1 Comparison with observations

To interpret the results from this model study, it is important to know how well the model performs and to be aware of certain biases. However, comparing all variables of interest with observations is beyond the scope of this paper. We restrict the comparison therefore to AOT, BC and SO_4 surface concentrations, cloud fraction/cover, LWP, IWP, and CREs at the surface and at the TOA (SW, LW, and net). We only consider data that covers a large part of our period of interest (July to October) and that reflects present-day conditions (between the years 1998 and 2015). The simulated values refer to the ensemble mean of the simulation **arctic_2004**.

Compared to CALIPSO measurements (60° to 82° N, 2006-2011; Sand et al. 2017, Fig. 6), the AOT in our model (0.037 in late summer and 0.033 in early autumn) is underestimated by a factor of 2 to 3. On the other hand, our simulated AOT over the Greenland/Barents Sea (10° W to 40° E, 75° to 82° N) is 0.034 in late summer and agrees very well with MODIS measurements from 2003 to 2011 (Sand et al. 2017, Fig. 5). The underestimation of AOT in our model compared to CALIPSO measurements can have various reasons, e.g. missing local aerosol sources in the model (e.g. marine organics or gas flaring emissions; Hawkins and Russell 2010; Chang et al. 2011; Stohl et al. 2013), an underestimation of aerosol transport from midlatitudes to the Arctic (Bourgeois and Bey 2011), uncertainties in the optical properties and emissions of aerosols (e.g. for BC, see Bond and Bergstrom 2006; Bond et al. 2013), and/or the neglection of spume drops in the sea salt parameterisation by Long et al. (2011).

Next to AOT, we also compare our simulated BC and SO_4 concentrations with observations. The impact of future BC and SO_4 emissions from Arctic shipping depends on the background concentration, which is mainly determined by long-range transport in the case of BC. We compare our simulated BC concentrations with the recent long-term surface observations at Zeppelin (78.92° N, 11.93° E) and Barrow (71° N, 156.6° W; Sinha et al. 2017). Note that earlier studies report higher BC concentrations at the same stations, since they neglected the effect of other aerosol components on the aerosol light absorption coefficient when deriving mass concentrations of BC (Sinha et al. 2017). From July to October, our monthly averages range from 4.3 to $9.7\,\mathrm{ng\,m^{-3}}$ at Zeppelin and from 6.1 to $12.3\,\mathrm{ng\,m^{-3}}$ at Barrow. This is in good agreement with the observations, which lie in the ranges 3 to $9\,\mathrm{ng\,m^{-3}}$ and 3 to $10\,\mathrm{ng\,m^{-3}}$ at Zeppelin and Barrow, respectively (depending on the averaging period and the derivation method).

We compare our simulated SO_4 surface concentrations with the values from Eckhardt et al. (2015). Averaged over the stations Alert (82.5° N, 62.5° W), Zeppelin, and Barrow and the years 2008 and 2009, they report a summer (July to September) value of $103.2\,\mathrm{ng\,m^{-3}}$. Our simulated value (averaged over the same locations and months) is approximately twice as large ($214\,\mathrm{ng\,m^{-3}}$). Concerning the large spread between models (Eckhardt et al. 2015), we consider this to be in reasonable agreement with the observations.

We used data from the SHEBA campaign, which took place from October 1997 to October 1998 in the region 180° W to 130°W, 70° to 80° N, to compare LWP and IWP in the Arctic (Shupe and Uttal 2007). Since only few data is available for October, we restrict the comparison to the months July, August, and September. The LWP measurements are based on microwave radiometer retrievals (Westwater et al. 2001). For the derivation of IWP, we use estimates from two different methods: an empirical technique that relates cloud ice water content to radar reflectivity, and the technique described in Matrosov et al. (2002), which relates ice particle size to radar Doppler velocities. Mixed-phase clouds have retrieval values for only the ice component. The uncertainties are $\approx$25 % for LWP and up to a factor of 2 for IWP (Matrosov et al. 2002; Shupe et al. 2004; Shupe et al. 2005).

Individual data points are unrealistically high (several thousands $\mathrm{g\,m^{-2}}$), which could be due to erroneous measurements. As an example, the LWP derived from the microwave radiometer becomes wet under rainy conditions, which results in overestimated brightness temperatures and, thus, overestimated LWP (Matthew Shupe, personal communication). In addition to the original data, we therefore also show results where values larger than 800 g m^{-2} are considered as NAN in Fig. 1. The value of 800 g m^{-2} is somewhat arbitrary, but comparing the original with the processed data helps to identify cases in which the mean is strongly biased by the outliers.

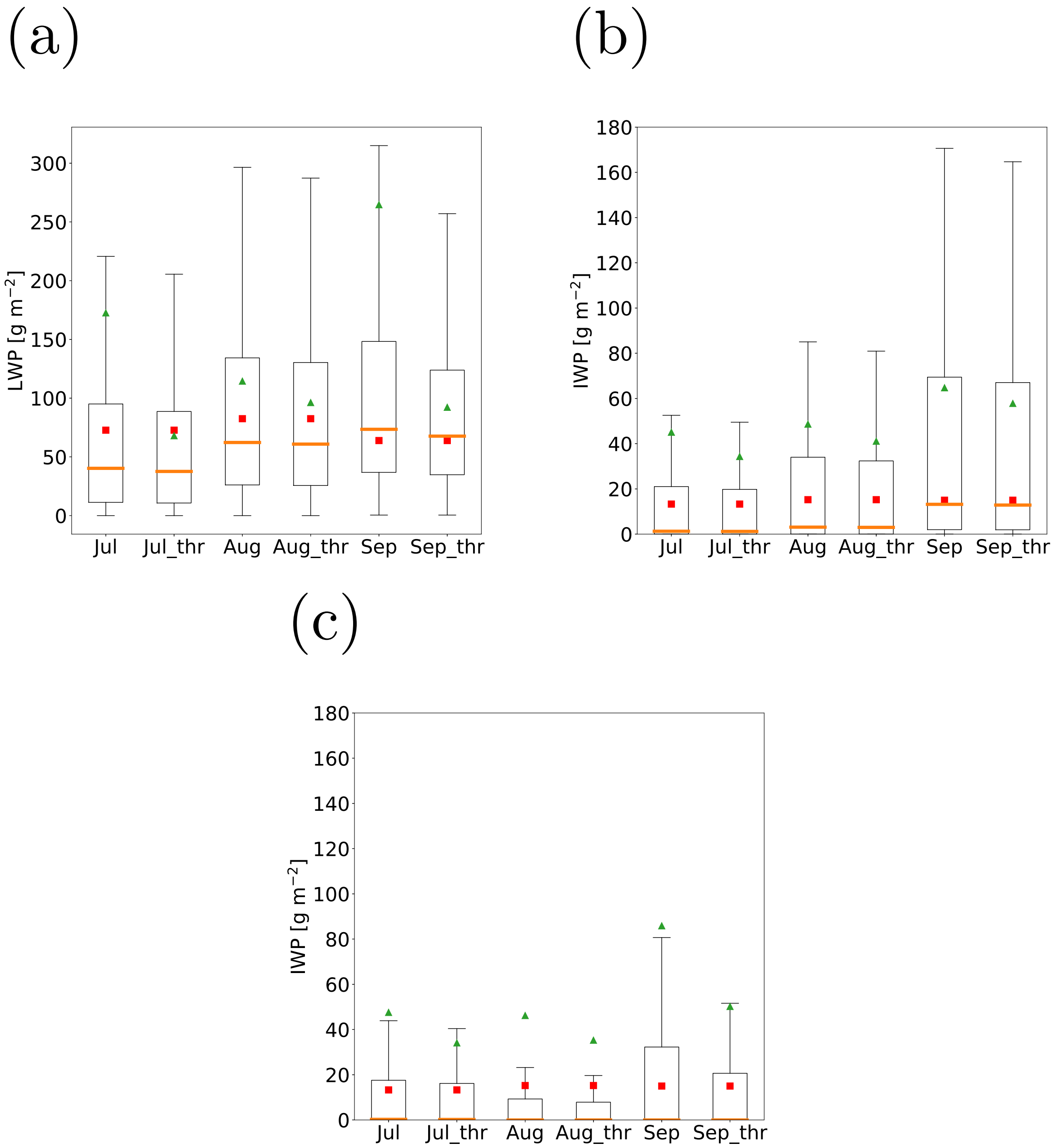

Figure 1: Monthly statistics of (a) LWP and (b)/(c) IWP from the SHEBA campaign for July, August, and September. The IWP in (b) is derived from an empirical technique using radar reflectivity, whereas the IWP in (c) is estimated with a technique based on radar Doppler velocities. The original data by Shupe and Uttal (2007) is shown (Jul, Aug, Sep) as well as data where values larger than 800 g m^{-2} are excluded (Jul_thr, Aug_thr, Sep_thr). Orange lines and green triangles show the observed medians and means of each month, respectively. For comparison, the monthly means from our simulations are shown as red squares. The length of the whiskers is restricted to 1.5 times the interquartile range.

The simulated LWP in ECHAM6-HAM2 (average over the SHEBA region) compares well with the observations (Fig. 1a). The simulated and the observed average are similar in July (ECHAM6-HAM2: 72.7 $g\,m^{-2}$; SHEBA: 68.0 $g\,m^{-2}$) and August (ECHAM6-HAM2: 82.5 $g\,m^{-2}$; SHEBA: 96.3 $g\,m^{-2}$), whereas ECHAM6-HAM2 underestimates the LWP in September (ECHAM6-HAM2: 63.9 $g\,m^{-2}$; SHEBA: 92.2 $g\,m^{-2}$). However, in September, the observed average is quite sensitive to the upper threshold we applied for LWP and should therefore be taken with caution.

While the probability distributions of IWP are different for the two different methods (Figs. 1b,c), the average values compare relatively well when the upper threshold is applied. The observed medians in Fig. 1c are (near) zero because cloud ice has a rather low occurrence frequency, i.e. the measured IWP is often zero. In all three months, ECHAM6-HAM2 severly underestimates the average IWP: 13.3 $g\,m^{-2}$ instead of 34.1 to 34.3 $g\,m^{-2}$ (depending on the method) in July, 15.2 $g\,m^{-2}$ instead of 35.3 to 41.1 $g\,m^{-2}$ in August, and 15.0 $g\,m^{-2}$ instead of 50.2 to 57.8 $g\,m^{-2}$ in September. This underestimation is a global phenomenon of ECHAM6-HAM2 (Lohmann and Neubauer 2018, in review) and previous model versions.

Furthermore, we compare our simulated cloud cover/fraction and surface CREs with values reported from the SHEBA campaign (Intrieri et al. 2002). The cloud fractions shown in Fig. 2a are determined from temporal and spatial averages of lidar and ceilometer measurements. Until mid-August 1998, data is derived from the lidar, after that from the ceilometer. The derived surface SW CRE from observations in Fig. 2b relies on model calculations for the clear-sky using single site albedos (ASFG). In Fig. 2c, the net surface CRE is also shown for calculations with line-averaged albedos (CRREL). The ASFG albedo was computed hourly by the Atmospheric Surface Flux Group radiometers. THE CRREL albedo was obtained by the Cold Regions Research and Engineering Laboratory group once per day around solar noon and includes many different ice types (e.g. melt ponds, open water). While the ASFG albedos are directly linked with the observed fluxes, the CRREL albedos are more representative of the SHEBA ice camp area.

The simulated cloud cover compares relatively well with the cloud fraction observed in the SHEBA campaign (Fig. 2a). We noted that the large values of observed cloud fraction (near 100 %) coincide with the time when the instrument switched from lidar to ceilometer. Data derived from ISCCP, averaged from 1982 to 1999 in summer, is smaller in the regions where the SHEBA campaign took place, namely 0.74 in the Beaufort and 0.79 in the Chucki Sea (Wang and Key 2005).

While the simulated LW CRE is in good agreement with the observations (Fig. 2b), the simulated SW CRE is consistently lower. However, the SW CRE depends strongly on the surface albedo. When using the CRREL instead of the ASFG albedo, the observed SW CRE is considerably stronger. This is evident in Fig. 2c, where the net CRE is shown with both surface albedo estimates. Overall, our simulated values compare reasonably with the estimates of net CRE from Intrieri et al. (2002).

Next to comparing the CREs at the surface, we also compare the simulated CREs at the TOA with recent satellite data from the Clouds and the Earth's Radiant Energy System (CERES; Loeb et al. 2018). The CRE values over the poles are more uncertain than over other parts of the world because clear-sky measurements over snow and ice are challenging (Loeb et al. 2018). However, the advantage of satellite products is that they provide data for a larger region and a longer time period than measurement campaigns such as SHEBA. In Fig. 3, interannual monthly means are shown for the period from June 2005 to July 2015. We averaged the data between a) 60° and 90° N and b) 75° and 90° N. The model compares well with the satellite data. Largest absolute deviations occur in July between 75° and 90° N for SW CRE, where the observed value is $-59.0\,W\,m^{-2}$ and the simulated value is $-49.6\,W\,m^{-2}$.

To summarise, ECHAM6-HAM2 has a low bias concerning cloud ice and AOT, whereas the simulated BC and sulphate surface concentrations, the LWP, the cloud cover, and the CREs at the surface and at the TOA compare well with observations.

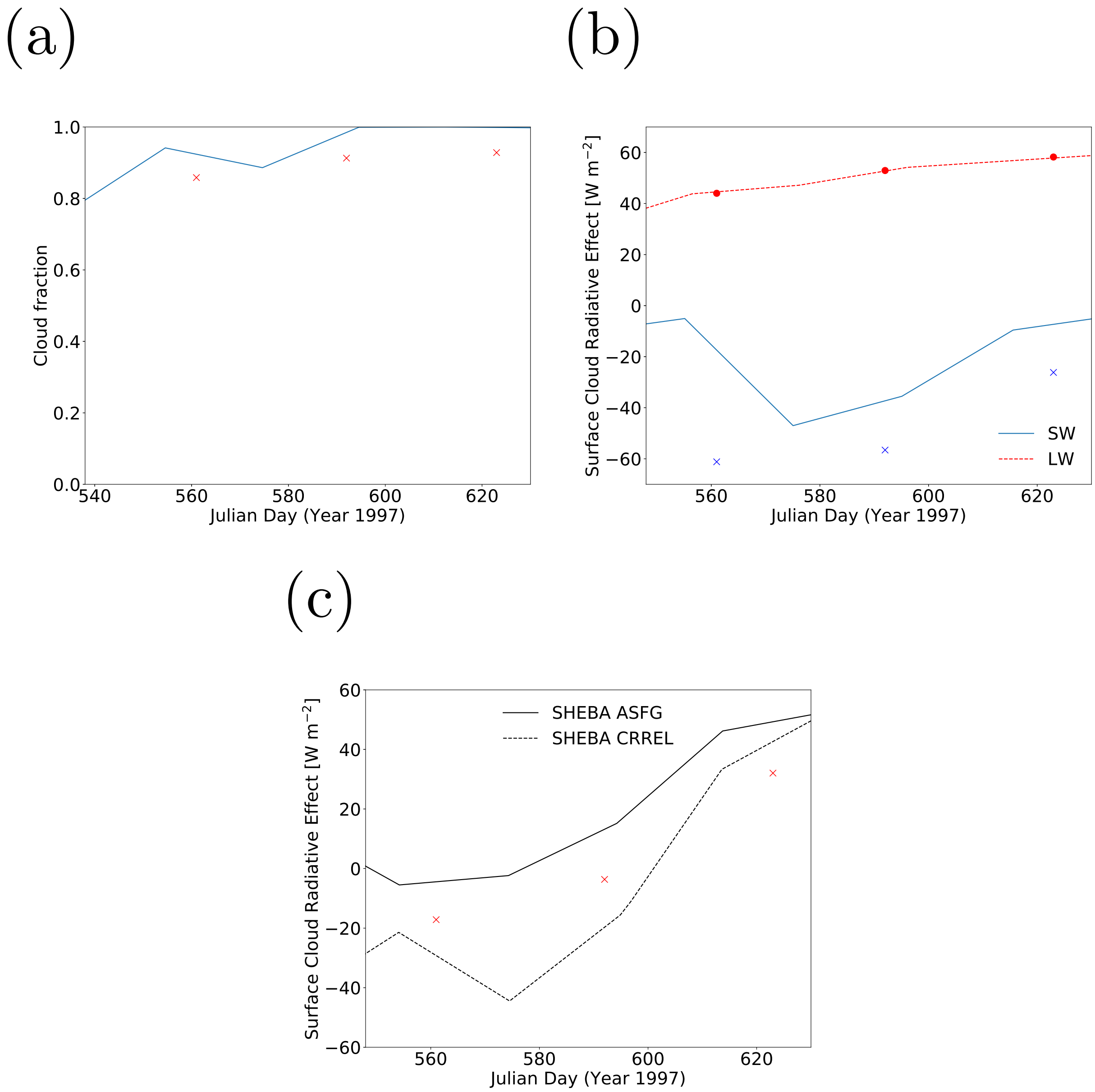

Figure 2: Comparison of ECHAM6-HAM2 (markers) with observations (lines) for a) cloud fraction, b) SW and LW surface cloud radiative effect, and c) net surface cloud radiative effect. All figures are adapted from Intrieri et al. (2002). The shown values for ECHAM6-HAM2 represent July, August, and September (placed at the 15th of each month).

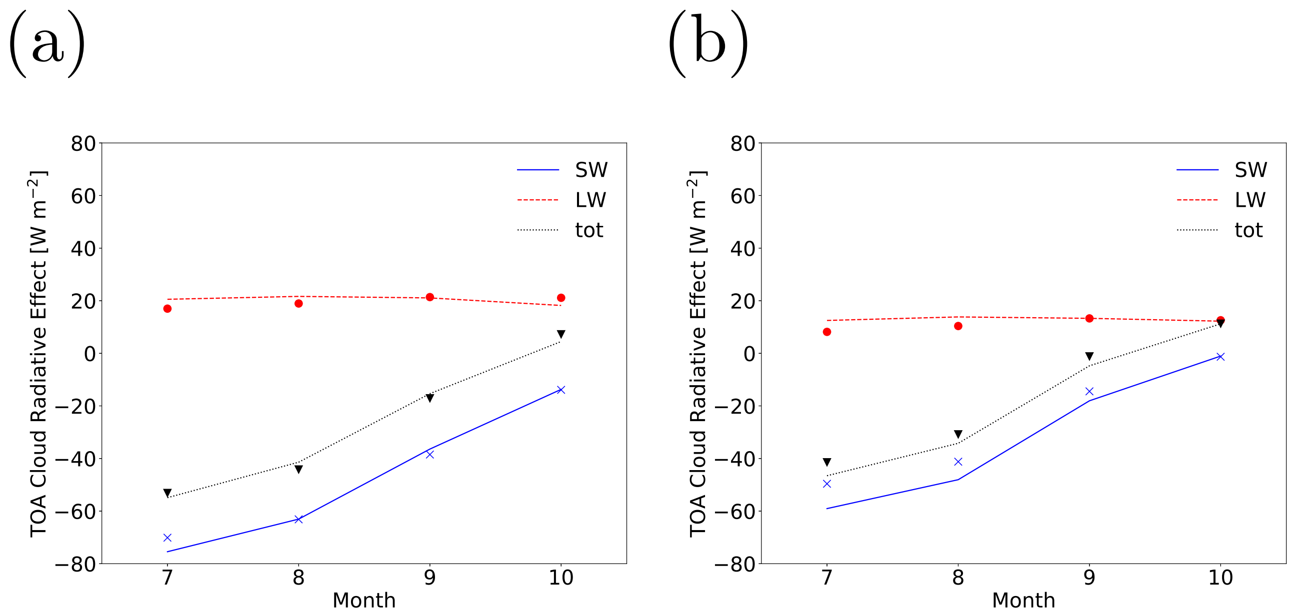

Figure 3: Comparison of ECHAM6-HAM2 (markers) with satellite-derived observations (lines) for SW, LW and net CRE at the TOA averaged between a) 60° and 90° N and b) 75° and 90° N.

2 Additional Figures

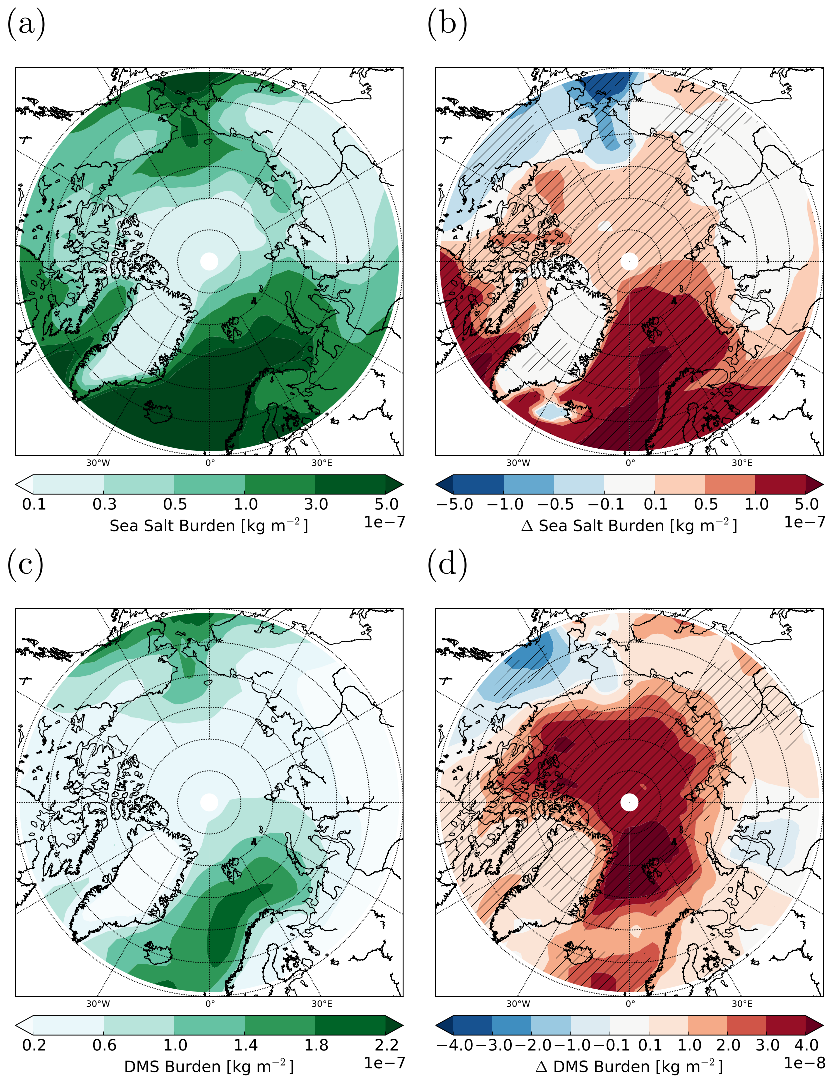

Figure 4: Sea salt and DMS burdens in 2004 in (a)/(c) and differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**) in (b)/(d) in early autumn (Sep/Oct). Hatched areas are significant at the 95% confidence level.

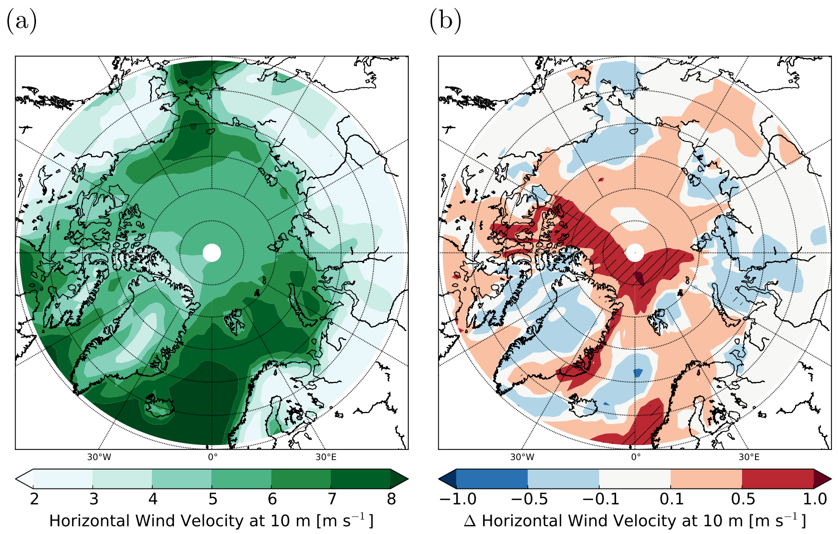

Figure 5: Wind speed at 10 m altitude in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level.

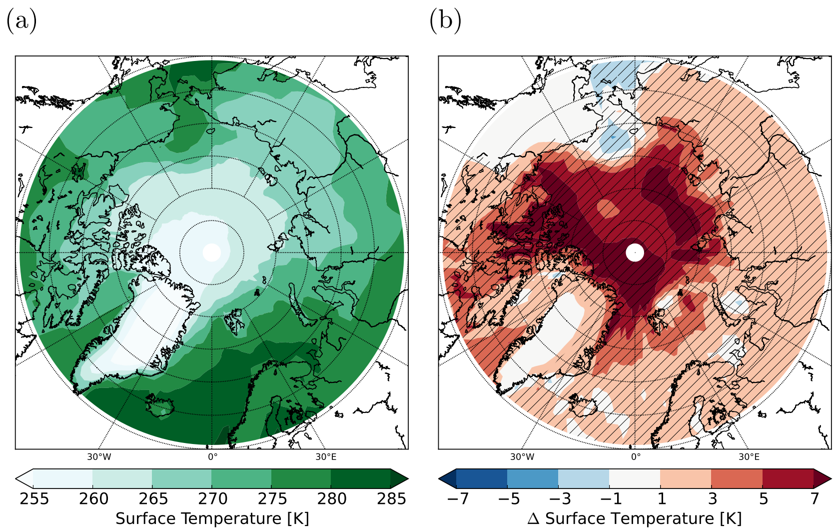

Figure 6: Surface temperature in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level. Note that the SST is prescribed in the simulations and shows no interannual variability.

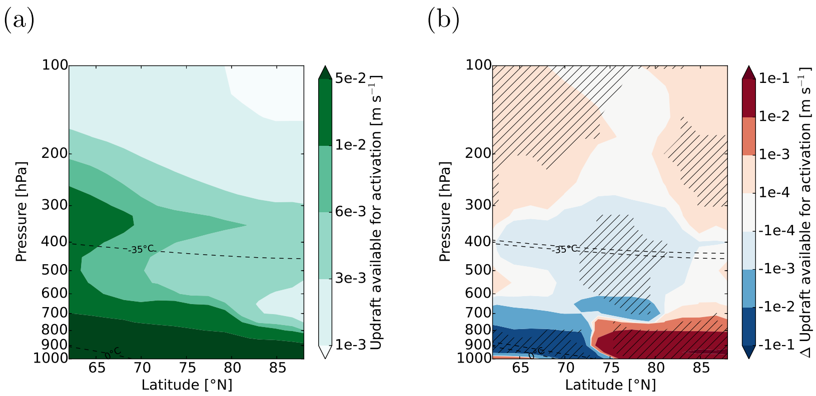

Figure 7: Updraft available for activation in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level.

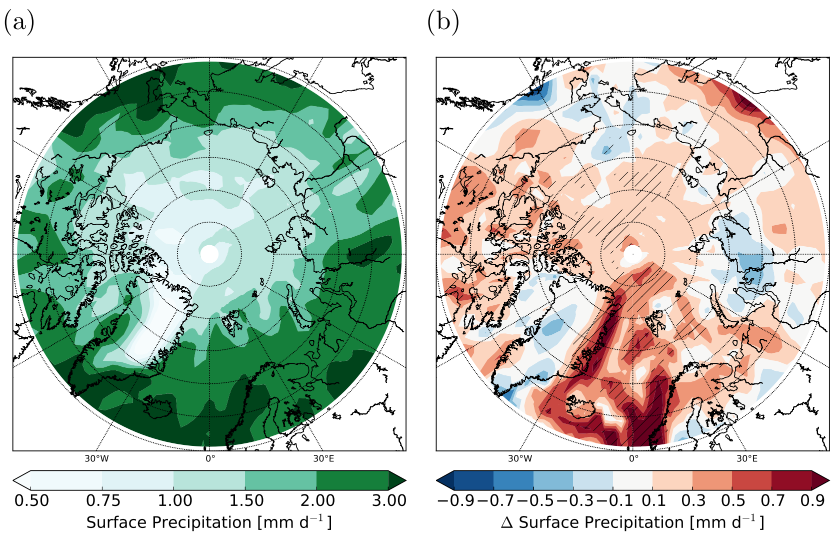

Figure 8: Surface precipitation rate in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level.

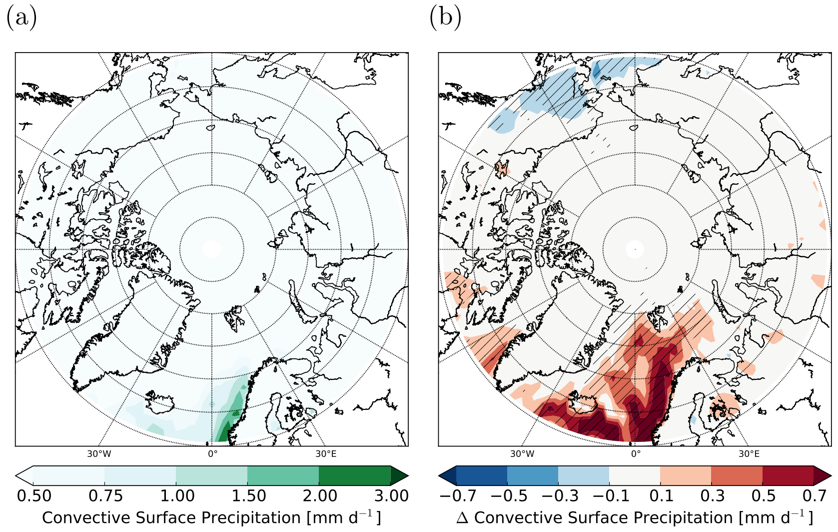

Figure 9: Convective surface precipitation rate in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level.

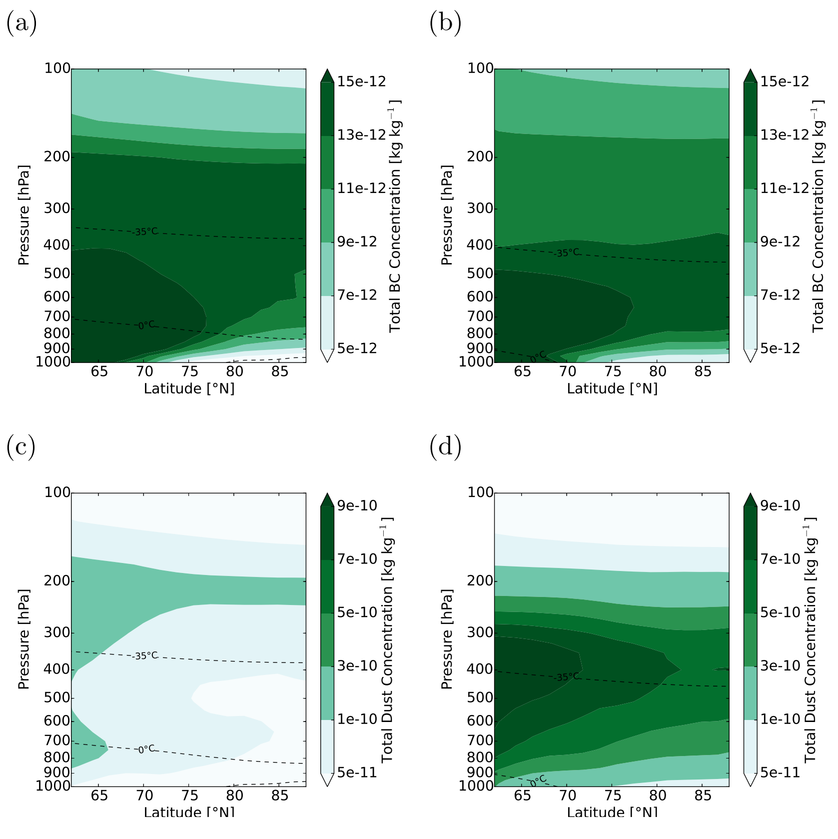

Figure 10: Total (a)/(b) BC and (c)/(d) dust concentrations in 2004. Figures on the left are averaged over late summer (Jul/Aug), figures on the right over early autumn (Sep/Oct).

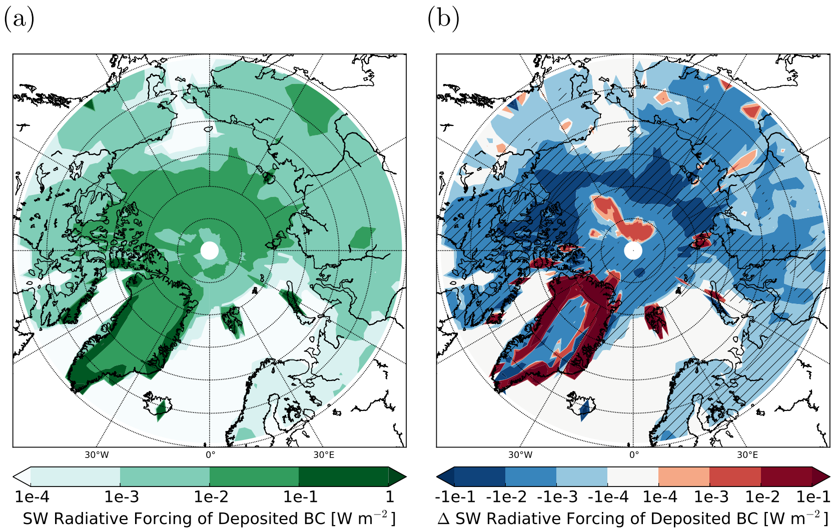

Figure 11: Radiative forcing of deposited BC in early autumn (Sep/Oct): a) absolute values in 2004 and b) differences between 2050 and 2004 (i.e. between simulations **arctic_2050_EM2004** and **arctic_2004**). Hatched areas are significant at the 95% confidence level. Note that the scale is logarithmic.

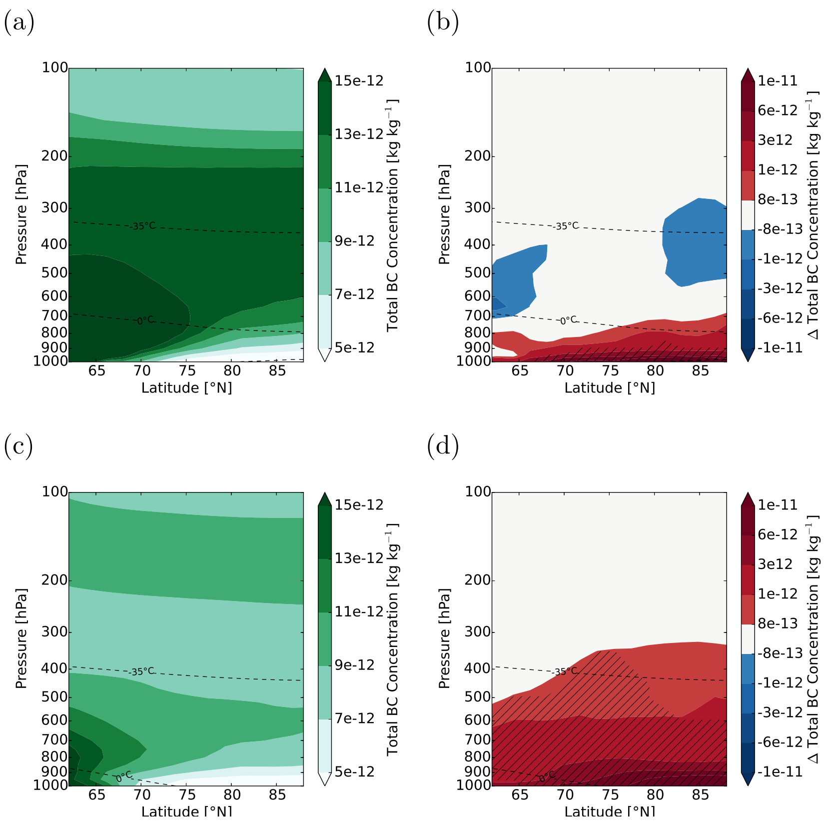

Figure 12: The impact of additional future ship emissions (**arctic_2050_shipping** versus **arctic_2050**) on the total BC concentration in b) late summer and d) early autumn. In (a)/(c), the reference without additional ship emissions is shown (**arctic_2050**). Hatched areas are significant at the 95% confidence level.

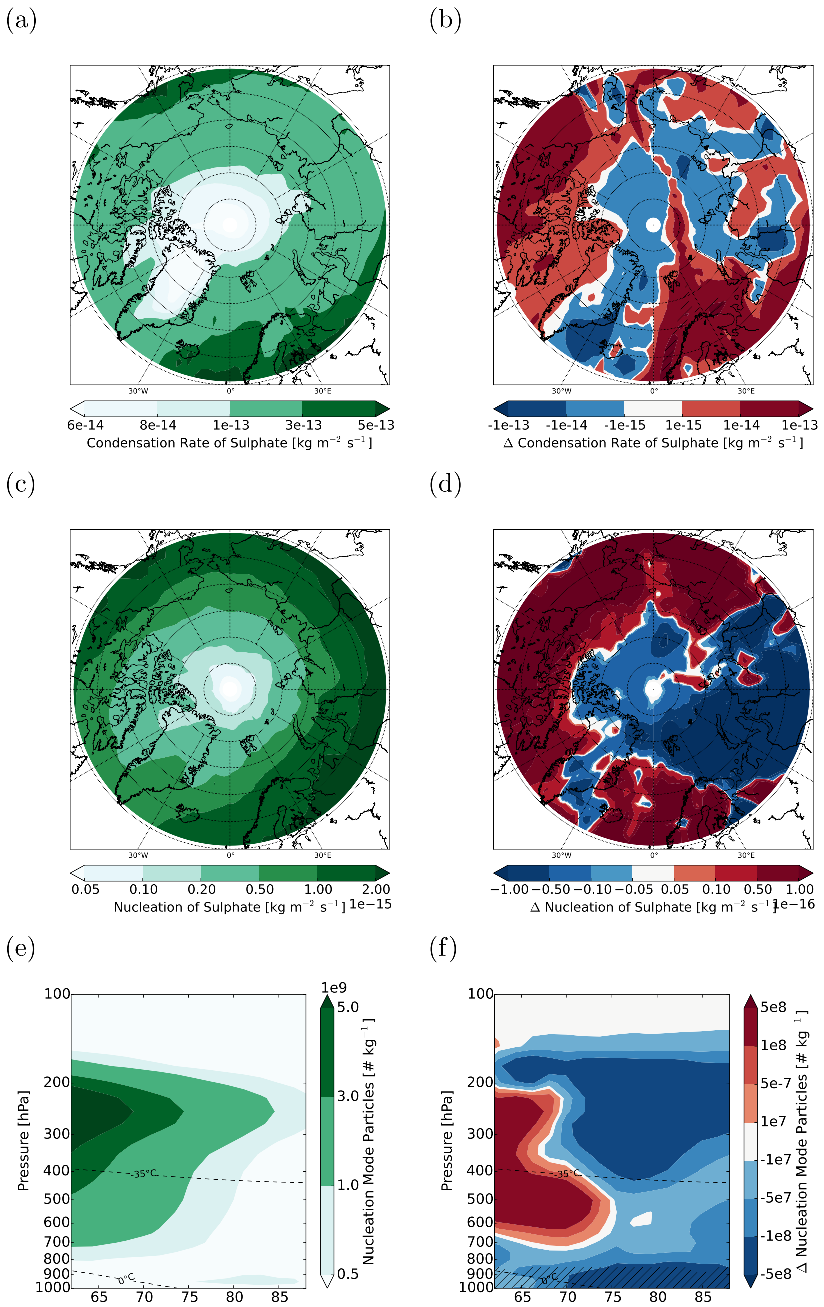

Figure 13: The impact of additional future ship emissions (**arctic_2050_shipping** versus **arctic_2050**) on: b) the vertically integrated condensation rate of sulphate on aerosol particles, d) the vertically integrated nucleation rate of sulphate, and f) the number of aerosol particles in the nucleation mode in early autumn (Sep/Oct). In (a)/(c)/(e), the reference without additional ship emissions is shown (**arctic_2050**). Hatched areas are significant at the 95% confidence level.

(a) (b)

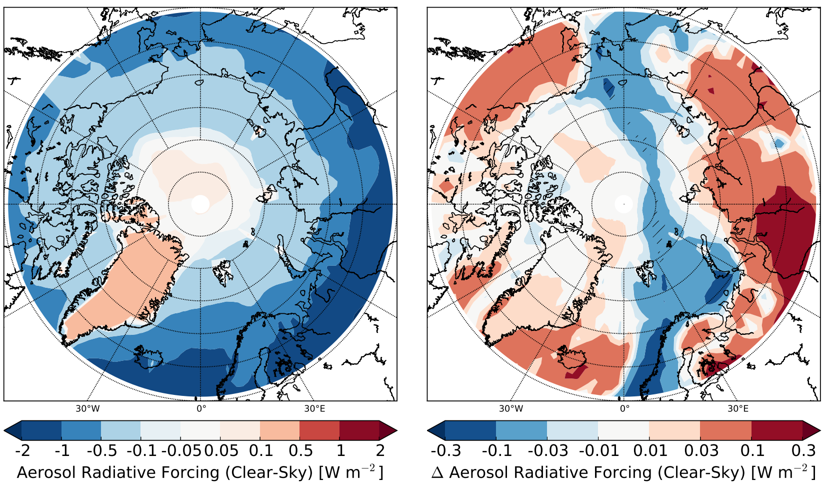

Figure 14: In (b), the impact of additional future ship emissions (**arctic_2050_shipping** versus **arctic_2050**) on the aerosol radiative forcing is shown under clear-sky conditions in early autumn (Sep/Oct). In (a), the reference without additional ship emissions is shown (**arctic_2050**). Hatched areas are significant at the 95% confidence level.

References

Bond, T. C., S. J. Doherty, D. W. Fahey, P. M. Forster, T. Berntsen, B. J. DeAngelo, M. G. Flanner, S. Ghan, B. Kärcher, D. Koch, S. Kinne, Y. Kondo, P. K. Quinn, M. C. Sarofim, M. G. Schultz, M. Schulz, C. Venkataraman, H. Zhang, S. Zhang, N. Bellouin, S. K. Guttikunda, P. K. Hopke, M. Z. Jacobson, J. W. Kaiser, Z. Klimont, U. Lohmann, J. P. Schwarz, D. Shindell, T. Storelvmo, S. G. Warren, and C. S. Zender (2013). "Bounding the role of black carbon in the climate system: A scientific assessment". In: *Geophy. Res. Lett.-Atmos.* 118.11, pp. 5380–5552. ISSN: 2169-8996. DOI: 10.1002/jgrd.50171.

Bond, T. C. and R. W. Bergstrom (2006). "Light Absorption by Carbonaceous Particles: An Investigative Review". In: *Aerosol Science and Technology* 40.1, pp. 27–67. DOI: 10.1080/02786820500421521.

Bourgeois, Q. and I. Bey (2011). "Pollution transport efficiency toward the Arctic: Sensitivity to aerosol scavenging and source regions". In: *J. Geophys. Res.* 116.D8. ISSN: 0148-0227. DOI: 10.1029/2010jd015096.

Chang, R. Y.-W., C. Leck, M. Graus, M. Müller, J. Paatero, J. F. Burkhart, A. Stohl, L. H. Orr, K. Hayden, S.-M. Li, A. Hansel, M. Tjernström, W. R. Leaitch, and J. P. D. Abbatt (2011). "Aerosol composition and sources in the central Arctic Ocean during ASCOS". In: *Atmos. Chem. Phys.* 11.20, pp. 10619–10636. DOI: 10.5194/acp-11-10619-2011.

Eckhardt, S., B. Quennehen, D. J. L. Olivié, T. K. Berntsen, R. Cherian, J. H. Christensen, W. Collins, S. Crepinsek, N. Daskalakis, M. Flanner, A. Herber, C. Heyes, Ø. Hodnebrog, L. Huang, M. Kanakidou, Z. Klimont, J. Langner, K. S. Law, M. T. Lund, R. Mahmood, A. Massling, S. Myriokefalitakis, I. E. Nielsen, J. K. Nøjgaard, J. Quaas, P. K. Quinn, J.-C. Raut, S. T. Rumbold, M. Schulz, S. Sharma, R. B. Skeie, H. Skov, T. Uttal, K. von Salzen, and A. Stohl (2015). "Current model capabilities for simulating black carbon and sulfate concentrations in the Arctic atmosphere: a multi-model evaluation using a comprehensive measurement data set". In: *Atmos. Chem. Phys.* 15.16, pp. 9413–9433. DOI: 10.5194/acp-15-9413-2015.

Hawkins, L. N. and L. M. Russell (2010). "Polysaccharides, Proteins, and Phytoplankton Fragments: Four Chemically Distinct Types of Marine Primary Organic Aerosol Classified by Single Particle Spectromicroscopy". In: *Advances in Meteorology* 2010.ID 612132. DOI: 10.1155/2010/612132.

Intrieri, J. M., C. Fairall, M. Shupe, P. Persson, A. E.L., P. Guest, and R. Moritz (2002). "An annual cycle of Arctic surface cloud forcing at SHEBA". In: *J. Geophys. Res.* 107.C10. ISSN: 0148-0227. DOI: 10.1029/2000jc000439.

Loeb, N. G., D. R. Doelling, H. Wang, W. Su, C. Nguyen, J. G. Corbett, L. Liang, C. Mitrescu, F. G. Rose, and S. Kato (2018). "Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product". In: *Journal of Climate* 31.2, pp. 895–918. DOI: 10.1175/JCLI-D-17-0208.1.

Lohmann, U. and D. Neubauer (2018). "The importance of mixed-phase clouds for climate sensitivity in the global aerosol-climate model ECHAM6-HAM2". In: *Atmos. Chem. Phys. Disc.* 2018, pp. 1–32. DOI: 10.5194/acp-2018-97.

Long, M. S., W. C. Keene, D. J. Kieber, D. J. Erickson, and H. Maring (2011). "A sea-state based source function for size- and composition-resolved marine aerosol production". In: *Atmos. Chem. Phys.* 11.3, pp. 1203–1216. DOI: 10.5194/acp-11-1203-2011.

Matrosov, S. Y., A. V. Korolev, and A. J. Heymsfield (2002). "Profiling Cloud Ice Mass and Particle Characteristic Size from Doppler Radar Measurements". In: *Journal of Atmospheric and Oceanic Technology* 19.7, pp. 1003–1018. DOI: 10.1175/1520-0426(2002)019<1003:PCIMAP>2.0.CO;2.

Sand, M., B. H. Samset, Y. Balkanski, S. Bauer, N. Bellouin, T. K. Berntsen, H. Bian, M. Chin, T. Diehl, R. Easter, S. J. Ghan, T. Iversen, A. Kirkevåg, J.-F. Lamarque, G. Lin, X. Liu,

G. Luo, G. Myhre, T. V. Noije, J. E. Penner, M. Schulz, Ø. Seland, R. B. Skeie, P. Stier, T. Takemura, K. Tsigaridis, F. Yu, K. Zhang, and H. Zhang (2017). "Aerosols at the poles: an AeroCom Phase II multi-model evaluation". In: *Atmos. Chem. Phys.* 17.19, pp. 12197–12218. DOI: 10.5194/acp-17-12197-2017.

Shupe, M. and T. Uttal (2007). *ETL Radar-based Cloud Microphysics Retrievals. Version 1.0. UCAR/NCAR - Earth Observing Laboratory.* https://doi.org/10.5065/D63R0R7P. DOI: doi.org/10.5065/D63R0R7P.

Shupe, M. D., P. Kollias, S. Y. Matrosov, and T. L. Schneider (2004). "Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra". In: *Journal of Atmospheric and Oceanic Technology* 21.4, pp. 660–670. DOI: 10.1175/1520-0426(2004)021<0660:DMCPFD>2.0.CO;2.

Shupe, M. D., T. Uttal, and S. Y. Matrosov (2005). "Arctic Cloud Microphysics Retrievals from Surface-Based Remote Sensors at SHEBA". In: *Journal of Applied Meteorology* 44.10, pp. 1544–1562. DOI: 10.1175/JAM2297.1.

Sinha, P. R., Y. Kondo, M. Koike, J. A. Ogren, A. Jefferson, T. E. Barrett, R. J. Sheesley, S. Ohata, N. Moteki, H. Coe, D. Liu, M. Irwin, P. Tunved, P. K. Quinn, and Y. Zhao (2017). "Evaluation of ground-based black carbon measurements by filter-based photometers at two Arctic sites". In: *Geophy. Res. Lett.-Atmos.* 122.6, pp. 3544–3572. ISSN: 2169-8996. DOI: 10.1002/2016JD025843.

Stohl, A., Z. Klimont, S. Eckhardt, K. Kupiainen, V. P. Shevchenko, V. M. Kopeikin, and A. N. Novigatsky (2013). "Black carbon in the Arctic: the underestimated role of gas flaring and residential combustion emissions". In: *Atmos. Chem. Phys.* 13.17, pp. 8833–8855. DOI: 10.5194/acp-13-8833-2013.

Wang, X. and J. R. Key (2005). "Arctic Surface, Cloud, and Radiation Properties Based on the AVHRR Polar Pathfinder Dataset. Part I: Spatial and Temporal Characteristics". In: *Journal of Climate* 18.14, pp. 2558–2574. DOI: 10.1175/JCLI3438.1.

Westwater, E. R., Y. Han, M. D. Shupe, and S. Y. Matrosov (2001). "Analysis of integrated cloud liquid and precipitable water vapor retrievals from microwave radiometers during the Surface Heat Budget of the Arctic Ocean project". In: *Geophy. Res. Lett.-Atmos.* 106.D23, pp. 32019–32030. ISSN: 2156-2202. DOI: 10.1029/2000JD000055.