Response to Referee 3

We thank the referee very much for his/her comments and suggestions, especially those highlighting that the uncertainty from this study must be communicated more clearly. They helped to improve the quality of the paper.

Note that we conducted new simulations because the ship emissions were shifted by two weeks in the old simulations. This especially affected the impact of transit ship emissions in the Arctic in late summer 2050. Furthermore, we increased the number of simulated years from 10 to 20 for better statistics. In some cases, the results have thus changed; as an example, the SW CRE now increases significantly with additional tenfold Arctic ship emissions in 2050.

Referee's comments in blue, our replies in *grey and italic*.

Gilgen et al present a set of sensitivity studies with the atmospheric GCM ECHAM-HAM. The control simulation is driven by conditions approximately representative for present-day, three sets of differing boundary conditions are then computed: (a) increased sea surface temperatures and decreased sea ice cover, (b), in addition, changed aerosol emissions, and (c) in addition, further ship emissions. Each of the simulations is run for a short period of ten years. A large set of results is presented. The study is to a large extent based solely on the results of the one model and thus the hypotheses developed are strongly dependent on the chosen parameterisations. Very little comparison to data (for the control simulation) is presented. In one paragraph, the cloud radiative effects are compared to SHEBA data – from this it seems that the model has a very large bias. I believe it would be necessary for an improved paper to at least show some evidence that the model performs satisfactorily in the Arctic in comparisons to observations, before the results from the sensitivity studies can be considered meaningful.

We agree with the reviewer that the results of this study are uncertain, especially because they were conducted with only one climate model. Therefore, more information about the performance of the model in the Arctic will help the reader to put the results of this study into context. Thus, we added a comparison of some key variables in the Supplementary Information: AOT; BC and sulphate concentrations at the surface; cloud cover; LWP and IWP; cloud radiative effects (SW, LW, and net) at the surface and at the TOA. The AOT and the IWP are underestimated, whereas all other variables compare well (the bias in the SW surface CRE mentioned in the old paper version was likely caused by the local surface albedo of the observations; see Supplementary Information for more details). Despite the model bias in AOT and IWP, we consider our results meaningful: in our opinion, model biases do not imply that the simulated changes are necessarily wrong. Vice versa, also a model that compares perfectly to present-day observations can make wrong projections for the future. Furthermore, we want to stress that observations in the Arctic are sparse and sometimes also uncertain.

Next to the detailed comparison that we provide in the Supplementary Information (the most important results of which are mentioned in the paper), we added the following paragraph to the results section:

"Heterogeneous freezing is still an active field of research, and contradictory evidence exists concerning the ability of combustion aerosols to act as INPs (Kanji et al., 2017). Laboratory results suggest that soot starts initiating freezing at temperatures ≤-30°C (Kanji et al., 2017, Fig. 1-7). On the other hand, Thomson et al. (2018) found an increase in INP concentrations in ship tracks at higher temperatures. The increases were small at temperatures around −20°C, moderate at −25°C (≈ 0.5 L^{-1}; saturation ratio of 1.22), and quite pronounced at -30°C (≈ 2 L^{-1}; saturation ratio of 1.32). The ship plumes were measured near the port of Gothenburg (57.7° N, 11.8° E) in 2013 and 2014, and the meteorology in general represented climate conditions of the late-autumn maritime North. If ship exhaust (not necessarily the BC particles) can indeed induce freezing at higher temperatures than in the laboratory-based BC-parameterisation used in our model, the impact on

cloud ice could be larger than in our simulations, especially in early autumn when temperatures are colder."

At two instances, the results are compared to previously-published results for similar scenarios. It is astonishing how different the results are. A key hypothesis is that sea salt emissions may substantially increase with decreasing sea ice coverage. Fundamentally, this is no surprise, so the question is how large this could be quantitatively. Unfortunately the two other model studies reported are much more different from the model presented here than the change due to sea ice retreat (one model has a factor of 3 more, the other, a factor of 1000 less emission flux in present-day conditions). Also the radiative forcing due to aerosol-radiation interactions is very different between models – the model presented here has a substantially positive forcing, the other model, a negative one. Since such results are easily available from multi-model ensembles (CMIP5 or AEROCOM), it would be easy to put the model the authors use into context, much beyond the two studies cited.

Our main goal was to compare how future sea salt emissions might change due to sea ice retreat. Therefore, we initally only compared our values with studies that looked specifically at this question. To our knowledge, nearly no CMIP5 model calculates interactively emissions of sea salt, but we agree that other model intercomparison studies can give insight into the spread in sea salt emissions between different models. Therefore, we included results from the study of de Leeuw et al. (2011) and added the following sentences:

"Note that at the present state, sea salt emissions are highly uncertain and differ considerably between models: a comparison of 11 chemical transport and global climate models shows that the global annual mass emissions of sea salt lie in the range between 2.2 and 9.9 × 10^{12} kg yr^{-1} for 9 of the models; 2 models calculate higher emissions of 22 × 10^{12} kg yr^{-1} and 118 × 10^{12} kg yr^{-1} (de Leeuw et al., 2011). For comparison, our simulated value (1.3 × 10^{12} kg yr^{-1}) is on the same order of magnitude as most of these models, but lower than in all of them because the parameterisation does not account for the contributions from spume drops (Long et al., 2011). Our simulated absolute increases in sea salt mass emissions might therefore be underestimated."

In the conclusions, we highlight again the uncertainty:
"Compared with observations, our model has a low bias in AOT and cloud ice, which could impact the absolute changes in the radiative aerosol forcing and the CREs. Furthermore, when we compare our results with other modelling studies that looked at natural aerosol changes with declining sea ice, we find large intermodel differences, e.g. concerning sea salt emissions. This highlights that the results from this study – as from any climate model study projecting the future – are uncertain."

When it comes to the interpretation of the results, much is left for speculation. If the authors choose to have a pure modelling study, why don't they at least precisely clarify the processes that change? Why not budgets for changes of CCN, INP? A table that lists all relevant numbers (e.g. for the entire region, and split for open ocean and sea ice surfaces) as simulated for the different scenarios would be useful (emission fluxes, CCN, INP, cloud particle concentrations, LWP, IWP).

We agree that our result's section is somehow speculative. It is not straightforward to understand 'what is going on' in the model, and understanding every single result in detail would be beyond the scope of this study. The large advantage of using a climate model is that it accounts for many processes and their couplings. In this case, the main goal was to look at aerosol-cloud interactions and radiative forcings with a complex aerosol-climate model, which has the drawback that the results stem from a combination of processes that cannot easily be separated.
Unfortunately, we do not calculate intermediate variables such as CCN or INP concentrations. The model calculates CDNC and ICNC based on the size and chemical composition of the aerosol particles, temperature, and supersaturation.
However, we agree with the referee that a table showing averages over the whole Arctic region, open ocean, sea ice, and regions where sea ice has melted is very interesting. Thus, we included additionally Tables 2 and 3 in the paper, which show averages for sea salt and DMS burdens, AOT,

LWP, IWP, cloud cover, CDNC burden, and surface temperature. The results of these tables are discussed in the text.

Specific comments:
p3 l30 – this is not "generally" true, e.g. not in summer (as the following sentence correctly acknowledges)
Changed to:
"Therefore, the LW absorption of clouds becomes more important and can dominate the total cloud radiative effect depending on the specific time and location."

p3 l32 – but it is likely a small effect (Pithan and Mauritsen). What is the reference for the following sentence ("generally...")?
Changed to:
"How Arctic clouds and their radiative effects will change in the future is still an open question. Generally, both the SW and the LW cloud radiative effect (CRE) are expected to become stronger when more CCN are available (Mauritsen et al., 2011). However, compared to other temperature feedbacks, the contribution of changes in Arctic clouds might be small (Pithan and Mauritsen, 2014)."

P5 l12 – it would be good to report already here whether this threshold is hit, and, if so, how often. It would further be good to analyse whether indeed the lack of nitrate or organics is a major problem of this model for the Arctic.
This threshold is occasionally hit, e.g. in the subtropics and over the Central Arctic Ocean. We now also provide a comparison with observations, as suggested by referee 2.

Changed to:
"In the standard ECHAM6-HAM2 setup, a minimum CDNC of 40 cm^{-3} is implemented. This ensures that the global CDNC is not unrealistically low due to missing aerosol species in the model such as nitrate or due to the simplistic model description of organics (no explicit treatment of secondary organic aerosols; neglection of marine organics). Without a lower threshold for CDNC, the model might underestimate the CDNC also in the Arctic, where organic aerosol particles are emitted from the surface microlayer (Hawkins and Russell, 2010; Bigg et al., 2004; Leck and Bigg, 2005; Chang et al., 2011). However, since the Arctic is a remote environment with low aerosol concentrations, observations show that the value 40 cm^{-3} is often undershot in this region: between July 15th and September 23th, Bigg and Leck (2001) measured daily median CCN concentrations between 15 and 50 cm^{-3} at a supersaturation of 0.25%. In July 2014, Leaitch et al. (2016) found a median CDNC of 10 cm^{-3} for low-altitude clouds (cloud top below 200 m) and of 101 cm^{-3} at higher altitudes. In October 2004, McFarquhar et al. (2007) conducted aircraft measurements in single-layer stratus clouds and found averaged cloud droplet number concentrations of 43.6±30.5 cm^{-3}. Applying the standard CDNC threshold of 40 cm^{-3} would drastically reduce the influence of changes in the CCN concentration and therefore impede aerosol-cloud interactions. Thus, we decided to use 10 cm^{-3} as a lower threshold for the CDNC everywhere and retuned this model version. The studies by Bigg and Leck (2001) and Leaitch et al. (2016) indicate that values even below this lower threshold can occur. While these measurements are representative for a specific point, our model represents average values over a larger area (1.875° × 1.875°), which should be less variable than a point measurement. Nevertheless, we acknowledge that the threshold of 10 cm^{-3} could still be too high under certain conditions. In the model, this threshold is occasionally hit, e.g. over the central Arctic Ocean or in the subtropics. "

P6 l1 – it would be good to comment on the results of Eckhardt et al. (ACP 2015)
Thank you for this reference. We included it in the Supplementary Material when comparing simulated and observed SO_4.

p8 l6 – 10 years seem very little for small forcings
We increased the number of years to 20.

p12 top paragraph – what do these discrepancies by a factor of about 3000 imply for

the fidelity of the results in terms of sea salt emission changes?
Thank you for this comment. We added a comparison with de Leeuw et al. (2011) and mention now the uncertainty in the conclusions (see text above).

P13 l8 increased
Corrected.

p13 l22 – i.e. homogeneous freezing of droplets?
This explanation was wrong. We thought that the increase in CDNC (as well as the increase in radius) increases the contact freezing rate, but this is only important in limited areas far north. The simulated increases in ICNC are due to enhanced convection.

Changed to:
"The increase of ICNC near the surface is mainly caused by enhanced convection, which leads to small but numerous simulated ice crystals following the temperature-dependent empirical parameterisation of Boudala et al. (2002)."

P13 l23 – at con-
stant ICNC?
Changed to:
"Between 500 hPa and 200 hPa, the enhanced ice water content is linked to the increase in ice crystal radius, while the ICNC slightly decreases."

L13 l27 – indeed surface fluxes? Or rather simply moist adiabat changes?
The explanation in Abe et al., 2016 is the following: "Because of the reduced sea ice, a more extended open ocean area increased the latent and sensible heat fluxes from the ocean to the atmosphere. Along with the seasonal march, the decreased atmospheric temperatures increased the temperature gradient between the air and sea surface in October. Therefore, the fluxes from the ocean to the atmosphere were enhanced in October rather than in September."
In our study, most pronounced increases in cloud cover occur where sea ice has melted (Table 3) and where also the changes in latent and sensible heat fluxes are most pronounced (significant changes; not shown in the paper).

P16 l2 – it would be important to clarify whether this section refers to the radiative forcing by aerosol-radiation interactions only, or to the effective radiative forcing due to aerosol-radiation interactions, or whether it includes aerosol-cloud interactions.
We calculate the radiative forcing by calling the radiation scheme once with and once without aerosols, i.e. it is the radiative forcing by aerosol-radiation interactions only. We now remind the reader of this in the results section and highlight in the methods that our radiative forcing refers to all aerosols (not only anthropogenic aerosols). Furthermore, we now apply abbreviations from the newest IPCC report in our paper (e.g. "RF_{ari}" or "ERF_{aci}") for clarification and avoid the term "Twomey effect".
We added this sentence:
"As mentioned previously, the aerosol radiative forcing refers to the instantaneous effect of all aerosols on radiation (RF_{ari})."

P16 l4
– what are the absorbing components, and why is the positive forcing so large? P16 l6
The absorbing components are BC and dust. Since the warming is only prevalent in the presence of clouds, we assume that the scattering of aerosol particles is less important in the presence of clouds and that the warming of aerosol particles might be enhanced by the higher SW radiation reflection by clouds.
Changed to:
"If the presence of clouds is considered, aerosol particles warm the atmosphere also over Alaska and northeast Siberia (late summer) and over the whole northern Russia (early autumn; shown in Fig. 7e). Part of this warming might be caused by BC and dust aerosols above clouds (Supplementary Fig. 10): the clouds reflect more SW radiation than the snow/ice-free surface and

part of the scattered SW radiation can also be absorbed by aerosol particles causing an increase in aerosol absorption as compared to clear-sky conditions (see e.g. Myhre et al. 1998). Moreover, the scattering of aerosol particles could become less important in the presence of clouds, which increases the relative importance of aerosol absorption to extinction.

– it would be useful to demonstrate this at least in the supplementary material (since the authors write "not shown" it seems hey have the analysis at hand)
We added Supplementary Fig. 10.

p17 l5 – how is the coincidence of approximately the same reduction by 0.2 Wm-2 explained? Is the same thing happening in both models?
Based on the comment of referee 2, we take this comparison out. In contrast to Struthers et al. (2011), we look at all aerosol particles, which makes a comparison difficult.

P18 l13 – i.e. the effect is twice as large as observed? The authors should report this analysis as a table or similar.
We provide now a more detailed analysis in the Supplementary Material.