Ship-based lidar evaluation of Southern Ocean low clouds in the storm-resolving general circulation model ICON and the ERA5 and MERRA-2 reanalyses

Peter Kuma[1,2,3,4], Frida A.-M. Bender[1,2], Adrian J. McDonald[3], Simon P. Alexander[5,6], Greg M. McFarquhar[7,8], John J. Cassano[9,10,11], Graeme E. Plank[3], Sean Hartery[3*], Simon Parsons[3†], Sally Garrett[12], Alex J. Schuddeboom[3], and Anna Possner[4]

[1]Department of Meteorology (MISU), Stockholm University, Stockholm, Sweden
[2]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[3]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, Aotearoa/New Zealand
[4]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Hesse, Germany
[5]Australian Antarctic Division, Kingston, Tasmania, Australia
[6]Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
[7]Cooperative Institute of Severe and High Impact Weather Research and Operations, University of Oklahoma, Norman, OK, USA
[8]School of Meteorology, University of Oklahoma, Norman, OK, USA
[9]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[10]National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA
[11]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA
[12]New Zealand Defence Force, Wellington, Aotearoa/New Zealand

Key Points:

- Large cloud biases are present in the global storm-resolving model but are lower in several key aspects than in the reanalyses.
- The models tend to underestimate total cloud fraction and very low-level cloud and fog, while overestimating cloud occurrence peak at 500 m.
- A “too few, too bright” problem of underestimated cloud fraction, compensated by overestimated cloud albedo, is present in MERRA-2.

*Current affiliation: Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Canada

†Current affiliation: New South Wales Department of Planning and Environment, Sydney, New South Wales, Australia

Corresponding author: Peter Kuma, `peter@peterkuma.net`

Abstract

Global storm resolving models (GSRMs) represent the next generation of global climate models. One of them is a 5-km Icosahedral Nonhydrostatic Weather and Climate Model (ICON). Its high resolution means that parameterizations of convection and clouds, including subgrid-scale clouds, are omitted, relying on explicit simulation but necessarily utilizing microphysics and turbulence parameterizations. Standard-resolution (10–100 km) models, which use convection and cloud parameterizations, have substantial cloud biases over the Southern Ocean (SO), adversely affecting radiation and sea surface temperature. The SO is dominated by low clouds, which cannot be observed accurately from space due to overlapping clouds, attenuation, and ground clutter. We evaluated SO clouds in ICON and the ERA5 and MERRA-2 reanalyses using approximately 2400 days of lidar observations and 2300 radiosonde profiles from 31 voyages and a Macquarie Island station during 2010–2021, compared to the models using a ground-based lidar simulator. We found that ICON and the reanalyses underestimate the total cloud fraction by about 10 and 20%, respectively. ICON and ERA5 overestimate the cloud occurrence peak at about 500 m, associated with underestimated lower tropospheric stability and overestimated lifting condensation level. The reanalyses strongly underestimate fog and very low-level clouds, and MERRA-2 underestimates cloud occurrence at almost all heights. Outgoing shortwave radiation is overestimated in MERRA-2, implying a "too few, too bright" cloud problem. SO cloud and fog biases are a substantial issue in the analyzed models and result in shortwave and longwave radiation biases.

Plain Language Summary

Global storm-resolving models are climate models with km-scale horizontal resolution, which are currently in development. Reanalyses are the best estimates of past meteorological conditions based on an underlying global model and observations. We evaluated clouds, temperature, and humidity profiles over the Southern Ocean in one such model, ICON and two reanalyses, based on 2400 days of ship and station observations. Thanks to the high resolution, ICON relies entirely on explicit simulation of clouds instead of subgrid-scale parameterizations. For the evaluation, we used ceilometer and radiosonde observations and a lidar simulator, which enables a fair comparison with ICON and reanalyses. We subset our results by cyclonic activity and stability. We found that ICON and reanalyses underestimate lidar-derived cloud fraction, and the reanalyses do so more strongly. Fog and very low-level clouds are especially underestimated in the reanalyses. However, ICON and one of the reanalyses also tend to overestimate the peak of cloud occurrence at 500 m above the ground, and it tends to be higher. This is linked to thermodynamic profiles, which show a higher lifting condensation level and lower stability. Southern Ocean cloud and fog biases are an important problem in the analyzed models and result in radiation balance biases.

1 Introduction

Increasing climate model spatial resolution is one way of improving the accuracy of the representation of the climate system in models (Mauritsen et al., 2022). It has been practiced since the advent of climate modeling as more computational power, memory, and storage capacity become available. It is, however, often not as easy as changing the grid size because of the complex interplay between model dynamics and physics, which necessitates adjusting and tuning all components together. Increasing resolution is, of course, limited by the available computational power and a trade-off with increasing parameterization complexity, which is another way of improving model accuracy. Current computational availability and acceleration from general-purpose computing on graphics processing units has progressed to enable km-scale (also called k-scale) Earth system models (ESMs) and coupled atmosphere–ocean general circulation models for research today and will become operational in the future. Therefore, it represents a natural advance in climate modeling. Global storm-resolving models (GSRMs) are emerging as a new front in the development of high-resolution global climate models, with horizontal grid resolutions of about 2–8 km (Satoh et al., 2019; Stevens et al., 2019). This resolution is enough to resolve mesoscale convective storms, but smaller-scale convective plumes and cloud structure remain unresolved. At an approximately 5-km scale, non-hydrostatic processes also become important (Weisman et al., 1997), and for this reason such models are generally non-hydrostatic. The terms global cloud-resolving models or global convection-permitting/-resolving models are also sometimes used interchangeably with GSRMs but imply that clouds or convection are resolved explicitly, which is not entirely true for GSRMs, as this would require an even higher horizontal resolution (Satoh et al., 2019). Representative of these efforts is the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) project (Stevens et al., 2019; DYAMOND author team, 2024), which is an intercomparison of nine global GSRMs over two 40-day time periods in summer (1 August–10 September 2016) and winter (20 January–1 March 2020). A new one-year GSRM intercomparison is currently proposed by Takasuka et al. (2024), with the hope of also evaluating the seasonal cycle and large-scale circulation. An alternative to using a computationally costly GSRM is to train an artificial neural network on GSRM output and use it for subgrid-scale clouds, as done with the GSRM ICON by Grundner et al. (2022) and Grundner (2023).

The main aim of this study is to evaluate the GSRM version of ICON developed by the nextGEMS project (nextGEMS authors team, 2024; Segura et al., 2025). ICON is developed and maintained jointly by Deutscher Wetterdienst, the Max-Planck-Institute for Meteorology, Deutsches Klimarechenzentrum (DKRZ), Karlsruhe Institute of Technology, and the Center for Climate Systems Modeling. Our aim is to quantify how well the GSRM ICON simulates clouds over the Southern Ocean (SO), particularly in light of the fact that subgrid-scale clouds and convection are not parameterized in this model. This region is mostly dominated by boundary layer clouds generated by shallow convection, and these are problematic to observe by spaceborne lidars and radars, which are affected by attenuation by overlapping and thick clouds (Mace et al., 2009; Medeiros et al., 2010) and ground clutter (Marchand et al., 2008), respectively. Specifically, the radar on CloudSat and lidar on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO), neither of which are operational any more, are affected by the above-mentioned issues, resulting in a strong underestimation of cloud occurrence below 2 km in a merged CloudSat–CALIPSO product relative to ground-based lidar observations at McMurdo Station (McErlich et al., 2021). Removing situations with higher overlapping clouds could enable a less biased comparison of low clouds. We hypothesize that this, in turn, can lead to systematic biases in low clouds in climate models, which are frequently evaluated against CloudSat–CALIPSO products. Reanalyses can also suffer from cloud biases, as these are usually parameterized in their atmospheric component and also in regions where input observations are sparse. This makes them a problematic reference for clouds over the SO, and any biases relative to a reanalysis should be interpreted with caution. Instead, we chose to use a large set of ship-based observations conducted with ceilometers and lidars on board the research vessel (RV) *Polarstern* and other ships and a station as a reference for the model evaluation. Altogether, we analyzed approximately 2400 days of data from 31 voyages and a sub-Antarctic station covering diverse longitudes and latitudes of the SO. To achieve a like-for-like comparison with the models (ICON, MERRA-2, and ERA5), we used a ground-based lidar simulator called the Automatic Lidar and Ceilometer Framework [ALCF; Kuma et al. (2021)]. We contrasted the results with the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5 [ERA5; ECMWF (2019)] and the Modern-Era Retrospective analysis for Research and Applications, Version 2 [MERRA-2; Gelaro et al. (2017)].

The nextGEMS project focuses on the research and development of GSRMs at multiple modeling centers and universities in Europe. The project also develops GSRM versions of the Icosahedral Nonhydrostatic Weather and Climate Model (ICON; Hohenegger et al. (2023)), the Integrated Forecasting System [IFS; ECMWF (2023)], and their ocean components at eddy-resolving resolutions: ICON-O (Korn et al., 2022) coupled with ICON and Finite-Element/volumE Sea ice-Ocean Model [FESOM; Q. Wang et al. (2014)] and Nucleus for European modeling of the Ocean [NEMO; Madec and the NEMO System Team (2023)] coupled with IFS. The project has so far produced ICON and IFS simulations with three development versions called Cycle 1–3 and a pre-final version, with a final production version planned by the end of the project. nextGEMS is not the only project developing GSRMs; other GSRMs (or GSRM versions of climate models) currently in development include: Convection-Permitting Simulations With the E3SM Global Atmosphere Model [SCREAM; Caldwell et al. (2021)], Non-hydrostatic Icosahedral Atmospheric Model [NICAM; Satoh et al. (2008)], Unified Model (UM), eXperimental System for High-resolution modeling for Earth-to-Local Domain [X-SHiELD; SHiELD authors team (2024)], Action de Recherche Petite Echelle Grande Echelle-NonHydrostatic version [ARPEGE-NH; Bubnová et al. (1995); Voldoire et al. (2017)], Finite-Volume Dynamical Core on the Cubed Sphere [FV3, Lin (2004)], the National Aeronautics and Space Administration (NASA) Goddard Earth Observing System global atmospheric model version 5 [GEOS5; Putman and Suarez (2011)], Model for Prediction Across Scales [MPAS; Skamarock et al. (2012)], and System for Atmospheric Modeling [SAM; Khairoutdinov and Randall (2003)].

Multiple cloud properties have an effect on shortwave (SW) and longwave (LW) radiation. To first order, the total cloud fraction, cloud phase, and the liquid and ice water path (LWP and IWP) are the most important cloud properties influencing SW and LW radiation. These properties are in turn influenced by the atmospheric thermodynamics, convection and circulation, and both the indirect and direct effects of aerosols. Second-order effects on SW and LW radiation are associated with the cloud droplet size distribution, ice crystal habit, cloud lifetime, and direct radiative interaction with aerosols (Boucher et al., 2013). In the 6$^{\text{th}}$ phase of the Coupled Model Intercomparison Project [CMIP6; Eyring et al. (2016)], the cloud feedback has increased relative to CMIP5 (Zelinka et al., 2020), especially in the Southern Hemisphere mid-to-high latitudes, which is one of the main reasons for the higher climate sensitivity of CMIP6 models.

The SO is known to be a problematic region for climate model biases (A. J. Schuddeboom & McDonald, 2021; Hyder et al., 2018; Cesana et al., 2022; Zhao et al., 2022) due to a lack of surface and in situ observations. This region has also long been a lower priority region for numerical weather prediction (NWP) and climate model development because of its distance from populated areas. Nevertheless, radiation biases and changes over an area of its size have a substantial influence on the global climate (Rintoul, 2011; Bodas-Salcedo et al., 2012), such as affecting the Earth's radiation balance, ocean heat, and carbon uptake (R. G. Williams et al., 2023), and the SO is also an important part of the global ocean conveyor belt (C. Wang et al., 2014). In general, marine clouds have a disproportionate effect on top-of-atmosphere (TOA) SW radiation due to the relatively low albedo of the sea surface. The relative longitudinal symmetry of the SO means that model cloud biases tend to be similar across longitudes.

In the following text, we refer to the SO as ocean regions south of 40°S, low-latitude SO as 40–55°S, and high-latitude SO as south of 55°S, all the way to the Antarctic coast. The reason for this dividing latitude is to split the SO into about two equal zones, as well as the results by A. J. Schuddeboom and McDonald (2021) (Fig. 2b) which show a contrast in CMIP model radiation biases. A. Schuddeboom et al. (2019) (Fig. 2) and Kuma et al. (2020) (Fig. 3) also show contrasting radiation biases in the Hadley Centre Global Environmental Model, which is also supported by Cesana et al. (2022), displaying contrasting cloud biases due to the 0°C isotherm reaching the surface at 55°S. The findings of Niu et al. (2024), however, support a different dividing line of 62°S based on cloud condensation nuclei concentration.

SO radiation biases have been relatively large and systematic compared to the rest of the globe since at least CMIP3 (Trenberth & Fasullo, 2010; Bodas-Salcedo et al., 2012), and the SO SW cloud radiative effect bias is still positive in eight CMIP6 models analyzed by A. J. Schuddeboom and McDonald (2021) over the high-latitude SO, whereas over the low-latitude SO it tends to be more neutral or negative in some models. Too much absorbed SW radiation over the SO was also identified in the GSRM SCREAM (Caldwell et al., 2021). Compensating biases are possible, such as the "too few too bright" cloud bias, characterized by too small a cloud fraction and too large a cloud albedo (Wall et al., 2017; Kuma et al., 2020), previously described by Webb et al. (2001), Weare (2004), M. H. Zhang et al. (2005), Karlsson et al. (2008), Nam et al. (2012), Klein et al. (2013), and Bender et al. (2017) in other regions and models, which means that a model can maintain a reasonable SW radiation balance by reflecting too much SW radiation from clouds, but these cover too small an area. A study by Konsta et al. (2022) showed that this type of bias is still present in six analyzed CMIP6 models in tropical marine clouds, using the General-circulation-model-Oriented CALIPSO Cloud Product [CALIPSO–GOCCP; Chepfer et al. (2010)] and Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar [PARASOL; Lier and Bach (2008)] as a reference. They suggest improper simulation of subgrid-scale cloud heterogeneity as a cause. Compensating cloud biases in the Australian Community Climate and Earth System Simulator (ACCESS) – Atmosphere-only model version 2 (AM2) over the SO were analyzed by Fiddes et al. (2022) and Fiddes et al. (2024). Possner et al. (2022) showed that over the SO, the DYAMOND GSRM ICON underestimates low-level cloud fraction on the order of 30% and overestimates net downward TOA SW radiation by approximately 10 $\mathrm{Wm^{-2}}$ in the highest model resolution run (2.5 km). Zhao et al. (2022) reported a similar SW radiation bias in five analyzed CMIP6 models over the high-latitude SO and an underestimation of the total cloud fraction on the order of 10% over the entire 40–60°S SO. Recently, Ramadoss et al. (2024) analyzed 48 hours of km-scale ICON limited-area model NWP simulations over an SO region adjacent to Tasmania against the Clouds, Aerosols, Precipitation, Radiation, and atmospherIc Composition Over the southeRn oceaN (CAPRICORN) voyage cloud and precipitation observations (McFarquhar et al., 2021). They found the ICON cloud optical thickness was underestimated relative to Himawari-8 satellite observations but also identified large differences in cloud top phase.

In general, sea surface temperature (SST) biases in the SO can originate either in the atmosphere (Hyder et al., 2018), caused by too much SW heating of the surface or too little LW cooling of the surface, such as in situations of too much cloud cover or cloud optical thickness, or in the ocean circulation. Interactions of both are also possible; for example, SST affecting clouds and clouds affecting the surface radiation. Using ERA5 as a reference, Q. Zhang et al. (2023) have shown that SST biases have improved in CMIP6 compared to CMIP5, with SST overall increasing in CMIP6. However, over the SO, this resulted in an even higher positive bias, especially in the Atlantic Ocean (AO) sector of the SO, increasing by up to 1°C. Luo et al. (2023) identified that the SO SST bias in an ensemble of 18 CMIP6 models originates not from the surface heat and radiation fluxes (using reanalyses as a reference) but from a warm bias in the Northern Atlantic Deep Water.

2 Methods

2.1 Voyage and Station Data

Together, we analyzed data from 31 voyages of RV *Polarstern*, the resupply vessel (RSV) *Aurora Australis*, RV *Tangaroa*, RV *Nathaniel B. Palmer*, Her (now His) Majesty's New Zealand Ship (HMNZS) *Wellington*, and one sub-Antarctic station (Macquarie Island) in the SO south of 40°S between 2010 and 2021. Fig. 1 shows a map of the campaigns, Table 1 lists the campaigns, and Table 2 lists references where available. The analyzed dataset comprised 2421 days of data south of 40°S, but the availability of ceilometer data was slightly shorter due to gaps in measurements.

The campaigns contained ceilometer observations captured by the Vaisala CL51, CT25K, and the Lufft CHM 15k, described in detail below (Sections 2.2 and 2.3). A ceilometer is a low-power, near-infrared, vertically pointing lidar principally designed to measure cloud base, but they also measure the full vertical structure of clouds as long as the laser signal is not attenuated by thick clouds, which can be used to infer additional information such as a cloud mask and cloud occurrence by height. We note that during the MICRE campaign, the ceilometers Vaisala CT25K and CL51 were installed at the Macquarie Island station concurrently, but in our analysis we only used the CT25K data obtained from the Atmospheric Radiation Measurement (ARM) data archive.

Apart from lidar observations, radiosondes were launched on weather balloons at regular synoptic times on the RV *Polarstern*, MARCUS, NBP17024, TAN1702, and TAN1802 campaigns, measuring pressure, temperature, relative humidity (RH), and the global navigation satellite system coordinates. In total, about 2300 radiosonde profiles south of 40°S

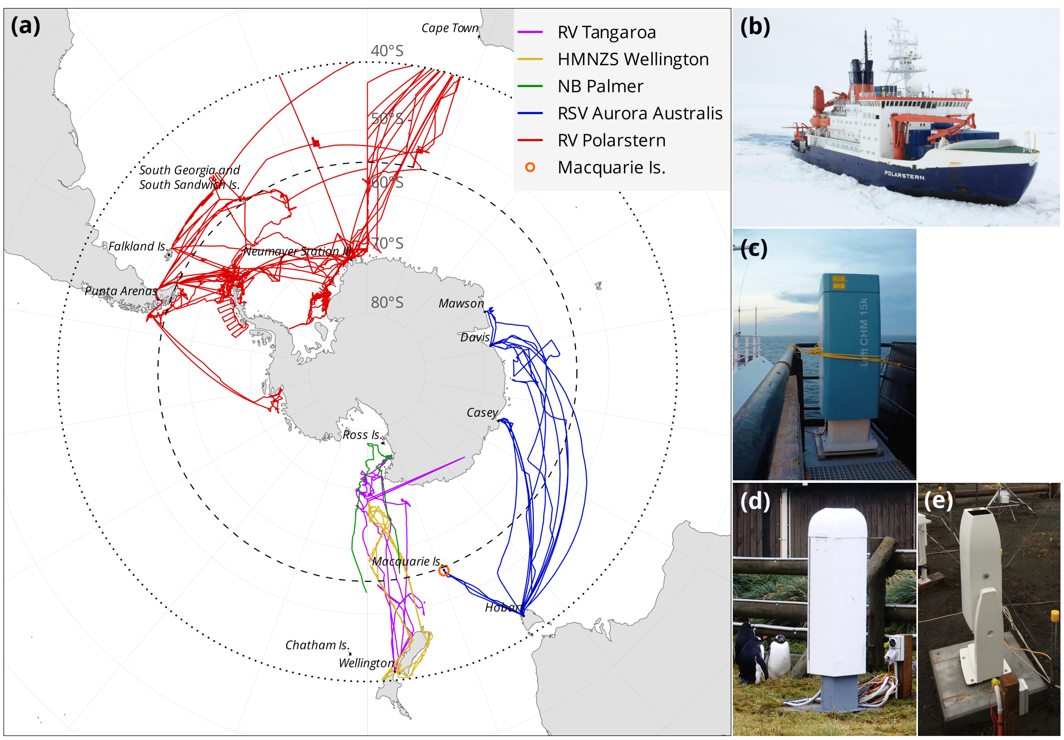

Figure 1. **(a)** A map showing the tracks of 31 voyages of RV *Polarstern*, RSV *Aurora Australis*, RV *Tangaroa*, RV *Nathaniel B. Palmer*, and HMNZS *Wellington* and one sub-Antarctic station (Macquarie Island) analyzed here. The tracks cover Antarctic sectors south of South America, the Atlantic Ocean, Africa, Australia, and New Zealand in the years 2010–2021 (inclusive). The dotted and dashed lines at 40°S and 55°S delineate the Southern Ocean area of our analysis and its partitioning into two subsets, respectively. A photo of **(b)** RV *Polarstern* (© Folke Mehrtens, Alfred-Wegener-Institut), **(c)** Lufft CHM 15k installed on RV *Tangaroa* (© Peter Kuma, University of Canterbury), **(d)** Vaisala CL51 (© Jeff Aquilina, Bureau of Meteorology), **(e)** Vaisala CT25K at Macquarie Island (© Simon P. Alexander, Australian Antarctic Division).

were available. Spatially and temporally collocated profiles were taken from the models. Because the time period covered by the ICON model output (2021–2024) was different from the time period covered by the observations (2010–2021), when comparing with ICON, we first had to remap the observation time to model time by taking the same time relative to the start of the year. Consequently, we also had four virtual/model profiles (one for each year from 2021 to 2024) for each observed profile. Derived thermodynamic [virtual potential temperature (θ_v), lifting condensation level (LCL), etc.] and dynamic physical quantities (wind speed and direction) for the measured vertical profiles were calculated with the program radiosonde tool [rstool; Kuma (2024d)]. Surface meteorological quantities were measured continuously by an onboard automatic weather station or individual instruments.

Some of the observational data were likely used in the assimilation of the reanalyses. The Macquarie Island station surface measurements and radiosonde profiles (not used in our analysis) were sent to the World Meteorological Organization Global Telecommunication System (GTS). The measurements on the RSV *Aurora Australis* and HMNZS *Wellington* were not used outside of research purposes. The AWS measurements, but not lidar or radiosonde measurements on the RV *Tangaroa* voyages, were collected by the New Zealand MetService and communicated to the GTS. The ceilometer measurements on NBP1704 were not used outside of research purposes.

2.2 Vaisala CL51 and CT25K

The Vaisala CL51 and CT25K (photos in Fig. 1d, e) are ceilometers operating at near-infrared wavelengths of 910 nm and 905 nm, respectively. The CL51 can also be configured to emulate the Vaisala CL31. The maximum range is 15.4 km (CL51), 7.7 km (CL31 emulation mode with 5 m vertical resolution), and 7.5 km (CT25K). The vertical resolution is 10 m (5 m configurable) in CL51 and 30 m in CT25K observations. The sampling (temporal) resolution is configurable, and in our datasets, it is approximately 6 s for CL51 on AA15-16, 16 s for CT25K on MARCUS and MICRE, 36 s for CL51 on RV *Polarstern*, and about 2.37 s for CL51 with CL31 emulation on TAN1502. The wavelengths of 905 and 910 nm are both affected by water vapor absorption of about 20% in the mid-latitudes (Wiegner & Gasteiger, 2015; Wiegner et al., 2019), with 910 nm affected more strongly, but we do not expect this to be a significant issue, as explained in Kuma et al. (2021). The instrument data files containing raw uncalibrated backscatter were first converted to the Network Common Data Form (NetCDF) with cl2nc (Kuma, 2024c) and then processed with the ALCF (Section 2.4) to produce absolutely calibrated attenuated volume backscattering coefficient (AVBC), cloud mask, cloud occurrence by height, and the total cloud fraction. Because the CT25K uses a very similar wavelength to the CL51, equivalent calculations as for the CL51 were done assuming a wavelength of 910 nm. The Vaisala CL51 and CT25K instruments were used on most of the voyages and stations analyzed here. Fig. 2a shows an example of AVBC derived from the CL51 instrument data.

2.3 Lufft CHM 15k

The Lufft CHM 15k (photo in Fig. 1c) ceilometer operates at a near-infrared wavelength of 1064 nm. The maximum range is 15.4 km; the vertical resolution is 5 m in the near range (up to 150 m) and 15 m above; the sampling (temporal) resolution is 2 s; and the number of vertical levels is 1024. NetCDF files containing uncalibrated backscatter produced by the instrument were processed with the ALCF (Section 2.4) to produce AVBC, cloud mask, cloud occurrence by height, and the total cloud fraction. The CHM 15k was used on four voyages (HMNZSW16, TAN1702, TAN1802, and NBP1704).

Table 1. An overview of the analyzed campaigns (voyages and stations). Start, end, and the number of days (UTC; inclusive) refer to the time period when the vessel was south of 40°S. Abbreviations: ceilometer (ceil.), Australia (AU), New Zealand (NZ), South America (SA), Atlantic Ocean (AO), and Africa (AF). The number of days is rounded to the nearest integer. CL51/31 indicates CL51 configured to emulate CL31. Missing days in the ceilometer data were HMNZSW16 (7 days): 24–27 November, 10 December, and 16–17 December 2016; MARCUS (3 days): 8, 10 November, and 10 December 2017; MICRE (9 days): 7–8, 29 June, 5, 16 July, 15 August, 17 October 2016, 11 February, and 21 March 2017; and TAN1502 (1 day): 24 January.

Name	**Vessel or station**	**Ceil.**	**Region**	**Start**	**End**	**Days**
AA15-16	RSV *Aurora Australis*	CL51	AU	2015-10-22	2016-02-22	124
HMNZSW16	HMNZS *Wellington*	CHM 15k	NZ	2016-11-23	2016-12-19	27
MARCUS	RSV *Aurora Australis*	CT25K	AU	2017-10-29	2018-03-26	149
MICRE	Macquarie Is. station	CT25K	AU/NZ	2016-04-03	2018-03-14	710
NBP1704	RV *Nathaniel B. Palmer*	CHM 15k	NZ	2017-04-14	2017-06-08	55
PS77/2	RV *Polarstern*	CL51	SA/AO/AF	2010-12-01	2011-02-04	65
PS77/3	RV *Polarstern*	CL51	SA/AO/AF	2011-02-07	2011-04-14	66
PS79/2	RV *Polarstern*	CL51	SA/AO/AF	2011-12-06	2012-01-02	27
PS79/3	RV *Polarstern*	CL51	SA/AO/AF	2012-01-10	2012-03-10	61
PS79/4	RV *Polarstern*	CL51	SA/AO/AF	2012-03-14	2012-04-08	26
PS81/2	RV *Polarstern*	CL51	SA/AO/AF	2012-12-02	2013-01-18	47
PS81/3	RV *Polarstern*	CL51	SA/AO/AF	2013-01-22	2013-03-17	55
PS81/4	RV *Polarstern*	CL51	SA/AO/AF	2013-03-18	2013-04-16	30
PS81/5	RV *Polarstern*	CL51	SA/AO/AF	2013-04-20	2013-05-23	33
PS81/6	RV *Polarstern*	CL51	SA/AO/AF	2013-06-10	2013-08-12	63
PS81/7	RV *Polarstern*	CL51	SA/AO/AF	2013-08-15	2013-10-14	60
PS81/8	RV *Polarstern*	CL51	SA/AO/AF	2013-11-12	2013-12-14	31
PS81/9	RV *Polarstern*	CL51	SA/AO/AF	2013-12-21	2014-03-02	71
PS89	RV *Polarstern*	CL51	SA/AO/AF	2014-12-05	2015-01-30	56
PS96	RV *Polarstern*	CL51	SA/AO/AF	2015-12-08	2016-02-14	68
PS97	RV *Polarstern*	CL51	SA/AO/AF	2016-02-15	2016-04-06	52
PS103	RV *Polarstern*	CL51	SA/AO/AF	2016-12-18	2017-02-02	46
PS104	RV *Polarstern*	CL51	SA/AO/AF	2017-02-08	2017-03-18	39
PS111	RV *Polarstern*	CL51	SA/AO/AF	2018-01-21	2018-03-14	52
PS112	RV *Polarstern*	CL51	SA/AO/AF	2018-03-18	2018-05-05	49
PS117	RV *Polarstern*	CL51	SA/AO/AF	2018-12-18	2019-02-07	51
PS118	RV *Polarstern*	CL51	SA/AO/AF	2019-02-18	2019-04-08	50
PS123	RV *Polarstern*	CL51	SA/AO/AF	2021-01-10	2021-01-31	21
PS124	RV *Polarstern*	CL51	SA/AO/AF	2021-02-03	2021-03-30	55
TAN1502	RV *Tangaroa*	CL51/31	NZ	2015-01-20	2015-03-12	51
TAN1702	RV *Tangaroa*	CHM 15k	NZ	2017-03-09	2017-03-31	23
TAN1802	RV *Tangaroa*	CHM 15k	NZ	2018-02-07	2018-03-20	41
Total						**2421**

Table 2. Campaign publication references.

Name	**References**
AA15-16	Klekociuk et al. (2020)
MARCUS	McFarquhar et al. (2021); Xia and McFarquhar (2024); Niu et al. (2024)
MICRE	McFarquhar et al. (2021)
NBP1704	Ackley et al. (2020)
PS77/2	König-Langlo (2011e, 2011a, 2011c, 2014h); Fahrbach and Rohardt (2011)
PS77/3	König-Langlo (2011d, 2011b, 2012g, 2014i); Knust and Rohardt (2011)
PS79/2	König-Langlo (2012h, 2012d, 2012a, 2014j); Kattner and Rohardt (2012)
PS79/3	König-Langlo (2012i, 2012b, 2012e, 2014k); Wolf-Gladrow and Rohardt (2012)
PS79/4	König-Langlo (2012j, 2012c, 2012f, 2014l); Lucassen and Rohardt (2012)
PS81/2	König-Langlo (2013l, 2013a, 2013f, 2014a); Boebel and Rohardt (2013)
PS81/3	König-Langlo (2013m, 2013g, 2013b, 2014b); Gutt and Rohardt (2013)
PS81/4	König-Langlo (2013n, 2013c, 2013h, 2014c); Bohrmann and Rohardt (2013)
PS81/5	König-Langlo (2013o, 2013d, 2013i, 2014d); Jokat and Rohardt (2013)
PS81/6	König-Langlo (2013p, 2013e, 2013j, 2014e); Lemke and Rohardt (2013)
PS81/7	König-Langlo (2013q, 2013k, 2014f, 2016c); Meyer and Rohardt (2013)
PS81/8	König-Langlo (2013r, 2014g, 2014n, 2014p); Schlindwein and Rohardt (2014)
PS81/9	König-Langlo (2014r, 2014m, 2014o, 2014q); Knust and Rohardt (2014)
PS89	König-Langlo (2015a, 2015d, 2015b, 2015c); Boebel and Rohardt (2016)
PS96	König-Langlo (2016h, 2016a, 2016d, 2016f); Schröder and Rohardt (2017)
PS97	König-Langlo (2016i, 2016e, 2016b, 2016g); Lamy and Rohardt (2017)
PS103	König-Langlo (2017f, 2017d, 2017a, 2017c); Boebel and Rohardt (2018)
PS104	König-Langlo (2017e, 2017g, 2017b); Gohl and Rohardt (2018); Schmithüsen (2021g)
PS111	Schmithüsen (2019a, 2020a, 2021h, 2021a); Schröder and Rohardt (2018)
PS112	Schmithüsen (2019b, 2020b, 2021b, 2021i); Meyer and Rohardt (2018)
PS117	Schmithüsen (2019c, 2020c, 2021j, 2021c); Boebel and Rohardt (2019)
PS118	Schmithüsen (2019d, 2020d, 2021d, 2021k); Dorschel and Rohardt (2019)
PS123	Schmithüsen (2021m, 2021e, 2021l); Schmithüsen, Jens, and Wenzel (2021); Hoppmann, Tippenhauer, and Heitland (2023)
PS124	Schmithüsen (2021n, 2021f); Schmithüsen, Rohleder, et al. (2021); Hoppmann, Tippenhauer, and Hellmer (2023)
TAN1802	Kremser et al. (2020, 2021)

2.4 ALCF

The Automatic Lidar and Ceilometer Framework (ALCF) is a ground-based lidar simulator and a tool for processing observed lidar data, supporting various instruments and models (Kuma et al., 2021). It performs radiative transfer calculations to derive equivalent lidar AVBC from an atmospheric model, which can then be compared with observed AVBC. For this purpose, it takes the cloud fraction, liquid and ice mass mixing ratio, temperature, and pressure model fields as an input and is run offline (on the model output rather than inside the model code). The lidar simulator in the ALCF is based on the instrument simulator Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011). After AVBC is calculated, a cloud mask, cloud occurrence by height, and the total cloud fraction are determined. The total cloud fraction is defined as the fraction of profiles with clouds at any height in the lidar cloud mask. The ALCF has in the past been used by several research teams for model and reanalysis evaluation (Kuma et al., 2020; Kremser et al., 2021; Guyot et al., 2022; Pei et al., 2023; Whitehead et al., 2023; McDonald, Kuma, et al., 2024).

Absolute calibration of the observed backscatter was performed by comparing the measured clear-sky molecular backscatter statistically with simulated clear-sky molecular backscatter. AVBC was resampled to 5 min temporal resolution and 50 m vertical resolution to increase the signal-to-noise ratio while having enough resolution to detect small-scale cloud variability. The noise standard deviation was calculated from AVBC at the highest range, where no clouds are expected. A cloud mask was calculated from AVBC using a fixed threshold of $2 \times 10^{-6}\mathrm{m}^{-1}\mathrm{sr}^{-1}$ after subtracting 5 standard deviations of range-scaled noise. Fig. 2b shows an example of simulated Vaisala CL51 backscatter from ERA5 data, corresponding to a day of measurements by the instrument on the PS81/3 voyage.

How attenuation of the lidar signal affects cloud detection is dependent on factors such as the optical thickness of the measured cloud and its backscattering phase function, as well as the range-dependent noise standard deviation (Kuma et al., 2021). A rough estimate can be made under an assumption of a relatively strongly backscattering cloud of $\beta = 100 \times 10^{-6}\mathrm{m}^{-1}\mathrm{sr}^{-1}$ at a height of $r_1 = 2$ km, range-dependent noise β_n at $r_2 = 8$ km of about $5 \times 10^{-6}\mathrm{m}^{-1}\mathrm{sr}^{-1}$, and cloud detection threshold $\beta_t = 2 \times 10^{-6}\mathrm{m}^{-1}\mathrm{sr}^{-1}$, noise multiplication factor $f = 5$. At full attenuation (relative to the detection threshold), the two-way attenuation factor A satisfies $A\beta = \beta_t + f \times \beta_n \left(\frac{r_1}{r_2}\right)^2$. This is equiv-

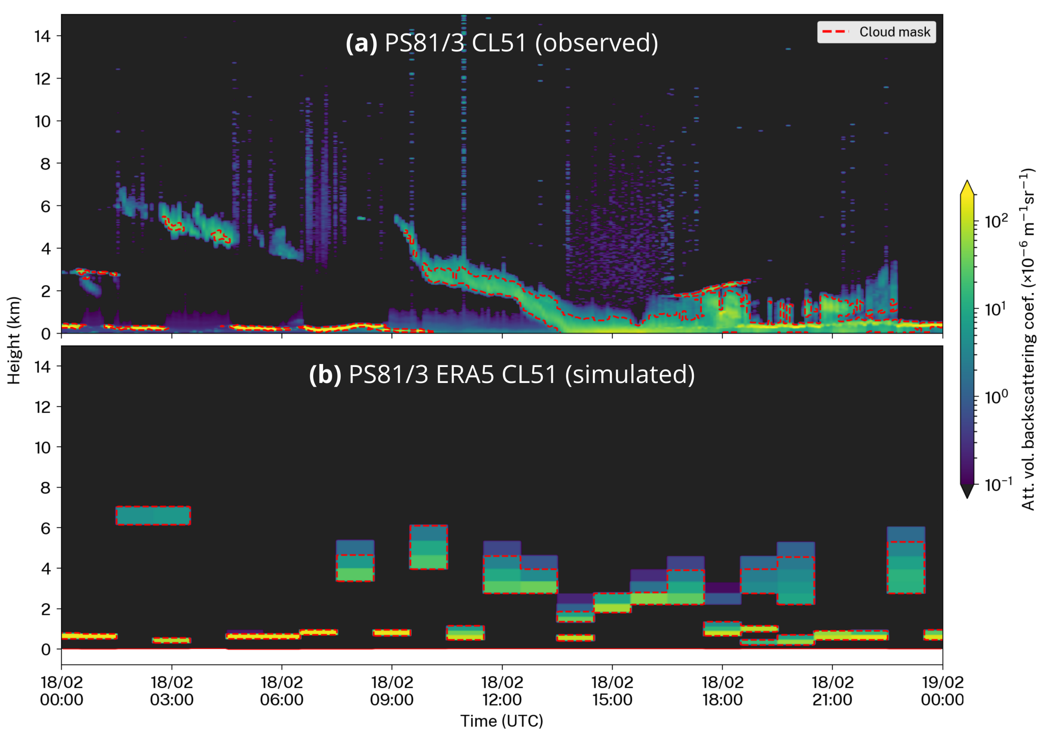

Figure 2. An example of the attenuated volume backscattering coefficient (AVBC) **(a)** measured by the CL51 during 24 hours on the PS81/3 voyage and **(b)** an equivalent AVBC simulated with the ALCF from ERA5 data during the same time period. The red line identifies the cloud mask determined by the ALCF.

alent to exponential decay ($A = e^{-2\delta}$) with optical depth δ (at the lidar wavelength) of about 1.7.

2.5 ICON

A coupled (atmosphere–ocean) GSRM version of the ICON model is in development as part of the nextGEMS project (Hohenegger et al., 2023). ICON is a very flexible model, allowing for simulations ranging from coarse-resolution ESM simulations, GSRM simulations, limited area model simulations, and large eddy simulations (LES) for both weather prediction and climate projections. ICON uses the atmospheric component ICON-A (Giorgetta et al., 2018), whose physics is derived from ECHAM6 (Stevens et al., 2013), and the ocean component ICON-O (Korn et al., 2022). Earlier runs of the GSRM ICON from DYAMOND were evaluated by Mauritsen et al. (2022).

Here, we use a free-running (i.e., the weather conditions in the model do not correspond to reality) coupled GSRM simulation made for the purpose of climate projection. nextGEMS has so far produced four cycles of model runs. We used a Cycle 3 run *ngc3028* produced in 2023 (Koldunov et al., 2023; nextGEMS authors team, 2023) for a model time period of 20 January 2020 to 22 July 2025, of which we analyzed the period 2021–2024 (inclusive). The horizontal resolution of ngc3028 is about 5 km. The model output is available on 90 vertical levels and 3-hourly instantaneous temporal resolution.

Unlike current general circulation models, the storm-resolving version of ICON does not use convective and cloud parameterization but relies on explicit simulation of convection and clouds on the model grid. Subgrid-scale clouds are not resolved, and the grid cell cloud fraction is always either 0 or 100%. While this makes the code development simpler without having to rely on uncertain parameterizations, it can miss smaller-scale clouds below the grid resolution. Turbulence and cloud microphysics have to be parameterized in this model as in other models, and aerosols are derived from a climatology. To account for the radiative effects of subgrid-scale clouds, a cloud inhomogeneity factor is introduced in the model, which scales down the cloud liquid water for radiative calculations. It ranges from 0.4 at lower tropospheric stability (LTS) of 0 K to 0.8 at 30 K. In addition, turbulent mixing in the Smagorinsky scheme was adjusted to allow mixing or entrainment in situations of no mixing under the traditional scheme, affecting stratocumulus clouds but not trade wind clouds (Segura et al., 2025).

Because the analyzed ICON simulation was free-running (years 2021–2024, inclusive), weather and climate oscillations [such as the El Niño–Southern Oscillation (ENSO) phase] are not expected to be equivalent to reality. To compare with the observations collected during a different time period (years 2010–2021, inclusive), we compared the model output with observations at the same time of year and geographical location, as determined for each data point, such as a lidar profile or a radiosonde launch. In the ALCF, this was done using the *override_year* option.

Due to our comparison being long-term and large-scale, it is expected that a comparison between the free-running model and observations is statistically robust, despite weather-related differences between the two. Furthermore, the results from multiple campaigns are combined in a way that equal statistical weight is given to each campaign, eliminating an outsize influence of longer campaigns, allowing us to estimate uncertainty ranges under the assumption of independence of weather conditions between the campaigns, and ensuring that the results are statistically representative over the whole area covered by the campaigns. Different approaches to a comparison would be possible. For example, one could use only the first several days of a free-running simulation initialized from observations (or a reanalysis) for a comparison, as done in the Transpose-AMIP experiments (K. D. Williams et al., 2013), thus being able to compare clouds and the physical drivers under the same weather conditions. Another possibility is the use of a model nudged to a reanalysis (Kuma et al., 2020), but this was not available for our ICON simulations. We discuss further the implications of comparing the observations with a free-running model in Section 4.

2.6 MERRA-2

The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) is a reanalysis produced by the Global Modeling and Assimilation Office at the NASA Goddard Space Flight Center (Gelaro et al., 2017). It uses version 5.12.4 of the Goddard Earth Observing System (GEOS) atmospheric model (Rienecker et al., 2008; Molod et al., 2015). Non-convective clouds (condensation, autoconversion, and evaporation) are parameterized using a prognostic scheme (Bacmeister et al., 2006), and subgrid cloud fraction is determined using total water distribution and a critical RH threshold. The reanalysis output analyzed here is available at a spatial resolution of 0.5° of latitude and 0.625° of longitude, which is about 56 km in the north–south direction and 35 km in the east–west direction at 60°S. The number of vertical model levels is 72. Here, we use the following products: 1-hourly instantaneous 2D single-level diagnostics (M2I1NXASM) for 2-m temperature and humidity; 3-hourly instantaneous 3D assimilated meteorological fields (M2I3NVASM) for cloud quantities, pressure, and temperature; 1-hourly average 2D surface flux diagnostics (M2T1NXFLX) for precipitation; and 1-hourly average 2D radiation diagnostics (M2T1NXRAD) for radiation quantities (Bosilovich et al., 2016). Vertically resolved fields in M2I3NVASM start at a height of about 60 m, which limits our analysis of fog and very low-level ($<$ 250 m) clouds in this reanalysis.

2.7 ERA5

ERA5 (ECMWF, 2019) is a reanalysis produced by the ECMWF. It is based on an NWP model IFS version CY41R2. It uses the Tiedtke (1993) prognostic cloud scheme and the Forbes and Ahlgrimm (2014) scheme for mixed-phase clouds. The horizontal resolution is 0.25° in latitude and longitude, which is about 28 km in the north–south direction and 14 km in the east–west direction at 60°S. Internally, the model uses 137 vertical levels. Here, we use output at 1-hourly instantaneous time intervals, except for radiation quantities, which are accumulations (from these we calculate daily means). Vertically resolved quantities are available on 37 pressure levels.

2.8 CERES

TOA radiation quantities are taken from the Clouds and the Earth's Radiant Energy System (CERES) instruments onboard the Terra and Aqua satellites (Wielicki et al., 1996; Loeb et al., 2018). In our analysis, we used the adjusted all-sky SW and LW upwelling fluxes at TOA, adjusted cloud LWP and IWP, and adjusted cloud amount from the synoptic TOA and surface fluxes and clouds 1-degree daily edition 4A product (CER_SYN1deg-Day_Terra-Aqua-MODIS_Edition4A) (Doelling et al., 2013, 2016). The water paths in the product are computed from optical depth and particle size from geostationary satellites and the Moderate Resolution Imaging Spectroradiometer [MODIS, Pagano and Durham (1993)] (CERES author team, 2025). The water paths were multiplied by the cloud amount to get the water path relative to the whole grid cell area, equivalent to the definition used in the models.

Radiation and water path calculations presented in the results (Section 3) were completed such that they always represent daily means in order to be consistent with the CERES SYN1deg data. Therefore, every instantaneous profile in the simulated lidar data was assigned a daily mean radiation and water path value corresponding to the day (in the Coordinated Universal Time; UTC). In turn, the average radiation and water paths during the entire voyage or station observation period were calculated as averages of the profile values. In the observed lidar data, the daily mean values were taken from the spatially and temporally co-located CERES SYN1deg data for the day (in UTC). The voyage and station averages were calculated in the same way.

2.9 Precipitation Identification Using Machine Learning

Precipitation can cause strong enough lidar backscattering to be recognized as clouds by the threshold-based cloud detection method used in the ALCF. This is undesirable if equivalent precipitation backscatter is not included in the simulated lidar profiles. It was not possible to include precipitation simulation in the ALCF due to the absence of required fields of liquid and ice precipitation mass mixing ratios in the model output. While the fields could in principle be calculated from surface fluxes, such a calculation would be highly uncertain. The required radiation calculations for precipitation are also currently not implemented in the ALCF, even though this is a planned future addition.

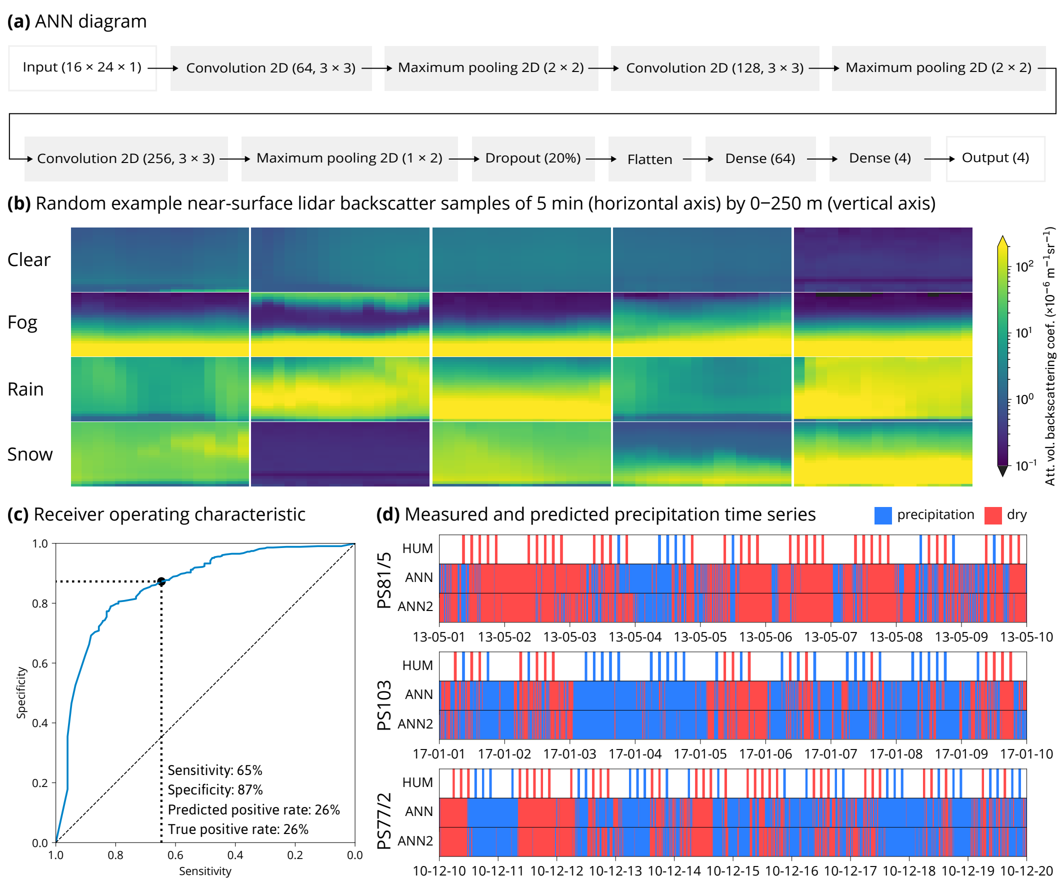

Figure 3. Artificial neural network (ANN) for prediction of precipitation in lidar backscatter. **(a)** Diagram showing the TensorFlow structure of the ANN, **(b)** randomly selected example samples of very low-level (0–250 m) backscatter in four categories (clear, fog, rain, and snow), as determined by coincident manual weather observations, **(c)** receiver operating characteristic diagram of the ANN, **(d)** examples of 10-day time series of human-observed ("HUM") and predicted precipitation based on an ANN trained on all voyages ("ANN") and all voyages except for the shown voyage ("ANN2") during three randomly selected voyages with the available data. Here, by "randomly selected," we mean selected from the top of a permutation generated by a pseudo-random number generator to prevent authors' bias in the selection.

In order to achieve a fair comparison of observations with model output, we exclude observed and simulated lidar profiles with precipitation, either manually or using an automated method. It is relatively difficult to distinguish precipitation backscatter from cloud backscatter in lidar observations, especially when only one wavelength channel and no polarized channel are available (Kim et al., 2020). In models, the same can be accomplished relatively easily by excluding profiles exceeding a certain surface precipitation flux. In the observations, using precipitation flux measurements from rain gauges can be very unreliable on ships due to ship movement, turbulence caused by nearby ship structures, and sea spray. Our analysis of rain gauge data from the RV *Tangaroa* showed large discrepancies between the rain gauge time series and human-performed synoptic observations, as well as large inconsistencies in the rain gauge time series. Human-performed observations of precipitation presence or absence are expected to be reliable but only cover a limited set of times. Therefore, it was desirable to implement a method of detecting precipitation from observed backscatter profiles alone.

On the RV *Polarstern* voyages, regular manual synoptic observations were available and included precipitation presence or absence and type. We used this dataset to train a convolutional artificial neural network (ANN) to recognize profiles with precipitation from lidar backscatter data (Fig. 3a), implemented in the TensorFlow ANN framework (Abadi et al., 2015). Samples of short time intervals (10 min) of very low-level lidar backscatter (0–250 m) were classified as clear, rain, snow, and fog, using the synoptic observations as a training dataset (Fig. 3b). From these, a binary, mutually exclusive classification of profiles as precipitating (rain or snow) or dry (clear or fog) was derived. For detecting model and reanalysis precipitation, we used a fixed threshold for surface precipitation flux of 0.1 mm h^{-1} (the ANN was not used).

The ANN achieved 65% sensitivity and 87% specificity when the true positive rate (26%) was made to match observations. The receiver operating characteristic curve is shown in Fig. 3c. We considered these rates satisfactory for the purpose of filtering precipitation profiles. Fig. 3d shows examples of the predicted precipitation compared to human-performed observations. The main ANN ('ANN' in Fig. 3) was trained on all data, and ancillary ANNs ('ANN2' in Fig. 3) were trained with portions of voyage data excluded to test the results for each voyage.

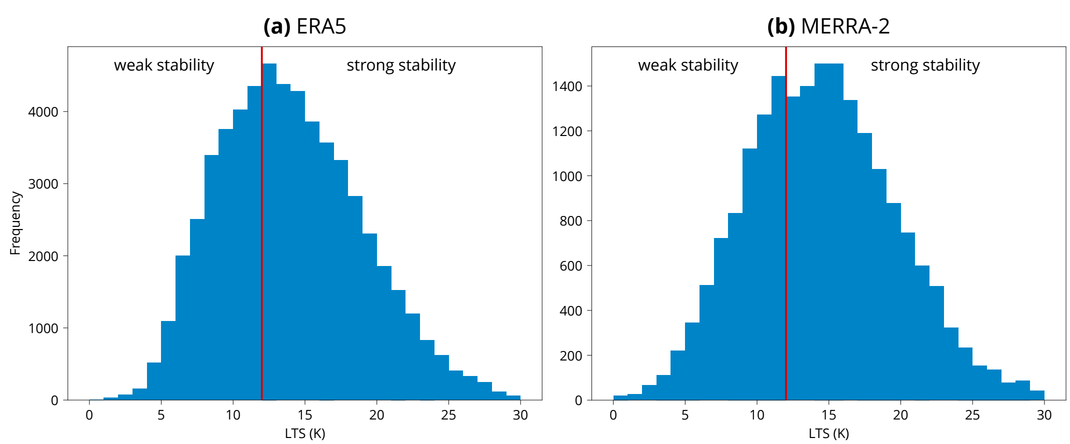

Figure 4. Lower tropospheric stability (LTS) distribution in **(a)** ERA5 and **(b)** MERRA-2 calculated for the 31 voyage tracks and one station from the highest instantaneous temporal resolution data available. Shown is also the chosen dividing threshold of 12 K for conditions of weak and strong stability.

2.10 Partitioning by Cyclonic Activity and Stability

In our analysis, we partitioned our dataset by cyclonic activity and stability into multiple subsets to evaluate cloud biases in the context of the main physical controlling processes. The SO is a region of the occurrence of both extratropical and polar cyclones. Cyclonic activity results in cloud formation at the air mass boundaries along the cold and warm fronts, as well as inside the cold sector, after a passing cold sector destabilizes the atmosphere relative to the surface temperature. In the cold front and cold sector, clouds are convectively driven, including deep convection, and the advection of colder air masses over warmer ocean surfaces can trigger convection and subsequent cloud formation. In contrast, warm advection can trigger fog or cloud formation by boundary layer air cooled by the ocean surface until it reaches saturation. More quiescent areas outside of cyclones can also be associated with clouds. These can be, for example, associated with clouds formed by warm or cold advection outside of cyclones, persistent clouds, clouds formed due to diurnal heating or cooling, or clouds formed due to ocean currents. Boundary layer stability can be expected to be associated with clouds by either allowing convection and turbulence under weak stability, inhibiting convection turbulence under strong stability, and by capping inversion controlling the cloud top height or trapping moist air near the surface and preventing fog dispersion. Therefore, dividing our dataset by these subsets allows us to quantify model biases associated with some of the main physical processes controlling cloud formation, persistence, and dissipation. Other methods of subsetting, such as using the International Satellite Cloud Climatology Project (ISCCP) pressure–optical thickness diagram (Rossow & Schiffer, 1991, 1999; Hahn et al., 2001) to separate profiles by cloud regimes and other cloud regime classifications (Oreopoulos et al., 2016; A. Schuddeboom et al., 2018), would be feasible.

We partitioned our data into two mutually exclusive subsets by cyclonic activity. For this purpose, we used a cyclone tracking algorithm to identify extratropical cyclones and polar cyclones over the SO in the reanalysis and ICON data. We used the open-source cyclone tracking package CyTRACK (Pérez-Alarcón et al., 2024). Generally, what constitutes an extratropical cyclone is considered relatively arbitrary due to the very large variability of the cyclones (Neu et al., 2013). The CyTRACK algorithm uses mean sea level pressure and wind speed thresholds as well as tracking across time steps to identify cyclone centers and their radii in each time step. With this information, we could classify every location at a given time as either cyclonic or non-cyclonic. Due to a relatively small total area covered by cyclones, as identified by the cyclone center and radius, for every time step and cyclone, we defined a cyclonic area as a circle of double the radius identified by CyTRACK centered at the cyclone center. All other areas were defined as non-cyclonic. For identifying cyclones in the observations and the reanalyses, ERA5 pressure and wind fields were used as the input to CyTRACK. This is justified by the fact that the large-scale pressure and wind fields in ERA5 are likely sufficiently close to reality. McErlich et al. (2023) have shown that wind is simulated well in ERA5 relative to the WindSat polarimetric microwave radiometer measurements (Meissner & Wentz, 2009). For identifying cyclones in ICON, its own pressure and wind fields were used as the input to CyTRACK because ICON is free-running, and thus the pressure and wind fields are different from reality. Subsetting by proximity to cyclones is a relatively crude measure because it does not take into account the different sectors of cyclones, which are commonly associated with different weather situations. However, this was a choice made for simplicity of the analysis, given the quantity of data. Konstali et al. (2024) performed a more complex attribution of precipitation to individual cyclone features.

In addition to the above, we partitioned our data into two mutually exclusive subsets based on LTS, which is derived as the difference between the potential temperature at 700 hPa and the surface. Based on a histogram of LTS in ERA5 and MERRA-2 calculated at all voyage tracks and stations (Fig. 4), we determined a statistically based dividing threshold of 12 K for weak stability (< 12 K) and strong stability ($\geq$ 12 K) conditions.

3 Results

3.1 Cyclonic Activity and Stability

Fig. 5a and b show the geographical distribution of the fraction of cyclonic days as determined by the cyclone tracking algorithm applied to the ERA5 reanalysis and ICON data (Section 2.10). As expected, the strongest cyclonic activity is in the high-latitude SO zone and is relatively zonally symmetric at all latitudes. The pattern matches reasonably well with Hoskins and Hodges (2005). While both reanalysis and ICON agree within about 8% in most areas, ICON is prevailingly more cyclonic by about 4%. There are clear differences, particularly in the highest occurrence rate regions, such as around Cape Adare, which is up to 20% more cyclonic in ICON, and the Weddell and Bellingshausen Seas, where ICON is less cyclonic by up to 10%. These differences might, however, stem from the relatively short time periods of comparison (4 years) and the fact that ICON is free-running.

Fig. 5c, d show the geographical distribution of the conditions of weak and strong stability as determined by the LTS (Section 2.10). Conditions of weak stability are prevalent in the mid-to-high SO (50–65°S), which might be explained by the relatively cold near-surface air overlying the relatively warm sea surface. Conditions of strong stability are common elsewhere over the SO. The distribution is also less zonally symmetric than the cyclonic activity. In the high-latitude SO, the presence of sea ice might have a substantial stabilizing effect (Knight et al., 2024). ICON is also substantially less stable than ERA5 across the whole region. In Section 3.5 we show that based on radiosonde observations, the bias is in ICON and not ERA5, and it is the result of underestimated temperature at heights corresponding to 700 hPa, as well as overestimated near-surface (2 m) air temperature, characterized by a higher frequency of occurrence in the 1–7°C range compared to observations at radiosonde launch locations (Fig. S1a). This may be related to large-scale circulation in ICON or radiative transfer biases.

3.2 Cloud Occurrence by Height

We used the ALCF to derive cloud occurrence by height and the total cloud fraction from observations, ICON, ERA5, and MERRA-2. The results for all campaigns individually are shown in Fig. S2. As shown in this figure, the biases are relatively consistent across the campaigns and longitudes. In addition, we aggregated the campaigns by calculating the averages and percentiles of all individual profiles, presented in Fig. 6. The analysis shows that the total cloud fraction is underestimated in ICON by about 10% and in the reanalyses by about 20%. When analyzed by height, ICON overestimates cloud occurrence below 1 km and underestimates it above; MERRA-2 underestimates cloud occurrence at all heights by up to 10%, especially near the surface; and ERA5 simulates cloud occurrence relatively well above 1 km but strongly underestimates it near the surface. We note that fog or very low-level clouds are strongly underestimated in the reanalyses (fog and clouds are both included in the cloud occurrence). We conclude that the ICON results match the observations better than the reanalyses in this metric.

For all observations considered (Fig. 6a), the data show cloud occurrence peaking near the surface, whereas the models show a higher peak (at about 500 m). The models generally underestimate the total cloud fraction by 10–30% and show a strong drop in cloud occurrence near the surface, which is not identified in the observations. ICON and ERA5 overestimate cloud occurrence at their peak (between 0 and 1 km). Above 1 km, ICON and MERRA-2 underestimate cloud occurrence, but ERA5 is accurate to about 3% or less. The exaggerated peak in models is partly explained by the LCL dis-

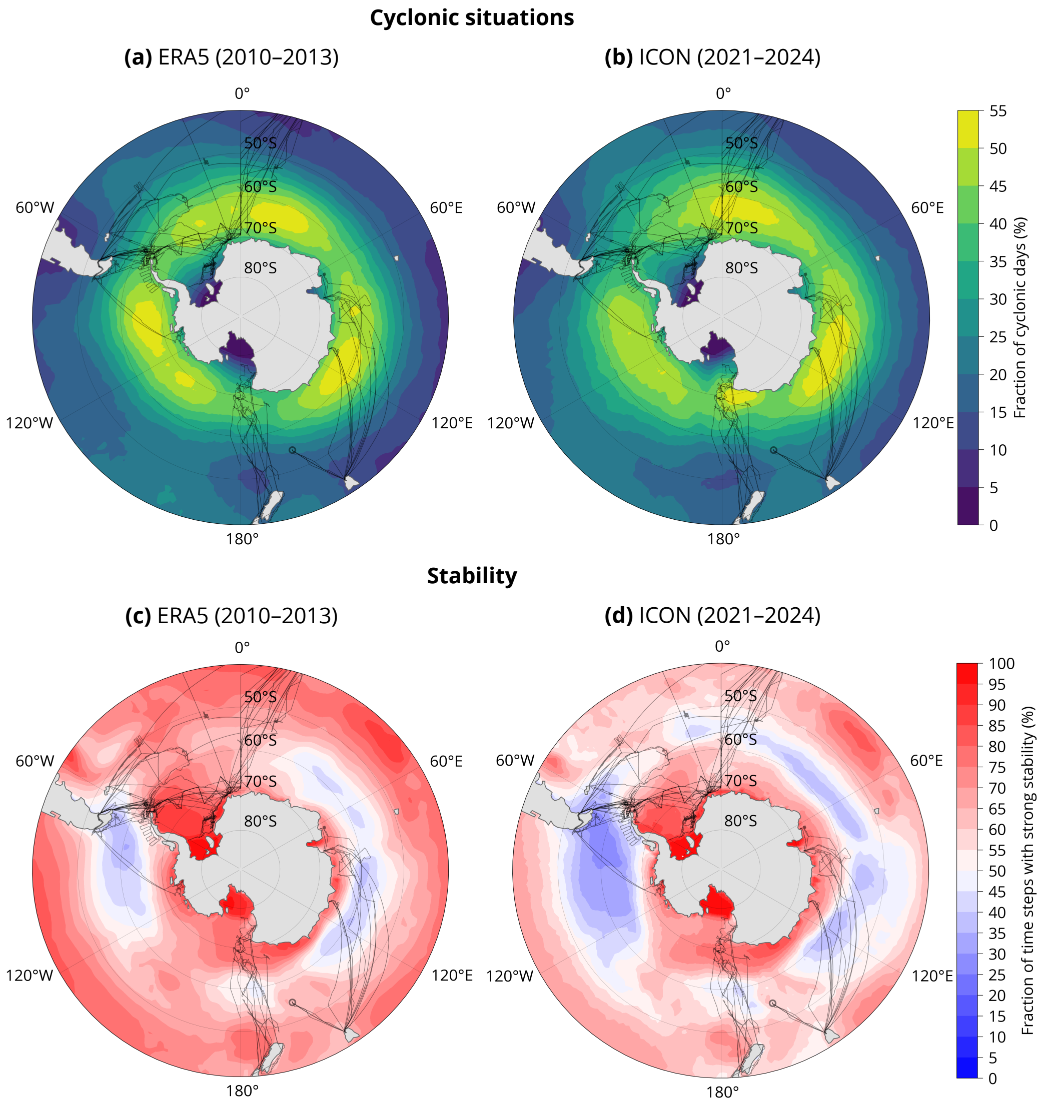

Figure 5. Geographical distribution of **(a, b)** cyclonic days and **(b, d)** strong stability (LTS $\geq$ 12 K) time steps in **(a, c)** ERA5 in years 2010–2013 (inclusive) and **(b, d)** ICON in model years 2021–2023 (free running). Cyclonic days are expressed as a fraction of the number of days with cyclonic activity, defined as grid points located within a double radius of any cyclone on a given day (UTC), as identified by CyTRACK. The voyage tracks and the point of the MI-CRE campaign are also shown.

tribution, which peaks about 300 m higher in the models than in the observations (near the surface), although this is not very pronounced. This is indicative of near-surface RH often being close to saturation in the observations but not in the models (Fig. S1b). There are multiple possible reasons for this bias, such as how the statistical distribution of RH within a grid cell is represented in the models, the air–sea moisture flux parameterization, or weaker stability in the models, which can cause more boundary mixing across heights and thus lower near-surface RH.

When the data are subset by latitude (Fig. 6b, c), we see that the low-latitude SO zone (40–55°S) displays a stronger peak of cloud occurrence near the surface than the high-latitude SO zone (between 55°S and the Antarctic coast), and this could be because higher latitudes have a greater prevalence of weakly stable profiles (Fig. 5c, d), although more stable profiles populate regions south of 65°S close to the Antarctic coast. Cyclonic activity is also stronger in high-latitude SO, which is typically associated with shallow or deep convection rather than the very stable stratification necessary for fog formation. The low- and high-latitude SO zones show similar biases in models as in the general case, but ERA5 does not overestimate the peak in the low-latitude SO zone (very low-level cloud occurrence is still strongly underestimated).

When the data are subset as either cyclonic or non-cyclonic situations (Fig. 6d, e), we see that the cyclonic situations have a larger amount of observed cloudiness, including peak and total cloud fraction, by about 10%. In the cyclonic situations, the model vertical profiles of cloud occurrence compare well with observations, but they peak higher by about 200 m and are larger by about 8%. The reanalyses tend to underestimate cloud occurrence above 1 km by about 5% and near the surface by about 15%. Non-cyclonic situations are similar to the general case, also because they form the majority of analyzed profiles (83%).

When the data are subset by stability (Fig. 6f, g), as defined in Section 2.10, we see that in situations of strong stability, cloud occurrence peaks strongly near the surface in observations, compared to situations of weak stability, where the peak is more diffuse between 0 and 1 km. Physically, conditions of strong stability are associated with the formation of advection fog, such as in situations of warm air advection from the north over a colder sea surface, thus inducing fog formation by cooling of the warm and humid air by the cold surface. In situations of strong stability, the models have smaller biases than in weak stability, with an overestimated peak of up to 12%, underestimated cloud occurrence above 1 km by up to 5%, and underestimated cloud occurrence near the surface by about 10% in the reanalyses but not ICON. In situations of weak stability, the bias in ICON is very pronounced, with a much larger peak in cloud occurrence at about 500 m; the reanalyses underestimate cloud occurrence below 1 km, especially near the surface; and MERRA-2 underestimates cloud occurrence more strongly at almost all heights.

In all subsets, even when the models overestimate cloud occurrence at some altitudes, they always substantially underestimate the total cloud fraction. ICON can be generally characterized as substantially overestimating cloud occurrence below 1 km and underestimating above, underestimating the total cloud fraction, and showing the greatest biases in conditions of weak stability and non-cyclonic conditions. ICON also has a peak cloud occurrence at higher altitudes than observations (500 m vs. near the surface), and correspondingly, its LCL tends to be higher. MERRA-2 can be generally characterized as underestimating cloud occurrence at nearly all altitudes as well as the total cloud fraction, but mostly above and below 500 m (the peak at 500 m is well represented). MERRA-2 displays the largest errors relative to observations in the low-latitude SO zone and under weak stability. ERA5 can be generally characterized as representing cloud occurrence correctly above about 1.5 km, overestimating between 500 m and 1 km, but underestimating very low-level cloud occurrence. The total cloud fraction is strongly underestimated in all subsets. ERA5 has a tendency towards greater cloud underestimation in the low-latitude SO zone and under weak stability; conversely, it overestimates the peak of cloud occurrence at 500 m in the high-latitude SO zone and under strong stability.

3.3 Daily Cloud Cover

We also analyzed the daily cloud cover (total cloud fraction) distribution. This is a measure of cloudiness, irrespective of height, calculated over the course of a day (UTC). A cloud detected at any height means that the lidar profile was classified as cloudy; otherwise, it was classified as a clear sky. When all profiles in a day are taken together, the cloud cover for the day is defined as the fraction of cloudy profiles in the total number of profiles. It is expressed in oktas (multiples of 1/8), reflecting the 3-hourly model output of MERRA-2 and ICON, i.e., 8 times per day. The same calculation is done for the lidar observations as for the simulated lidar profiles. We use the term "okta" independently of its use in instantaneous synoptic observations, and here it simply means 1/8 (0.125) of the daily cloud cover.

In Fig. 7 we show the results for the same subsets of data as in Section 3.2. Observations display the highest proportion of high cloud cover values (5–8 oktas), peaking at 7 oktas. This pattern is not represented by ICON or either reanalysis. While ICON is closest to matching the observed distribution, it tends to be 1 okta clearer than the observations, peaking at 6 oktas, and substantially underestimating days with 8 oktas. Overall, the reanalyses show results similar to each other, underestimating cloud cover by about 2 oktas and strongly underestimating days with 7 and 8 oktas. Of the two reanalyses, MERRA-2 has slightly higher cloud cover than ERA5, by about 6% at 6 octas, which makes it more consistent with observations.

When analyzed by subsets, observations in the cyclonic subset show the highest cloud cover, with 8 oktas occurring on one half of such days (Fig. 7d). This sensitivity to cyclonic conditions is not observed in ICON or the reanalyses. Interestingly, clear sky days (0 oktas) also have a local maximum peaking at about 15% in this subset. When we contrast the low- and high-latitude zones, we see that the high-latitude zone tends to have greater cloud cover, peaking at 8 oktas (Fig. 7c). The high-latitude zone also has almost no clear sky or small cloud cover cases (0–4 oktas). ICON and the reanalyses represent this characteristic of the distribution well for 0–3 oktas, but otherwise show biases similar to the general case. One of the greatest biases is present in ERA5 in the subset of weak stability, in which ERA5 peaks at 3 oktas, while the observations peak at 7 oktas and show negligible cloud cover below 5 oktas.

3.4 Top of Atmosphere Radiation, Liquid and Ice Water Path

In Fig. 6, we also show the mean outgoing SW and LW TOA radiation, whose calculation is described in Section 2.8. In observations, these come from daily mean CERES measurements averaged over the voyage tracks or a station location, whereas in the models they come from daily means of TOA radiation in the model output averaged over the same location and time periods.

In the general case (Fig. 6a), ICON and ERA5 underestimate the outgoing SW radiation by 22 and 20 $\mathrm{Wm^{-2}}$ (respectively), and MERRA-2 overestimates it by 6 $\mathrm{Wm^{-2}}$. While in ICON and ERA5, this is in line with the underestimated total cloud fraction of 10% and 22% (respectively); in MERRA-2, the opposite result is expected from the underestimated total cloud fraction of about 20%. Neglecting the direct radiative effects of sea and aerosol, this is only possible if the albedo of cloudy areas is overestimated, compensating for the lack of cloudy areas.

We note that the radiative transfer calculations used in the lidar simulator mean that the impact of both cloud phase and cloud fraction are convolved to produce the cloud

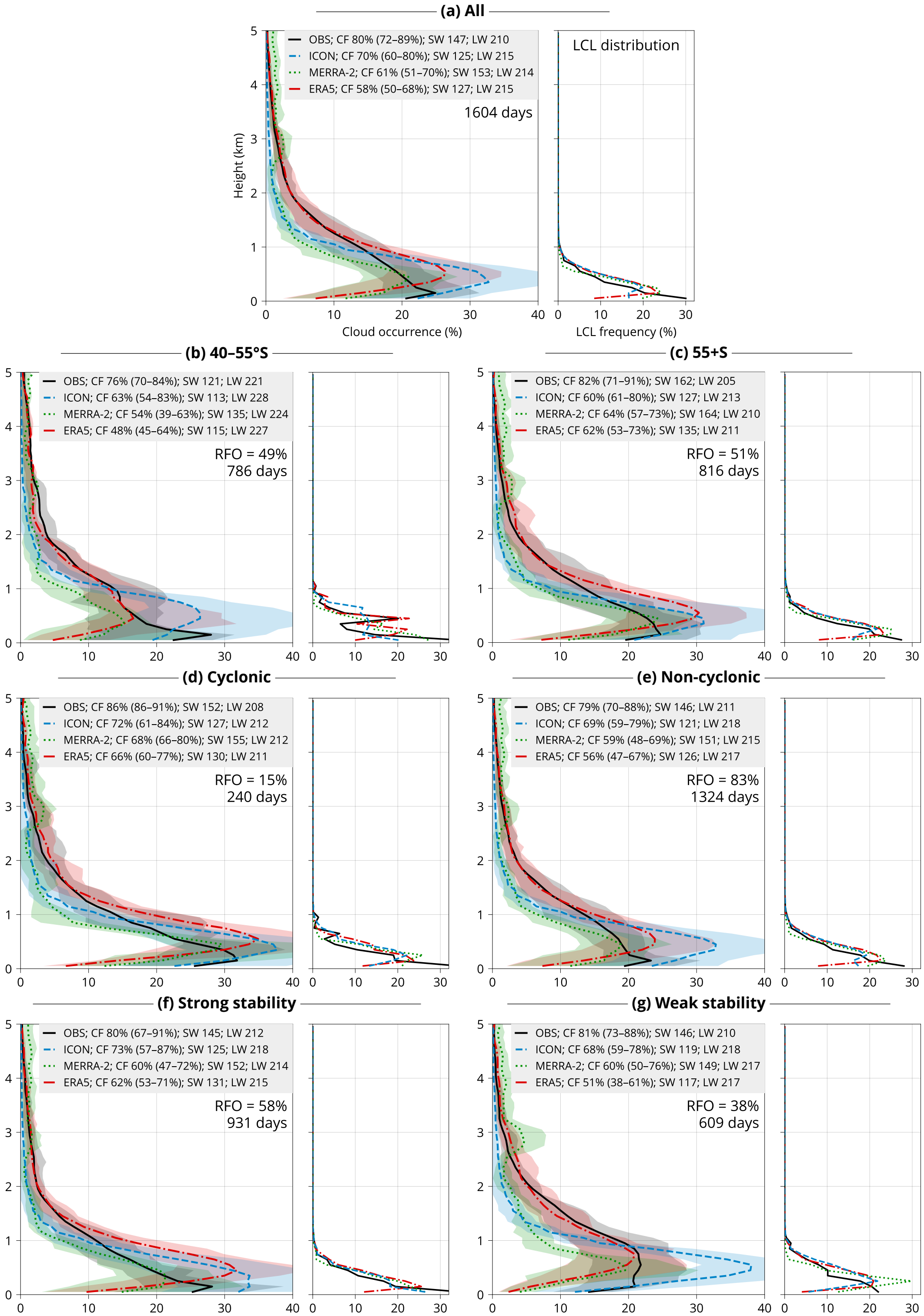

Figure 6. Cloud occurrence by height calculated as the average of all voyages and stations and lifting condensation level (LCL) distribution. The LCL is derived from radiosonde profiles and equivalent model profiles, which were not available for all voyages and times. The total cloud fraction (CF), average outgoing shortwave (SW) and longwave (LW) radiation, and the relative frequency of occurrence (RFO) are shown. The bands are the 16$^{\text{th}}$–84$^{\text{th}}$ percentile, calculated from the set of all voyages and stations.

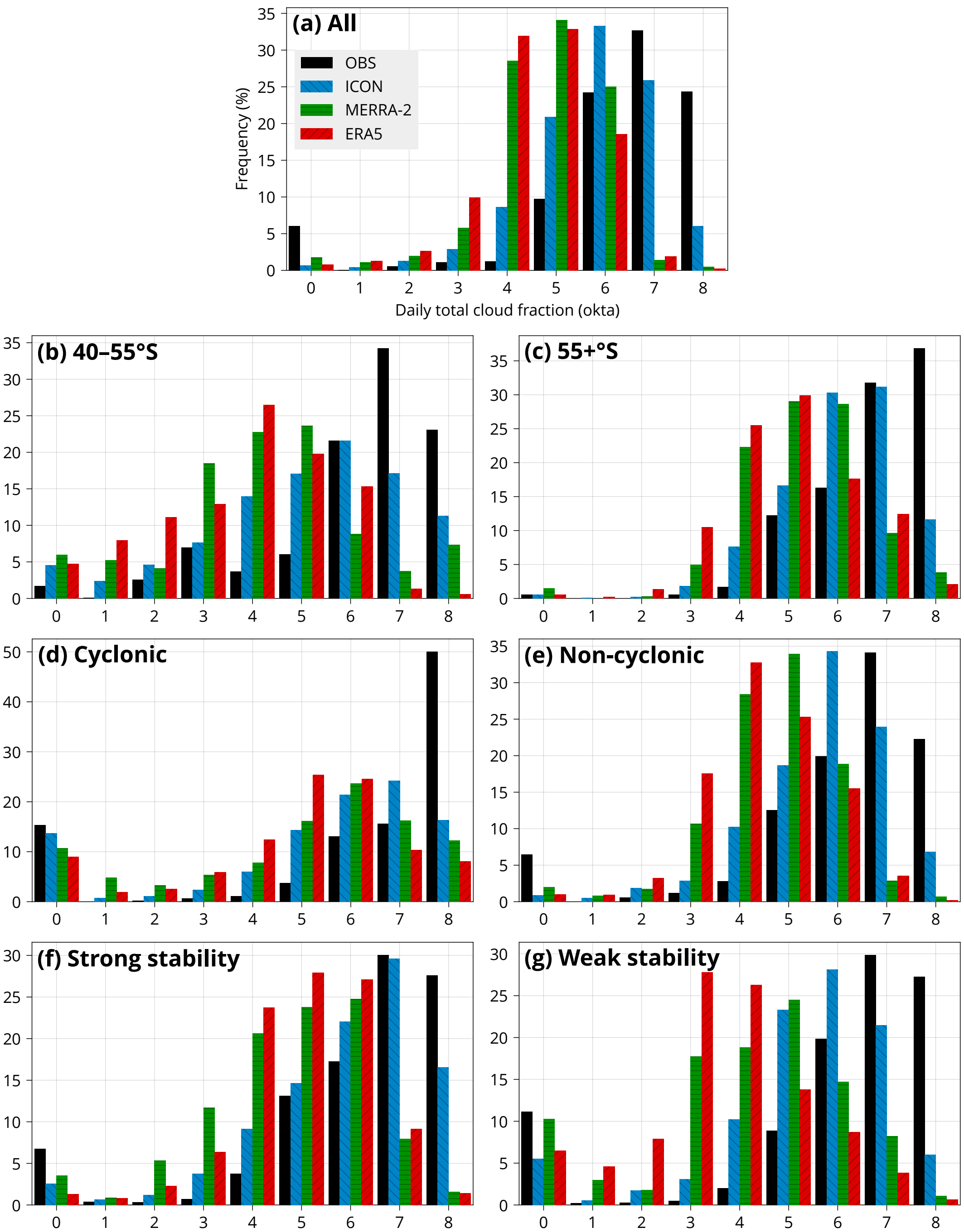

Figure 7. Daily total cloud fraction histograms calculated as the average of all voyage and station histograms. The total cloud fraction of a day (UTC) is calculated as a fraction of cloudy (based on the cloud mask) observed (OBS) or simulated lidar profiles. The models and subsets are as in Fig. 6.

mask. Therefore, the cloud occurrence is not affected by any cloud phase biases as long as the cloud is optically thick enough to be detected and the laser signal is not too attenuated. A combination of underestimated total cloud fraction and overestimated outgoing SW at TOA is indicative of an overestimated cloud albedo (in cloudy areas) due to either cloud liquid and ice water content, cloud phase, droplet or ice crystal size distribution, shape or orientation of ice crystals, cloud overlap, or their combination. The influence of cold clouds is likely second-order due to the much larger typical effective radius of ice crystals than cloud droplets.

In contrast to SW radiation, the models have much smaller LW radiation biases, which is expected due to the prevailing low-level clouds having similar temperatures as the surface. Roh et al. (2021) also found LW biases to be much lower than SW biases in DYAMOND models over the tropical Atlantic Ocean. In ICON, the outgoing LW radiation is overestimated by 5% (Fig. 6a). This is likely caused by an underestimated total cloud fraction exposing a larger sea surface area to cooling to space, which is typically warmer than the atmospheric temperature at 0–2 km, where most of the clouds are located. In the MERRA-2 and ERA5 reanalyses, the LW biases are also slightly positive, 4 and 5 Wm^{-2}, respectively. This is again in line with the underestimated total cloud fraction by about 20%. However, if the clouds are too thick, as expected from the SW results, this might also provide a compensating effect, in which too small a cloud area is counteracted by greater optical thickness in the LW spectrum, thus reducing the outgoing LW radiation more in thick relative to thinner clouds. For thin clouds, the outgoing TOA LW radiation originates both from the warmer surface (partly blocked by the clouds) and the clouds, whereas for thick clouds, the outgoing TOA LW radiation originates mostly from the colder-than-surface clouds.

In all the subsets (Fig. 6b–g), the same type of biases are observed, namely the outgoing SW radiation is underestimated in ICON and ERA5 and overestimated in MERRA-2, and the outgoing LW radiation is overestimated in all the models. Even though the total cloud fraction is higher by 6% over the high-latitude SO than the low-latitude SO, the outgoing SW radiation is much greater by 41 Wm^{-2}, implying a much greater cloud albedo (of cloudy areas) over the high-latitude SO. ICON has little difference in the total cloud fraction between low- and high-latitude SO, but greater outgoing SW radiation by 14 Wm^{-2} over the high-latitude SO, likely due to thicker clouds under deeper convection in less stable and more cyclonic conditions relative to the low-latitude SO. In contrast, the reanalyses showed both greater total cloud fraction and outgoing SW radiation over the high-latitude SO compared to the low-latitude SO.

Fig. 8 shows the SW and LW radiation as histograms and their corresponding averages. ERA5 and ICON overestimate outgoing SW near 80 Wm^{-2} (Fig. 8a), which probably relates to clear sky situations, as expected from the underestimated cloud fraction. They also underestimate the highly reflective situations above 200 Wm^{-2}. MERRA-2 exhibits the too-few-too-bright problem in terms of overestimating SW reflectivity around 290 Wm^{-2}, given that the total cloud fraction in MERRA-2 is strongly underestimated. The LW distribution shows that all of the models overestimate outgoing LW (Fig. 8b), which is expected from the underestimated cloud fraction, exposing more of the warmer ocean surface relative to colder clouds.

Fig. S3 shows the LWP and IWP distributions as histograms and their corresponding averages. The LWP and IWP are calculated from the mass of water in the column divided by the area of the column, i.e., not just the area of the cloudy portion of the column, as in some definitions. The available observational satellite reference for the LWP and IWP over high latitudes is unfortunately very uncertain due to a high solar zenith angle and the inability of passive visible and infrared retrievals to detect phase below the cloud top of mixed-phase clouds (Huang et al., 2006; Greenwald, 2009; Seethala & Horváth, 2010; Eliasson et al., 2011; Duncan & Eriksson, 2018; Khanal et al., 2020), and this limits our comparison. The LWP distribution shows that all models overestimate

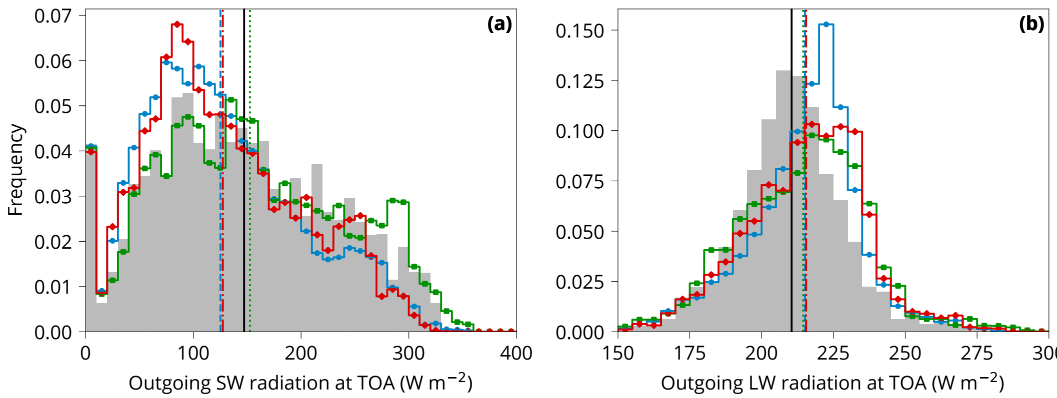

Figure 8. Histograms and averages of outgoing **(a)** SW and **(b)** LW radiation at TOA in CERES SYN1deg observations (OBS), ICON, MERRA-2, and ERA5. All campaigns are weighted equally. The statistics are calculated from daily mean values corresponding to each time step and geographical location of the voyage tracks and stations.

cases with a near-zero LWP (Fig. S3a), which relates to the underestimated total cloud fraction. MERRA-2 shows quite overestimated high-LWP situations, which is most likely related to the too-few-too-bright problem of simulating lower total cloud fraction but clouds with a higher LWP to compensate. The IWP (Fig. S3b) is somewhat less important radiatively than LWP because of the typically larger and less numerous hydrometeors. Similarly to the LWP, the models overestimate situations with a near-zero IWP. ERA5 is otherwise simulating the IWP distribution well, but ICON and MERRA-2 underestimate the IWP. In the cloudy situations (Fig. S3c, d), it can be seen more distinctly that MERRA-2 overestimates moderate (0.05–0.15 kg m^{-2}) and high LWP (over 0.15 kg m^{-2}), and ERA5 and ICON underestimate moderate LWP. ICON also overestimates high LWP, resulting in overestimated average LWP.

3.5 Relative humidity and potential temperature profiles

In order to examine the potential link in the cloud biases to the local physical conditions, we analyzed the radiosonde profiles available from the campaigns (Section 2.1). The profiles were partitioned into the same subsets as above (Sections 3.2 and 3.3). We focus on comparing θ_v and RH, being one of the primary factors affecting shallow convection and the associated low-level cloud formation and dissipation. The observed and model profiles of θ_v and RH are shown in Fig. 9.

Overall, the mean θ_v is accurate to within 0.5 K in ICON and MERRA-2, except for ICON being colder by up to 2.5 K in the mid-to-high troposphere (less stable) (Fig. 9a). Larger differences exist, however, in the 40–55°S zone, where ICON is colder by about 5 K at 5 km (Fig. 9b). In other subsets, the bias is relatively small. MERRA-2 and ERA5 are very close to the observations, possibly due to a high accuracy of assimilation of this quantity. Notably, the variability of θ_v (as represented by the percentiles) is much smaller in ICON than in the observations. This indicates that this model's internal variability in the lower-tropospheric thermodynamic conditions in the SO is smaller than in reality.

RH displays much larger biases. In all subsets, ICON is too humid in the first 1 km by about 5%, but very accurate above, except for the 40–55°S zone and conditions of weak stability (Fig. 9b, g), where it is too dry between about 1 and 3 km. Even though RH measured by radiosondes in the first 100 m is not very different between the observations

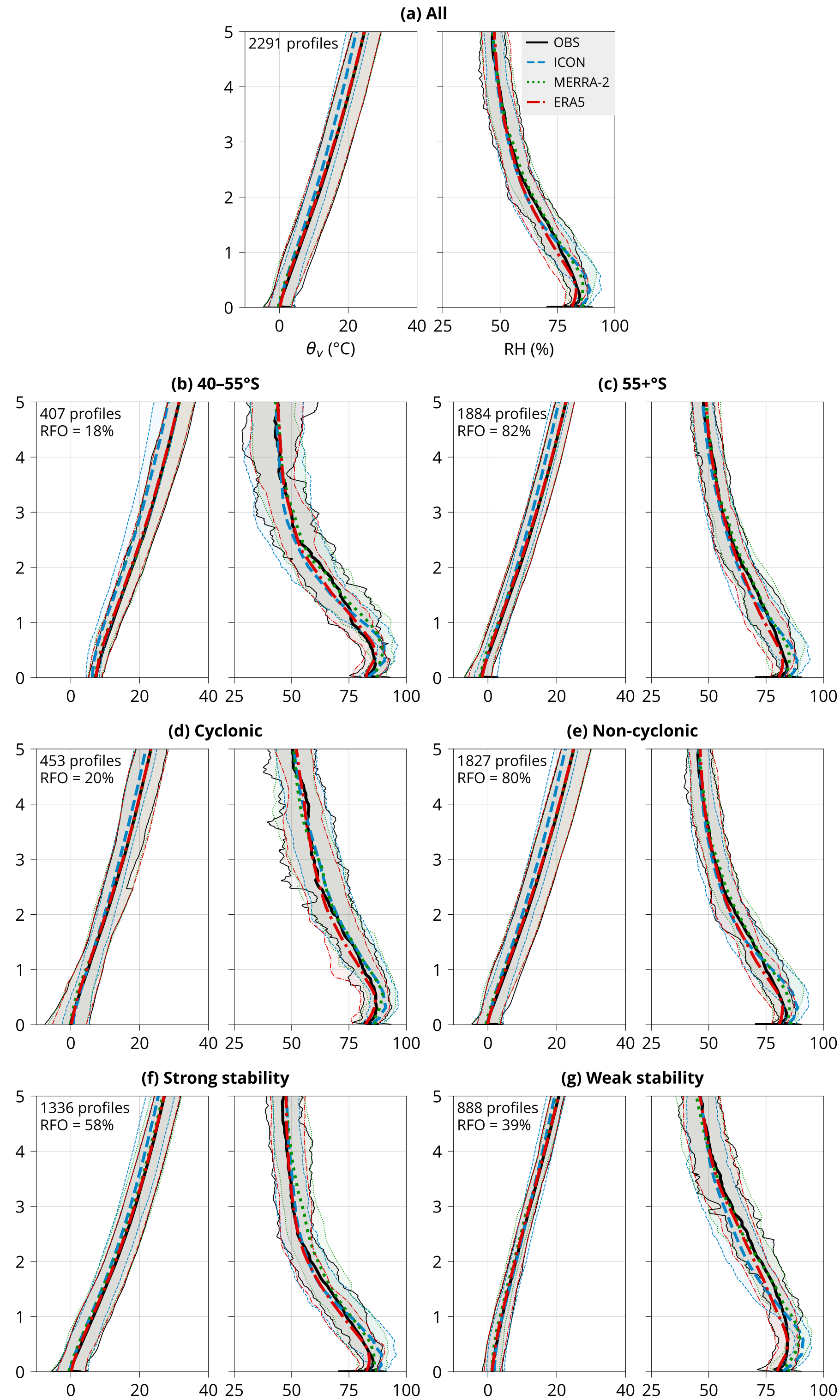

Figure 9. Virtual potential temperature (θ_v) and relative humidity (RH) determined from radiosonde launches and co-located profiles in ICON, ERA5, and MERRA-2 in subsets as in Fig. 6. The solid lines are the average calculated from the averages of every individual voyage and station. The bands span the 16$^{\text{th}}$–84$^{\text{th}}$ percentiles, calculated from the distribution of the voyage and station averages. Shown is also the relative frequency of occurrence and the number of profiles in each subset.

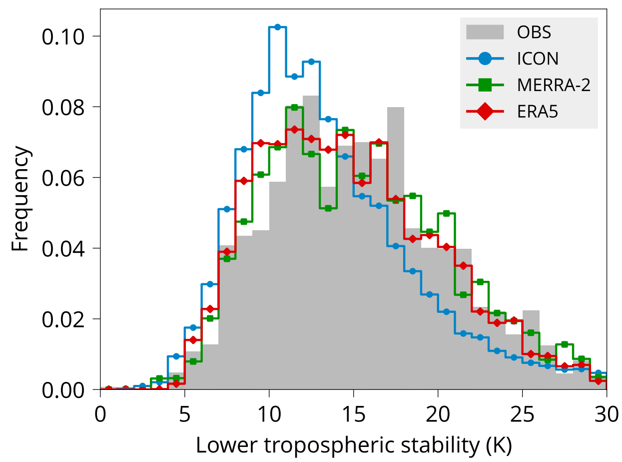

Figure 10. Histogram of lower tropospheric stability calculated from the observed radiosonde profiles and the corresponding model profiles. All campaigns are weighted equally.

and the models (Fig. 9a), near-surface (2-m) RH at the radiosonde launch locations is much greater in the observations, most often close to 100%, unlike in the models, where 85% tends to be the most common (Fig. S1b). This also explains why LCL is much more frequently located at the surface in the observations than in the models (Fig. 6a). LCL is fully determined by near-surface temperature, near-surface RH, and surface pressure.

Fig. S4 shows θ_v and RH profiles for profiles containing fog, cloud at 500 m, and cloud at 1.5 km. These situations are characterized by particular cloud biases as identified in the lidar cloud occurrence analysis. The rationale is to examine θ_v and RH associated with these situations. Foggy situations are characterized by a rapid increase of θ_v with height and an observed average RH of about 90% near the surface (Fig. S4a). In contrast, the models simulate higher RH in the first 100 m under foggy conditions by several percentage points. In situations with clouds occurring at 500 m, θ_v is relatively constant between the surface and 500 m (Fig. S4b), as expected for convectively driven clouds. The observed RH peaks at 500 m at about 90%. The models, however, simulated higher RH between the surface and 500 m under these conditions. ICON and ERA5 show a stronger decrease of RH above this height than observations, and ERA5 shows more strongly stable stratification. Unlike the foggy and 500-m cloud situations, situations with clouds at 1.5 km do not have a flat θ_v with height. This indicates that, unlike the former, clouds at 1.5 km are not (or not as strongly) convectively driven. As expected, RH in these situations peaks at 1.5 km at about 85% in observations. In the models, this peak is much less pronounced.

Fig. 10 shows the histogram of LTS calculated from all radiosonde profiles and the corresponding profiles in the models. It can be seen that ICON substantially underestimates the occurrence of cases of strong stability above 16 K while overestimating the cases of moderate stability (8 to 16 K). When considered together with the cloud occurrence results presented in Fig. 6, we see that since ICON is biased towards weak stability, it overrepresents cloud profiles strongly peaking at 500 m (Fig. 6g) over cloud profiles with fog or very low-level cloud (Fig. 6f). This can be a physical reason for its overall positive bias in cloud at 500 m (Fig. 6a) instead of the observed cloud occurrence profile peaking near the surface. The reanalyses simulate the LTS distribution well except for a slight underestimation of LTS.

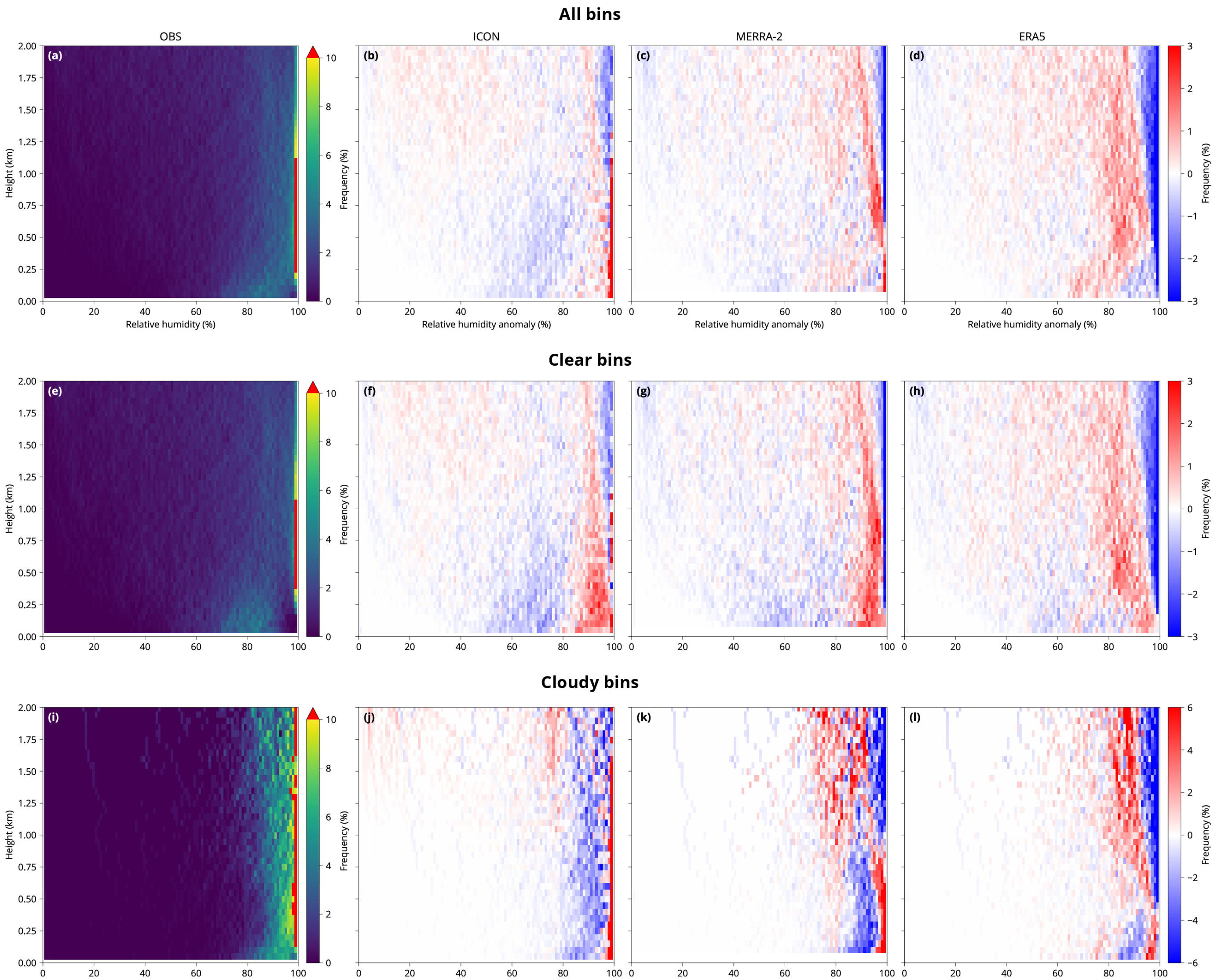

Figure 11. Relative humidity histograms calculated from the observed radiosonde profiles and the equivalent model profiles for **(a)** all bins, **(b)** clear bins, and **(c)** cloudy bins, determined from the lidar cloud mask. Model histogram values are relative to observations. The histogram values are normalized to 100% for each level separately. All campaigns are weighted equally.

Fig. 11 shows RH histograms calculated from the radiosonde observations and equivalent profiles in the models (shown as anomalies relative to the observations), calculated for all, clear, and cloudy bins, based on the lidar observations and the simulated lidar backscatter in the models. Here, we show only the first 2 km to concentrate on the identified cloud biases seen at these heights. We can see several notable features. The models simulate progressively fewer high-RH (>90%) bins above the ground (Fig. 11b–d). This can be related to either ice nucleation happening in the models, which requires smaller RH for saturation, or the grid cell size in the models, which requires lower grid cell average RH than 100% for saturation to occur in a fraction of the grid cell. The models also tend to simulate more clear bins than observations for RH between 80 and 100% between the ground and about 1 km (Fig. 11f–h). In the observations, these values of RH are associated with cloudy bins (Fig. 11i). Conversely, the models predominantly associate only RH very close to 100% with cloudy bins at these heights (Fig. 11j–l). This may be one of the main reasons for the identified cloud or fog biases near the ground. A possible explanation is that cloud droplets are able to form or persist at RH between 90 and 100% at these heights over the SO. This could be due to abundant hygroscopic nuclei such as sea salt (Zieger et al., 2017; Kong et al., 2018) or droplet generation from sea spray in the common high swell and high wind speed conditions over the SO (Revell et al., 2019; Hartery et al., 2020). Stratus fractus or other broken clouds could also lead to less than 100% RH when averaged over the size of the vertical bins (up to 30 m in some of the radiosonde profiles).

Fig. S5 shows histograms the same as the previous figure, but for θ_v. They show a more complex picture, characterized by a central peak at about 0°C near the surface, increasing to about 5°C at 2 km (Fig. S5a). For cloudy bins, the central peak is generally more constant with height and even shows a minimum in θ_v at about 500 m (Fig. S5i). This is indicative of convection being associated with clouds at these heights, which results in flat θ_v profiles. In the reanalyses, in the first 200 m, values slightly above 0°C are associated with more clear bins than in observations, and values slightly below 0°C with fewer (Fig. S5g–h). Conversely, the opposite is true for cloudy bins (Fig. S5k–l). Situations with 0°C near-surface air temperature might occur predominantly when an open ocean surface keeps the near-surface air temperature close to 0°C under otherwise colder air mass conditions, such as under cold advection. ICON displays a notable bias above about 1 km, where the central peak is strongly underestimated (Fig. S5j). Instead, these heights and values of θ_v are more associated with clear bins (Fig. S5f). This might be related to the strong underestimation of cloud occurrence at these heights.

4 Limitations of this Study

Let us consider the main limitations of the presented results. The spatial coverage of our dataset does not include most parts of the Indian Ocean and Pacific Ocean sectors of the SO. Even though climatological features of the SO are typically relatively uniform zonally, variations exist, such as those related to the Antarctic Peninsula and the southern tip of South America. The voyages were mostly undertaken in the Austral summer months and only rarely in the winter months, due to the poor accessibility of this region during winter. Therefore, our results are likely representative of summer and, to a lesser extent, spring and autumn conditions. Ship access to sea-ice-covered areas of the SO is also limited. Cloud regimes and phases in the region are seasonally variable (Danker et al., 2022).

The time period of ICON is relatively short, with only four full years of simulation available. Moreover, the simulation is free-running and ocean-coupled, which means that observations had to be temporally mapped to this time period (at the same time relative to the start of the year) for the comparison. For these reasons, one can expect the results to be slightly different due to reasons unrelated to model biases, such as different weather conditions, partially accounted for by the cyclone and stability subsetting, and the phase of climate oscillations, such as the ENSO in the observations and ICON. The interannual variability in cloud occurrence in ICON can be seen in Fig. S2, where each year in ICON is represented by a separate line. As could be expected, the interannual variability tends to be substantially smaller than the biases and thus is unlikely to have a strong impact on the main findings.

It would be possible to use short-term ICON simulations for almost one-to-one comparison to observations. However, here we focus on long-term biases, which are statistically more robust. Our analysis is, therefore, complementary to shorter process-level studies. The reanalyses pose the difficulty of determining how much assimilated observations impact the results. While one might expect temperature and RH profiles to be well represented in the reanalyses due to assimilation of satellite data, we see that this is not always the case in comparison with the radiosonde profiles and near-surface meteorological observations. This could be due to the limited vertical accuracy of satellite sounding measurements and obscuration by clouds. Despite the assimilation, the cloud and radiation biases are often comparable to or greater than in the free-running model.

Ground-based lidar observations are affected by attenuation by thick cloud layers, and for this reason the results are most representative of boundary layer clouds, while higher-level clouds are only occasionally visible to the lidar when boundary layer clouds are not present. Ground-based lidar observations can be regarded as complementary to satellite lidar observations for the evaluation of low-level clouds, which are predominant in this region, while mid- and high-level clouds are likely better sampled by satellite observations (McErlich et al., 2021). Ground-based observations are, however, complicated by precipitation, and satellite observations can also be used if the effect of overlapping clouds is carefully eliminated. Lidar retrievals close to the surface ($\sim$100 m) are affected by uncertainties related to incomplete overlap, signal saturation (dead time), and after-pulse effect corrections (Kuma et al., 2021).

Supercooled liquid clouds (liquid clouds under subzero temperature) commonly occur over the SO. In our analysis of the LWP and IWP, we see that both phases are abundant. Because liquid water droplets are typically smaller and more numerous than ice crystals in cold clouds, they attenuate a greater amount of the lidar radiation. Clouds with a relatively modest optical thickness of 1.7 can attenuate the lidar signal for a detection at 2 km using an instrument with noise properties like the Vaisala CL31 (Section 2.4). While supercooled liquid clouds and their attenuation are accounted for by the lidar simulator, they can strongly attenuate the signal and cause artificially low values of cloud occurrence at higher altitudes. For example, we found that cloud occurrence at 1.5 km is underestimated in ICON and underlying clouds are overestimated. However, this can also mean that clouds at 1.5 km are present in the model, but the signal is too attenuated by the lower clouds in the model, but not in the observations, where the underlying clouds are not as pronounced.

We have attempted to remove lidar profiles with precipitation (about 26% of all profiles), which could not be properly simulated with the lidar simulator (Section 2.9). However, the approach was limited by the relatively low sensitivity of the ANN (65%) and the fact that we had to choose a fixed threshold for surface precipitation flux in the models, which might not correspond to detection by the ANN applied to observations. We also made no attempt to remove profiles with precipitation that did not reach the surface. The above reasons may result in an artificial bias in the comparison, though we expect this to be much smaller than the identified model biases.

Subsetting by cyclonic activity and stability is done based on the ERA5 data. As we have shown, the reanalyses also suffer from biases in near-surface and upper-level quantities. Therefore, the subsetting is limited by the accuracy of the ERA5 pressure field, near-surface temperature, and temperature at 700 hPa. Near-surface ship observations are affected by the ship structures as well as the variable height above sea level at which the measurements are taken. The accuracy of radiosonde measurements in the first tens of meters from the surface is also likely affected by the ship environment, such as turbulence generated by ship structures and the ship exhaust. Vertical averaging of the radiosonde data can result in lower RH near saturation due to averaging of drier and moister layers together. For example, some of the RV *Polarstern* radiosondes are available in vertical resolution of about 20–30 m. As mentioned in Section 3.4, the satellite retrieval of the LWP and IWP is affected by large biases, especially over high latitudes, which limits our comparison with the models.

Table 3. Summary of the main biases. Values are relative to observations and rounded to the nearest multiple of 5, except for daily cloud cover and RH, which are rounded to the nearest integer. The best-performing value is marked in **bold**. Abbreviations: boundary layer (BL), relative humidity (RH), shortwave (SW), longwave (LW), liquid water path (LWP), ice water path (IWP), and lifting condensation level (LCL).

	ICON	MERRA-2	ERA5
Total cloud fraction (%)	**-10**	-20	-20
Daily cloud cover (okta)	**-1**	-2	-2
Fog (%)	**0**	-10	-10
BL clouds (at ∼500 m)	15	**0**	5
Mid-lev. clouds (at ∼1.5 km)	-5	-5	**0**
RH at 500 m	2	2	**0**
SW (W m^{-2})	-25	**5**	-20
LW (W m^{-2})	5	5	5
LWP (g m^{-2})	**10**	20	-15
IWP (g m^{-2})	-30	-30	**-15**
LCL distribution peak (m)	300	300	300

5 Discussion and Conclusions

We analyzed a total of about 2400 days of lidar and 2300 radiosonde observations from 31 campaigns and the Macquarie Island sub-Antarctic station, covering the Atlantic, Australian, and New Zealand sectors of the SO over 10 years. This dataset, together with the use of a ground-based lidar simulator, provided a comprehensive basis for evaluating SO cloud and thermodynamic profile biases in the GSRM ICON and the ERA5 and MERRA-2 reanalyses. Our analysis provides a unique evaluation perspective, complementary to satellite observations for evaluating boundary layer clouds and fog, which are predominant in this region. We did not, however, analyze the cloud phase based on ground-based observations. Cloud phase can have a strong impact on the SW radiative transfer due to larger and therefore less numerous hydrometeors in cold and mixed-phase clouds (for the same amount of water), scattering much less SW radiation. Especially, the underestimation of fog or very low-level clouds is very substantial in the reanalyses, and we showed that this relates to cloud and fog formation or persistence at RH between 80 and 100% in the boundary layer in the observations, while in models RH values less than 100% are associated with clear bins. We subset the dataset by low and high latitude SO bands, cyclonic activity, and stability in order to identify how these conditions influence the biases. The main identified biases are summarized in Table 3 and discussed below.

Our main finding corroborates previous findings of large boundary layer cloud biases in models and their subsequent effect on the radiative transfer. For example, low- and mid-level clouds in the cold-air sector of cyclones were identified as being responsible for most of the SW bias by Bodas-Salcedo et al. (2012). Precipitation in intense extratropical oceanic cyclones is projected to increase with future warming (Kodama et al., 2019). The understanding of radiation biases was refined by Bodas-Salcedo et al. (2014), who highlighted that the SW bias was associated with an incorrectly simulated mid-level cloud regime, which occurred in regions where clouds with tops at mid-level and low levels occurred. Ramadoss et al. (2024) have shown that in precipitating conditions, km-scale ICON has SW radiative biases associated with the overrepresentation of the liquid phase at the cloud top in low stratocumulus clouds in a short (48-h) simulation over the SO. Fiddes et al. (2024) suggested that biases in the LWP are the largest contributor to the cloud radiative bias over the SO. Our general finding applies to the new GSRM ICON, but the biases are lower than in the reanalyses in several aspects, namely the total cloud fraction, daily cloud cover, fog, and the LWP (Table 3), despite the reanalyses having the advantage of assimilation of the observed meteorological conditions. ICON, on the other hand, performs worse than the reanalyses in clouds and RH at 500 m, mid-level clouds (here defined as 1.5 km), outgoing SW radiation, and the IWP. ICON has the advantage of a much higher spatial resolution and, to a limited extent, explicit calculation of traditionally subgrid-scale processes such as convection. These are incomplete due to the lack of sub-grid scale convection parameterization below the km scale. The lack of parameterized subgrid-scale convection in ICON was a pragmatic choice in the model development, but it can be a source of substantial cloud biases even at the 5-km resolution.

We show that relative to ERA5, the distribution and strength of cyclonic activity over the SO is well represented in ICON, but it displays lower values of LTS. The latter is also manifested in the radiosonde profile comparison (Fig. 10), showing that the θ_v profiles in ICON are less stable than in the observations. It is also manifested in near-surface air temperature, which is overestimated in the 1–7°C range at the radiosonde launch locations (Fig. S1a). The underestimated LTS is linked to the overestimated cloud peak at 500 m in the lidar cloud occurrence comparison (Fig 6f–g). It might also be interacting with the cloud inhomogeneity factor employed in ICON (Section 2.5), resulting in lower cloud liquid water used in radiative calculations, hence decreased outgoing SW radiation. Based on the θ_v profile analysis, clouds at 500 m are predominantly convectively driven, and it is therefore expected that a model bias towards weak stability results in an increased cloud formation at this level. The underestimation of clouds above 1 km in ICON does not have a clear physical reason in our analysis and is likely partially or fully caused by stronger obscuration of the simulated lidar signal by the underlying and overestimated clouds in ICON at around 500 m.

The campaigns show remarkably similar biases in cloud occurrence by height in the lidar comparison (Fig. S2), which indicates that common underlying causes for the biases exist regardless of longitude and season. ICON underestimates the total cloud fraction by about 10%, with an overestimation of clouds below 1 km and an underestimation of clouds above 1 km. The reanalyses underestimate the total cloud fraction by about 20%. ERA5 overestimates clouds below 1 km but underestimates very low-level clouds and fog. ICON strongly overestimates the peak of cloud occurrence at about 500 m. This can be explained by the radiosonde comparison, showing that it is too moist at around this height (Fig. 9a); has underestimated LTS (Fig. 5 and 10), permitting shallow convection to this height; and has underestimated near-surface RH (Fig. S1), resulting in higher LCL (Fig. 6). Similar to our results for mid-level clouds, Cesana et al. (2022) showed that CMIP6 models also tend to underestimate cloud occurrence above 2 km over the SO, although their analysis in this case was limited to liquid clouds.

The inability of the models to simulate fog can be linked to various biases identified in our analysis. Near-surface RH is too low in the models (Fig. S1), potentially due to low moisture flux from the surface and too effective boundary layer mixing. Near-surface temperature is also too high in ICON, and it can be expected that fog formation occurs in low near-surface temperature conditions when a warm and moist air mass is cooled by the surface to the saturation point. Fig. S4 shows that fog occurs under highly stratified conditions. The underestimated LTS in ICON (and to a lesser extent in the reanalyses; Fig. 10) indicates that the models are biased to weaker stability, thus having less favorable conditions for fog formation and persistence. The RH distribution in cloudy bins (Fig. 11) also suggests that in observations, very low-level hydrometeors can occur under lower RH in observations than in the models. This could be due to high availability of cloud condensation nuclei (CCN) or ice nucleating particles (INPs) or due to hydrometeors and aerosols formed via sea spray under high swell and wind conditions. These parametrizations are likely very uncertain in the models in the SO due to the sparsity of reference data. Kawai et al. (2016) have shown that marine fog has some of the highest concentrations globally over the SO, and SO marine fog has a greater occurrence in winter. They conclude that marine fog is related to large-scale circulation and warm advection, and this is expected to change in a warming climate.

Compared to lidar observations, the daily cloud cover tends to be about 1 okta lower in ICON and 2 oktas lower in the reanalyses. Conditions of weak stability are associated with some of the greatest biases, especially in ERA5. The models also underestimate the cloud cover very strongly in cyclonic conditions, which are very cloudy in the observations (8 oktas) but much less so in the models. Similarly, McErlich et al. (2023) found a 40% underestimation of cloud liquid water in cyclones over the SO in ERA5, despite total column water vapor being simulated much more accurately (5% underestimation).

The radiosonde observations indicate that the LCL is too high in ICON and reanalyses, which is probably responsible for the higher peak of clouds in the models and the lack of very low-level clouds and fog. Notably, ICON exhibits smaller internal variability in θ_v than the radiosonde observations. The analysis of the LWP and IWP (Fig. S3) shows that both phases are present in observations in about equal amounts. The models show diverse biases, the most pronounced being overestimation of high-LWP values in MERRA-2 and overestimation of cases with a near-zero LWP and IWP in all models. All models tend to compensate for the overestimated cases of a near-zero LWP with more high-LWP values to get a mean LWP that is either less (but close) to the observations (ERA5) or higher than the observations (ICON and MERRA-2). The IWP is underestimated in all of the models. In the case of ICON and MERRA-2, the mean IWP was underestimated and LWP overestimated, indicating that the models produce too much liquid and not enough ice phase. This is in contrast with previous findings of the lack of supercooled liquid over the SO in other models. If the liquid phase is overestimated relative to the ice phase, one would expect underestimated cloud SW reflectivity due to a larger number of smaller hydrometeors for the same amount of water. Cloudy areas would then appear brighter in the SW spectrum. This can contribute to the too few, too bright bias, i.e., the overestimated brightness of cloudy areas compensates for the lower total cloud fraction in the models. As mentioned in Section 3.4, the LWP and IWP are, however, affected by the high uncertainty of the satellite retrievals.

The relationship between cloud biases and radiation has a number of notable features. MERRA-2 exhibits the too-few-too-bright bias previously identified in models. In our results, this is characterized by overestimated outgoing TOA SW radiation, while at the same time total cloud fraction is underestimated based on the ground-based lidar observations. On the other hand, this relationship is not present in ICON or ERA5. ICON predicts smaller outgoing TOA SW radiation and smaller total cloud fraction than observations, and the deficit of outgoing TOA SW radiation is approximately proportional to the deficit of the total cloud fraction. While this might be a welcome feature and an improvement over previous models, it does mean that the outgoing TOA SW radiation is overall underestimated instead of being compensated by a higher cloud albedo. This can, of course, lead to undesirable secondary effects such as overestimated solar heating of the sea surface, among other factors responsible for SO SST biases in climate models (Q. Zhang et al., 2023; Luo et al., 2023; Hyder et al., 2018). In contrast with our results, A. J. Schuddeboom and McDonald (2021) showed that CMIP6 models tend to overestimate a stratocumulus cloud regime over the SO.

Our results imply that SO cloud biases are a substantial issue even in the km-scale resolution ICON and the reanalyses. More effort is therefore needed to improve the model cloud simulations in this understudied region. We see that while the ICON is superior to the coarser reanalyses in some aspects (Table 3), it is affected by cloud biases large enough to cause important radiative biases. Parts of the GSRM relevant to low clouds, however, do not benefit from the higher resolution, such as cloud microphysics, unresolved clouds smaller than the grid cell, and turbulence. Cloud biases have also been shown to be a persistent issue in other GSRM models (Seiki et al., 2022).

We suggest the following avenues for future research. Evaluation of ocean–atmosphere heat, moisture, and momentum fluxes with in-situ observations over the SO and comparison of GSRM simulations with large-eddy simulations in process-oriented studies; evaluation of the DYAMOND project simulations in a similar manner as performed here (for models that provide the necessary fields); and combining active satellite sensors such as the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIOP) on CALIPSO and Atmospheric Lidar [ATLID; Hélière et al. (2017)] on the Earth Clouds, Aerosols and Radiation Explorer [EarthCARE; Illingworth et al. (2015)] satellite with ground-based remote sensing could provide a more complete understanding of the cloud biases across the whole troposphere. Cloud phase could be analyzed in more detail using the CALIPSO data, as was done by Roh et al. (2020) in a cloud-resolving model, or using ground-based observations with the dual-polarization Mini Micro Pulse Lidar [MiniMPL; Spinhirne (1993); Campbell et al. (2002); Flynn et al. (2007)] data available from the TAN1802 voyage. Guyot et al. (2022) and Whitehead et al. (2024) have developed a machine learning method for identifying cloud phase from ceilometer data, and this could be used with our ground-based lidar dataset to analyze the cloud phase. However, their method would require a careful calibration with reference data coming from this region.

Open Research Section

The RV *Polarstern* datasets are openly available on Pangaea (`https://pangaea.de`), as listed in Table 2. The MARCUS and MICRE datasets are openly available from ARM (`https://www.arm.gov`). The MERRA-2 data are openly available from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) (`https://disc.gsfc.nasa.gov/datasets?project=MERRA-2`). The ERA5 data are openly available from the Copernicus Climate Data Store (CDS) (`https://cds.climate.copernicus.eu`). The ICON data are available on the Levante cluster of the DKRZ (`https://www.dkrz.de/en/systems/hpc/hlre-4-levante`) after registration at `https://luv.dkrz.de/register/`. The CERES products are openly available from the project website (`https://ceres.larc.nasa.gov`) and the NASA Atmospheric Science Data Centre (`https://asdc.larc.nasa.gov/project/CERES`). The TAN1802 data are openly available on Zenodo (Kremser et al., 2020). The remaining voyage data (AA15-16, HMNZSW16, NBP1704, TAN1502, and TAN1702) are openly available on Zenodo (McDonald, Alexander, et al., 2024). The Natural Earth dataset is openly available from `https://www.naturalearthdata.com`. The code used in our analysis is open-source and available on Zenodo: the code for performing the presented analysis (Kuma, 2024c), precipitation detection (Kuma, 2024a), cl2nc (Kuma, 2024b), and a custom version of the ALCF used in our analysis (Kuma et al., 2024).

Acknowledgments

This research was supported by the European Union's Horizon 2020 research and innovation program projects nextGEMS (grant number 101003470) and CleanCloud (grant number 101137639), the Hans Ertel Centre for Weather Research (grant number 4823DWDP7), and the Swedish e-Science Research Centre (SeRC). FB received funding from the Wenner-Gren Foundation. The work of GM was supported by the United States (U.S.) Department of Energy Award DE-SC0021159. Supercomputing resources were provided by the DKRZ (project 1125 ICON-development) and the National Academic Infrastructure for Supercomputing in Sweden (allocation 2023/22-202). The data collection by the University of Canterbury was funded by the Deep South National Science Challenge Clouds and Aerosols project. Data collection on the AA15-16 voyages was funded by the Australian Antarctic Science project (grant no. 4292). We acknowledge the contribution of Thorsten Mauritsen to funding acquisition and project management. We acknowledge the RV *Polarstern* datasets provided by the Alfred Wegener Institute and Pangaea; the

AA15-16 dataset provided by the Australian Antarctic Division (AAD) and University of Canterbury (UC); the RV *Tangaroa* datasets provided by the National Institute of Water and Atmospheric Research and UC; the NBP1704 dataset provided by the National Science Foundation, Cooperative Institute for Research in Environmental Sciences, University of Colorado, and UC; the HMNZSW16 dataset provided by the Royal New Zealand Navy and UC; the MARCUS dataset provided by ARM and AAD; and the MICRE dataset provided by ARM, the Australian Bureau of Meteorology, and AAD. Technical, logistical, and ship support for MARCUS and MICRE were provided by the AAD through Australian Antarctic Science projects 4292 and 4387, and we thank Steven Whiteside, Lloyd Symonds, Rick van den Enden, Peter de Vries, Chris Young, Chris Richards, Andrew Klekociuk, John French, Terry Egan, Nick Cartwright, and Ken Barrett for all of their assistance. We thank the scientific staff, the crew, and everyone involved in collecting data on the voyages and stations, especially Gert König-Langlo, Holger Schmithüsen, Roger Marchand, Peter Guest, Kelly Schick, Jamie Halla, and Mike J. Harvey (†). We thank Loretta Preis for providing additional RV *Polarstern* data. We acknowledge the ICON model output provided by the nextGEMS project, Deutscher Wetterdienst, Max-Planck-Institute for Meteorology, DKRZ, Karlsruhe Institute of Technology, and Center for Climate Systems Modeling; the reanalysis dataset ERA5 provided by the Copernicus Climate Change Service; MERRA-2 provided by the Global Modeling and Assimilation Office; CERES datasets provided by the NASA Langley Atmospheric Science Data Center Distributed Active Archive Center; and the Natural Earth dataset provided by naturalearthdata.com. Last but not least, we acknowledge the use of open-source software: Python, Cython (Behnel et al., 2011), TensorFlow (Abadi et al., 2016), Devuan GNU+Linux, parallel (Tange, 2011), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), cartopy (Met Office, 2010), pyproj, Inkscape, Bash, GNU Fortran, HDF (Folk et al., 1999), and NetCDF (Rew & Davis, 1990). We dedicate this study to the memory of Mike J. Harvey, who very substantially contributed to obtaining the atmospheric observations on the RV *Tangaroa* voyages used in this study.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems [Software]*. Retrieved from `https://www.tensorflow.org`

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th usenix conference on operating systems design and implementation* (pp. 265–283). USA: USENIX Association.

Ackley, S. F., Stammerjohn, S., Maksym, T., Smith, M., Cassano, J., Guest, P., ... Parno, J. (2020). Sea-ice production and air/ice/ocean/biogeochemistry interactions in the Ross Sea during the PIPERS 2017 autumn field campaign. *Annals of Glaciology*, *61*(82), 181–195. doi: 10.1017/aog.2020.31

Bacmeister, J. T., Suarez, M. J., & Robertson, F. R. (2006). Rain reevaporation, boundary layer–convection interactions, and pacific rainfall patterns in an AGCM. *Journal of the Atmospheric Sciences*, *63*(12), 3383–3403. doi: 10.1175/JAS3791.1

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. S., & Smith, K. (2011). Cython: The best of both worlds. *Computing in Science Engineering*, *13*(2), 31–39. doi: 10.1109/MCSE.2010.118

Bender, F. A.-M., Engström, A., Wood, R., & Charlson, R. J. (2017). Evaluation of hemispheric asymmetries in marine cloud radiative properties. *Journal of Climate*, *30*(11), 4131–4147. doi: 10.1175/JCLI-D-16-0263.1

Bodas-Salcedo, A., Webb, M., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S., ... others (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. doi:

10.1175/2011BAMS2856.1

Bodas-Salcedo, A., Williams, K. D., Field, P. R., & Lock, A. P. (2012). The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: The role of midlatitude cyclone clouds. *Journal of Climate*, *25*(21), 7467–7486. doi: 10.1175/jcli-d-11-00702.1

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J. L., … Yokohata, T. (2014). Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models. *Journal of Climate*, *27*(1), 41–56. doi: 10.1175/jcli-d-13-00169.1

Boebel, O., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.808838

Boebel, O., & Rohardt, G. (2016). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS89 (ANT-XXX/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.858884

Boebel, O., & Rohardt, G. (2018). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS103 (ANT-XXXII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.889513

Boebel, O., & Rohardt, G. (2019). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS117 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.905562

Bohrmann, G., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.810678

Bosilovich, M. G., Lucchesi, R., & Suarez, M. (2016). MERRA-2: File specification [Computer software manual]. Greenbelt, Maryland 20771, USA. Retrieved from `http://gmao.gsfc.nasa.gov/pubs/office_notes` (GMAO Office Note No. 9 (Version 1.1))

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., … Zhang, X. Y. (2013). Clouds and aerosols. In T. F. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Bubnová, R., Hello, G., Bénard, P., & Geleyn, J.-F. (1995). Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Monthly Weather Review*, *123*(2), 515–535. doi: 10.1175/1520-0493(1995)123⟨0515:IOTFEE⟩2.0.CO;2

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., … Zender, C. S. (2021). Convection-permitting simulations with the E3SM global atmosphere model. *Journal of Advances in Modeling Earth Systems*, *13*(11), e2021MS002544. doi: 10.1029/2021MS002544

Campbell, J. R., Hlavka, D. L., Welton, E. J., Flynn, C. J., Turner, D. D., Spinhirne, J. D., … Hwang, I. (2002). Full-time, eye-safe cloud and aerosol lidar observation at atmospheric radiation measurement program sites: Instruments and data processing. *Journal of Atmospheric and Oceanic Technology*, *19*(4), 431–442. doi: 10.1175/1520-0426(2002)019⟨0431:FTESCA⟩2.0.CO;2

CERES author team. (2025). *CERES_SYN1deg_Ed4A data quality summary.* Retrieved from `https://ceres.larc.nasa.gov/documents/DQ_summaries/CERES_SYN1deg_Ed4A_DQS.pdf` (last access: 26 June 2025)

Cesana, G. V., Khadir, T., Chepfer, H., & Chiriaco, M. (2022). Southern Ocean solar reflection biases in CMIP6 models linked to cloud phase and vertical structure representations. *Geophysical Research Letters*, *49*(22), e2022GL099777. doi: 10.1029/2022GL099777

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., … Zeng, S. (2010). The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP). *Journal of Geophysical Research: Atmospheres*, *115*(D4), D00H16. doi: 10.1029/2009JD012251

Danker, J., Sourdeval, O., McCoy, I. L., Wood, R., & Possner, A. (2022). Exploring relations between cloud morphology, cloud phase, and cloud radiative properties in Southern Ocean's stratocumulus clouds. *Atmospheric Chemistry and Physics*, *22*(15), 10247–10265. doi: 10.5194/acp-22-10247-2022

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., … Sun, M. (2013). Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072–1090. doi: 10.1175/JTECH-D-12-00136.1

Doelling, D. R., Sun, M., Nguyen, L. T., Nordeen, M. L., Haney, C. O., Keyes, D. F., & Mlynczak, P. E. (2016). Advances in geostationary-derived longwave fluxes for the CERES synoptic (SYN1deg) product. *Journal of Atmospheric and Oceanic Technology*, *33*(3), 503–521. doi: 10.1175/JTECH-D-15-0147.1

Dorschel, B., & Rohardt, G. (2019). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS118 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.905608

Duncan, D. I., & Eriksson, P. (2018). An update on global atmospheric ice estimates from satellite observations and reanalyses. *Atmospheric Chemistry and Physics*, *18*(15), 11205–11219. doi: 10.5194/acp-18-11205-2018

DYAMOND author team. (2024). *DYAMOND Initiative.* Retrieved from `https://www.esiwace.eu/the-project/past-phases/dyamond-initiative` (Accessed on 19 June 2024)

ECMWF. (2019, 11). *Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.* Copernicus Climate Change Service Climate Data Store (CDS). doi: 10.24381/cds.bd0915c6

ECMWF. (2023). *IFS documentation CY48R1.* Shinfield Park, Reading, RG2 9AX, United Kingdom: Author. Retrieved from `https://www.ecmwf.int/en/publications/ifs-documentation` (Accessed on 4 December 2024)

Eliasson, S., Buehler, S. A., Milz, M., Eriksson, P., & John, V. O. (2011). Assessing observed and modelled spatial distributions of ice water path using satellite data. *Atmospheric Chemistry and Physics*, *11*(1), 375–391. doi: 10.5194/acp-11-375-2011

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fahrbach, E., & Rohardt, G. (2011). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760120

Fiddes, S. L., Mallet, M. D., Protat, A., Woodhouse, M. T., Alexander, S. P., & Furtado, K. (2024). A machine learning approach for evaluating Southern Ocean cloud radiative biases in a global atmosphere model. *Geoscientific Model Development*, *17*(7), 2641–2662. doi: 10.5194/gmd-17-2641-2024

Fiddes, S. L., Protat, A., Mallet, M. D., Alexander, S. P., & Woodhouse, M. T. (2022). Southern Ocean cloud and shortwave radiation biases in a nudged climate model simulation: does the model ever get it right? *Atmospheric Chemistry and Physics*, *22*(22), 14603–14630. doi: 10.5194/acp-22-14603-2022

Flynn, C. J., Mendozaa, A., Zhengb, Y., & Mathurb, S. (2007). Novel polarization-sensitive micropulse lidar measurement technique. *Optics express*, *15*(6), 2785–2790. doi: 10.1364/OE.15.002785

Folk, M., McGrath, R., & Yeager, N. (1999). HDF: an update and future directions.

In *Ieee 1999 international geoscience and remote sensing symposium. igarss'99 (cat. no.99ch36293)* (Vol. 1, pp. 273–275). doi: 10.1109/IGARSS.1999.773469

Forbes, R. M., & Ahlgrimm, M. (2014). On the representation of high-latitude boundary layer mixed-phase cloud in the ECMWF global model. *Monthly Weather Review*, *142*(9), 3425–3445. doi: 10.1175/MWR-D-13-00325.1

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... others (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., ... Stevens, B. (2018). ICON-A, the atmosphere component of the ICON Earth system model: I. model description. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1613–1637. doi: 10.1029/2017MS001242

Gohl, K., & Rohardt, G. (2018). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS104 (ANT-XXXII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.889515

Greenwald, T. J. (2009). A 2 year comparison of amsr-e and modis cloud liquid water path observations. *Geophysical Research Letters*, *36*(20). doi: 10.1029/2009GL040394

Grundner, A. (2023). *Data-driven cloud cover parameterizations for the ICON Earth system model using deep learning and symbolic regression* (Doctoral dissertation, University of Bremen, Bremen, Germany). doi: 10.26092/elib/2821

Grundner, A., Beucler, T., Gentine, P., Iglesias-Suarez, F., Giorgetta, M. A., & Eyring, V. (2022). Deep learning based cloud cover parameterization for ICON. *Journal of Advances in Modeling Earth Systems*, *14*(12), e2021MS002959. doi: 10.1029/2021MS002959

Gutt, J., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.809727

Guyot, A., Protat, A., Alexander, S. P., Klekociuk, A. R., Kuma, P., & McDonald, A. (2022). Detection of supercooled liquid water containing clouds with ceilometers: development and evaluation of deterministic and data-driven retrievals. *Atmospheric Measurement Techniques*, *15*(12), 3663–3681. doi: 10.5194/amt-15-3663-2022

Hahn, C. J., Rossow, W. B., & Warren, S. G. (2001). ISCCP cloud properties associated with standard cloud types identified in individual surface observations. *Journal of Climate*, *14*(1), 11–28. doi: 10.1175/1520-0442(2001)014⟨0011:ICPAWS⟩2.0.CO;2

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, *585*(7825), 357–362. doi: 10.1038/s41586-020-2649-2

Hartery, S., Toohey, D., Revell, L., Sellegri, K., Kuma, P., Harvey, M., & McDonald, A. J. (2020). Constraining the surface flux of sea spray particles from the Southern Ocean. *Journal of Geophysical Research: Atmospheres*, *125*(4), e2019JD032026. doi: 10.1029/2019JD032026

Hélière, A., Gelsthorpe, R., Hors, L. L., & Toulemont, Y. (2017). ATLID, the atmospheric lidar on board the Earthcare satellite. In B. Cugny, E. Armandillo, & N. Karafolas (Eds.), *International conference on space optics — icso 2012* (Vol. 10564, p. 105642D). SPIE. doi: 10.1117/12.2309095

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., ... Stevens, B. (2023). ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, *16*(2), 779–811. doi: 10.5194/gmd-16-779-2023

Hoppmann, M., Tippenhauer, S., & Heitland, T. (2023). *Continuous thermosalinograph oceanography along RV POLARSTERN cruise track PS123 [Dataset].*

PANGAEA. doi: 10.1594/PANGAEA.938753

Hoppmann, M., Tippenhauer, S., & Hellmer, H. H. (2023). *Continuous thermosalinograph oceanography along RV POLARSTERN cruise track PS124 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.938767

Hoskins, B. J., & Hodges, K. I. (2005). A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, *18*(20), 4108–4129. doi: 10.1175/JCLI3570.1

Huang, J., Minnis, P., Lin, B., Yi, Y., Fan, T.-F., Sun-Mack, S., & Ayers, J. K. (2006). Determination of ice water path in ice-over-water cloud systems using combined MODIS and AMSR-E measurements. *Geophysical Research Letters*, *33*(21). doi: 10.1029/2006GL027038

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., … Belcher, S. E. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*(1), 3625. doi: 10.1038/s41467-018-05634-2

Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., … van Zadelhoff, G.-J. (2015). The EarthCARE satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation. *Bulletin of the American Meteorological Society*, *96*(8), 1311–1332. doi: 10.1175/BAMS-D-12-00227.1

Jokat, W., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/5 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.816055

Karlsson, J., Svensson, G., & Rodhe, H. (2008). Cloud radiative forcing of subtropical low level clouds in global models. *Climate Dynamics*, *30*(7), 779–788. doi: 10.1007/s00382-007-0322-1

Kattner, G., & Rohardt, G. (2012). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.776596

Kawai, H., Koshiro, T., Endo, H., Arakawa, O., & Hagihara, Y. (2016). Changes in marine fog in a warmer climate. *Atmospheric Science Letters*, *17*(10), 548–555. doi: https://doi.org/10.1002/asl.691

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*(4), 607–625. doi: 10.1175/1520-0469(2003)060⟨0607:CRMOTA⟩2.0.CO;2

Khanal, S., Wang, Z., & French, J. R. (2020). Improving middle and high latitude cloud liquid water path measurements from MODIS. *Atmospheric Research*, *243*, 105033. doi: 10.1016/j.atmosres.2020.105033

Kim, Y.-H., Moon, S.-H., & Yoon, Y. (2020). Detection of precipitation and fog using machine learning on backscatter data from lidar ceilometer. *Applied Sciences*, *10*(18). doi: 10.3390/app10186452

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Gleckler, P. J. (2013). Are climate model simulations of clouds improving? an evaluation using the ISCCP simulator. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1329–1342. doi: 10.1002/jgrd.50141

Klekociuk, A. R., French, W. J. R., Alexander, S. P., Kuma, P., & McDonald, A. J. (2020). The state of the atmosphere in the 2016 southern Kerguelen Axis campaign region. *Deep Sea Research Part II: Topical Studies in Oceanography*, *174*. (Ecosystem drivers of food webs on the Kerguelen Axis of the Southern Ocean) doi: 10.1016/j.dsr2.2019.02.001

Knight, C. L., Mallet, M. D., Alexander, S. P., Fraser, A. D., Protat, A., & McFarquhar, G. M. (2024). Cloud properties and boundary layer stability above

Southern Ocean sea ice and coastal Antarctica. *Journal of Geophysical Research: Atmospheres*, *129*(10), e2022JD038280. doi: https://doi.org/10.1029/2022JD038280

Knust, R., & Rohardt, G. (2011). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760121

Knust, R., & Rohardt, G. (2014). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS82 (ANT-XXIX/9) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.839407

Kodama, C., Stevens, B., Mauritsen, T., Seiki, T., & Satoh, M. (2019). A new perspective for future precipitation change from intense extratropical cyclones. *Geophysical Research Letters*, *46*(21), 12435–12444. doi: 10.1029/2019GL084001

Koldunov, N., Kölling, T., Pedruzo-Bagazgoitia, X., Rackow, T., Redler, R., Sidorenko, D., ... Ziemen, F. A. (2023). *nextGEMS: output of the model development cycle 3 simulations for ICON and IFS.* World Data Center for Climate (WDCC) at DKRZ. doi: 10.26050/WDCC/nextGEMS_cyc3

Kong, X., Wolf, M. J., Roesch, M., Thomson, E. S., Bartels-Rausch, T., Alpert, P. A., ... Cziczo, D. J. (2018). A continuous flow diffusion chamber study of sea salt particles acting as cloud nuclei: deliquescence and ice nucleation. *Tellus B: Chemical and Physical Meteorology*, *70*(1), 1–11. doi: 10.1080/16000889.2018.1463806

König-Langlo, G. (2011a). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760388

König-Langlo, G. (2011b). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760389

König-Langlo, G. (2011c). *Meteorological observations during POLARSTERN cruise ANT-XXVII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760392

König-Langlo, G. (2011d). *Meteorological observations during POLARSTERN cruise ANT-XXVII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.760393

König-Langlo, G. (2011e). *Upper air soundings during POLARSTERN cruise ANT-XXVII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.849045

König-Langlo, G. (2012a). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784453

König-Langlo, G. (2012b). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784454

König-Langlo, G. (2012c). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784455

König-Langlo, G. (2012d). *Meteorological observations during POLARSTERN cruise ANT-XXVIII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784458

König-Langlo, G. (2012e). *Meteorological observations during POLARSTERN cruise ANT-XXVIII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784459

König-Langlo, G. (2012f). *Meteorological observations during POLARSTERN cruise ANT-XXVIII/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.784460

König-Langlo, G. (2012g). *Upper air soundings during POLARSTERN cruise ANT-XXVII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.849044

König-Langlo, G. (2012h). *Upper air soundings during POLARSTERN cruise ANT-XXVIII/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844866

König-Langlo, G. (2012i). *Upper air soundings during POLARSTERN cruise ANT-XXVIII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844865

König-Langlo, G. (2012j). *Upper air soundings during POLARSTERN cruise ANT-XXVIII/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844859

König-Langlo, G. (2013a). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815472

König-Langlo, G. (2013b). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815473

König-Langlo, G. (2013c). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815723

König-Langlo, G. (2013d). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/5 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815474

König-Langlo, G. (2013e). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/6 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.820733

König-Langlo, G. (2013f). *Meteorological observations during POLARSTERN cruise ANT-XXIX/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815476

König-Langlo, G. (2013g). *Meteorological observations during POLARSTERN cruise ANT-XXIX/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815477

König-Langlo, G. (2013h). *Meteorological observations during POLARSTERN cruise ANT-XXIX/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815724

König-Langlo, G. (2013i). *Meteorological observations during POLARSTERN cruise ANT-XXIX/5 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.815478

König-Langlo, G. (2013j). *Meteorological observations during POLARSTERN cruise ANT-XXIX/6 (AWECS) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.819610

König-Langlo, G. (2013k). *Meteorological observations during POLARSTERN cruise ANT-XXIX/7 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.820843

König-Langlo, G. (2013l). *Upper air soundings during POLARSTERN cruise ANT-XXIX/2 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844856

König-Langlo, G. (2013m). *Upper air soundings during POLARSTERN cruise ANT-XXIX/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844855

König-Langlo, G. (2013n). *Upper air soundings during POLARSTERN cruise ANT-XXIX/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844854

König-Langlo, G. (2013o). *Upper air soundings during POLARSTERN cruise ANT-XXIX/5 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844853

König-Langlo, G. (2013p). *Upper air soundings during POLARSTERN cruise ANT-XXIX/6 (AWECS) to the Antarctic in 2013 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.842810

König-Langlo, G. (2013q). *Upper air soundings during POLARSTERN cruise ANT-XXIX/7 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844852

König-Langlo, G. (2013r). *Upper air soundings during POLARSTERN cruise ANT-XXIX/8 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844809

König-Langlo, G. (2014a). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/2, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834526

König-Langlo, G. (2014b). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/3, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834527

König-Langlo, G. (2014c). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/4, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834528

König-Langlo, G. (2014d). *Ceilometer CL51 raw data measured during PO-*

LARSTERN cruise ANT-XXIX/5, links to files [Dataset]. PANGAEA. doi: 10.1594/PANGAEA.834529

König-Langlo, G. (2014e). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/6 (AWECS), links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.833801

König-Langlo, G. (2014f). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/7, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.833802

König-Langlo, G. (2014g). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/8, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834530

König-Langlo, G. (2014h). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVII/2, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834511

König-Langlo, G. (2014i). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVII/3, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834512

König-Langlo, G. (2014j). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/2, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834518

König-Langlo, G. (2014k). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/3, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834519

König-Langlo, G. (2014l). *Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/4, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834520

König-Langlo, G. (2014m). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS82 (ANT-XXIX/9), links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.834531

König-Langlo, G. (2014n). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/8 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.832602

König-Langlo, G. (2014o). *Continuous meteorological surface measurement during POLARSTERN cruise PS82 (ANT-XXIX/9) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.832603

König-Langlo, G. (2014p). *Meteorological observations during POLARSTERN cruise ANT-XXIX/8 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.832605

König-Langlo, G. (2014q). *Meteorological observations during POLARSTERN cruise PS82 (ANT-XXIX/9) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.832606

König-Langlo, G. (2014r). *Upper air soundings during POLARSTERN cruise PS82 (ANT-XXIX/9) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844805

König-Langlo, G. (2015a). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS89, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844075

König-Langlo, G. (2015b). *Continuous meteorological surface measurement during POLARSTERN cruise PS89 (ANT-XXX/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.849502

König-Langlo, G. (2015c). *Meteorological observations during POLARSTERN cruise PS89 (ANT-XXX/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844571

König-Langlo, G. (2015d). *Upper air soundings during POLARSTERN cruise PS89 (ANT-XXX/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.844791

König-Langlo, G. (2016a). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS96, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.860516

König-Langlo, G. (2016b). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS97, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.860517

König-Langlo, G. (2016c). *Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/7 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.858532

König-Langlo, G. (2016d). *Continuous meteorological surface measurement during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861441

König-Langlo, G. (2016e). *Continuous meteorological surface measurement during POLARSTERN cruise PS97 (ANT-XXXI/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861442

König-Langlo, G. (2016f). *Meteorological observations during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861438

König-Langlo, G. (2016g). *Meteorological observations during POLARSTERN cruise PS97 (ANT-XXXI/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861439

König-Langlo, G. (2016h). *Upper air soundings during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861658

König-Langlo, G. (2016i). *Upper air soundings during POLARSTERN cruise PS97 (ANT-XXXI/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.861659

König-Langlo, G. (2017a). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS103, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.871722

König-Langlo, G. (2017b). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS104, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.874154

König-Langlo, G. (2017c). *Continuous meteorological surface measurement during POLARSTERN cruise PS103 (ANT-XXXII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.873016

König-Langlo, G. (2017d). *Meteorological observations during POLARSTERN cruise PS103 (ANT-XXXII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.871721

König-Langlo, G. (2017e). *Meteorological observations during POLARSTERN cruise PS104 (ANT-XXXII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.874156

König-Langlo, G. (2017f). *Upper air soundings during POLARSTERN cruise PS103 (ANT-XXXII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.871788

König-Langlo, G. (2017g). *Upper air soundings during POLARSTERN cruise PS104 (ANT-XXXII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.874224

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., … Ogura, T. (2022). Low-level marine tropical clouds in six CMIP6 models are too few, too bright but also too compact and too homogeneous. *Geophysical Research Letters*, *49*(11), e2021GL097593. doi: 10.1029/2021GL097593

Konstali, K., Spengler, T., Spensberger, C., & Sorteberg, A. (2024). Linking future precipitation changes to weather features in CESM2-LE. *Journal of Geophysical Research: Atmospheres*, *129*(16), e2024JD041190. doi: https://doi.org/10.1029/2024JD041190

Korn, P., Brüggemann, N., Jungclaus, J. H., Lorenz, S. J., Gutjahr, O., Haak, H., … Marotzke, J. (2022). ICON-O: The ocean component of the ICON Earth system model—global simulation characteristics and local telescoping capability. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2021MS002952. doi: 10.1029/2021MS002952

Kremser, S., Harvey, M., Kuma, P., Hartery, S., Saint-Macary, A., McGregor, J., ... Parsons, S. (2020). *Southern Ocean Cloud and Aerosol data set: a compilation of measurements from the 2018 Southern Ocean Ross Sea Marine Ecosystems and Environment voyage [Dataset].* Zenodo. doi: 10.5281/zenodo.4060237

Kremser, S., Harvey, M., Kuma, P., Hartery, S., Saint-Macary, A., McGregor, J., ... Parsons, S. (2021). Southern Ocean cloud and aerosol data: a compilation of measurements from the 2018 Southern Ocean Ross Sea Marine Ecosystems and Environment voyage. *Earth System Science Data*, *13*(7), 3115–3153. doi: 10.5194/essd-13-3115-2021

Kuma, P. (2024a). *ALCF precipitation detection, version 1.0.0 [Software].* Zenodo. doi: 10.5281/zenodo.14548242

Kuma, P. (2024b). *cl2nc 3.7.1 [Software].* Zenodo. doi: 10.5281/zenodo.14548122

Kuma, P. (2024c). *Code for a manuscript "Ship and ground-based lidar and radiosonde evaluation of Southern Ocean clouds in the storm-resolving general circulation model ICON and the ERA5 and MERRA-2 reanalyses", version 1.0.0 [Software].* Zenodo. doi: 10.5281/zenodo.14548174

Kuma, P. (2024d). *rstool 2.0.0 [Software].* Zenodo. doi: 10.5281/zenodo.14548537

Kuma, P., McDonald, A. J., Morgenstern, O., Alexander, S. P., Cassano, J. J., Garrett, S., ... Williams, J. (2020). Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations. *Atmospheric Chemistry and Physics*, *20*(11), 6607–6630. doi: 10.5194/acp-20-6607-2020

Kuma, P., McDonald, A. J., Morgenstern, O., Querel, R., Silber, I., & Flynn, C. J. (2021). Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0). *Geoscientific Model Development*, *14*(1), 43-72. doi: 10.5194/gmd-14-43-2021

Kuma, P., McDonald, A. J., Morgenstern, O., Querel, R., Silber, I., & Flynn, C. J. (2024). *Automatic Lidar and Ceilometer Framework (ALCF), custom version for a manuscript "Ship and ground- based lidar and radiosonde evaluation of Southern Ocean clouds in the storm-resolving general circulation model ICON and the ERA5 and MERRA-2 reanalyses".* Zenodo. doi: 10.5281/zenodo.14548400

Lamy, F., & Rohardt, G. (2017). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS97 (ANT-XXXI/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.873147

Lemke, P., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/6 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.819831

Lier, P., & Bach, M. (2008). PARASOL a microsatellite in the A-Train for Earth atmospheric observations. *Acta Astronautica*, *62*(2), 257–263. doi: 10.1016/j.actaastro.2006.12.052

Lin, S.-J. (2004). A "vertically Lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, *132*(10), 2293–2307. doi: 10.1175/1520-0493(2004)132⟨2293:AVLFDC⟩2.0.CO;2

Loeb, N., Su, W., Doelling, D., Wong, T., Minnis, P., Thomas, S., & Miller, W. (2018). 5.03 - Earth's top-of-atmosphere radiation budget. In S. Liang (Ed.), *Comprehensive remote sensing* (p. 67-84). Oxford: Elsevier. doi: https://doi.org/10.1016/B978-0-12-409548-9.10367-7

Lucassen, M., & Rohardt, G. (2012). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/4 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.802809

Luo, F., Ying, J., Liu, T., & Chen, D. (2023). Origins of Southern Ocean warm sea surface temperature bias in CMIP6 models. *npj Climate and Atmospheric Science*, *6*(1), 127. doi: 10.1038/s41612-023-00456-6

Mace, G. G., Zhang, Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., &

Winker, D. (2009). A description of hydrometeor layer occurrence statistics derived from the first year of merged Cloudsat and CALIPSO data. *Journal of Geophysical Research: Atmospheres*, *114*(D8). doi: 10.1029/2007JD009755

Madec, G., & the NEMO System Team. (2023). *NEMO ocean engine reference manual.* Zenodo. doi: 10.5281/zenodo.8167700

Marchand, R., Mace, G. G., Ackerman, T., & Stephens, G. (2008). Hydrometeor detection using Cloudsat—an Earth-orbiting 94-GHz cloud radar. *Journal of Atmospheric and Oceanic Technology*, *25*(4), 519–533. doi: 10.1175/2007JTECHA1006.1

Mauritsen, T., Redler, R., Esch, M., Stevens, B., Hohenegger, C., Klocke, D., ... Schnur, R. (2022). Early development and tuning of a global coupled cloud resolving model, and its fast response to increasing CO2. *Tellus A: Dynamic Meteorology and Oceanography*. doi: 10.16993/tellusa.54

McDonald, A. J., Alexander, S. P., McFarquhar, G. M., Cassano, J. J., Plank, G. E., Parsons, S., ... Schick, K. (2024). *Voyage data for the manuscript "Ship-based lidar evaluation of Southern Ocean clouds in the storm-resolving general circulation model ICON, and the ERA5 and MERRA-2 reanalyses" [Dataset].* Zenodo. doi: 10.5281/zenodo.14422428

McDonald, A. J., Kuma, P., Pannell, M., Petterson, O., Plank, G. E., Rozliaiani, M. A. H., & Whitehead, L. E. (2024). Evaluating cloud properties at Scott Base: Comparing ceilometer observations with ERA5, JRA55, and MERRA2 reanalyses using an instrument simulator. *ESS Open Archive*. (Preprint) doi: 10.22541/essoar.171820795.52152814/v1

McErlich, C., McDonald, A., Renwick, J., & Schuddeboom, A. (2023). An assessment of Southern Hemisphere extratropical cyclones in ERA5 using WindSat. *Journal of Geophysical Research: Atmospheres*, *128*(22), e2023JD038554. doi: 10.1029/2023JD038554

McErlich, C., McDonald, A., Schuddeboom, A., & Silber, I. (2021). Comparing satellite- and ground-based observations of cloud occurrence over high southern latitudes. *Journal of Geophysical Research: Atmospheres*, *126*(6), e2020JD033607. doi: 10.1029/2020JD033607

McFarquhar, G. M., Bretherton, C. S., Marchand, R., Protat, A., DeMott, P. J., Alexander, S. P., ... McDonald, A. (2021). Observations of clouds, aerosols, precipitation, and surface radiation over the Southern Ocean: An overview of CAPRICORN, MARCUS, MICRE, and SOCRATES. *Bulletin of the American Meteorological Society*, *102*(4), E894–E928. doi: 10.1175/BAMS-D-20-0132.1

Medeiros, B., Nuijens, L., Antoniazzi, C., & Stevens, B. (2010). Low-latitude boundary layer clouds as seen by CALIPSO. *Journal of Geophysical Research: Atmospheres*, *115*(D23), D23207. doi: 10.1029/2010JD014437

Meissner, T., & Wentz, F. J. (2009). Wind-vector retrievals under rain with passive satellite microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(9), 3065–3083. doi: 10.1109/TGRS.2009.2027012

Met Office. (2010). Cartopy: a cartographic python library with a Matplotlib interface [Software] [Computer software manual]. Exeter, Devon. Retrieved from `https://scitools.org.uk/cartopy`

Meyer, B., & Rohardt, G. (2013). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/7 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.823259

Meyer, B., & Rohardt, G. (2018). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS112 (ANT-XXXIII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.895580

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development*, *8*(5), 1339–1356. doi: 10.5194/gmd-8-1339-2015

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The 'too few, too bright' tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*(21), L21801. doi: 10.1029/2012GL053421

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., … Wernli, H. (2013). IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, *94*(4), 529–547. doi: 10.1175/BAMS-D-11-00154.1

nextGEMS authors team. (2023). *nextGEMS Cycle 3.* Retrieved from `https://easy.gems.dkrz.de/DYAMOND/NextGEMS/cycle3.html` (Accessed on 12 June 2024)

nextGEMS authors team. (2024). *nextGEMS website.* Retrieved from `https://nextgems-h2020.eu` (Accessed on 17 June 2024)

Niu, Q., McFarquhar, G. M., Marchand, R., Theisen, A., Cavallo, S. M., Flynn, C., … Hill, T. C. J. (2024). 62°S witnesses the transition of boundary layer marine aerosol pattern over the Southern Ocean (50°S–68°S, 63°E–150°E) during the spring and summer: Results from MARCUS (I). *Journal of Geophysical Research: Atmospheres*, *129*(9), e2023JD040396. doi: 10.1029/2023JD040396

Oreopoulos, L., Cho, N., Lee, D., & Kato, S. (2016). Radiative effects of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *121*(5), 2299–2317. doi: 10.1002/2015JD024502

Pagano, T. S., & Durham, R. M. (1993). Moderate Resolution Imaging Spectroradiometer (MODIS). In W. L. Barnes (Ed.), *Sensor systems for the early earth observing system platforms* (Vol. 1939, pp. 2–17). SPIE. doi: 10.1117/12.152835

Pei, Z., Fiddes, S. L., French, W. J. R., Alexander, S. P., Mallet, M. D., Kuma, P., & McDonald, A. (2023). Assessing the cloud radiative bias at Macquarie Island in the ACCESS-AM2 model. *Atmospheric Chemistry and Physics*, *23*(23), 14691–14714. doi: 10.5194/acp-23-14691-2023

Possner, A., Danker, J., & Gryspeerdt, E. (2022). Resolution dependence of Southern Ocean mixed-phase clouds in ICON. *ESS Open Archive*. (Preprint) doi: 10.1002/essoar.10511442.1

Putman, W. M., & Suarez, M. (2011). Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophysical Research Letters*, *38*(16), L16809. doi: 10.1029/2011GL048438

Pérez-Alarcón, A., Coll-Hidalgo, P., Trigo, R. M., Nieto, R., & Gimeno, L. (2024). CyTRACK: An open-source and user-friendly python toolbox for detecting and tracking cyclones. *Environmental Modelling & Software*, *176*, 106027. doi: 10.1016/j.envsoft.2024.106027

Ramadoss, V., Pfannkuch, K., Protat, A., Huang, Y., Siems, S., & Possner, A. (2024). An evaluation of cloud-precipitation structures in mixed-phase stratocumuli over the southern ocean in kilometer-scale ICON simulations during CAPRICORN. *Journal of Geophysical Research: Atmospheres*, *129*(18), e2022JD038251. doi: https://doi.org/10.1029/2022JD038251

Revell, L. E., Kremser, S., Hartery, S., Harvey, M., Mulcahy, J. P., Williams, J., … Schuddeboom, A. (2019). The sensitivity of Southern Ocean aerosols and cloud microphysics to sea spray and sulfate aerosol production in the HadGEM3-GA7.1 chemistry–climate model. *Atmospheric Chemistry and Physics*, *19*(24), 15447–15466. doi: 10.5194/acp-19-15447-2019

Rew, R., & Davis, G. (1990). NetCDF: an interface for scientific data access. *IEEE Computer Graphics and Applications*, *10*(4), 76–82. doi: 10.1109/38.56302

Rienecker, M. M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., … Nielsen, J. E. (2008). *The GEOS-5 data assimilation system—documentation of versions 5.0.1, 5.1.0, and 5.2.0* (Tech. Rep.). NASA Center for AeroSpace Information, 7115 Standard Drive, Hanover, MD 21076-1320, USA. Retrieved from `https://gmao.gsfc.nasa.gov/pubs/docs/`

Rienecker369.pdf (Technical Report Series on Global Modeling and Data Assimilation, Volume 27)

Rintoul, S. R. (2011). The Southern Ocean in the Earth system. In P. A. Berkman, M. A. Lang, D. W. H. Walton, & O. R. Young (Eds.), *Science diplomacy: Antarctica, science, and the governance of international spaces* (pp. 175–187). Washington, DC, USA: Smithsonian Institution Scholarly Press. doi: 10.5479/si.9781935623069.175

Roh, W., Satoh, M., Hashino, T., Okamoto, H., & Seiki, T. (2020). Evaluations of the thermodynamic phases of clouds in a cloud-system-resolving model using CALIPSO and a satellite simulator over the Southern Ocean. *Journal of the Atmospheric Sciences*, *77*(11), 3781–3801. doi: 10.1175/JAS-D-19-0273.1

Roh, W., Satoh, M., & Hohenegger, C. (2021). Intercomparison of cloud properties in DYAMOND simulations over the Atlantic Ocean. *Journal of the Meteorological Society of Japan. Ser. II*, *99*(6), 1439-1451. doi: 10.2151/jmsj.2021-070

Rossow, W. B., & Schiffer, R. A. (1991). ISCCP cloud data products. *Bulletin of the American Meteorological Society*, *72*(1), 2–20. doi: 10.1175/1520-0477(1991)072⟨0002:ICDP⟩2.0.CO;2

Rossow, W. B., & Schiffer, R. A. (1999). Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, *80*(11), 2261–2288. doi: 10.1175/1520-0477(1999)080⟨2261:AIUCFI⟩2.0.CO;2

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., & Iga, S. (2008). Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, *227*(7), 3486–3514. doi: 10.1016/j.jcp.2007.02.006

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184. doi: 10.1007/s40641-019-00131-0

Schlindwein, V., & Rohardt, G. (2014). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/8 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.839406

Schmithüsen, H. (2019a). *Radiosonde measurements during POLARSTERN cruise PS111 (ANT-XXXIII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.903862

Schmithüsen, H. (2019b). *Radiosonde measurements during POLARSTERN cruise PS112 (ANT-XXXIII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.903863

Schmithüsen, H. (2019c). *Radiosonde measurements during POLARSTERN cruise PS117 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.903916

Schmithüsen, H. (2019d). *Radiosonde measurements during POLARSTERN cruise PS118 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.903922

Schmithüsen, H. (2020a). *Meteorological observations during POLARSTERN cruise PS111 (ANT-XXXIII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.913625

Schmithüsen, H. (2020b). *Meteorological observations during POLARSTERN cruise PS112 (ANT-XXXIII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.913626

Schmithüsen, H. (2020c). *Meteorological observations during POLARSTERN cruise PS117 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.913632

Schmithüsen, H. (2020d). *Meteorological observations during POLARSTERN cruise PS118 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.913633

Schmithüsen, H. (2021a). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS111, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929484

Schmithüsen, H. (2021b). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS112, links to files [Dataset].* PANGAEA. doi:

10.1594/PANGAEA.929488
Schmithüsen, H. (2021c). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS117, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929505
Schmithüsen, H. (2021d). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS118, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929510
Schmithüsen, H. (2021e). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS123, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935273
Schmithüsen, H. (2021f). *Ceilometer CL51 raw data measured during POLARSTERN cruise PS124, links to files [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935274
Schmithüsen, H. (2021g). *Continuous meteorological surface measurement during POLARSTERN cruise PS104 (ANT-XXXII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929235
Schmithüsen, H. (2021h). *Continuous meteorological surface measurement during POLARSTERN cruise PS111 (ANT-XXXIII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929241
Schmithüsen, H. (2021i). *Continuous meteorological surface measurement during POLARSTERN cruise PS112 (ANT-XXXIII/3) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.929242
Schmithüsen, H. (2021j). *Continuous meteorological surface measurement during POLARSTERN cruise PS117 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935214
Schmithüsen, H. (2021k). *Continuous meteorological surface measurement during POLARSTERN cruise PS118 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935216
Schmithüsen, H. (2021l). *Continuous meteorological surface measurement during POLARSTERN cruise PS123 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935226
Schmithüsen, H. (2021m). *Radiosonde measurements during POLARSTERN cruise PS123 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935189
Schmithüsen, H. (2021n). *Radiosonde measurements during POLARSTERN cruise PS124 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935190
Schmithüsen, H., Jens, H., & Wenzel, J. (2021). *Meteorological observations during POLARSTERN cruise PS123 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935269
Schmithüsen, H., Rohleder, C., Otte, F., & Schröter, S. (2021). *Meteorological observations during POLARSTERN cruise PS124 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.935270
Schröder, M., & Rohardt, G. (2017). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS96 (ANT-XXXI/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.873151
Schröder, M., & Rohardt, G. (2018). *Continuous thermosalinograph oceanography along POLARSTERN cruise track PS111 (ANT-XXXIII/2) [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.895579
Schuddeboom, A., McDonald, A. J., Morgenstern, O., Harvey, M., & Parsons, S. (2018). Regional regime-based evaluation of present-day general circulation model cloud simulations using self-organizing maps. *Journal of Geophysical Research: Atmospheres*, *123*(8), 4259–4272. doi: 10.1002/2017JD028196
Schuddeboom, A., Varma, V., McDonald, A. J., Morgenstern, O., Harvey, M., Parsons, S., … Furtado, K. (2019). Cluster-based evaluation of model compensating errors: A case study of cloud radiative effect in the Southern Ocean. *Geophysical Research Letters*, *46*(6), 3446-3453. doi: 10.1029/2018GL081686

Schuddeboom, A. J., & McDonald, A. J. (2021). The Southern Ocean radiative bias, cloud compensating errors, and equilibrium climate sensitivity in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, *126*(22), e2021JD035310. doi: 10.1029/2021JD035310

Seethala, C., & Horváth, Á. (2010). Global assessment of AMSR-E and MODIS cloud liquid water path retrievals in warm oceanic clouds. *Journal of Geophysical Research: Atmospheres*, *115*(D13). doi: 10.1029/2009JD012662

Segura, H., Pedruzo-Bagazgoitia, X., Weiss, P., Müller, S. K., Rackow, T., Lee, J., ... Stevens, B. (2025). nextGEMS: entering the era of kilometer-scale Earth system modeling. *EGUsphere*, *2025*, 1–39. doi: 10.5194/egusphere-2025-509

Seiki, T., Roh, W., & Satoh, M. (2022). Cloud microphysics in global cloud resolving models. *Atmosphere-Ocean*, *60*(3–4), 477–505. doi: 10.1080/07055900.2022.2075310

SHiELD authors team. (2024). *SHiELD: System for High-resolution prediction on Earth-to-Local Domains.* Retrieved from `https://www.gfdl.noaa.gov/shield/` (Accessed on 18 June 2024)

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., & Ringler, T. D. (2012). A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Monthly Weather Review*, *140*(9), 3090–3105. doi: 10.1175/MWR-D-11-00215.1

Spinhirne, J. D. (1993). Micro pulse lidar. *IEEE Transactions on Geoscience and Remote Sensing*, *31*(1), 48–55.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., ... Roeckner, E. (2013). Atmospheric component of the MPI-M Earth system model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. doi: 10.1002/jame.20015

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... Zhou, L. (2019). DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, *6*(1), 61. doi: 10.1186/s40645-019-0304-z

Takasuka, D., Satoh, M., Miyakawa, T., Kodama, C., Klocke, D., Stevens, B., ... Terai, C. R. (2024). A protocol and analysis of year-long simulations of global storm-resolving models and beyond. *Research Square*. (Preprint) doi: 10.21203/rs.3.rs-4458164/v1

Tange, O. (2011). GNU parallel: The command-line power tool. *The USENIX Magazine*, *36*(1), 42–47.

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, *121*(11), 3040–3061. doi: 10.1175/1520-0493(1993)121⟨3040:ROCILS⟩2.0.CO;2

Trenberth, K. E., & Fasullo, J. T. (2010). Simulation of present-day and twenty-first-century energy budgets of the Southern Oceans. *Journal of Climate*, *23*(2), 440–454. doi: 10.1175/2009JCLI3152.1

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... and, Y. V.-B. (2020, feb). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. doi: 10.1038/s41592-019-0686-2

Voldoire, A., Decharme, B., Pianezze, J., Lebeaupin Brossier, C., Sevault, F., Seyfried, L., ... Riette, S. (2017). SURFEX v8.0 interface with OASIS3-MCT to couple atmosphere with hydrology, ocean, waves and sea-ice models, from coastal to global scales. *Geoscientific Model Development*, *10*(11), 4207–4227. doi: 10.5194/gmd-10-4207-2017

Wall, C. J., Hartmann, D. L., & Ma, P.-L. (2017). Instantaneous linkages between clouds and large-scale meteorology over the Southern Ocean in observations and a climate model. *Journal of Climate*, *30*(23), 9455–9474. doi: 10.1175/JCLI-D-17-0156.1

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. doi: 10.1038/nclimate2118

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., ... Schröter, J. (2014). The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model. *Geoscientific Model Development*, *7*(2), 663–693. doi: 10.5194/gmd-7-663-2014

Weare, B. C. (2004). A comparison of AMIP II model cloud layer properties with ISCCP D2 estimates. *Climate Dynamics*, *22*(2), 281–292. doi: 10.1007/s00382-003-0374-9

Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dynamics*, *17*(12), 905–922. doi: 10.1007/s003820100157

Weisman, M. L., Skamarock, W. C., & Klemp, J. B. (1997). The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, *125*(4), 527–548. doi: 10.1175/1520-0493(1997)125⟨0527:TRDOEM⟩2.0.CO;2

Whitehead, L. E., McDonald, A. J., & Guyot, A. (2023). Supercooled liquid water cloud classification using lidar backscatter peak properties. *EGUsphere*, *2023*, 1–30. doi: 10.5194/egusphere-2023-1085

Whitehead, L. E., McDonald, A. J., & Guyot, A. (2024). Supercooled liquid water cloud classification using lidar backscatter peak properties. *Atmospheric Measurement Techniques*, *17*(19), 5765–5784. doi: 10.5194/amt-17-5765-2024

Wiegner, M., & Gasteiger, J. (2015). Correction of water vapor absorption for aerosol remote sensing with ceilometers. *Atmospheric Measurement Techniques*, *8*(9), 3971–3984. doi: 10.5194/amt-8-3971-2015

Wiegner, M., Mattis, I., Pattantyús-Ábrahám, M., Bravo-Aranda, J. A., Poltera, Y., Haefele, A., ... Pönitz, K. (2019). Aerosol backscatter profiles from ceilometers: validation of water vapor correction in the framework of CeiLinEx2015. *Atmospheric Measurement Techniques*, *12*(1), 471–490. doi: 10.5194/amt-12-471-2019

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System experiment. *Bulletin of the American Meteorological Society*, *77*(5), 853–868. doi: 10.1175/1520-0477(1996)077⟨0853:CATERE⟩2.0.CO;2

Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., ... Williamson, D. L. (2013). The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models. *Journal of Climate*, *26*(10), 3258–3274. doi: 10.1175/JCLI-D-12-00429.1

Williams, R. G., Ceppi, P., Roussenov, V., Katavouta, A., & Meijers, A. J. S. (2023). The role of the Southern Ocean in the global climate response to carbon emissions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *381*(2249). doi: 10.1098/rsta.2022.0062

Wolf-Gladrow, D. A., & Rohardt, G. (2012). *Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/3 [Dataset].* PANGAEA. doi: 10.1594/PANGAEA.780004

Xia, Z., & McFarquhar, G. M. (2024). Dependence of cloud macrophysical properties and phase distributions on environmental conditions over the North Atlantic and Southern Ocean: Results from COMBLE and MARCUS. *Journal of Geophysical Research: Atmospheres*, *129*(12), e2023JD039869. doi: 10.1029/2023JD039869

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., … Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. doi: 10.1029/2019GL085782

Zhang, M. H., Lin, W. Y., Klein, S. A., Bacmeister, J. T., Bony, S., Cederwall, R. T., … Zhang, J. H. (2005). Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *Journal of Geophysical Research: Atmospheres*, *110*(D15), D15S02. doi: 10.1029/2004JD005021

Zhang, Q., Liu, B., Li, S., & Zhou, T. (2023). Understanding models' global sea surface temperature bias in mean state: From CMIP5 to CMIP6. *Geophysical Research Letters*, *50*(4), e2022GL100888. doi: 10.1029/2022GL100888

Zhao, L., Wang, Y., Zhao, C., Dong, X., & Yung, Y. L. (2022). Compensating errors in cloud radiative and physical properties over the Southern Ocean in the CMIP6 climate models. *Advances in Atmospheric Sciences*, *39*(12), 2156–2171. doi: 10.1007/s00376-022-2036-z

Zieger, P., Väisänen, O., Corbin, J. C., Partridge, D. G., Bastelberger, S., Mousavi-Fard, M., … Salter, M. E. (2017). Revising the hygroscopicity of inorganic sea salt particles. *Nature Communications*, *8*(1), 15883. doi: 10.1038/ncomms15883