSH field shows a better represented Beaufort Gyre. The CAA resolution not only impacts the FW content pattern, but also the circulation and export pathways of water masses. For example, as illustrated by passive tracers (Fig. 11), in simulation LOW the Pacific Water penetrates more into the Canadian Basin, and has a higher concentration at the Fram Strait than in HIGH. And the better resolved CAA channels in HIGH allow more Atlantic Water from BSO to be released through the CAA.

To access the simulated mean state of FW transport through main Arctic gateways, we compare the model results for the period of 1980-2000 with the synthesized values by *Serreze et al.* (2006, see Table 1). Observations suggested that more FW is released through the CAA than through the Fram Strait. This is reproduced in simulation HIGH, while the FW transports through the two export gateways are nearly the same in LOW. Although the simulated CAA FW export in both simulations is lower than the synthesized value, the CAA FW export in HIGH is significantly higher than in LOW, and still within the observational uncertainty range. At the Fram Strait both the ocean volume and FW transports in LOW are higher than in HIGH, as expected from the impact of resolution in the CAA discussed above. Although using higher resolution reduces the Fram Strait FW export, the mean value is still close to the lower bound of the observational range. At the Bering Strait, the FW import is underestimated in the two simulations, with simulation HIGH obtaining a slightly higher value, very close to the lower bound of the observational range. As the Bering Strait ocean volume transports in the two simulations are within the range suggested by observations, the underestimation of FW transports is due to biases in the Pacific Water salinity, which could be still in a phase of large scale spin-up within the model integration period.

4.4 Variability of liquid freshwater

The simulated liquid FW contents do not show significant interannual variability, but rather large decadal changes (Fig. 12a). In both simulations the FW content decreases from the beginning of 1980s until the mid-1990s, and then increases afterwards. The descending trend of observed FW content (*Polyakov et al.*, 2013b) before the mid-1990s is much lower (Fig. 12a). Most of the models used in the CORE-II model intercomparison obtained a significant descending trend before the mid-1990s (Figure 8 of *Wang et al.*, 2016b), as in the two simulations presented here. Compared to the period of 1980-2000, the mean Arctic FW content averaged over the 2000s has increased by about $4500\,\mathrm{km}^3$ based on observations (*Polyakov et al.*, 2013b; *Haine et al.*, 2015), while the increase is only about $1700\,\mathrm{km}^3$ in our two simulations (Table 1).

The linear trend in the FW content for the period 1996–2009 based on the data set of *Polyakov et al.* (2013b) shown in Fig. 12a is $844\,\mathrm{km}^3/\mathrm{yr}$. The upward trends in the two simulations are lower, having half of this value in LOW ($423\,\mathrm{km}^3/\mathrm{yr}$) and 60% of it in HIGH ($521\,\mathrm{km}^3/\mathrm{yr}$). On average the 13 CORE-II models analyzed in *Wang et al.* (2016b) underestimated the observed upward trend also by half. Although the total Arctic liquid FW content increases nearly linearly after the mid-1990s, the situation is quite different in the individual Arctic basins. In both simulations, during the last 5 years of the integration, the upward trend strengthens in the Canadian Basin, while the trend almost stops in the Eurasian Basin, and a descending trend is non-negligible over the continental shelves (Fig. 9). The model result is consistent to the observed scenario of changes in

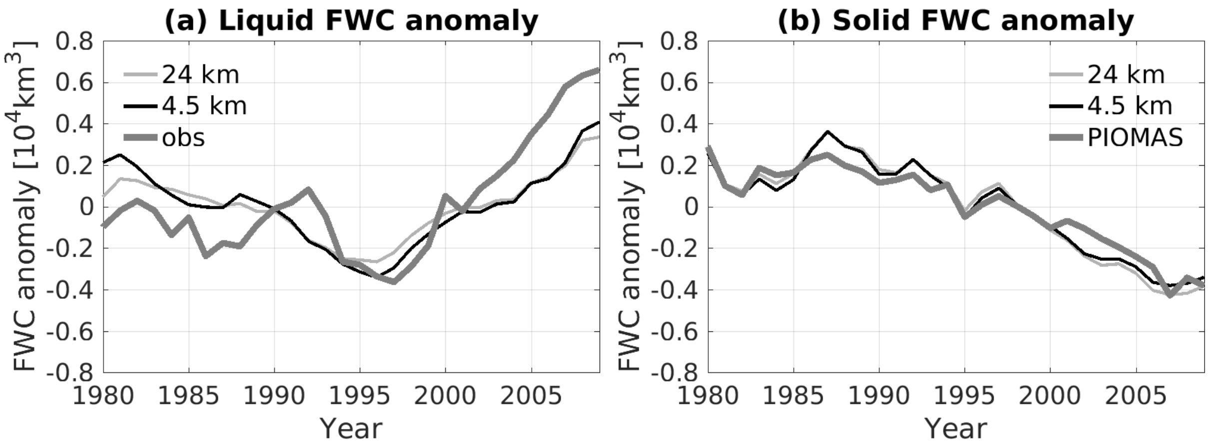

Figure 12. Anomalies of annual mean (a) liquid freshwater (FW) content and (b) solid FW content of the Arctic Ocean relative to the mean of the plotted period. The FW content is calculated using a reference salinity of 34.8. The liquid FW content observation is provided by *Polyakov et al.* (2013a), and the solid FW content is compared to the data derived from PIOMAS reanalysis (*Schweiger et al.*, 2011). The time period from 1980 to 2009 is shown.

FW distribution in the two Arctic basins described by *Morison et al.* (2012). They explained that the changes were due to a cyclonic shift in the ocean pathway of Eurasian runoff associated with an increased Arctic Oscillation index.

We are also interested in the model representation of temporal variation of FW content spatial distribution. In Fig. 10e-h the difference in FW content between the periods of 2003-2007 and 1993-2002 is shown. The observation indicates that the most significant increase in FW content between the two periods occurs along the Chukchi Sea continental slope and on the periphery of the Beaufort Gyre. At the latter location the simulations did not obtain a similar pattern of positive changes. The FW content increases on both side of the Lomonosov Ridge in the observation. Simulation HIGH consistently obtains positive changes in the Eurasian Basin with a larger magnitude. It has negative values north of Greenland, which is not present in the observation. Further efforts are required to understand the reason.

The spatial pattern of positive changes in FW content in HIGH is very similar to that obtained in a model with about 12 km resolution in the Arctic Ocean shown by *Wang et al.* (2016b). In their study most other models show a quite different pattern because of too coarse model resolution used (about 1 degree resolution). Besides the difference in FW content between the two periods, we also calculated the linear trend of vertically integrated FW content from 1996 to 2009 (Fig. 10i-k). The two methods of diagnosing the temporal variation of the FW content provide similar conclusions on the impact of model resolution (compare Fig. 10e-g with Fig. 10i-k). As mentioned above, the resolution inside the CAA plays an important role in representing the mean state of the Arctic Ocean FW content. Here the additional sensitivity experiment, where high resolution is only applied in the CAA channels, helps to illustrate that the high resolution inside the Arctic Ocean does have some impact on the representation of FW content spatial variation, for example, in the Beaufort Gyre and the central Eurasian Basin.

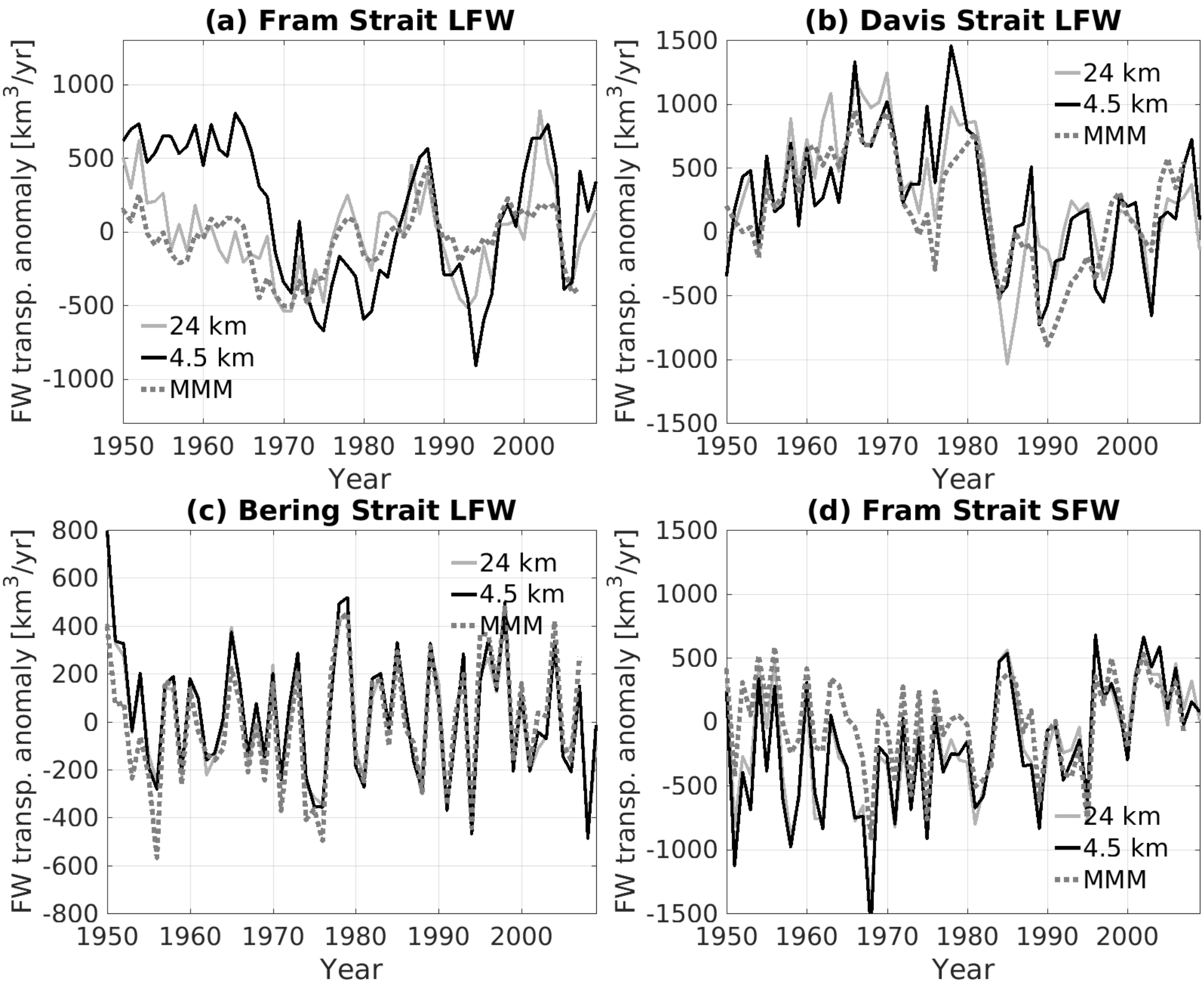

Figure 13. Anomalies of annual mean freshwater (FW) transport through main Arctic gateways. Liquid FW transport through (a) Fram Strait, (b) Davis Strait, (c) Bering Strait and solid FW transport through (d) Fram Strait. The dotted lines show the multi-model means (MMM) obtained from 13 CORE-II models (*Wang et al.*, 2016a, b). The whole integration period 1950–2009 is shown.

The interannual variability of FW transport through the Arctic gateways shows large similarity between the two simulations after the spin-up phase (Fig. 13a-c). The correlation coefficients between the FW transports from the two simulations are similar at the Davis and Fram Straits (0.75 and 0.78, respectively for the period of 1980–2009, Table 1). The correlation is lower than the correlation for ocean volume transports, indicating that the simulated interannual variability of salt transport changes between the two simulations and leads to reduced inter-simulation correlation for FW transports. The current model results are largely similar to the multi-model mean result analyzed by *Wang et al.* (2016b, also plotted in Fig. 13). The most significant difference is in the Fram Strait FW transport. For example, the changes of FW transport from the mid-1990s to the beginning of 2000s is more pronounced in our two simulations (Fig. 13a). The variability of FW transport at the Fram Strait has been found to be the least consistently simulated among both AOMIP and CORE-II models (*Jahn et al.*, 2012; *Wang et al.*, 2016b). At the Bering Strait the variability is nearly not distinguishable between the two simulations and the multi-model mean obtained in the past model study (Fig. 13c).

On decadal time scales, the observed FW export through the Davis Strait in the 2000s is about 10% lower than the climatology of 1980-2000 (*Haine et al.*, 2015). Both simulations reproduce the reduction in the Davis Strait FW export, but the magnitude of reduction is less significant than the observed (Table 1). In simulation HIGH the reduction (about 5%) is larger than in LOW. At the Fram Strait the FW export is suggested to be slightly higher in the 2000s than in the period of 1980-2000 (*Haine et al.*, 2015), while the two simulations similarly show an opposite result, obtaining a reduction of $\sim 300\,\mathrm{km}^3/\mathrm{yr}$ in the 2000s. The Bering Strait FW transport remains nearly at the same level after the 2000s, which is reproduced by the simulations. Note that the uncertainty in observations is large due to the sparseness of measurements, and both the observed and simulated changes in FW transports through the Arctic gateways between the two periods are smaller than the magnitude of respective observational uncertainty.

4.5 Sea ice and solid freshwater

The sea ice volume (and corresponding solid FW content) in the two simulations is nearly the same (Table 1), because both the sea ice thickness and concentration are not significantly influenced by the model resolution (Fig. 14). At 4.5 km the sea ice model starts to capture some small scale features (sea ice leads) with reasonable spatial and temporal variability (*Wang et al.*, 2016c). However, the mean sea ice thickness and concentration is not impacted by whether those small scale features are represented or not in the model. Note that much higher model resolution is required in order to simulate sea ice leads with realistic width, because they are typically narrower than 1 km in reality (*Tschudi et al.*, 2002).

The summer sea ice area along the sea ice edge on the Eurasian side is slightly overestimated in both simulations, and the simulated sea ice thickness is about half a meter thicker than the satellite observation in the last few model years (Fig. 14). Because of lacking sufficient long-term sea ice thickness observations, we compare our simulated solid FW content with the estimate from the PIOMAS Arctic sea ice volume reanalysis (*Schweiger et al.*, 2011). The simulated mean solid FW content in the period of 1980-2000 is about 20% higher than the PIOMAS estimate (Table 1).

The time series of annual mean solid FW content show that the two simulations obtain a descending trend very similar to that from the PIOMAS estimate (Fig. 12b). Compared to the mean value before 2000, the solid FW content decreases by about $4000\,\mathrm{km}^3$ averaged over the 2000s in the simulations, similar to the PIOMAS result (Table 1). Because our simulated sea ice thickness is underestimated before 2000 compared to the submarine observations (as shown in figure 13 of *Wang et al.*, 2016a) and overestimated in later years compared to satellite observations (Fig. 14a-c), the descending trend of solid FW content over the last 3 decades in our simulation and in the PIOMAS estimate as well might be lower than reality.

Arctic sea ice is mainly exported through the Fram Strait. The two simulations produced very similar solid FW transports through the Fram Strait, well representing the observed value (Table 1). Although the sea ice area export through the Fram

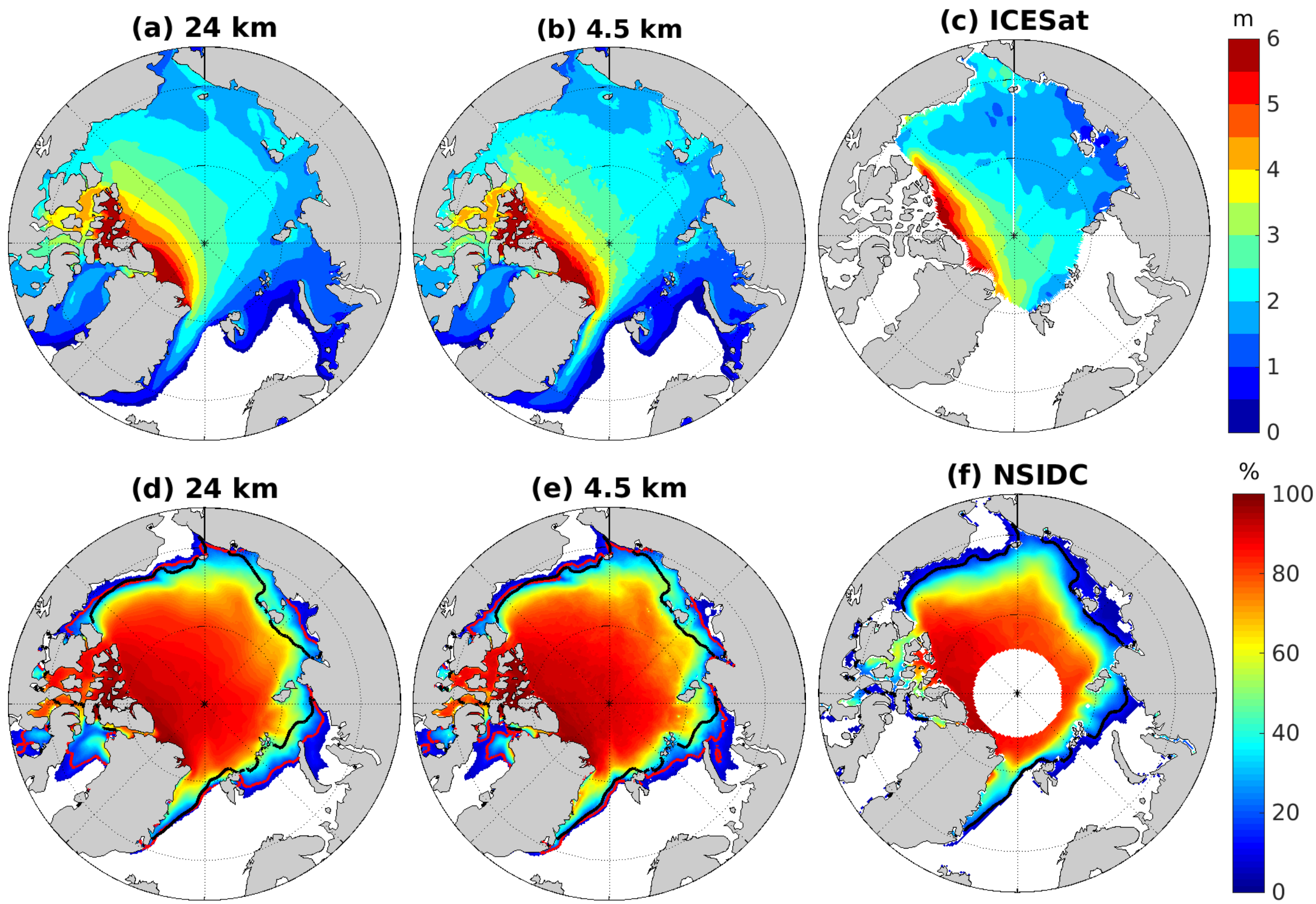

Figure 14. Spring sea ice thickness averaged from 2004 to 2007 for (a) simulation LOW, (b) HIGH, and (c) the ICESat observation (*Kwok et al.*, 2009). September sea ice concentration averaged from 1979 to 2009 for (d) simulation LOW, (e) HIGH, and (f) the NSIDC observation (*Fetterer et al.*, 2016). In (d),(e) the black curves show the 15% contour lines of the observed sea ice concentration, while the red curves show those of simulations. The periods when both observations and model results are available are chosen for calculating the means.

Strait has been increasing in recent decades due to increasing sea ice drift (e.g., *Smedsrud et al.*, 2017), sea ice volume and thus solid FW export has been decreasing due to the thinning of Arctic sea ice. Compared to the estimate of the 1980-2000 period, the solid FW export flux decreased by $400\,\mathrm{km}^3/\mathrm{yr}$ in the 2000s (*Haine et al.*, 2015). The two simulations similarly produce a decrease in the Fram Strait solid FW export of about $300\,\mathrm{km}^3/\mathrm{yr}$ between the two periods (Table 1). On interannual time scales the two simulated solid FW transports are well correlated (Table 1 and Fig. 13d). As shown in *Wang et al.* (2016a, b), ocean climate models can more consistently simulate the interannual variability of solid FW transports through Arctic gateways than the liquid FW transports.

5 Discussion

5.1 Atlantic Water

(a) Heat content and water mass sources

We found that at the end of the simulations the Arctic heat content is higher than the climatology in both simulations, but it is about 4×10^{21} J higher in simulation LOW than in HIGH. This difference in heat content requires an additional heat flux of 2 TW over 60 years. Due to inaccuracy in diagnosing heat budget terms (e.g., caused by interpolation) and missing heat diffusion terms in our model output, the mismatch between the ocean heat content changing rate and Arctic net heat flux can have the same order of magnitude as this value. Therefore it is hard to carry out analysis of closed heat budget in this and previous modeling studies (e.g., *Lique and Steele*, 2013). In the following we try to better understand the difference of ocean heat content between the two simulations by analyzing AW passive tracers.

The temperature and heat content in the AW layer is influenced by both the warm Fram Strait and the cold BSO AW branches, the latter of which joins the former mainly through the St. Anna Trough (*Schauer et al.*, 2002). Using passive tracers we can obtain the spatial distribution of the two water sources (Fig. 15). The locations of the maxima of the Fram Strait passive tracer coincide with the maxima of temperature in both basins (cf. Fig. 15a and Fig. 5a). The maxima of the Fram Strait passive tracer are located deeper in simulation LOW than in HIGH, consistent to the deepening of the AW layer shown by its temperature maxima. Below about 350 m depth the concentration of the Fram Strait passive tracer in LOW is higher in both Arctic basins than in HIGH (Fig. 15a,c). In HIGH the Fram Strait passive tracer has weaker penetration into the Canadian Basin, and a stronger cyclonic circulation inside the Eurasian Basin. At the end of 2000 the Fram Strait passive tracer averaged over the whole Arctic volume in LOW is about 14% higher than in HIGH. As the volume import of the Fram Strait AW in HIGH is larger (calculated at 79°N in the Fram Strait), a lower passive tracer storage implies that the export of Fram Strait branch AW is stronger in HIGH, either via direct recirculation north of the Fram Strait or after cyclonic circulation in the Eurasian Basin,

The BSO passive tracer indicates that cold AW (lower than 0°C) from the BSO has a lower concentration in the Eurasian Basin in simulation LOW than in HIGH, and the situation is opposite in the Canadian Basin (Fig. 15b,d). AW from both branches have replenished the Canadian Basin more intensively in simulation LOW. BSO AW has the effect to reduce the temperature of the AW layer, so the higher temperature and heat content in the Canadian Basin in LOW should be attributed to the larger amount of warm Fram Strait AW. Because the temperature of the BSO branch is similar after the atmospheric cooling over the continental shelves, the slightly lower volume transport through BSO in LOW (Table 1) has a positive contribution to the overall AW layer heat content.

(b) Future work related to simulating AW

Fram Strait is the main pathway of oceanic heat flux from the North Atlantic into the Arctic basins. It is very challenging for numerical models to simulate the complex AW circulation in the Fram Strait. In the few degree latitude band the AW loses heat due to surface cooling and starts to subduct under cold Polar Water, and a fraction of AW recirculates to the west and then

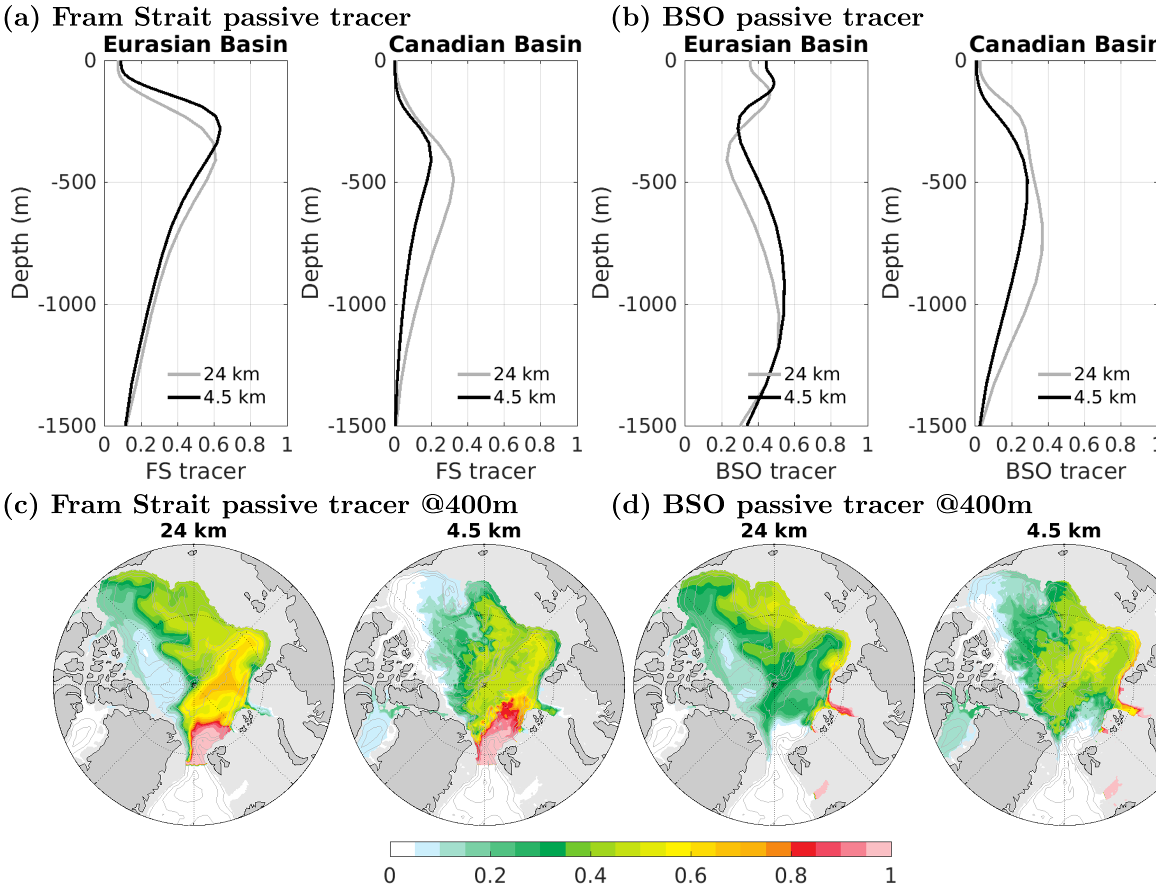

Figure 15. (a) Mean Fram Strait passive tracer concentration in the two Arctic basins averaged over the year 2000. (b) The same as (a) but for the Barents Sea Opening (BSO) passive tracer. (c) The Fram Strait passive tracer at 400 m depth averaged over the year 2000. (d) The same as (c) but for the BSO passive tracer.

southwards in different paths (*Quadfasel et al.*, 1987; *Gascard et al.*, 1988; *Saloranta and Haugan*, 2001; *Marnela et al.*, 2013; *de Steur et al.*, 2014). Strong variability associated with mesoscale eddies was observed in the Fram Strait (*von Appen et al.*, 2016), which may play an important role in setting the AW recirculation (*Hattermann et al.*, 2016). The first baroclinic Rossby radius in the Fram Strait is very small (about 2 km in winter), thus our high resolution (4.5 km grid size) simulation cannot resolve mesoscale eddies. At this resolution the simulated warm AW is confined to the strong boundary current and does not reach the central Fram Strait, presenting a cold bias in the center of the strait (*Wekerle et al.*, 2017). As in other high resolution, but not eddy-resolving models (e.g., *Fieg et al.*, 2010), our simulated AW temperature in the boundary current is too high at Fram Strait and north of Svalbard (Fig. 4a). The deficiency indicates a clear requirement for eddy resolving resolution in the Fram Strait region in order to faithfully simulate the amount and property of AW that enters the Arctic basins through the Fram Strait (as shown by *Wekerle et al.* (2017b)). In long climate simulations, however, it is hardly possible to afford 1 km model resolution in the near future. Accordingly, further effort on parameterizing mesoscale eddy effects is required to represent AW circulation in the Fram Strait.

The AW is located at intermediate depths in the Arctic Ocean and is separated from surface water and sea ice by a strong halocline. However, recent pan-Arctic microstructure measurements of turbulent kinetic energy dissipation reveal that tides can significantly enhance vertical mixing and bring up substantial heat in some areas (*Rippeth et al.*, 2015), implying an impact of AW heat on Arctic sea ice. It was shown that tides can explain a non-negligible part of the sea ice volume reduction in numerical simulations (*Luneva et al.*, 2015). Tides are not simulated in our model, so their potential impact on sea ice and AW characteristics is not explicitly considered. If tides were present in the simulations, and indeed have significant impact on heat uptaken, the influence of AW on sea ice would be different in the two simulations, because the temperature and depth of the AW layer are different between them. Dedicated studies are required to investigate such effect.

After the AW warming in the Arctic basins in the 1990s, unprecedented warming has been observed in the 2000s (*Polyakov et al.*, 2013b). However, no warming as strong as observed was obtained in the latter period in the two model simulations (Fig. 3). The AW transport calculated in the northern Fram Strait was found to decrease in the 2000s in the simulations. As the warming in 1990s is reasonably represented in the model, the discrepancy between the observed and simulated temperature variation in recent years could be attributed to model deficiency in representing ocean processes under the condition of sea ice decline, or to the quality of the atmospheric forcing data used. Furthermore, the AW layer temperature in the Arctic interior is not only determined by the amount of warm AW through the Fram Strait and cold AW from the Barents Sea, but also the circulation details of the two branches inside the basins. Research on these subjects is required in future work.

5.2 Freshwater

(a) Freshwater content drift and sea surface salinity restoring

In both simulations the Arctic liquid FW content increases rapidly during the first 20-30 years, the same as in other ocean climate models participating in the CORE-II intercomparison project analyzed by *Wang et al.* (2016b). They showed that the source of excessive FW is sea surface salinity (SSS) restoring. We repeated the low resolution simulation with SSS restoring switched off. In this simulation, the salinity in the Canadian Basin has a positive bias instead of a negative one, most pronounced at the surface (Fig. 5c). In the Eurasian Basin the salinity bias is still negative, but becomes smaller. The spin-up in this sensitivity run also takes about 20 to 30 years (Fig. 9). The FW content in Canadian Basin decreases in the spin-up phase, with a magnitude similar to that of FW content increase in the two simulations with SSS restoring (Fig. 9c). In the Eurasian Basin the FW content remains lower than in LOW by nearly a constant offset after 30 model years (Fig. 9b). The total Arctic FW content does not have a significant model drift (Fig. 9a), because the opposite drifts in the two basins largely cancel each other. In the last 30 model years, the variability of FW content in both basins in the sensitivity simulation is similar to that in simulation LOW and HIGH.

In the sensitivity simulation without SSS restoring, the salinity has a positive bias at the surface and negative bias in the lower halocline in the Canadian Basin (Fig. 5c). This implies that too much vertical mixing has taken place, which could be linked to the fact that brine rejection induced convection on very small spatial scales is neither resolved nor properly parameterized in the model: If salt rejected during ice formation is added to the ocean surface, the static instability on the model grid may initialize strong vertical mixing and weaken the vertical salinity gradient, resulting in negative salinity anomaly in the halocline and positive salinity anomaly near the ocean surface. The ocean temperature profile in this depth range is also smoothed out. This issue was discussed by, for example, *Duffy et al.* (1999) and *Nguyen et al.* (2009), who proposed to distribute rejected salt in the ocean column with some vertical distribution function, thus preventing static instability. By doing so they got significantly improved salinity profiles. We have implemented this parameterization for brine rejection in the model and are able to achieve improvement on the salinity representation in the Canadian Basin. However, it is not easy to define one particular salt vertical distribution function that can satisfy different Arctic basins and the Southern Ocean at the same time. Some research is required before we can suggest a default scheme for brine rejection in our global model simulations. The background vertical diffusivity was suggested to be one of the key parameters controlling the simulated Arctic Ocean hydrography and circulation, especially in the Canadian Basin (*Zhang and Steele*, 2007; *Nguyen et al.*, 2009). In our next model tuning phase, FESOM sensitivity to such model parameters should be carefully examined.

The salinity bias and overestimated FW content in the Eurasian Basin is very possibly caused by inaccurate representation of the pathways of upper ocean circulation (Fig. 8e,f). The Transpolar Drift carrying fresh Pacific Water and river water is located too much to the Eurasian side of the Lomonosov Ridge, and the anticyclonic surface circulation in the Canadian Basin occupies a too large spatial range compared to the observation (Fig. 10a-d). The low resolution inside CAA in simulation LOW causes more FW to release through the Fram Strait, which further increases the FW content in the Eurasian Basin. In the sensitivity simulation without SSS restoring, the SSS is still well represented in the Eurasian Basin. The Eurasian Basin salinity bias in the halocline becomes smaller in this sensitivity simulation, because the FW content is lower and the anticyclonic circulation shrinks in the Canadian Basin, with less FW penetrating into the Eurasian Basin. The upper ocean circulations are mainly driven by surface wind stress, so it is required to investigate the wind forcing fields and the impact of sea ice on the ocean surface stress in order to better understand the Eurasian Basin salinity drift.

(b) Basinwise and Beaufort Gyre freshwater content variability

In this work we have assessed the total Arctic FW content and its distribution between the Eurasian and Canadian Basins. It was found that the increase of FW storage in the Canadian Basin in recent years behaves nearly identically in different simulations (Fig. 9c). On the contrary, the trend of FW content in the Beaufort Gyre region indicates difference among the simulations (Fig. 10i-k). Recent research indicates that mesoscale eddy fluxes counteract Ekman pumping, thus playing a crucial role in Beaufort Gyre FW content variability (e.g., *Manucharyan et al.*, 2016; *Yang et al.*, 2016). In model simulations eddy parameterization (applied on coarse meshes) and the effect of implicit numerical mixing will certainly influence the dynamical balance and the Beaufort Gyre FW content. Further effort is required to investigate the model representation of Beaufort Gyre FW content and more importantly its relationship to Arctic FW release to the North Atlantic.

5.3 Unstructured-mesh modeling

The variable-resolution functionality provided by unstructured-mesh models offers new possibility in ocean modeling. One can increase model resolution locally where research interest is located, without the necessity of using traditional nesting. On the mesh the resolution can vary in space conveniently according to given functions chosen for particular applications. Many ocean process studies have been carried out taking use of FESOM in global and regional simulations, for example, with focus on overflows (*Wang et al.*, 2012), ice shelf cavities (*Timmermann et al.*, 2012), deep water formation (*Scholz et al.*, 2013), polynyas (*Haid and Timmermann*, 2013), and Arctic sea ice and ocean dynamics (*Wekerle et al.*, 2013; *Wang et al.*, 2016c; *Wekerle et al.*, 2017b). In global ocean climate simulations, the value of unstructured meshes can be more outstanding. One can design meshes with resolution varying continuously in space according to the strength of ocean variability, for example, by considering observed sea surface height variability (*Sein et al.*, 2016) and/or Rossby radius (*Sein et al.*, 2017), to permit or resolve mesoscale eddies in mid to low latitudes. It would be interesting to use this kind of global meshes together with specific mesh refinement in the Arctic Ocean for the purpose of Arctic Ocean studies, as the lower latitude ocean will be better resolved with acceptable increase of computational cost and provide more faithful oceanic linkage with the Arctic Ocean. Developing such a model configuration is aligned with our strategic plan for Arctic Ocean modeling using FESOM and the coupled climate model. It will facilitate us to study and predict not only Arctic changes, but also large scale linkage between high and lower latitudes. Towards this goal, we need to understand, for example, the impact of regional resolution in the Arctic region, using economy configurations as reported in this paper.

With an unstructured-mesh model like FESOM, one can locally increase model resolution to accurately resolve the narrow channels in the CAA and faithfully simulate the FW export (*Wekerle et al.*, 2013). However, if the finest grid size is just used in narrow straits, the model time step and the overall model throughput can be constrained by this grid size (the Courant–Friedrichs–Lewy (CFL) constraint). In ocean climate simulations, therefore, it is not preferable to design resolution in narrow straits to be much higher than the highest resolution used in large ocean basins. Table 2 shows the computational performance of the three simulations studied in this work. The number of grid points in HIGH-CAA is similar to that in LOW, but its time step is one third of LOW. The consequence is that the throughput in HIGH-CAA is three times lower than LOW and the CPU cost is three times higher. Therefore, meshes like HIGH-CAA are mainly used in process studies (one example is this work where we use it to isolate and understand the role of better simulating the CAA throughflow) [1]. Because of good scalability of FESOM (*Biastoch et al.*, 2018), simulation HIGH has a throughput similar to that of HIGH-CAA (Table 2). We usually try to use as many CPUs as possible until the computational performance (in terms of simulated years per day, SYPD) does not further increase effectively[2]. In this case, the model throughput is mainly determined by the time step.

[1] In long climate simulations we often modify the geometry of the CAA channels to allow adequate CAA throughflow, instead of locally increasing the resolution in the very small area of the CAA. However, geometry adjustment is not trivial as shown by the large model spread in CAA FW transports among the ocean climate models analyzed in *Wang et al.* (2016b). When developing global climate models, the modeling groups certainly need more efforts to better adjust the CAA representation.

[2] In our practice, how many CPUs to use can be conveniently decided by considering the number of surface grid nodes. Our recommendation is to have 250 to 350 surface grid points per CPU in FESOM 1.4 applications.

Table 2. Summary of computational cost and performance.

	Mesh size	Time step	Number of CPU used	CPU hours per model year	Throughput (SYPD)
LOW	130K(2D), 3.7M(3D)	36 min	384	340	27
HIGH-CAA	130K(2D), 3.7M(3D)	12 min	384	1040	9
HIGH	640K(2D), 14M(3D)	12 min	2400	7200	8

SYPD means simulated years per day. The computational cost is estimated in case of monthly model output, CORE-II forcing input, and simulations performed on the Cray XC40, equipped with Intel Xeon Haswell processors, of the North-German Supercomputing Alliance ('Norddeutscher Verbund zur Förderung des Hoch- und Höchstleistungsrechnens'; HLRN).

One way to overcome the drawback of meshes where only very few grid cells have increased resolution (like mesh HIGH-CAA) is to use different time steps in different parts of the mesh. This functionality is not available in FESOM yet. In most of FESOM applications, grid cells with increased resolution take a dominant share of the total number of grid cells (like mesh HIGH), therefore developing such a functionality has not been put to a high priority.

In the structured-mesh model community, global and near-global ocean models with mesoscale-eddy resolving resolutions have been developed in many groups (e.g., *Chassignet et al.*, 2009; *Storch et al.*, 2012; *Oke et al.*, 2013; *Metzger et al.*, 2014; *Dupont et al.*, 2015; *Iovino et al.*, 2016), and coupled climate models with eddy resolving ocean have also been used in practice (e.g., *Griffies et al.*, 2015). Most of the models analyzed in past CORE-II model intercomparison studies have relatively coarse resolution. For developing our unstructured-mesh model system with regional focus, it would be helpful to communicate experience with the large structured-mesh model community in future high resolution climate model intercomparison projects (for example, through the future CMIP projects where increasing model resolution will be pursued (*Haarsma et al.*, 2016)).

6 Summary

A faithful model representation of the ocean circulation, water mass property and sea ice state in the Arctic Ocean is still challenging, not only for its mean state, but also for the variability of some of the key diagnostics, in state-of-the-art ocean-sea ice models (e.g., *Jahn et al.*, 2012; *Wang et al.*, 2016a, b; *Ilicak et al.*, 2016). With the development of computing resources and model technology, high resolution Arctic Ocean modeling starts to become affordable even in ocean climate simulations. In this work we explored the impact of high horizontal resolution on the circulation of the Atlantic Water (AW) in the intermediate layer and freshwater (FW) in the upper layer of the Arctic Ocean. In particular, the mean state and variability of the AW layer and the Arctic FW budget are assessed, for which previous model intercomparison studies have provided basic knowledge on common model issues.

The simulations of the unstructured-mesh ocean-sea ice model FESOM (*Wang et al.*, 2014) with two global meshes differing in resolution in the Arctic Ocean are evaluated. The coarse resolution mesh has been used in previous CORE-II model intercomparison studies (e.g., *Griffies et al.*, 2014; *Danabasoglu et al.*, 2014). Its resolution in the Arctic Ocean is 24 km. On the high resolution mesh the Arctic resolution is increased to 4.5 km. As our intention is to provide information for developing model configurations that can be used for ocean climate simulations, a reasonably high model throughput is a prerequisite. With 4.5 km resolution in the Arctic Ocean we can run FESOM for about 8 model years per day. Using further higher resolution, though preferable for the Arctic region due to very small Rossby radius, would prevent us and groups working on other ocean climate models from carrying out long simulations at the current stage. For ocean process studies, we certainly can use the variable resolution functionality of FESOM to even better resolve local dynamics (for example, using 1 km horizontal grid size locally to resolve mesoscale eddies in the Fram Strait, *Wekerle et al.* (2017b)). This aspect of Arctic Ocean modeling is beyond the scope of this paper. As we kept the same model resolution outside the Arctic region, we are able to attribute the difference in the two simulations to the Arctic Ocean resolution. Note that we did not try to tune the two model setups separately in this paper. We used a model configuration (schemes and parameters for ocean and sea ice) similar to what has been used in the CORE-II studies (e.g., *Danabasoglu et al.*, 2014), except that eddy diffusivity is scaled by the resolution.

At 24 km resolution the simulated AW layer is unrealistically deep and thick, which currently is a common issue in coarse resolution models (*Ilicak et al.*, 2016). Such a model bias was found to be caused by numerical mixing in past AOMIP studies (*Holloway et al.*, 2007). When using 4.5 km resolution, the AW in both Arctic basins is located at the observed depth with a very reasonable thickness. The tracer advection scheme (a second order FCT scheme) used in our simulations is the one suggested for large scale applications in FESOM, because it enforces monotonicity and has decent computational cost. Idealized 2D test cases clearly indicate that numerical smoothing associated with this scheme can be significantly reduced with increasing resolution (*Wang*, 2007), which can explain the obtained improvement of the AW layer in the high resolution simulation. As we kept the vertical resolution the same in the two simulations, which needs separate investigation, the reduction in numerical mixing is only due to the change in horizontal resolution.

With higher resolution the cyclonic AW boundary current becomes narrower and more energetic. Moreover, the topographic steering on the current is stronger, causing more AW to recirculate along the Lomonosov Ridge in the Eurasian Basin. The resulting constrained penetration of AW into the Canadian Basin in the high resolution simulation helps to eliminate the intensive warming and deepening trend of the AW in the Canadian Basin present in the coarse resolution setup. More AW recirculates in the Eurasian Basin and leaves the Arctic Ocean, which can partly explain that the increase in Arctic heat content is much lower in the high resolution simulation. The strength of interannual and seasonal variability of AW temperature, especially in the boundary current along the continental slope and Lomonosov Ridge, becomes significantly higher with increasing resolution.

The impact of horizontal resolution on ocean surface circulation and FW cycle is limited to the spatial pattern of liquid FW content and pathways of different water masses. It mainly stems from the difference in the representation of the Canadian Arctic Archipelago (CAA) channels, not the resolution in the Arctic basins. The CAA channels are often treated very differently in different ocean climate models, for example, for the number of CAA channels and number of active grid points across the channels, as shown in the model intercomparison study by *Wang et al.* (2016b). They found that the spread in simulated CAA and Fram Strait FW transports is considerably large. Therefore, inspecting and tuning CAA representation is one of the important tasks in future development of ocean climate models.

The mean state and variability of total and basinwise liquid FW content, the variability of liquid FW transports through Arctic gateways and the characteristics of Arctic sea ice volume and export do not change significantly with increasing resolution.

The recent upward trend of FW content in the Beaufort Gyre shows some sensitivity to the resolution inside the Arctic basin. How well mesoscale eddies are resolved in the Canadian Basin in the high resolution simulation, and how realistic the effect of eddies is parameterized in the low resolution simulation, both need to be assessed in the context of the interplay with Ekman pumping in future studies. Here it is important to note that the variability of both the Arctic FW storage and release to North Atlantic is insensitive to the model resolution applied in our simulations. We also found that only better resolving the CAA channels (in the simulation where only the CAA is resolved with 4.5 km) did not significantly impact the representation of the AW layer.

Besides identifying the impact of horizontal resolution on the Arctic Ocean circulation, we also discussed scientific questions and model issues that need to be explored in future work, and some of the illustrated model issues are common in many other ocean-sea ice models. Overall, increasing model resolution does considerably improve the performance of the Arctic Ocean simulation, while further efforts are necessary to solve remaining issues that are not linked to applied model resolution, and to develop/improve parameterizations that are still required even with best resolution affordable now.

Code and data availability. FESOM v1.4 can be downloaded from https://swrepo1.awi.de/projects/fesom after registration. For the sake of the journal requirement, the configuration used, together with the mesh information, is archived at doi.org/10.5281/zenodo.1116851. Mesh partitioning in FESOM is based on a METIS Version 5.1.0 package developed at the Department of Computer Science and Engineering at the University of Minnesota (http://glaros.dtc.umn.edu/gkhome/views/metis). METIS and the solver pARMS (*Li et al.*, 2003) present separate libraries which are freely available subject to their licenses. The Polar Science Center Hydrographic Climatology (*Steele et al.*, 2001) used for model initialization and the CORE-II atmospheric forcing data (*Large and Yeager*, 2009) are freely available online. The simulation results can be obtained from the authors upon request.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The public availability of different observational data sets and reanalysis data used in this work is a great help for model development, so the efforts of respective working groups are appreciated. We would like to thank I. Polyakov and B. Rabe for providing us their data compiled from large data sets. Q. Wang is funded by the Helmholtz Climate Initiative REKLIM (Regional Climate Change) project. C. Wekerle is funded by the FRontiers in Arctic marine Monitoring program (FRAM). The model simulations were performed at the North-German Supercomputing Alliance (HLRN). We thank the two reviewers and the editor for their helpful comments.

References

Aagaard, K., and E. C. Carmack (1989), The role of sea ice and other fresh-water in the Arctic circulation, *J. Geophys. Res.*, *94*, 14,485–14,498.

Aagaard, K., J. H. Swift, and E. Carmack (1985), Thermohaline circulation in the Arctic mediterranean seas, *Journal of Geophysical Research-oceans*, *90*, 4833–4846.

Aksenov, Y., V. V. Ivanov, A. J. G. Nurser, S. Bacon, I. V. Polyakov, A. C. Coward, A. C. Naveira-Garabato, and A. Beszczynska-Moeller (2011), The Arctic Circumpolar Boundary Current, *Journal of Geophysical Research-oceans*, *116*, C09,017.

Aksenov, Y., M. Karcher, A. Proshutinsky, R. Gerdes, B. de Cuevas, E. Golubeva, F. Kauker, A. T. Nguyen, G. A. Platov, M. Wadley, E. Watanabe, A. C. Coward, and A. J. G. Nurser (2016), Arctic pathways of Pacific Water: Arctic Ocean model intercomparison experiments, *J. Geophys. Res. - Oceans*, *121*, 27–59.

Arrigo, K. R., and G. L. van Dijken (2015), Continued increases in Arctic Ocean primary production, *Progress in Oceanography*, *136*, 60–70.

Arzel, O., T. Fichefet, H. Goosse, and J.-L. Dufresne (2008), Causes and impacts of changes in the Arctic freshwater budget during the 20th and 21st centuries in an AOGCM, *Climate Dynamics*, *30*, 37–58.

Beszczynska-Moeller, A., E. Fahrbach, U. Schauer, and E. Hansen (2012), Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997-2010, *ICES J. Mar. Science*, *69*, 852–863.

Biastoch, A., D. Sein, J.V. Durgadoo, Q. Wang, and S. Danilov (2018), Simulating the Agulhas system in global ocean models – nesting vs. multi-resolution unstructured meshes, *Ocean Modelling*, *in press*, DOI:10.1016/j.ocemod.2017.12.002.

Carmack, E. C., M. Yamamoto-Kawai, T. W. N. Haine, S. Bacon, B. A. Bluhm, C. Lique, H. Melling, I. V. Polyakov, F. Straneo, M.-L. Timmermans, and W. J. Williams (2016), Freshwater and its role in the Arctic Marine System: Sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans, *J. Geophys. Res. Biogeosci.*, *121*, 675–717.

Cavalieri, D. J., and C. L. Parkinson (2012), Arctic sea ice variability and trends, 1979-2010, *Cryosphere*, *6*, 881–889.

Chassignet, E. P., and others (2009), U.S. GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM), *Oceanography*, *22*, 64–75.

Comiso, J. C. (2012), Large decadal decline in the Arctic multiyear ice cover, *J. Clim.*, *25*, 1176–1193.

Curry, B., C. M. Lee, B. Petrie, R. E. Moritz, and R. Kwok (2014), Multiyear volume, liquid freshwater, and sea ice transports through Davis Strait, 2004-2010, *Journal of Physical Oceanography*, *44*, 1244–1266.

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman (2009), Changes in continental freshwater discharge from 1948 to 2004, *J. Climate*, *22*, 2773–2792.

Danabasoglu, G., S. G. Yeager, D. Bailey, E. Behrens, M. Bentsen, D. Bi, A. Biastoch, C. Böning, A. Bozec, V. M. Canuto, C. Cassou, E. Chassignet, A. C. Coward, S. Danilov, N. Diansky, H. Drange, R. Farneti, E. Fernandez, P. G. Fogli, G. Forget, Y. Fujii, S. M. Griffies, A. Gusev, P. Heimbach, A. Howard, T. Jung, M. Kelley, W. G. Large, A. Leboissetier, J. Lu, G. Madec, S. J. Marsland, S. Masina, A. Navarra, A. G. Nurser, A. Pirani, D. S. y Melia, B. L. Samuels, M. Scheinert, D. Sidorenko, A.-M. Treguier, H. Tsujino, P. Uotila, S. Valcke, A. Voldoire, and Q. Wang (2014), North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase {II} (CORE-II). part i: Mean states, *Ocean Modelling*, *73*, 76 – 107.

Danilov, S., G. Kivman, and J. Schroeter (2004), A finite-element ocean model: principles and evaluation, *Ocean Model.*, *6*, 125–150.

Danilov, S., Q. Wang, R. Timmermann, N. Iakovlev, D. Sidorenko, M. Kimmritz, T. Jung, and J. Schroeter (2015), Finite-Element Sea Ice Model (FESIM), version 2, *Geoscientific Model Development*, *8*, 1747–1761.

de Steur, L., E. Hansen, C. Mauritzen, A. Beszczynska-Möller, and E. Fahrbach (2014), Impact of recirculation on the East Greenland Current in Fram Strait: Results from moored current meter measurements between 1997 and 2009, *Deep Sea Research Part I: Oceanographic Research Papers*, *92*, 26 – 40.

Dickson, R., B. Rudels, S. Dye, M. Karcher, J. Meincke, and I. Yashayaev (2007), Current estimates of freshwater flux through Arctic and subarctic seas, *Prog. Oceanogr.*, *73*, 210–230.

Dmitrenko, I. A., I. V. Polyakov, S. A. Kirillov, L. A. Timokhov, I. E. Frolov, V. T. Sokolov, H. L. Simmons, V. V. Ivanov, and D. Walsh (2008), Toward a warmer arctic ocean: Spreading of the early 21st century Atlantic Water warm anomaly along the Eurasian Basin margins, *J. Geophys. Res.-Oceans*, *113*, C05,023.

Dmitrenko, I. A., V. V. Ivanov, S. A. Kirillov, E. L. Vinogradova, S. Torres-Valdes, and D. Bauch (2011), Properties of the Atlantic derived halocline waters over the Laptev Sea continental margin: Evidence from 2002 to 2009, *J. Geophys. Res. - Oceans*, *116*, C10,024.

Dmitrenko, I. A., B. Rudels, S. A. Kirillov, Y. O. Aksenov, V. S. Lien, V. V. Ivanov, U. Schauer, I. V. Polyakov, A. Coward, and D. G. Barber (2015), Atlantic water flow into the Arctic Ocean through the St. Anna Trough in the northern Kara Sea, *J. Geophys. Res. - Oceans*, *120*, 5158–5178.

Duffy, P. B., M. Eby, and A. J. Weaver (1999), Effects of sinking of salt rejected during formation of sea ice on results of an ocean-atmosphere-sea ice climate model, *Geophysical Research Letters*, *26*, 1739–1742.

Dupont, F., S. Higginson, R. Bourdallé-Badie, Y. Lu, F. Roy, G. Smith, J.-F. Lemieux, G. Garric, and F. Davidson (2015), A high-resolution ocean and sea-ice modelling system for the Arctic and North Atlantic oceans, *Geosci. Model Dev.*, *8*, 1577–1594.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2016), Sea Ice Index. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media, p. updated daily.

Fieg, K., R. Gerdes, E. Fahrbach, A. Beszczynska-Möller, and U. Schauer (2010), Simulation of oceanic volume transports through Fram Strait 1995–2005, *Ocean Dynamics*, *60*, 491–502.

Gascard, J.-C., C. Kergomard, P.-F. Jeannin, and M. Fily (1988), Diagnostic study of the Fram Strait marginal ice zone during summer from 1983 and 1984 Marginal Ice Zone Experiment Lagrangian observations, *J. Geophys. Res. Oceans*, *93*, 3613–3641.

Gent, P. R., and J. C. McWilliams (1990), Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, *20*, 150–155.

Gerdes, R., M. J. Karcher, F. Kauker, and U. Schauer (2003), Causes and development of repeated Arctic Ocean warming events, *Geophysical Research Letters*, *30*, 1980.

Giles, K. A., S. W. Laxon, A. L. Ridout, D. J. Wingham, and S. Bacon (2012), Western Arctic Ocean freshwater storage increased by wind-driven spin-up of the Beaufort Gyre, *Nature Geoscience*, *5*, 194–197.

Goosse, H., T. Fichefet, and J. M. Campin (1997), The effects of the water flow through the Canadian Archipelago in a global ice-ocean model, *Geophysical Research Letters*, *24*, 1507–1510.

Griffies, S., M. Winton, B. Samuels, G. Danabasoglu, S. Yeager, S. Marlsand, H. Drange, and M. Bentsen (2012), Datasets and protocol for the CLIVAR WGOMD Coordinated Ocean-sea ice Reference Experiments (COREs), *Tech. Rep. 21*, WCRP Report.

Griffies, S. M. (1998), The Gent-Mcwilliams skew flux, *J. Phys. Oceanogr.*, *28*, 831–841.

Griffies, S. M., A. Biastoch, C. Böning, F. Bryan, G. Danabasoglu, E. P. Chassignet, M. H. England, R. Gerdes, H. Haak, R. W. Hallberg, W. Hazeleger, J. Jungclaus, W. G. Large, G. Madec, A. Pirani, B. L. Samuels, M. Scheinert, A. S. Gupta, C. A. Severijns, H. L. Simmons, A. M. Treguier, M. Winton, S. Yeager, and J. Yin (2009), Coordinated Ocean-ice Reference Experiments (COREs), *Ocean Modell.*, *26*, 1–46.

Griffies, S. M., J. Yin, P. J. Durack, P. Goddard, S. C. Bates, E. Behrens, M. Bentsen, D. Bi, A. Biastoch, C. W. Boening, A. Bozec, E. Chassignet, G. Danabasoglu, S. Danilov, C. M. Domingues, H. Drange, R. Farneti, E. Fernandez, R. J. Greatbatch, D. M. Holland, M. Ilicak, W. G. Large, K. Lorbacher, J. Lu, S. J. Marsland, A. Mishra, A. J. G. Nurser, D. Salas y Melia, J. B. Palter, B. L. Samuels, J. Schroeter, F. U. Schwarzkopf, D. Sidorenko, A. M. Treguier, Y.-h. Tseng, H. Tsujino, P. Uotila, S. Valcke, A. Voldoire, Q. Wang, M. Winton, and X. Zhang (2014), An assessment of global and regional sea level for years 1993-2007 in a suite of interannual CORE-II simulations, *Ocean Modelling*, *78*, 35–89.

Griffies, S.M., and others (2015), Impacts on Ocean Heat from Transient Mesoscale Eddies in a Hierarchy of Climate Models, *Journal of Climate*, *28*, 952–977.

Haid, V., and R. Timmermann (2013), Simulated heat flux and sea ice production at coastal polynyas in the southwestern Weddell Sea, *J. Geophys. Res. - Oceans*, *118*, 2640–2652.

Haine, T., B. Curry, R. Gerdes, E. Hansen, M. Karcher, C. Lee, B. Rudels, G. Spreen, L. de Steur, K. Stewart, and R. Woodgate (2015), Arctic freshwater export: Status, mechanisms, and prospects, *Global and Planetary Change*, *125*, 13–35.

Hakkinen, S. (1999), A simulation of thermohaline effects of a great salinity anomaly, *Journal of Climate*, *12*, 1781–1795.

Haarsma, R. J., M. Roberts, P. Vidale, C. Senior, A. Bellucci, Q. Bao, P. Chang, S. Corti, N. Fučkar, V. Guemas, J. von Hardenberg, W. Hazeleger, C. Kodama, T. Koenigk, L. Leung, J. Lu, J. Luo, J. Mao, M. Mizielinski, R. Mizuta, P. Nobre, M. Satoh, E. Scoccimarro, T. Semmler, J. Small, J. and J.-S. von Storch (2016), High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geosci. Model Dev.*, *9*, 4185–4208.

Hattermann, T., P. E. Isachsen, W.-J. von Appen, J. Albretsen, and A. Sundfjord (2016), Eddy-driven recirculation of Atlantic Water in Fram Strait, *Geophys. Res. Lett.*, *43*, 3406–3414.

Holland, M. M., M. B. Cecilia, M. Eby, and A. J. Weavr (2001), The role of ice–ocean interactions in the variability of the North Atlantic thermohaline circulation, *J. Clim.*, *14*, 656–675.

Holloway, G., F. Dupont, E. Golubeva, S. Haekkinen, E. Hunke, M. Jin, M. Karcher, F. Kauker, M. Maltrud, M. A. M. Maqueda, W. Maslowski, G. Platov, D. Stark, M. Steele, T. Suzuki, J. Wang, and J. Zhang (2007), Water properties and circulation in Arctic Ocean models, *Journal of Geophysical Research-oceans*, *112*, C04S03.

Hunke, E., and J. Dukowicz (1997), An elastic-viscous-plastic model for sea ice dynamics, *J. Phys. Oceanogr.*, *27*, 1849–1867.

Ilicak, M., H. Drange, Q. Wang, R. Gerdes, Y. Aksenov, D. Bailey, M. Bentsen, A. Biastoch, A. Bozec, C. Böning, C. Cassou, E. Chassignet, A. C. Coward, B. Curry, G. Danabasoglu, S. Danilov, E. Fernandez, P. G. Fogli, Y. Fujii, S. M. Griffies, D. Iovino, A. Jahn, T. Jung, W. G. Large, C. Lee, C. Lique, J. Lu, S. Masina, A. J. George Nurser, C. Roth, D. Salas y Mélia, B. L. Samuels, P. Spence, H. Tsujino, S. Valcke, A. Voldoire, X. Wang, and S. G. Yeager (2016), An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part III: Hydrography and fluxes, *Ocean Modelling*, *100*, 141–161.

Iovino, D., and others (2016), A 1/16^o eddying simulation of the global NEMO sea-ice-ocean system, *Geosci. Model Dev.*,, *9*, 2665–2684.

Jahn, A., and M. M. Holland (2013), Implications of arctic sea ice changes for North Atlantic deep convection and the meridional overturning circulation in CCSM4-CMIP5 simulations, *Geophys. Res. Lett.*, *40*, 1206–1211.

Jahn, A., Y. Aksenov, B. A. de Cuevas, L. de Steur, S. Hakkinen, E. Hansen, C. Herbaut, M. . N. Houssais, M. Karcher, F. Kauker, C. Lique, A. Nguyen, P. Pemberton, D. Worthen, and J. Zhang (2012), Arctic Ocean freshwater: How robust are model simulations?, *J. Geophys. Res. - Oceans*, *117*, C00D16.

Jakobsson, M., R. Macnab, L. Mayer, R. Anderson, M. Edwards, J. Hatzky, H. W. Schenke, and P. Johnson (2008), An improved bathymetric portrayal of the Arctic Ocean: Implications for ocean modeling and geological, geophysical and oceanographic analyses, *Geophys. Res. Lett.*, *35*, L07,602.

Johnson, M., A. Proshutinsky, Y. Aksenov, A. T. Nguyen, R. Lindsay, C. Haas, J. Zhang, N. Diansky, R. Kwok, W. Maslowski, S. Haekkinen, I. Ashik, and B. de Cuevas (2012), Evaluation of Arctic sea ice thickness simulated by Arctic Ocean Model Intercomparison Project models, *Journal of Geophysical Research-oceans*, *117*, C00D13.

Jungclaus, J. H., H. Haak, M. Latif, and U. Mikolajewicz (2005), Arctic-North Atlantic interactions and multidecadal variability of the meridional overturning circulation, *Journal of Climate*, *18*, 4013–4031.

Karcher, M., F. Kauker, R. Gerdes, E. Hunke, and J. Zhang (2007), On the dynamics of Atlantic Water circulation in the Arctic Ocean, *J. Geophys. Res. - Oceans*, *112*, C04S02.

Karcher, M., A. Beszczynska-Moeller, F. Kauker, R. Gerdes, S. Heyen, B. Rudels, and U. Schauer (2011), Arctic Ocean warming and its consequences for the Denmark Strait overflow, *Journal of Geophysical Research-oceans*, *116*, C02,037.

Karcher, M. J., and J. M. Oberhuber (2002), Pathways and modification of the upper and intermediate waters of the Arctic Ocean, *Journal of Geophysical Research-oceans*, *107*, 3049.

Karcher, M. J., R. Gerdes, F. Kauker, and C. Koberle (2003), Arctic warming: Evolution and spreading of the 1990s warm event in the Nordic seas and the Arctic Ocean, *Journal of Geophysical Research-oceans*, *108*, 3034.

Kawasaki, T., and H. Hasumi (2015), The inflow of Atlantic water at the Fram Strait and its interannual variability, *J. Geophys. Res. Oceans*, *121*, 502–519.

Koldunov, N. V., N. Serra, A. Köhl, D. Stammer, O. Henry, A. Cazenave, P. Prandi, P. Knudsen, O. B. Andersen, Y. Gao, and J. Johannessen (2014), Multimodel simulations of Arctic Ocean sea surface height variability in the period 1970–2009, *Journal of Geophysical Research: Oceans*, *119*(12), 8936–8954.

Kwok, R., G. F. Cunningham, M. Wensnahan, I. Rigor, H. J. Zwally, and D. Yi (2009), Thinning and volume loss of the Arctic Ocean sea ice cover: 2003-2008, *Journal of Geophysical Research-oceans*, *114*, C07,005.

Large, W. G., and S. G. Yeager (2009), The global climatology of an interannually varying air-sea flux data set, *Climate Dynamics*, *33*, 341–364.

Large, W. G., J. C. Mcwilliams, and S. C. Doney (1994), Oceanic vertical mixing - a review and a model with a nonlocal boundary-layer parameterization, *Reviews of Geophysics*, *32*, 363–403.

Laxon, S. W., K. A. Giles, A. L. Ridout, D. J. Wingham, R. Willatt, R. Cullen, R. Kwok, A. Schweiger, J. Zhang, C. Haas, S. Hendricks, R. Krishfield, N. Kurtz, S. Farrell, and M. Davidson (2013), CryoSat-2 estimates of Arctic sea ice thickness and volume, *Geophys. Res. Lett.*, *40*, 732–737.

Li, Z., Y. Saad, and M. Sosonkina (2003), pARMS: a parallel version of the algebraic recursive multilevel solver, *Numer. Linear Algebr.*, *10*, 485–509.

Lique, C., and M. Steele (2012), Where can we find a seasonal cycle of the Atlantic water temperature within the Arctic Basin?, *J. Geophys. Res. - Oceans*, *117*, C03,026.

Lique, C., and M. Steele (2013), Seasonal to decadal variability of Arctic Ocean heat content: A model-based analysis and implications for autonomous observing systems, *J. Geophys. Res. - Oceans*, *118*, 1673–1695.

Löhner, R., K. Morgan, J. Peraire, and M. Vahdati (1987), Finite-element flux-corrected transport (FEM-FCT) for the Euler and Navier-Stokes equations, *Internat. J. Numer. Methods Fluids*, *7*, 1093–1109.

Luneva, M. V., Y. Aksenov, J. D. Harle, and J. T. Holt (2015), The effects of tides on the water mass mixing and sea ice in the Arctic Ocean, *J. Geophys. Res. - Oceans*, *120*, 6669–6699.

Manucharyan, G. E., M. A. Spall, and A. F. Thompson (2016), A theory of the wind-driven Beaufort Gyre variability, *Journal of Physical Oceanography*, *46*, 3263–3278.

Marnela, M., B. Rudels, M.-N. Houssais, A. Beszczynska-Möller, and P. B. Eriksson (2013), Recirculation in the Fram Strait and transports of water in and north of the Fram Strait derived from CTD data, *Ocean Sci.*, *8*, 499–519.

Maslowski, W., D. Marble, W. Walczowski, U. Schauer, J. L. Clement, and A. J. Semtner (2004), On climatological mass, heat, and salt transports through the Barents sea and Fram strait from a pan-Arctic coupled ice-ocean model simulation, *J. Geophys. Res. - Oceans*, *109*, C03,032.

McPhee, M. G., A. Proshutinsky, J. H. Morison, M. Steele, and M. B. Alkire (2009), Rapid change in freshwater content of the Arctic Ocean, *Geophysical Research Letters*, *36*, L10,602.

Melling, H. (2000), Exchanges of freshwater through the shallow straits of the North American Arctic, in *The Freshwater Budget of the Arctic Ocean*, edited by E. L. e. a. Lewis, pp. 479–502, Springer, New York.

Metzger, E. J., and others (2014), US Navy operational global ocean and Arctic ice prediction systems, *Oceanography*, *27*, 32–43.

Morison, J., R. Kwok, C. Peralta-Ferriz, M. Alkire, I. Rigor, R. Andersen, and M. Steele (2012), Changing Arctic Ocean freshwater pathways, *Nature*, *481*, 66–70.

Nguyen, A. T., D. Menemenlis, and R. Kwok (2009), Improved modeling of the Arctic halocline with a subgrid-scale brine rejection parameterization, *J. Geophys. Res. - Oceans*, *114*, C11,014.

Nurser, A. J. G., and S. Bacon (2014), The Rossby radius in the Arctic Ocean, *Ocean Sci.*, *10*, 967–975.

Oke, P.R., and others (2013), Evaluation of a near-global eddy-resolving ocean model, *Geosci. Model Dev.*, *6*, 591–615.

Orvik, K. A., and P. Niiler (2002), Major pathways of Atlantic water in the northern North Atlantic and Nordic Seas toward Arctic, *Geophys. Res. Lett.*, *29*, 1896.

Parkinson, C., and W. Washington (1979), A large-scale numerical model of sea ice, *J. Geophys. Res. - Oceans*, *84*, 311–337.

Polyakov, I., L. Timokhov, V. Alexeev, S. Bacon, I. Dmitrenko, L. Fortier, I. Frolov, J. Gascard, E. Hansen, V. Ivanov, S. Laxon, C. Mauritzen, D. Perovich, K. Shimada, H. Simmons, V. Sokolov, M. Steele, and J. Toole (2010), Arctic ocean warming contributes to reduced polar ice cap, *Journal of Physical Oceanography*, *40*(12), 2743–2756.

Polyakov, I. V., A. V. Pnyushkov, and L. A. Timokhov (2012), Warming of the intermediate Atlantic Water of the Arctic Ocean in the 2000s, *Journal of Climate*, *25*, 8362–8370.

Polyakov, I. V., A. V. Pnyushkov, R. Rember, L. Padman, E. C. Carmack, and J. M. Jackson (2013a), Winter convection transports Atlantic Water heat to the surface layer in the eastern Arctic Ocean, *Journal of Physical Oceanography*, *43*, 2142–2155.

Polyakov, I. V., U. S. Bhatt, J. E. Walsh, E. P. Abrahamsen, A. V. Pnyushkov, and P. F. Wassmann (2013b), Recent oceanic changes in the Arctic in the context of long-term observations, *Ecological Applications*, *23*, 1745–1764.

Proshutinsky, A., and Z. Kowalik (2007), Preface to special section on Arctic Ocean Model Intercomparison Project (AOMIP) studies and results, *J. Geophys. Res. - Oceans*, *112*, C04S01.

Proshutinsky, A., R. H. Bourke, and F. A. McLaughlin (2002), The role of the Beaufort Gyre in Arctic climate variability: Seasonal to decadal climate scales, *Geophysical Research Letters*, *29*, 2100.

Proshutinsky, A., R. Krishfield, M.-L. Timmermans, J. Toole, E. Carmack, F. McLaughlin, W. J. Williams, S. Zimmermann, M. Itoh, and K. Shimada (2009), Beaufort Gyre freshwater reservoir: State and variability from observations, *Journal of Geophysical Research-oceans*, *114*, C00A10.

Proshutinsky, A., Y. Aksenov, J. C. Kinney, R. Gerdes, E. Golubeva, D. Holland, G. Holloway, A. Jahn, M. Johnson, E. Popova, M. Steele, and E. Watanabe (2011), Recent Advances in Arctic Ocean Studies Employing Models from the Arctic Ocean Model Intercomparison Project, *Oceanography*, *24*, 102–113.

Quadfasel, D., J.-C. Gascard, and K.-P. Koltermann (1987), Large-scale oceanography in Fram Strait during the 1984 Marginal Ice Zone Experiment, *J. Geophys. Res. Oceans*, *92*, 6719–6728.

Rabe, B., M. Karcher, U. Schauer, J. M. Toole, R. A. Krishfield, S. Pisarev, F. Kauker, R. Gerdes, and T. Kikuchi (2011), Assessment of Arctic Ocean freshwater content changes from the 1990s to the 2006-2008 period, *Deep-sea Research Part I-oceanographic Research Papers*, *58*, 173–185.

Rabe, B., M. Karcher, F. Kauker, U. Schauer, J. M. Toole, R. A. Krishfield, S. Pisarev, T. Kikuchi, and J. Su (2014), Arctic ocean basin liquid freshwater storage trend 1992-2012, *Geophysical Research Letters*, *41*, 961–968.

Redi, M. H. (1982), Oceanic isopycnal mixing by coordinate rotation, *J. Phys. Oceanogr.*, *12*, 1154–1158.

Rippeth, T. P., B. J. Lincoln, Y.-D. Lenn, J. A. M. Green, A. Sundfjord, and S. Bacon (2015), Tide-mediated warming of Arctic halocline by Atlantic heat fluxes over rough topography, *Nature Geoscience*, *8*, 191–194.

Roach, A. T., K. Aagaard, C. Pease, S. A. Salo, T. Weingartner, V. Pavlov, and M. Kulakov (1995), Direct measurements of transport and water properties through the Bering Strait, *Journal of Geophysical Research-Oceans*, *100*, 18,443–18,457.

Rudels, B. (2015), Arctic Ocean circulation, processes and water masses: A description of observations and ideas with focus on the period prior to the International Polar Year 2007–2009, *Progress in Oceanography*, *132*, 22–67.

Rudels, B., and H. Friedrich (2000), The transformation of the Atlantic Water in the Arctic Ocean and their significance for the freshwaer budget, in *The Freshater Budget of the Arctic Ocean*, edited by E. L. e. a. Lewis, pp. 503–532, Kluwer.

Rudels, B., E. P. Jones, L. G. Anderson, and G. Kattner (1994), On the intermediate depth waters of the Arctic Ocean, in *The Polar Oceans and Their Role in Shaping the Global Environment*, edited by O. M. Johannessen, R. D. Muench, and J. E. Overland, pp. 33–46, American Geophysical Union.

Rudels, B., L. G. Anderson, and E. P. Jones (1996), Formation and evolution of the surface mixed layer and halocline of the Arctic Ocean, *J. Geophys. Res.-Oceans*, *101*, 8807–8821.

Saloranta, T. M., and P. M. Haugan (2001), Interannual variability in the hydrography of Atlantic water northwest of Svalbard, *J. Geophys. Res. - Oceans*, *106*, 13,931–13,943.

Schauer, U., B. Rudels, E. P. Jones, L. G. Anderson, R. D. Muench, G. Bjork, J. H. Swift, V. Ivanov, and A. M. Larsson (2002), Confluence and redistribution of Atlantic Water in the Nansen, Amundsen and Makarov Basins, *Annales Geophysicae*, *20*, 257–273.

Schauer, U., A. Beszczynska-Moeller, W. Walczowski, E. Fahrbach, J. Piechura, and E. Hansen (2008), Variation of measured heat flow through the Fram Strait between 1997 and 2006, in *Arctic-Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate*, edited by R. e. a. Dickson, pp. 65–85, Springer.

Scholz, P., G. Lohmann, Q. Wang, and S. Danilov (2013), Evaluation of a finite-element sea-ice ocean model (fesom) set-up to study the interannual to decadal variability in the deep-water formation rates, *Ocean Dynamics*, *63*, 347–370.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok (2011), Uncertainty in modeled Arctic sea ice volume, *Journal of Geophysical Research-oceans*, *116*, C00D06.

Sein, D. V., S. Danilov, A. Biastoch, J. V. Durgadoo, D. Sidorenko, S. Harig, and Q. Wang (2016), Designing variable ocean model resolution based on the observed ocean variability, *Journal of Advances in Modeling Earth Systems*, *8*, 904–916.

Sein, D.V., N.V. Koldunov, S. Danilov, Q. Wang, D. Sidorenko, I. Fast, T. Rackow, W. Cabos, and T. Jung (2017), Ocean modeling on a mesh with resolution following the local Rossby radius, *Journal of Advances in Modeling Earth Systems*, *9*, 2601–2614.

Serreze, M. C., and R. G. Barry (2011), Processes and impacts of Arctic amplification: A research synthesis, *Global and Planetary Change*, *77*, 85–96.

Serreze, M. C., A. P. Barrett, A. G. Slater, R. A. Woodgate, K. Aagaard, R. B. Lammers, M. Steele, R. Moritz, M. Meredith, and C. M. Lee (2006), The large-scale freshwater cycle of the Arctic, *J. Geophys. Res. - Oceans*, *111*, C11,010.

Skagseth, T., O.and Furevik, R. Ingvaldsen, H. L. K. Mork, K. Orvik, and V. Ozhigi (2008), Volume and heat transports to the Arctic Ocean via the Norwegian and Barents Seas, in *Arctic-Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate*, edited by R. e. a. Dickson, pp. 45–64, Springer.

Smagorinsky, J. (1963), General circulation experiments with the primitive equations: I. the basic experiment, *Mon. Wea. Rev.*, *91*, 99–164.

Smedsrud, L. H., R. Ingvaldsen, J. E. O. Nilsen, and O. Skagseth (2010), Heat in the Barents Sea: transport, storage, and surface fluxes, *Ocean Science*, *6*(0), 219–234.

Smedsrud, L. H., I. Esau, R. B. Ingvaldsen, T. Eldevik, P. M. Haugan, C. Li, V. S. Lien, A. Olsen, A. M. Omar, O. H. Ottera, B. Risebrobakken, A. B. Sando, V. A. Semenov, and S. A. Sorokina (2013), The role of the Barents Sea in the Arctic climate system, *Reviews of Geophysics*, *51*, 415–449.

Smedsrud, L. H., M. H. Halvorsen, J. C. Stroeve, R. Zhang, and K. Kloster (2017), Fram Strait sea ice export variability and September Arctic sea ice extent over the last 80 years, *The Cryosphere*, *11*, 65–79.

Steele, M., R. Morley, and W. Ermold (2001), PHC: A global ocean hydrography with a high quality Arctic Ocean, *J. Climate*, *14*, 2079–2087.

Storch, J.-S. von, and others (2012), An Estimate of the Lorenz Energy Cycle for the World Ocean Based on the STORM/NCEP Simulation, *Journal of Physical Oceanography*, *42*, 2185–2205.

Stroeve, J. C., V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier (2012), Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophysical Research Letters*, *39*, L16,502.

Timmermann, R., S. Danilov, J. Schröter, C. Böning, D. Sidorenko, and K. Rollenhagen (2009), Ocean circulation and sea ice distribution in a finite element global sea ice-ocean model, *Ocean Modell.*, *27*, 114–129.

Timmermann, R., Q. Wang, and H. Hellmer (2012), Ice-shelf basal melting in a global finite-element sea-ice/ice-shelf/ocean model, *Annals of Glaciology*, *53*, 303–314.

Tremblay, J.-É., L. G. Anderson, P. Matrai, P. Coupel, S. Bélanger, C. Michel, and M. Reigstad (2015), Global and regional drivers of nutrient supply, primary production and CO2 drawdown in the changing Arctic Ocean, *Progress in Oceanography*, *139*, 171–196.

Tschudi, M. A., J. A. Curry, and J. A. Maslanik (2002), Characterization of springtime leads in the Beaufort/Chukchi Seas from airborne and satellite observations during FIRE/SHEBA, *J. Geophys. Res. - Oceans*, *107*, 8034.

Vihma, T. (2014), Effects of Arctic sea ice decline on weather and climate: A review, *Surveys in Geophysics*, *35*, 1175–1214.

von Appen, W.-J., U. Schauer, T. Hattermann, and A. Beszczynska-Möller (2016), Seasonal cycle of mesoscale instability of the West Spitsbergen Current, *Journal of Physical Oceanography*, *46*, 1231–1254.

Wadley, M. R., and G. R. Bigg (2002), Impact of flow through the Canadian Archipelago and Bering Strait on the north Atlantic and Arctic circulation: An ocean modelling study, *Quarterly Journal of the Royal Meteorological Society*, *128*, 2187–2203.

Wallace, J. M., I. M. Held, D. W. J. Thompson, K. E. Trenberth, and J. E. Walsh (2014), Global warming and winter weather, *Science*, *343*, 729–730.

Wang, Q. (2007), The Finite Element Ocean Model and its aspect of vertical discretization, Ph.D. thesis, Bremen University.

Wang, Q., S. Danilov, and J. Schröter (2008), Finite Element Ocean circulation Model based on triangular prismatic elements, with application in studying the effect of vertical discretization, *J. Geophys. Res. - Oceans*, *113*, C05,015.

Wang, Q., S. Danilov, E. Fahrbach, J. Schröter, and T. Jung (2012), On the impact of wind forcing on the seasonal variability of Weddell Sea Bottom Water transport, *Geophys. Res. Lett.*, *39*, L06,603.

Wang, Q., S. Danilov, D. Sidorenko, R. Timmermann, C. Wekerle, X. Wang, T. Jung, and J. Schröter (2014), The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model, *Geosci. Model Dev.*, *7*, 663–693.

Wang, Q., M. Ilicak, R. Gerdes, H. Drange, Y. Aksenov, D. Bailey, M. Bentsen, A. Biastoch, A. Bozec, C. Böning, C. Cassou, E. Chassignet, A. Coward, B. Curry, G. Danabasoglu, S. Danilov, E. Fernandez, P. Fogli, Y. Fujii, S. Griffies, D. Iovino, A. Jahn, T. Jung, W. Large, C. Lee, C. Lique, J. Lu, S. Masina, A. Nurser, B. Rabe, C. Roth, D. Salas y Mélia, B. Samuels, P. Spence, H. Tsujino, S. Valcke, A. Voldoire, X. Wang, and S. Yeager (2016a), An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater, *Ocean Modell.*, *99*, 110–132.

Wang, Q., M. Ilicak, R. Gerdes, H. Drange, Y. Aksenov, D. Bailey, M. Bentsen, A. Biastoch, A. Bozec, C. Böning, C. Cassou, E. Chassignet, A. Coward, B. Curry, G. Danabasoglu, S. Danilov, E. Fernandez, P. Fogli, Y. Fujii, S. Griffies, D. Iovino, A. Jahn, T. Jung, W. Large, C. Lee, C. Lique, J. Lu, S. Masina, A. Nurser, B. Rabe, C. Roth, D. Salas y Mélia, B. Samuels, P. Spence, H. Tsujino, S. Valcke, A. Voldoire, X. Wang, and S. Yeager (2016b), An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part II: Liquid freshwater, *Ocean Modell.*, *99*, 65–90.

Wang, Q., S. Danilov, T. Jung, L. Kaleschke, and A. Wernecke (2016c), Sea ice leads in the Arctic Ocean: Model assessment, interannual variability and trends, *Geophysical Research Letters*, *43*, 7019–7027.

Wekerle, C., Q. Wang, S. Danilov, T. Jung, and J. Schröter (2013), The Canadian Arctic Archipelago throughflow in a multiresolution global model: Model assessment and the driving mechanism of interannual variability, *J. Geophys. Res. - Oceans*, *118*, 4525–4541.

Wekerle, C., Q. Wang, S. Danilov, V. Schourup-Kristensen, W.-J. von Appen, and T. Jung (2017), Atlantic Water in the Nordic Seas: locally eddy-permitting ocean simulation in a global setup, *J. Geophys. Res. Oceans*, *122*, 914–940.

Wekerle, C., Q. Wang, W.-J. von Appen, S. Danilov, V. Schourup-Kristensen, and T. Jung (2017b), Eddy-resolving simulation of the Atlantic Water circulation in the Fram Strait with focus on the seasonal cycle, *J. Geophys. Res. Oceans*, *122*, 8385–8405.

Woodgate, R. A., and K. Aagaard (2005), Revising the Bering Strait freshwater flux into the Arctic Ocean, *Geophysical Research Letters*, *32*, L02,602.

Woodgate, R. A., K. Aagaard, R. D. Muench, J. Gunn, G. Björk, B. Rudels, A. T. Roach, and U. Schauer (2001), The Arctic Ocean boundary current along the Eurasian slope and the adjacent Lomonosov Ridge: Water mass properties, transports and transformations from moored instruments, *Deep Sea Research Part I: Oceanographic Research Papers*, *48*, 1757–1792.

Yang, J., A. Proshutinsky, and X. Lin (2016), Dynamics of an idealized Beaufort Gyre: 1. the effect of a small beta and lack of western boundaries, *J. Geophys. Res. - Oceans*, *121*, 1249–1261.

Zhang, J., and M. Steele (2007), Effect of vertical mixing on the Atlantic Water layer circulation in the Arctic Ocean, *J. Geophys. Res. - Oceans*, *112*, C04,S04.