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

Peter Kuma[1,2,*], Frida A.-M. Bender[1,2], Adrian J. McDonald[3], Simon P. Alexander[4,5], Greg M. McFarquhar[6,7], John J. Cassano[8,9,10], Graeme E. Plank[3], Sean Hartery[11], Simon Parsons[12], Sally Garrett[13], and Alex J. Schuddeboom[3]

[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]Australian Antarctic Division, Kingston, Tasmania, Australia
[5]Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
[6]Cooperative Institute of Severe and High Impact Weather Research and Operations, University of Oklahoma, Norman, OK, USA
[7]School of Meteorology, University of Oklahoma, Norman, OK, USA
[8]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[9]National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA
[10]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA
[11]Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Canada
[12]New South Wales Department of Planning and Environment, Sydney, New South Wales, Australia
[13]New Zealand Defence Force, Wellington, New Zealand
*Corresponding author: Peter Kuma (peter.kuma@misu.su.se)

November 11, 2024

Abstract

Global storm-resolving models (GSRMs) are the next avenue of climate modelling. Among them is the 5-km Icosahedral Nonhydrostatic Weather and Climate Model (ICON). The high resolution allows for parameterizations of convection and clouds to be avoided. Standard-resolution models have substantial cloud biases over the Southern Ocean (SO), affecting radiation and sea surface temperature. We evaluated SO clouds in ICON and the ERA5 and MERRA-2 reanalyses. The SO is dominated by low clouds, which cannot be observed accurately from space due to overlapping clouds, attenuation, and ground clutter. Instead, we analysed about 2400 days of lidar observations from 31 voyages and a station using a ground-based lidar simulator. 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, potentially explained by their lifting condensation levels being too high. The reanalyses strongly underestimate near-surface clouds or fog. MERRA-2 tends to underestimate cloud occurrence at all heights. Less stable conditions are the most problematic for ICON and the reanalyses. In daily cloud cover, ICON and the reanalyses tend to be about 1 and 2 oktas clearer, respectively. Compared to radiosondes, potential temperature is accurate in the reanalyses, but ICON underestimates stability over the low-latitude SO and too humid in the boundary layer. MERRA-2 is too humid at all heights. SO cloud biases remain a substantial issue in the GSRM, but are an improvement over the lower-resolution reanalyses. Explicitly resolved convection and cloud processes were not enough to address the model cloud biases.

1 Introduction

Increasing climate model resolution is one way of improving model accuracy of representation of the climate system (Mauritsen et al., 2022). It has been practiced since the advent of climate modelling 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 (GPUs) is progressing to enable km-scale (also called k-scale) Earth system models (ESMs) and coupled atmosphere–ocean general circulation models (AOGCMs) in research conditions today and operationally in the forthcoming years. Therefore, it represents a natural advance in climate modelling. 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 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).

nextGEMS is a European Union–funded project (nextGEMS authors team, 2024) focused on the research and development of GSRMs at multiple modelling centres and universities in Europe. The project also develops GSRM versions of the Icosahedral Nonhydrostatic Weather and Climate Model (ICON), the Integrated Forecasting System (IFS), and their ocean components at eddy-resolving resolutions: ICON-O coupled with ICON and Finite-Element/volumE Sea ice-Ocean Model (FESOM) and Nucleus for European Modelling of the Ocean (NEMO) coupled with IFS. The project has so far produced ICON and IFS simulations in three cycles (Cycle 1–3) and pre-final simulation, with a final production simulation 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)], 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 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 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. In the 6^{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), which is one of the main reasons for the higher climate sensitivity of CMIP6 models.

The Southern Ocean (SO) is known to be a problematic region for climate model biases due to a lack of surface and in situ observations and being 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, and the SO is an important part of the global ocean conveyor belt. 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. Here, we conventionally 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.

SO radiation biases have been relatively large and systematic compared to the rest of the globe since at least CMIP3 (Trenberth and Fasullo, 2010), and the SO SW cloud radiative effect (CRE) bias is still positive in eight analysed CMIP6 models analysed by 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, characterised by too small cloud fraction and too large cloud albedo (Wall et al., 2017; Kuma et al., 2020), previously described by Webb et al. (2001), Weare (2004), 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. That is, a model maintains a reasonable SW radiation balance by reflecting too much SW radiation from clouds, but has too little cloud area overall. A study by Konsta et al. (2022) showed that this type of bias is still present in six analysed CMIP6 models in tropical marine clouds, using the GCM-Oriented Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (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 analysed 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% [relative to Moderate Resolution Imaging Spectroradiometer (MODIS) data] and overestimates downwelling TOA SW radiation on by approximately 10 $\mathrm{Wm^{-2}}$ [relative to the Clouds and the Earth's Radiant Energy System (CERES) data] in the highest model resolution run (2.5 km). Zhao et al. (2022) reported a similar SW radiation bias in five analysed CMIP6 models over high-latitude SO and the total cloud fraction underestimation on the order of 10% over the entire 40–60°S SO. Recently, Ramadoss et al. (2024) analysed 48 hours of km-scale ICON limited area model NWP simulations over a 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, caused by too much shortwave heating of the surface, too little longwave cooling of the surface, or in the ocean circulation. Interactions of both are also possible, for example SST affecting clouds and clouds affecting the surface radiation. Zhang et al. (2023) has shown that SST biases have improved in CMIP6 compared to CMIP5 [relative to the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5; ERA5], with SST overall increasing in the later CMIP phase. 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.

The main aim of this study is to evaluate the GSRM version of ICON, developed jointly by nextGEMS, Deutscher Wetterdienst, Max-Planck-Institute for Meteorology, Deutsches Klimarechenzentrum (DKRZ), Karlsruhe Institute of Technology, and the Center for Climate Systems Modeling. Previous studies have identified substantial large-scale biases in climate model clouds over the SO, affecting sea surface temperature and the Earth's albedo. Our aim is to quantify how well the GSRM ICON is simulating clouds in this region, 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 and ground clutter, respectively. Specifically, the radar on CloudSat and lidar on CALIPSO (neither of which are now operational) are affected by the abovementioned issues, resulting in a strong underestimation of cloud occurrence below 2 km relative to ground-based lidar observations (McErlich et al., 2021). 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 parametrised 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 of the RV *Polarstern* and other voyages and stations as a reference for the model evaluation.

Altogether, we analysed about 2400 days of data from 31 voyages, and one sub-Antarctic station covering diverse longitudes and latitudes of the SO. To achieve a like-for-like comparison with the model, we used a ground-based lidar simulator called the Automatic Lidar and Ceilometer Framework [ALCF; Kuma et al. (2021)]. We contrasted the results with ERA5 (ECMWF, 2019) and the Modern-Era Retrospective analysis for Research and Applications, Version 2 [MERRA-2; Gelaro et al. (2017)].

2 Methods

2.1 Voyage and station data

Together, we analysed data from 31 voyages of RV *Polarstern*, the resupply vessel (RSV) *Aurora Australis*, RV *Tangaroa*, RV *Nathaniel B. Palmer*, Her 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 voyages, Table 1 list the voyages, campaigns, and stations, and Table 2 lists references where available. Altogether, the voyages and station dataset comprised 2421 days of data south of 40°S, but the availability of ceilometer data was slightly smaller due to gaps in measurements.

Missing days in the ceilometer data were HMNZSW16 (7 days): 24–27 November, 10 December, 16–17 December 2016; Measurements of Aerosols, Radiation, and CloUds over the Southern Ocean (MARCUS; 3 days): 8, 10 November, 10 December 2017; Macquarie Island Cloud Radiation Experiment (MICRE; 9 days): 7–8, 29 June, 5, 16 July, 15 August, 17 October 2016, 11 February, 21 March 2017; TAN1502 (1 day): 24 January.

The data sources contained ceilometer observations captured by the Vaisala CL51 operating at a wavelength of 910 nm, the Vaisala CT25K operating at 905 nm, and the Lufft CHM 15k operating at 1064 nm, 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.

Apart from lidar observations, radiosondes were launched on weather balloons at regular synoptic times on the RV *Polarstern*, MARCUS, NBP17024, TAN1702, and TAN1802 voyages and campaigns, measuring pressure, temperature, relative humidity, and the global navigation satellite sys-

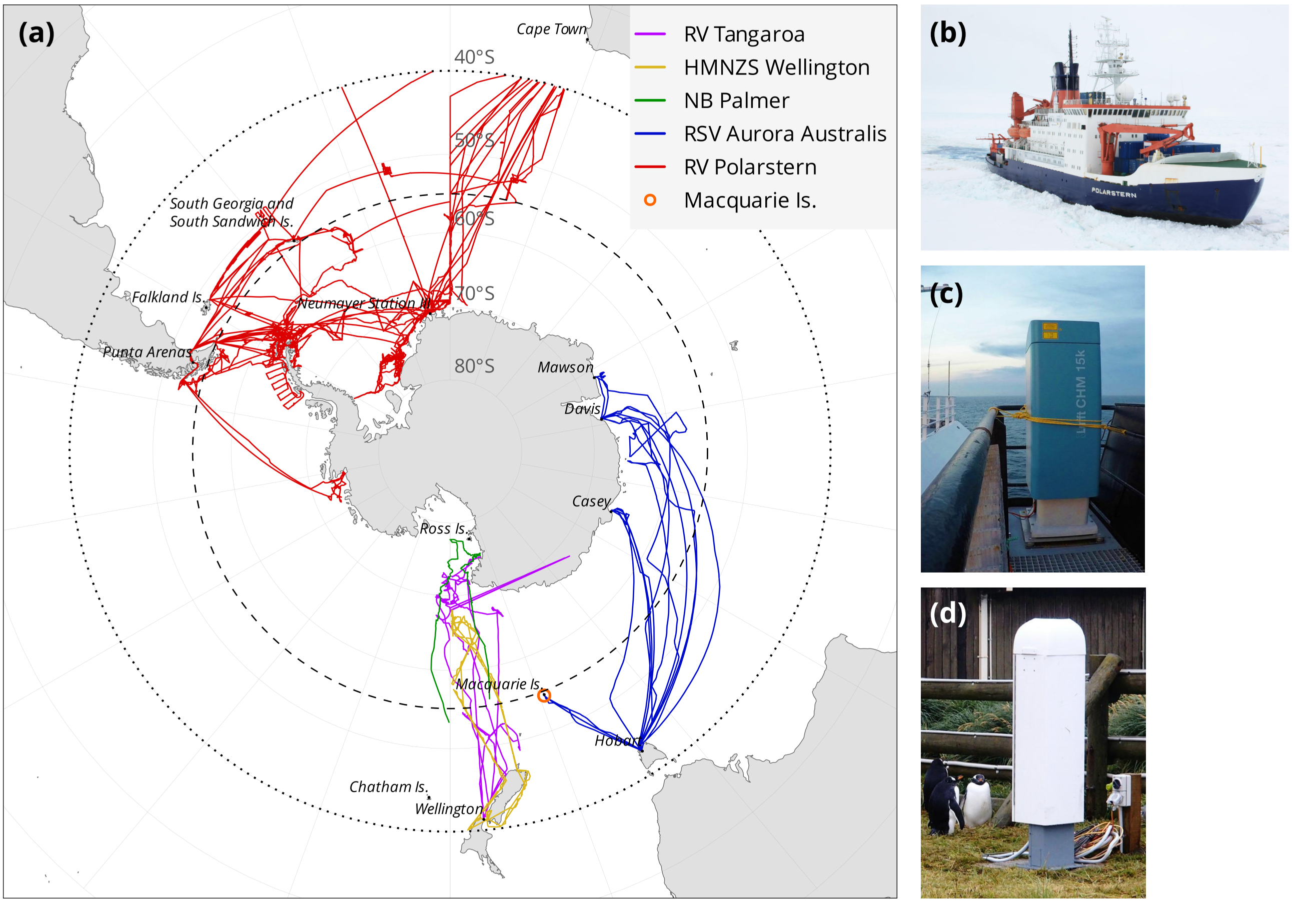

Figure 1 | (a) A map showing the tracks of 31 voyages of RV *Polarstern*, RSV *Aurora Australis*, RV *Tangaroa*, RV *Nathaniel B. Palmer*, and HMNZS *Wellington* and one sub-Antarctic station (Macquarie Island) analysed here. The tracks cover Antarctic sectors south of South America, the Atlantic Ocean, 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).

tem coordinates. Derived thermodynamic (virtual potential temperature, lifting condensation level, etc.) and dynamic physical quantities (wind speed and direction) for the measured vertical profiles were calculated with rstool (Kuma, 2024). Surface meteorological quantities were measured continuously by an onboard automatic weather station or individual instruments.

2.2 Vaisala CL51 and CT25K

The Vaisala CL51 (photo in Fig. 1d) and CT25K are ceilometers operating at a near-infrared wavelength 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 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 wavelength of 910 nm is affected by water vapour absorption of about 20% in the mid-latitudes (Wiegner and Gasteiger, 2015; Wiegner et al., 2019), 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 Network Common Data Form (NetCDF) with cl2nc (`https://github.com/peterkuma/cl2nc`) 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 CL51, equivalent calculations as for CL51 were done assuming a wavelength of 910 nm. The Vaisala CL51 and CT25K instruments were used on most of the voyages and stations analysed 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) is a ceilometer operating 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 again 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).

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 in 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 ALCF has 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 et al., 2024).

Table 1 | An overview of the analysed voyages, campaigns, 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.

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 | Voyage, campaign and station 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 (2011a,b,c, 2014a); Fahrbach and Rohardt (2011)
PS77/3	König-Langlo (2011d,e, 2012a, 2014b); Knust and Rohardt (2011)
PS79/2	König-Langlo (2012b,c,d, 2014c); Kattner and Rohardt (2012)
PS79/3	König-Langlo (2012e,f,g, 2014d); Wolf-Gladrow and Rohardt (2012)
PS79/4	König-Langlo (2012h,i,j, 2014e); Lucassen and Rohardt (2012)
PS81/2	König-Langlo (2013a,b,c, 2014f); Boebel and Rohardt (2013)
PS81/3	König-Langlo (2013d,e,f, 2014g); Gutt and Rohardt (2013)
PS81/4	König-Langlo (2013g,h,i, 2014f); Bohrmann and Rohardt (2013)
PS81/5	König-Langlo (2013j,k,l, 2014g); Jokat and Rohardt (2013)
PS81/6	König-Langlo (2013m,n,o, 2014h); Lemke and Rohardt (2013)
PS81/7	König-Langlo (2013p,q, 2014i, 2016a); Meyer and Rohardt (2013)
PS81/8	König-Langlo (2013r, 2014j,k,l); Schlindwein and Rohardt (2014)
PS81/9	König-Langlo (2014m,n,o,p); Knust and Rohardt (2014)
PS89	König-Langlo (2015a,b,c,d); Boebel and Rohardt (2016)
PS96	König-Langlo (2016b,c,d,e); Schröder and Rohardt (2017)
PS97	König-Langlo (2016f,g,h,i); Lamy and Rohardt (2017)
PS103	König-Langlo (2017a,b,c,d); Boebel and Rohardt (2018)
PS104	König-Langlo (2017e,f,g); Gohl and Rohardt (2018); Schmithüsen (2021a)
PS111	Schmithüsen (2019a, 2020a, 2021b,c); Schröder and Rohardt (2018)
PS112	Schmithüsen (2019b, 2020b, 2021d,e); Meyer and Rohardt (2018)
PS117	Schmithüsen (2019c, 2020c, 2021f,g); Boebel and Rohardt (2019)
PS118	Schmithüsen (2019d, 2020d, 2021h,i); Dorschel and Rohardt (2019)
PS123	Schmithüsen (2021j,k,l); Schmithüsen et al. (2021a); Hoppmann et al. (2023)
PS124	Schmithüsen (2021m,n); Schmithüsen et al. (2021b); Hoppmann et al. (2023)
TAN1802	Kremser et al. (2020, 2021)

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 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-scale 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.

2.5 ICON

A coupled (atmosphere–ocean) GSRM version of the ICON model is in development at the nextGEMS project (Hohenegger et al., 2023). ICON is an exceptionally versatile model, allowing for simulations ranging from coarse-resolution ESM simulations, GSRM simulations, limited area model simulations, to 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., *not* nudged or using prescribed SST) 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 analysed the full years 2021–2024 (inclusive). While a Cycle 4 run was available, we could not use it due to the unavailability of the necessary variables. 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 (GCMs), 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. 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 are still parameterized in this model.

Because the model is free-running, weather and climate oscillations (such as the El Niño–Southern Oscillation) are not expected to be equivalent to reality at the same time and place. To compare with the observations collected in different 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.

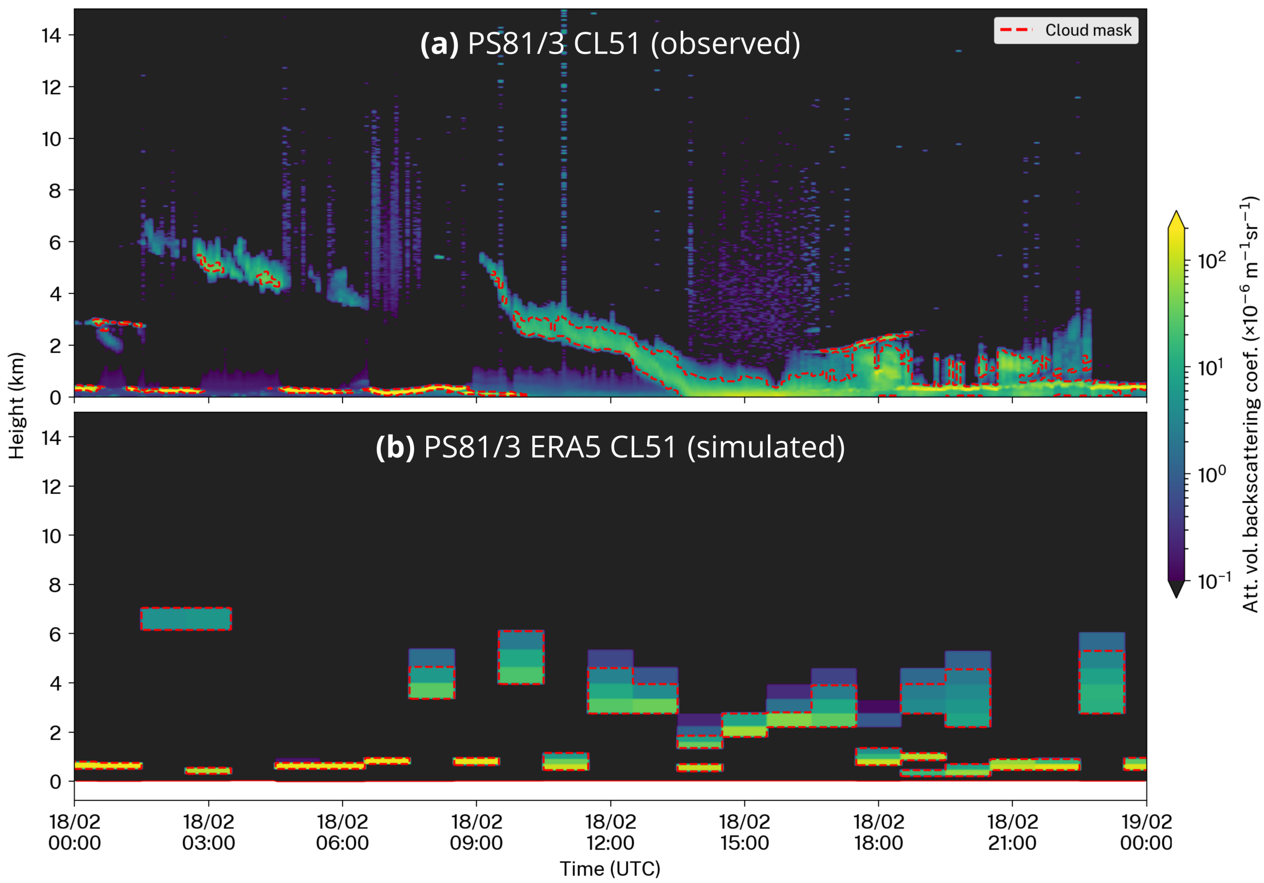

Figure 2 | An example of attenuated volume backscattering coefficient (AVBC) **(a)** measured by CL51 during 24 hours on the PS81/3 voyage and **(b)** an equivalent AVBC simulated with the ALCF from ERA5 data during the same time period. The red line identifies the cloud mask determined by the ALCF.

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). The reanalysis output analysed 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).

2.7 ERA5

ERA5 (ECMWF, 2019) is a reanalysis produced by the ECMWF. It is based on a numerical weather prediction model IFS version CY41R2. 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 made available on 37 pressure levels.

2.8 CERES

TOA radiation quantities are taken from the CERES instruments on board 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 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).

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

2.9 Precipitation identification using machine learning

Precipitation can cause strong enough lidar backscattering to be recognised 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 in the ICON model output and the reanalysis data (the liquid and ice precipitation mass mixing ratios). The required radiation calculations for precipitation are also currently not implemented in the ALCF, even though this is a planned feature. In order to achieve a fair comparison of observations with models, one should 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 polarised channel are available. In models, the same can be accomplished relatively easily by excluding profiles exceeding a certain amount of 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 time instants. Therefore, it was desirable to implement a method of detecting precipitation from observed backscatter profiles alone.

On the RV *Polarstern* voyages, regular human-performed synoptic observations were available and

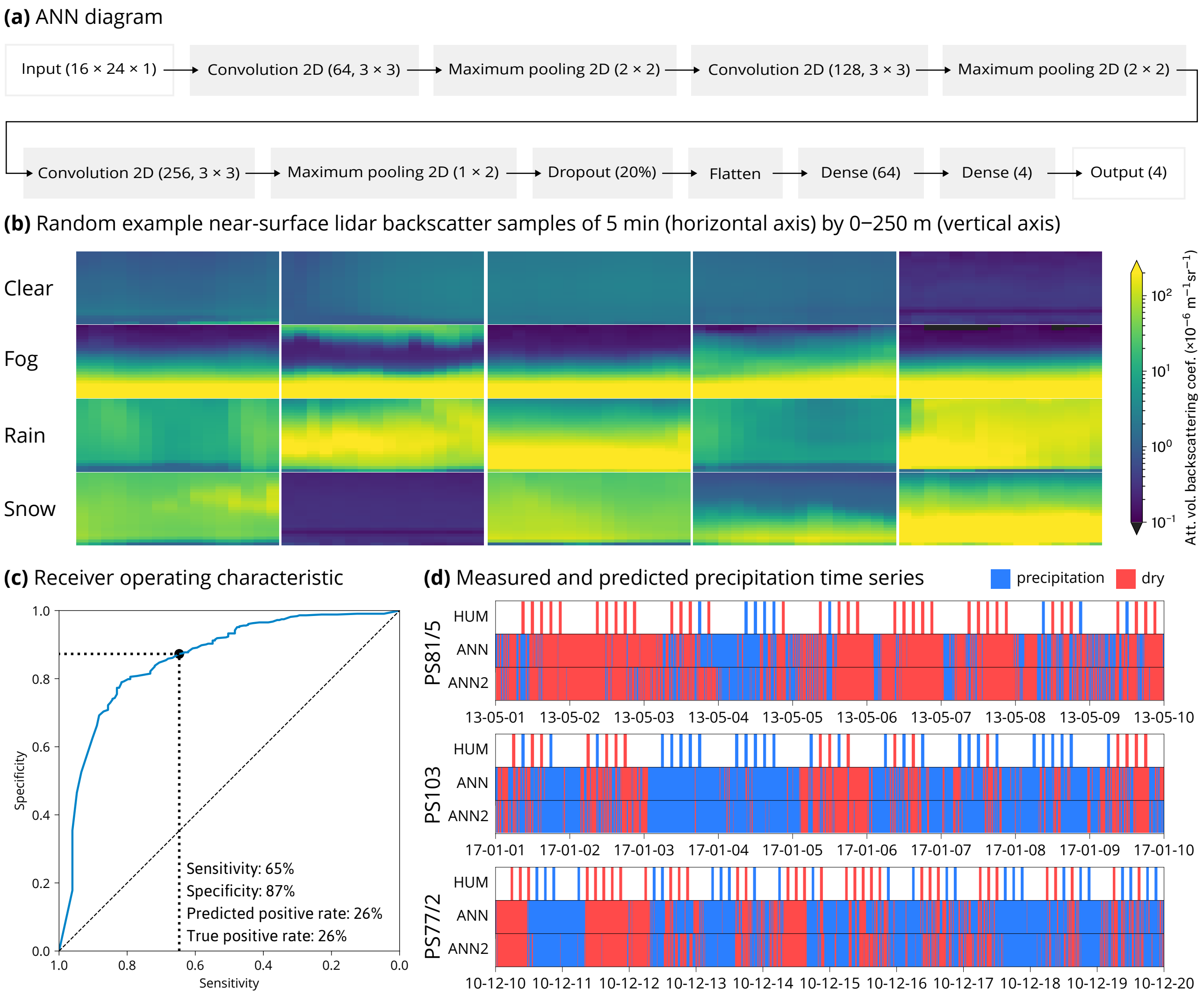

Figure 3 | Artificial neural network (ANN) for prediction of precipitation in lidar backscatter. **(a)** Diagram showing the TensorFlow structure of the ANN, **(b)** randomly selected example samples of near-surface backscatter in four categories (clear, fog, rain, and snow), as determined by coincident human-performed 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.

included precipitation presence or absence and type. We used this dataset to train a convolutional artificial neural network (ANN) of the U-Net type (Ronneberger et al., 2015) to recognise profiles with precipitation from lidar backscatter (Fig. 3a), implemented in the TensorFlow ANN framework (Abadi et al., 2015). Samples of short time intervals (10 min) of near-surface 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.

2.10 Partitioning by cyclonic activity and stability

We partitioned our data into two mutually exclusive subsets by cyclonic activity. For this purpose, we used a cyclone tracking algorithm to identify extratropical and polar cyclones (ECs and PCs) 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 EC is considered relatively arbitrary due to the very large variability of ECs (Neu et al., 2013). We used the mean sea level pressure field and horizontal wind speed fields as input to the CyTRACK algorithm. The algorithm uses pressure and wind speed thresholds as well as tracking across time steps to identify cyclone centres and radii. With this information, we could classify geographical areas as either cyclonic or non-cyclonic. Due to a relatively small total area covered, we chose a circle of a double radius (relative to one identified by CyTRACK) centred at the cyclone centre as a cyclonic area for every time step and cyclone. All other areas were identified 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. For identifying cyclones in ICON, its own pressure and wind fields were used as the input to CyTRACK, because the model is free-running, and thus the pressure and wind fields are different from reality.

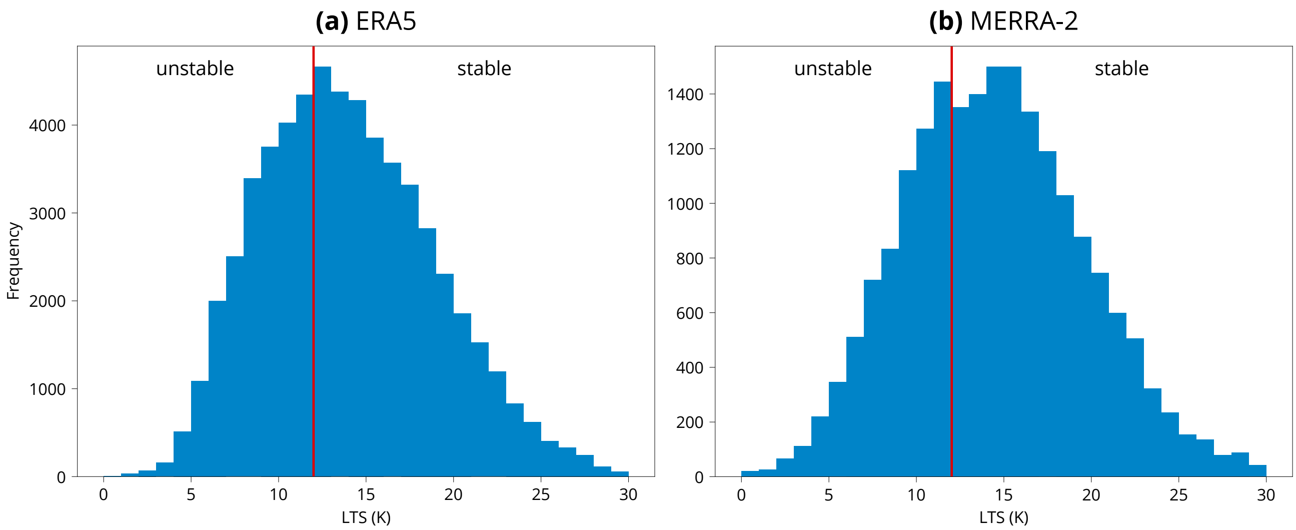

Figure 4 | Lower tropospheric stability (LTS) distribution in **(a)** ERA5 and **(b)** MERRA-2 calculated for the 31 voyage tracks and one station from the highest instantaneous temporal resolution data available. Shown is also the chosen dividing threshold of 12 K for relatively stable and unstable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

In addition to the above, we partitioned our data into two mutually exclusive subsets by stability. We determined this by calculating lower tropospheric stability (LTS) 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 dividing threshold of 12 K for relatively unstable (< 12 K) and relatively stable (>= 12 K) conditions.

3 Results

3.1 Cyclonic activity and stability

Here, we briefly describe the results of the cyclonic activity and stability distribution, which is relevant for the subsequent analysis, because these conditions are used for subsetting our dataset. Fig. 5a, b show a geographical distribution of the fraction of cyclonic days as determined by the cyclone tracking algorithm applied on the ERA5 reanalysis and ICON data (Section 2.10). As expected, the strongest cyclonic activity is in the high-latitude SO zone, and it is relatively zonally symmetric at all latitudes. While both reanalysis and the model agree relatively well, differences in the strength of the local extremes of occurrence are notable, especially over the Amundsen Sea, which is more cyclonic in the reanalysis, and around Cape Adare, which is more cyclonic in ICON. These differences might, however, stem from the relatively short time periods of comparison (4 years) and the fact that the model is free-running.

Fig. 5c, d show a geographical distribution of the relatively stable and unstable conditions as determined by the LTS (Section 2.10). Relatively unstable conditions are prevalent in the middle SO (50–65°S), which might be explained by the relatively cold near-surface air overlying the relatively warm sea surface. Relatively stable conditions are prevalent 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 substantial stabilising effect (Knight et al., 2024). The ERA5 reanalysis is also substantially more stable than ICON across the whole region.

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 (Fig. 6). In addition, we aggregated the data sources (voyages and stations) by calculating the averages and percentiles of all individual profiles, presented in Fig. 7. The analysis shows that the total cloud fraction (determined as the fraction of profiles with clouds at any height in the lidar cloud mask) is underestimated in ICON and the reanalyses by about 10% and 20%, respectively. When analysed by height, ICON overestimates cloud occurrence below 1 km and underestimates it above, MERRA-2 underestimates cloud occurrence at all heights, 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 near-surface clouds are strongly lacking in the reanalyses (fog and clouds are both included in the cloud occurrence). As shown in Fig. 6, the biases are relatively consistent across the data sources and longitudes. We conclude that the ICON results are overall better matching the observations than the reanalyses in this metric.

In the general case (Fig. 7a), the observations show cloud occurrence peaking at the surface, whereas models show a higher peak (at about 500 m). The models underestimate the total cloud fraction by 10–20% and show a strong drop in cloud occurrence near the surface, but this is not supported by the observations. ICON and ERA5 overestimate cloud occurrence at the peak (between 0–1 km). Above 1 km, ICON and MERRA-2 underestimate cloud occurrence, but ERA5 is very accurate. The exaggerated peak in models is partly supported by the lifting condensation level (LCL) distri-

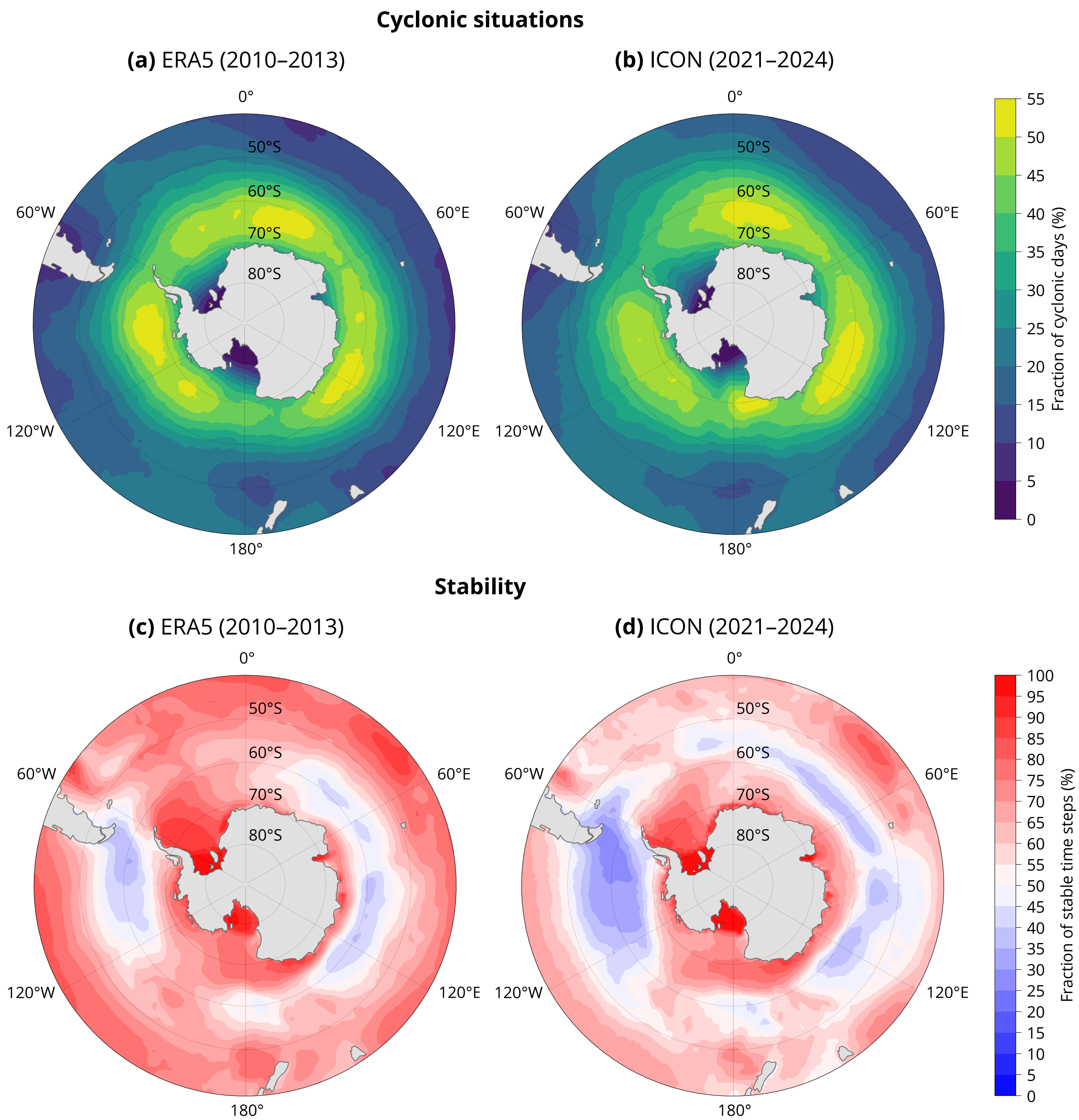

Figure 5 | Geographical distribution of **(a, b)** cyclonic days and **(b, d)** relatively stable (LTS > 12 K) time steps in **(a, c)** ERA5 in years 2010–2013 (inclusive) and **(b, d)** ICON in model years 2021–2023 (free running). Cyclonic days are expressed as a fraction of the number of days with cyclonic activity, defined as grid points located within a double radius of any cyclone on a given day (UTC), as identified by CyTRACK.

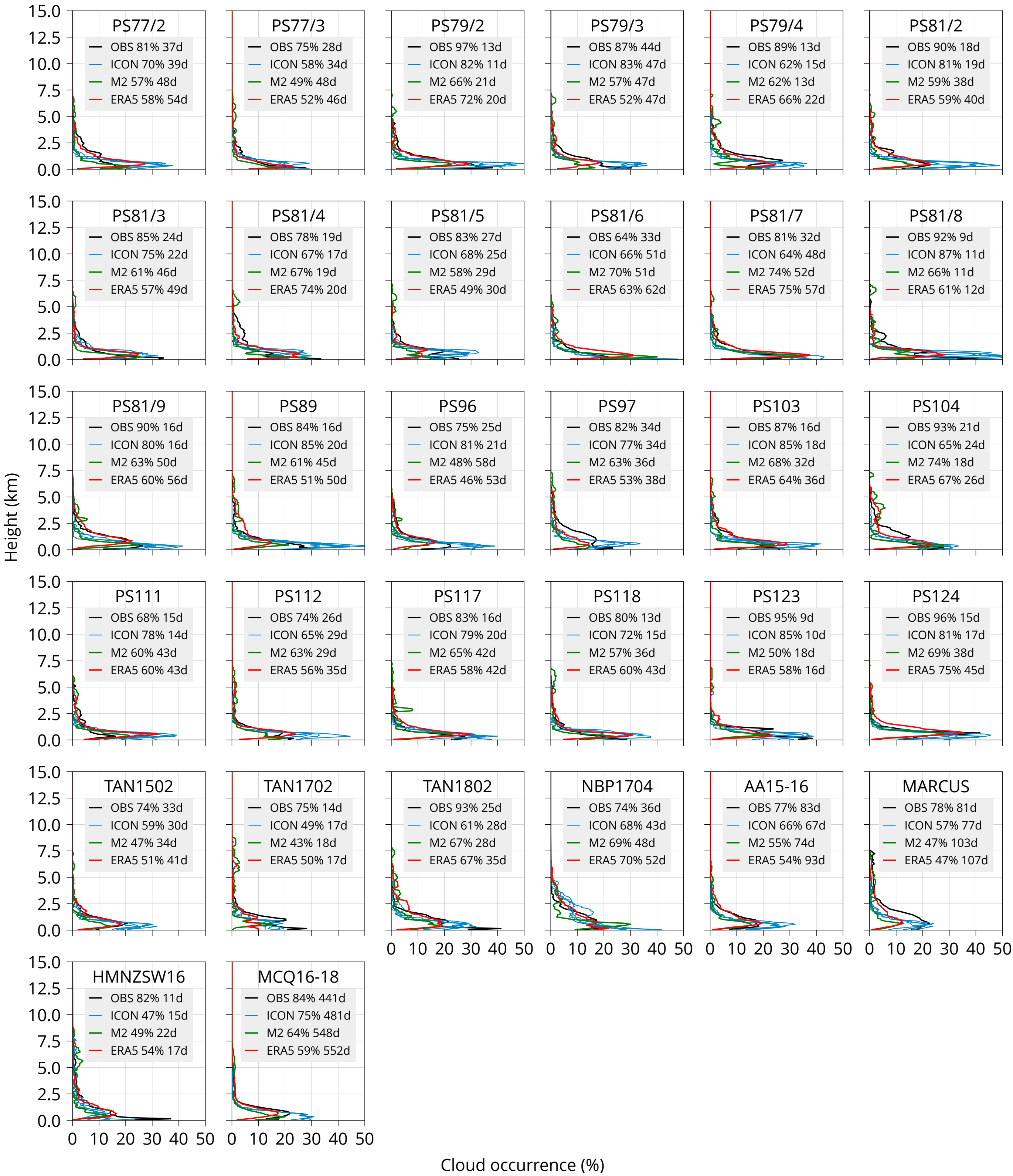

Figure 6 | Cloud occurrence by height for the 31 voyages and one sub-Antarctic station (MICRE) in observations (OBS) and simulated by the ALCF from the ICON model, MERRA-2 (M2), and ERA5 reanalysis data. The numbers in the legend indicate the total cloud fraction and the number of days of data.

bution, which peaks higher in the models than in the observations (at the surface), although this is not very pronounced.

When subsetted by low- and high-latitude zones (Fig. 7b, c), we see that the low-latitude SO zone shows a stronger peak of cloud occurrence near the surface than the high-latitude SO zone, and this could be because higher latitudes have more unstable atmospheric profiles. 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 (near-surface cloud occurrence is still strongly underestimated).

When subsetted by cyclonic and non-cyclonic situations (Fig. 7d, e), we see that the cyclonic situations have a larger amount of observed cloudiness, including the peak and total cloud fraction. In these situations, the models are doing a relatively good job of getting the vertical profile of cloud occurrence right, but still tend to underestimate cloud occurrence above 1 km and near the surface. Non-cyclonic situations are similar to the general case.

When subsetted by relatively stable and unstable conditions (Fig. 7f, g), as defined in Section 2.10, we see that in relatively stable situations cloud occurrence peaks strongly at the surface in observations, compared to relatively unstable situations, where the peak is more obtuse and spread across the altitudes of 0–1 km. In relatively stable situations, the models are doing a fairly good job, but overestimate cloud occurrence at the peak below 1 km; above 1 km, they show similar biases as in the general case. In relatively unstable situations, the bias in ICON is very pronounced, with a much stronger peak at about 500 m, ERA5 is underestimating cloud occurrence below 1 km (especially near the surface), and MERRA-2 is underestimating it even more strongly.

In all situations, even when the models overestimate cloud occurrence at some altitudes, they always substantially underestimate the total cloud fraction. ICON can be generally characterised as substantially overestimating cloud occurrence below 1 km and underestimating above, underestimating the total cloud fraction, and showing greatest biases in relatively unstable and non-cyclonic conditions. It also shows a peak of cloud occurrence at higher altitude than observations (500 m vs. near the surface), and correspondingly, its LCL tends to be also higher. MERRA-2 can be generally characterised 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). It struggles the most in the low-latitude SO zone and in the relatively unstable situations. ERA5 can be generally characterised as representing cloud occurrence correctly above about 1–1.5 km, overestimating below, but underestimating near-surface cloud occurrence (0–500 m). The total cloud fraction is strongly underestimated in all situations. It has a tendency towards underestimation in the low-latitude SO zone and relatively unstable situations; conversely, overestimating in the high-latitude SO zone and the relatively stable conditions.

3.3 Cloud cover

We analysed 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, expressed in oktas (multiples of 1/8).

In Fig. 8 we show the results for the same subsets of data as in the previous section. Observations have the greatest representation of high cloud cover (5–8 oktas), peaking at 7 oktas. This is not represented by ICON or the reanalyses. While ICON is the closest, it tends to be 1 okta clearer than the observations, peaking at 6 oktas, and highly underestimating days with 8 oktas. Overall,

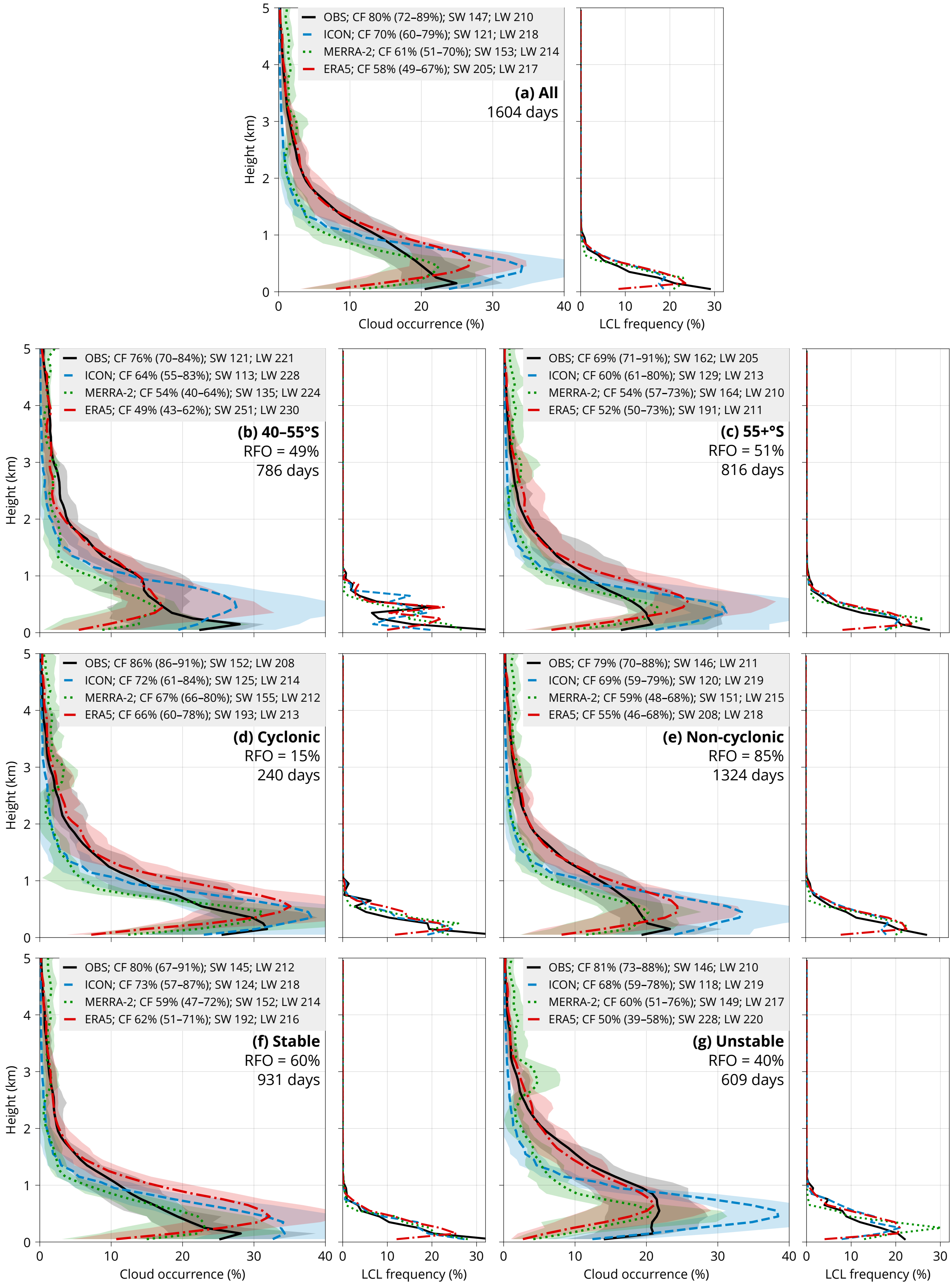

Figure 7 | Cloud occurrence by height calculated as the average of all voyages and stations for the observed (OBS) and simulated lidar cloud mask, and lifting condensation level (LCL) distribution. The total cloud fraction (CF), average shortwave (SW), and longwave (LW) are shown in the legend, and the relative frequency of occurrence (RFO) of the subset is shown below. The bands span the 16th–84th percentile calculated from the set of all voyages and stations. The subsets **(d–g)** are defined in Section 2.10.

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 shows slightly higher cloud cover and thus is more consistent with observations.

When analysed by subsets, observations in the cyclonic subset show the highest cloud cover, with 8 oktas occurring on one half of such days (Fig. 8d). This is not represented by ICON or the reanalyses at all. 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. 8c). The high-latitude zone also has almost no clear sky or small cloud cover cases (0–4 oktas). ICON and the reanalyses represent at least 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 relatively unstable subset, in which ERA5 peaks at 3 oktas, whereas the observations peak at 7 oktas and show negligible cloud cover below 5 oktas.

3.4 Thermodynamic profiles

We analysed about 2300 radiosonde profiles south of 40°S from the 24 RV *Polarstern* voyages, MARCUS, NBP1704, TAN1702, and TAN1802. Spatially and temporally colocated profiles were taken from ICON and the reanalyses. Because the time period of the ICON model output was different from the observations, model time was chosen to be the same as the radiosonde launch time relative to the start of the year. The profiles were partitioned into the same subsets as above (Sections 3.2 and 3.3). Apart from relative humidity, we focus on comparing virtual potential temperature (θ_v) due to its role in low-level tropospheric stability, being one of the primary factors affecting shallow convection and the associated low-level cloud formation and dissipation. The observed and model profiles of virtual potential temperature are shown in Fig. 9.

Overall, the mean θ_v is relatively well represented in ICON and MERRA-2, being only slightly colder in the mid-to-high troposphere (less stable) in ICON than in observations (Fig. 9a). Large differences exist, however, in the 40–55°S zone, where ICON is colder in θ_v by up to about 5 K and more so at higher altitudes (Fig. 9b). In other subsets, the bias is relatively small. MERRA-2 is very close to the observations, possibly due to a high accuracy of assimilation of this quantity. Notably, the variability of virtual potential temperature (as represented by the percentiles) is much smaller in ICON than in the observations. This indicates that the model's internal variability in the lower-tropospheric thermodynamic conditions in the SO is smaller than in reality.

Relative humidity displays much larger biases. In all subsets, ICON is too humid in the first 1 km but very accurate above, except for the 40–55°S zone and unstable conditions (Fig. 9b, g), where it is too dry between about 1 and 3 km. MERRA-2, on the other hand, is much more humid than observations at all altitudes and in all subsets, by up to about 20% at 5 km.

4 Limitations

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 mostly representative of summer and to a lesser extent, spring and autumn conditions.

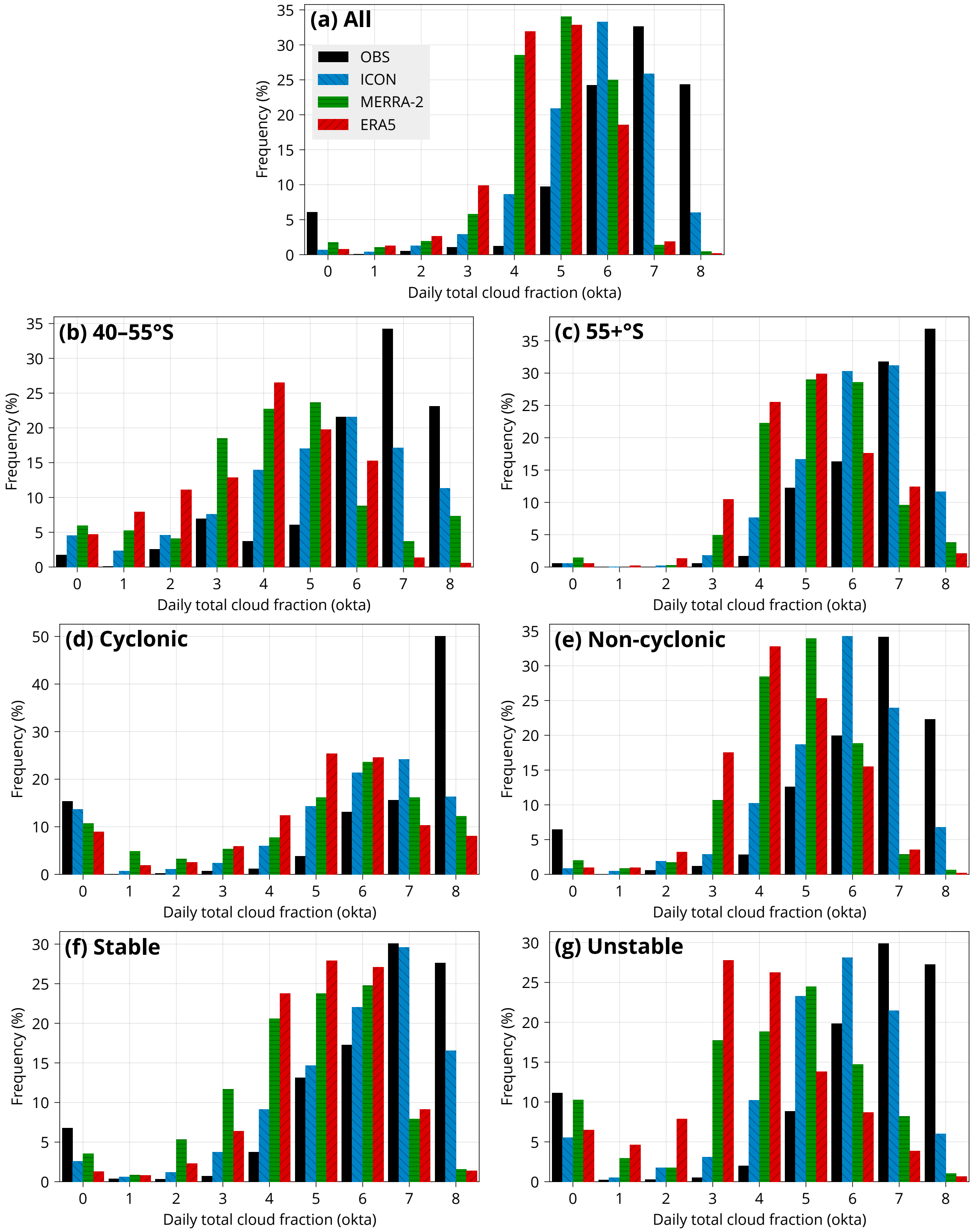

Figure 8 | Daily total cloud fraction histograms calculated as the average of all voyage and station histograms. The total cloud fraction of a day (UTC) is calculated as a fraction of cloudy (based on the cloud mask) observed (OBS) or simulated lidar profiles. The models and subsets are as in Fig. 7.

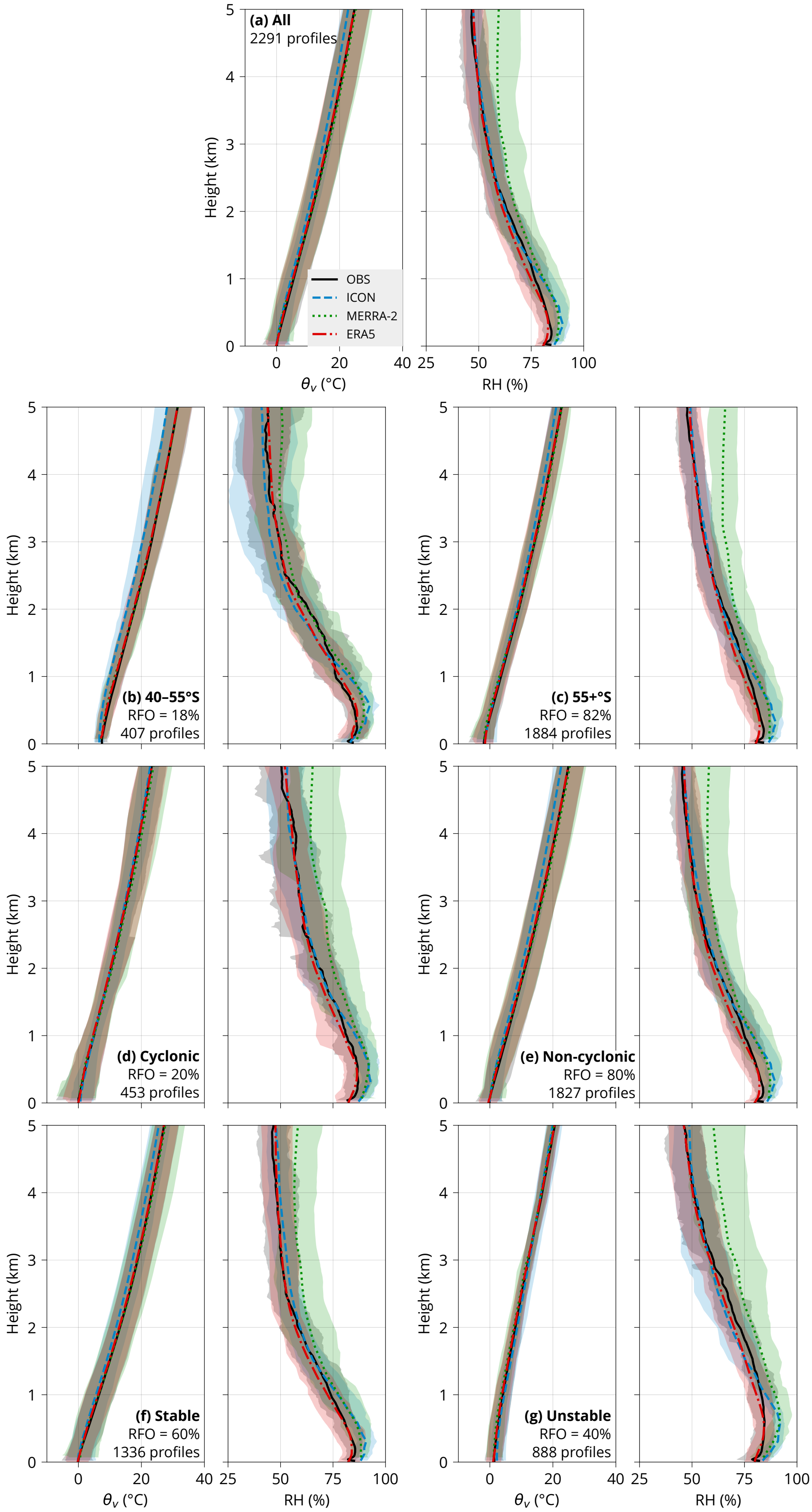

Figure 9 | Virtual potential temperature (θ_v) and relative humidity (RH) determined from radiosonde launches and co-located profiles in ICON, ERA5, and MERRA-2 in subsets as in Fig. 7. The solid lines are the average calculated from the averages of every individual voyage and station. The bands span 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.

The time period of ICON is relatively short, with only four full years of simulation available. Moreover, the simulation is free-running, 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 and the phase of climate oscillations such as the ENSO in the observations and the model.

Ground-based lidar observations are affected by attenuation by thick cloud layers, and for this reason the results are mostly 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 superior to satellite lidar observations for low-level clouds, which are predominant in this region, while mid- and high-level clouds are better represented in satellite observations (McErlich et al., 2021).

We attempted to remove lidar profiles with precipitation, 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 model and reanalyses, which might not exactly correspond to detection by the ANN applied to observations. Also, we did not make an attempt to remove profiles with precipitation not reaching the surface. The above reasons can result in an artificial bias in the comparison, even though we expect this to be much smaller than the identified model biases.

5 Discussion and conclusions

We analysed a total of about 2400 days of lidar and 2300 radiosonde observations from 31 voyages/campaigns and a subantarctic station, covering the Atlantic, Australian, and New Zealand sectors of the SO over the span of 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 different from satellite observations – one that is more suitable for evaluating boundary layer clouds, which are predominant in this region. Furthermore, we subsetted our dataset by low and high latitude bands, cyclonic activity, and stability in order to identify how these conditions relate to the biases.

Our main finding corroborates previous findings of large boundary layer cloud biases in models and their subsequent effect on the radiative transfer. This also applies to the new GSRM ICON, but the biases are lower than in the reanalyses, despite the reanalyses having the advantage of assimilation of the observed meteorological conditions. The GSRM has, on the other hand, the advantage of a much higher spatial resolution and, to a limited extent, explicit calculation of traditionally subgrid-scale processes such as convection.

We show that relative to ERA5, the distribution and strength of cyclonic activity over the SO is well represented in ICON, but it is substantially less stable in terms of LTS. The latter is also manifested in the radiosonde profile comparison, showing that the virtual potential temperature profiles in ICON are less stable than in the observations over low-latitude SO.

The 31 voyages and a station show fairly similar biases in cloud occurrence by height in the lidar comparison, 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 2 km and an underestimation of clouds above 2 km. The reanalyses also underestimate the total cloud fraction by about 20%. ERA5 overestimates cloud below 1 km but underestimates near-surface cloud or fog. ICON strongly overestimates the peak of cloud occurrence at about 500 m, which might be explained by the radiosonde comparison, showing that it is too moist at around this height. Similar to our results, 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.

Compared to lidar observations, the daily cloud cover tends to be about 1 okta lower in ICON and 2 oktas lower in the reanalyses. Unstable conditions 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.

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 near-surface clouds or fog. The radiosonde comparison, however, does not seem to explain cloud biases at higher altitudes. MERRA-2 is too moist at all heights. ICON also exhibits smaller internal variability than the radiosonde observations. Overall, the radiosonde comparison is only partially explaining the identified cloud biases, and other physical causes are likely contributing. This warrants further investigation, especially of ocean–atmosphere fluxes, shallow convection, and boundary layer turbulence. The lack of parametrised subgrid-scale convection in ICON might be a substantial issue even at the 5-km resolution.

The relationship between cloud biases and radiation has a number of notable features. Perhaps unsurprisingly, the reanalyses exhibit the too few, too bright bias previously identified in models. In our results, this is characterised by outgoing TOA SW radiation similar to or higher than in the satellite observations, while at the same time total cloud fraction is substantially underestimated relative to the ground-based lidar observations. This feature seems to be much more pronounced in ERA5 than in MERRA-2. On the other hand, this type of relationship is not present in ICON. This model mostly 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 (Zhang et al., 2023; Luo et al., 2023).

The results imply that SO cloud biases are still a substantial issue in the km-scale resolution ICON model, even though an improvement over the lower-resolution reanalyses is notable. More effort is needed to improve the model cloud simulations in this fast-changing and understudied region. The transition from models with parametrised convection and clouds to storm-resolving models might not solve these biases without additional effort. Evaluation of ocean–atmosphere heat, moisture, and momentum fluxes against in situ observations over the SO and comparison of GSRM simulations against large-eddy simulations are two potential avenues for future research that could elucidate the physical mechanisms behind the biases, in addition to the more common efforts in SO cloud microphysics and precipitation evaluation.

Acknowledgements

PK and FB, and the nextGEMS project received funding from the European Union's Horizon 2020 research and innovation program under a grant agreement no. 101003470. FB received funding from the Wenner-Gren foundation and the Swedish e-Science Research Centre. 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), 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 the Atmospheric Radiation Measurement (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, 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; 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 and 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.

Code availability

The ALCF, cl2nc, rstool, lidar precipitation detection code, and our data processing and plotting code are open source and available at https://alcf.peterkuma.net, https://github.com/peterkuma/cl2nc, https://github.com/peterkuma/rstool, and https://github.com/peterkuma/alcf-precip, https://github.com/peterkuma/icon-so-2024, respectively. CyTRACK is available at https://github.com/apalarcon/CyTRACK. The ICON model is available at https://gitlab.dkrz.de/icon/icon-mpim.

Data availability

The RV *Polarstern* datasets are openly available on Pangaea (https://pangaea.de). 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 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 available on Zenodo (https://doi.org/10.5281/zenodo.4060237).

Author contributions

The authors have made the following contributions based on the CRediT taxonomy (https://credit.niso.org). PK: conceptualization, data curation, formal analysis, investigation, methodology, software, writing – original draft; FB: conceptualization, funding acquisition, project administration, supervision, and writing – review & editing; AM, SA, GM, JC, GP, SH, SP, SG, and AS: data curation, investigation, resources, writing – review & editing.

References

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org.

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, pages 265–283, USA, 2016. USENIX Association. ISBN 9781931971331.

S. F. Ackley, S. Stammerjohn, T. Maksym, M. Smith, J. Cassano, P. Guest, J.-L. Tison, B. Delille, B. Loose, P. Sedwick, L. DePace, L. Roach, and J. Parno. 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, 2020. doi:10.1017/aog.2020.31.

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

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

A. Bodas-Salcedo, M. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. Klein, Y. Zhang, R. Marchand, J. Haynes, R. Pincus, et al. COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, 92(8):1023–1043, 2011. doi:10.1175/2011BAMS2856.1.

O. Boebel and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/2, 2013. URL https://doi.org/10.1594/PANGAEA.808838.

O. Boebel and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS89 (ANT-XXX/2), 2016. URL https://doi.org/10.1594/PANGAEA.858884.

O. Boebel and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS103 (ANT-XXXII/2), 2018. URL https://doi.org/10.1594/PANGAEA.889513.

O. Boebel and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS117, 2019. URL https://doi.org/10.1594/PANGAEA.905562.

G. Bohrmann and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/4, 2013. URL https://doi.org/10.1594/PANGAEA.810678.

M. G. Bosilovich, R. Lucchesi, and M. Suarez. *MERRA-2: File Specification*. Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA, 2016. URL http://gmao.gsfc.nasa.gov/pubs/office_notes. GMAO Office Note No. 9 (Version 1.1).

R. Bubnová, G. Hello, P. Bénard, and J.-F. Geleyn. 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, 1995. doi:10.1175/1520-0493(1995)123<0515:IOTFEE>2.0.CO;2.

P. M. Caldwell, C. R. Terai, B. Hillman, N. D. Keen, P. Bogenschutz, W. Lin, H. Beydoun, M. Taylor, L. Bertagna, A. M. Bradley, T. C. Clevenger, A. S. Donahue, C. Eldred, J. Foucar, J.-C. Golaz, O. Guba, R. Jacob, J. Johnson, J. Krishna, W. Liu, K. Pressel, A. G. Salinger, B. Singh, A. Steyer, P. Ullrich, D. Wu, X. Yuan, J. Shpund, H.-Y. Ma, and C. S. Zender. Convection-permitting simulations with the E3SM global atmosphere model. *Journal of Advances in Modeling Earth Systems*, 13(11):e2021MS002544, 2021. doi:10.1029/2021MS002544.

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

2022. doi:10.1029/2022GL099777.

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

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

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

B. Dorschel and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS118, 2019. URL `https://doi.org/10.1594/PANGAEA.905608`.

DYAMOND author team. DYAMOND Initiative, 2024. URL `https://www.esiwace.eu/the-project/past-phases/dyamond-initiative`. Accessed on 19 June 2024.

ECMWF. 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), 11 2019. URL `https://cds.climate.copernicus.eu/cdsapp#!/home`.

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

E. Fahrbach and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVII/2, 2011. URL `https://doi.org/10.1594/PANGAEA.760120`.

S. L. Fiddes, A. Protat, M. D. Mallet, S. P. Alexander, and M. T. Woodhouse. 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, 2022. doi:10.5194/acp-22-14603-2022.

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

M. Folk, R. McGrath, and N. Yeager. HDF: an update and future directions. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*, volume 1, pages 273–275, 1999. doi:10.1109/IGARSS.1999.773469.

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

M. A. Giorgetta, R. Brokopf, T. Crueger, M. Esch, S. Fiedler, J. Helmert, C. Hohenegger, L. Kornblueh, M. Köhler, E. Manzini, T. Mauritsen, C. Nam, T. Raddatz, S. Rast, D. Reinert, M. Sakradzija, H. Schmidt, R. Schneck, R. Schnur, L. Silvers, H. Wan, G. Zängl, and B. Stevens. 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, 2018. doi:10.1029/2017MS001242.

K. Gohl and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS104 (ANT-XXXII/3), 2018. URL `https://doi.org/10.1594/PANGAEA.889515`.

A. Grundner. *Data-Driven Cloud Cover Parameterizations for the ICON Earth System Model Using Deep Learning and Symbolic Regression*. PhD thesis, University of Bremen, Bremen, Germany, 10 2023. URL `https://doi.org/10.26092/elib/2821`.

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

J. Gutt and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/3, 2013. URL `https://doi.org/10.1594/PANGAEA.809727`.

A. Guyot, A. Protat, S. P. Alexander, A. R. Klekociuk, P. Kuma, and A. McDonald. 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, 2022. doi:10.5194/amt-15-3663-2022.

C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi:10.1038/s41586-020-2649-2.

C. Hohenegger, P. Korn, L. Linardakis, R. Redler, R. Schnur, P. Adamidis, J. Bao, S. Bastin, M. Behravesh, M. Bergemann, J. Biercamp, H. Bockelmann, R. Brokopf, N. Brüggemann, L. Casaroli, F. Chegini, G. Datseris, M. Esch, G. George, M. Giorgetta, O. Gutjahr, H. Haak, M. Hanke, T. Ilyina, T. Jahns, J. Jungclaus, M. Kern, D. Klocke, L. Kluft, T. Kölling, L. Kornblueh, S. Kosukhin, C. Kroll, J. Lee, T. Mauritsen, C. Mehlmann, T. Mieslinger, A. K. Naumann, L. Paccini, A. Peinado, D. S. Praturi, D. Putrasahan, S. Rast, T. Riddick, N. Roeber, H. Schmidt, U. Schulzweida, F. Schütte, H. Segura, R. Shevchenko, V. Singh, M. Specht, C. C. Stephan, J.-S. von Storch, R. Vogel, C. Wengel, M. Winkler, F. Ziemen, J. Marotzke, and B. Stevens. ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, 16(2):779–811, 2023. doi:10.5194/gmd-16-779-2023.

M. Hoppmann, S. Tippenhauer, and T. Heitland. Continuous thermosalinograph oceanography along RV POLARSTERN cruise track PS123, 2023. URL `https://doi.org/10.1594/PANGAEA.938753`.

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

W. Jokat and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/5, 2013. URL `https://doi.org/10.1594/PANGAEA.816055`.

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

G. Kattner and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/2, 2012. URL `https://doi.org/10.1594/PANGAEA.776596`.

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

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

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

C. L. Knight, M. D. Mallet, S. P. Alexander, A. D. Fraser, A. Protat, and G. M. McFarquhar. Cloud properties and boundary layer stability above Southern Ocean sea ice and coastal Antarctica. *Journal of Geophysical Research: Atmospheres*, 129(10):e2022JD038280, 2024. doi:https://doi.org/10.1029/2022JD038280.

R. Knust and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVII/3, 2011. URL `https://doi.org/10.1594/PANGAEA.760121`.

R. Knust and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS82 (ANT-XXIX/9), 2014. URL `https://doi.org/10.1594/PANGAEA.839407`.

N. Koldunov, T. Kölling, X. Pedruzo-Bagazgoitia, T. Rackow, R. Redler, D. Sidorenko, K.-H. Wieners, and F. A. Ziemen. nextGEMS: output of the model development cycle 3 simulations for ICON and IFS, 2023.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXVII/2, 2011a. URL `https://doi.org/10.1594/PANGAEA.849045`.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVII/2, 2011b. URL `https://doi.org/10.1594/PANGAEA.760388`.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXVII/2, 2011c. URL `https://doi.org/10.1594/PANGAEA.760392`.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXVII/3, 2011d. URL `https://doi.org/10.1594/PANGAEA.760393`.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVII/3, 2011e. URL `https://doi.org/10.1594/PANGAEA.760389`.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXVII/3, 2012a. URL `https://doi.org/10.1594/PANGAEA.849044`.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXVIII/2, 2012b. URL `https://doi.org/10.1594/PANGAEA.844866`.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXVIII/2, 2012c. URL `https://doi.org/10.1594/PANGAEA.784458`.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/2, 2012d. URL `https://doi.org/10.1594/PANGAEA.784453`.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXVIII/3, 2012e. URL https://doi.org/10.1594/PANGAEA.844865.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/3, 2012f. URL https://doi.org/10.1594/PANGAEA.784454.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXVIII/3, 2012g. URL https://doi.org/10.1594/PANGAEA.784459.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXVIII/4, 2012h. URL https://doi.org/10.1594/PANGAEA.844859.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXVIII/4, 2012i. URL https://doi.org/10.1594/PANGAEA.784455.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXVIII/4, 2012j. URL https://doi.org/10.1594/PANGAEA.784460.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/2, 2013a. URL https://doi.org/10.1594/PANGAEA.844856.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/2, 2013b. URL https://doi.org/10.1594/PANGAEA.815472.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/2, 2013c. URL https://doi.org/10.1594/PANGAEA.815476.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/3, 2013d. URL https://doi.org/10.1594/PANGAEA.844855.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/3, 2013e. URL https://doi.org/10.1594/PANGAEA.815477.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/3, 2013f. URL https://doi.org/10.1594/PANGAEA.815473.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/4, 2013g. URL https://doi.org/10.1594/PANGAEA.844854.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/4, 2013h. URL https://doi.org/10.1594/PANGAEA.815723.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/4, 2013i. URL https://doi.org/10.1594/PANGAEA.815724.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/5, 2013j. URL https://doi.org/10.1594/PANGAEA.844853.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/5, 2013k. URL https://doi.org/10.1594/PANGAEA.815474.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/5, 2013l. URL https://doi.org/10.1594/PANGAEA.815478.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/6 (AWECS) to the Antarctic in 2013, 2013m. URL https://doi.org/10.1594/PANGAEA.842810.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/6, 2013n. URL https://doi.org/10.1594/PANGAEA.820733.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/6 (AWECS), 2013o. URL https://doi.org/10.1594/PANGAEA.819610.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/7, 2013p. URL https://doi.org/10.1594/PANGAEA.844852.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/7, 2013q. URL https://doi.org/10.1594/PANGAEA.820843.

G. König-Langlo. Upper air soundings during POLARSTERN cruise ANT-XXIX/8, 2013r. URL https://doi.org/10.1594/PANGAEA.844809.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVII/2, links to files, 2014a. URL https://doi.org/10.1594/PANGAEA.834511.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVII/3, links to files, 2014b. URL https://doi.org/10.1594/PANGAEA.834512.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/2, links to files, 2014c. URL https://doi.org/10.1594/PANGAEA.834518.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/3, links to files, 2014d. URL https://doi.org/10.1594/PANGAEA.834519.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXVIII/4, links to files, 2014e. URL https://doi.org/10.1594/PANGAEA.834520.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/2, links to files, 2014f. URL https://doi.org/10.1594/PANGAEA.834526.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/3, links to files, 2014g. URL https://doi.org/10.1594/PANGAEA.834527.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/6 (AWECS), links to files, 2014h. URL https://doi.org/10.1594/PANGAEA.833801.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/7, links to files, 2014i. URL https://doi.org/10.1594/PANGAEA.833802.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise ANT-XXIX/8, links to files, 2014j. URL https://doi.org/10.1594/PANGAEA.834530.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/8, 2014k. URL https://doi.org/10.1594/PANGAEA.832602.

G. König-Langlo. Meteorological observations during POLARSTERN cruise ANT-XXIX/8, 2014l. URL https://doi.org/10.1594/PANGAEA.832605.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS82 (ANT-XXIX/9), 2014m. URL https://doi.org/10.1594/PANGAEA.844805.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS82 (ANT-XXIX/9), links to files, 2014n. URL https://doi.org/10.1594/PANGAEA.834531.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise PS82 (ANT-XXIX/9), 2014o. URL https://doi.org/10.1594/PANGAEA.832603.

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS82 (ANT-XXIX/9), 2014p. URL https://doi.org/10.1594/PANGAEA.832606.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS89, links to files, 2015a. URL https://doi.org/10.1594/PANGAEA.844075.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS89 (ANT-XXX/2), 2015b. URL https://doi.org/10.1594/PANGAEA.844791.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise PS89 (ANT-XXX/2), 2015c. URL https://doi.org/10.1594/PANGAEA.849502.

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS89 (ANT-XXX/2), 2015d. URL https://doi.org/10.1594/PANGAEA.844571.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise ANT-XXIX/7, 2016a. URL https://doi.org/10.1594/PANGAEA.858532.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN), 2016b. URL https://doi.org/10.1594/PANGAEA.861658.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS96, links to files, 2016c. URL https://doi.org/10.1594/PANGAEA.860516.

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

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS96 (ANT-XXXI/2 FROSN), 2016e. URL https://doi.org/10.1594/PANGAEA.861438.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS97 (ANT-XXXI/3), 2016f. URL https://doi.org/10.1594/PANGAEA.861659.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise PS97 (ANT-XXXI/3), 2016g. URL https://doi.org/10.1594/PANGAEA.861442.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS97, links to files, 2016h. URL https://doi.org/10.1594/PANGAEA.860517.

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS97 (ANT-XXXI/3), 2016i. URL https://doi.org/10.1594/PANGAEA.861439.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS103 (ANT-XXXII/2), 2017a. URL https://doi.org/10.1594/PANGAEA.871788.

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS103 (ANT-XXXII/2), 2017b. URL https://doi.org/10.1594/PANGAEA.871721.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS103, links to files, 2017c. URL https://doi.org/10.1594/PANGAEA.871722.

G. König-Langlo. Continuous meteorological surface measurement during POLARSTERN cruise PS103 (ANT-XXXII/2), 2017d. URL https://doi.org/10.1594/PANGAEA.873016.

G. König-Langlo. Meteorological observations during POLARSTERN cruise PS104 (ANT-XXXII/3), 2017e. URL https://doi.org/10.1594/PANGAEA.874156.

G. König-Langlo. Upper air soundings during POLARSTERN cruise PS104 (ANT-XXXII/3), 2017f. URL https://doi.org/10.1594/PANGAEA.874224.

G. König-Langlo. Ceilometer CL51 raw data measured during POLARSTERN cruise PS104, links to files, 2017g. URL https://doi.org/10.1594/PANGAEA.874154.

D. Konsta, J.-L. Dufresne, H. Chepfer, J. Vial, T. Koshiro, H. Kawai, A. Bodas-Salcedo, R. Roehrig, M. Watanabe, and T. Ogura. 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, 2022. doi:10.1029/2021GL097593.

P. Korn, N. Brüggemann, J. H. Jungclaus, S. J. Lorenz, O. Gutjahr, H. Haak, L. Linardakis, C. Mehlmann, U. Mikolajewicz, D. Notz, D. A. Putrasahan, V. Singh, J.-S. von Storch, X. Zhu, and J. Marotzke. 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, 2022. doi:10.1029/2021MS002952.

S. Kremser, M. Harvey, P. Kuma, S. Hartery, A. Saint-Macary, J. McGregor, A. Schuddeboom, M. von Hobe, S. T. Lennartz, A. Geddes, R. Querel, A. McDonald, M. Peltola, K. Sellegri, I. Silber, C. Law, C. J. Flynn, A. Marriner, T. C. Hill, P. J. DeMott, C. C. Hume, G. Plank, G. Graham, and S. Parsons. Southern Ocean Cloud and Aerosol data set: a compilation of measurements from the 2018 Southern Ocean Ross Sea Marine Ecosystems and Environment voyage, 2020. URL https://doi.org/10.5281/zenodo.4060237.

S. Kremser, M. Harvey, P. Kuma, S. Hartery, A. Saint-Macary, J. McGregor, A. Schuddeboom, M. von Hobe, S. T. Lennartz, A. Geddes, R. Querel, A. McDonald, M. Peltola, K. Sellegri, I. Silber, C. S. Law, C. J. Flynn, A. Marriner, T. C. J. Hill, P. J. DeMott, C. C. Hume, G. Plank, G. Graham, and S. Parsons. 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, 2021. doi:10.5194/essd-13-3115-2021.

P. Kuma. rstool version 2.0.0, 2024. URL https://github.com/peterkuma/rstool. Accessed on 26 August 2024.

P. Kuma, A. J. McDonald, O. Morgenstern, S. P. Alexander, J. J. Cassano, S. Garrett, J. Halla, S. Hartery, M. J. Harvey, S. Parsons, G. Plank, V. Varma, and J. Williams. 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, 2020. doi:10.5194/acp-20-6607-2020.

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

F. Lamy and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS97 (ANT-XXXI/3), 2017. URL https://doi.org/10.1594/PANGAEA.873147.

P. Lemke and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/6, 2013. URL https://doi.org/10.1594/PANGAEA.819831.

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

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

N. Loeb, W. Su, D. Doelling, T. Wong, P. Minnis, S. Thomas, and W. Miller. 5.03 - Earth's top-of-atmosphere radiation budget. In S. Liang, editor, *Comprehensive Remote Sensing*, pages 67–84. Elsevier, Oxford, 2018. ISBN 978-0-12-803221-3. doi:https://doi.org/10.1016/B978-0-12-409548-9.10367-7.

M. Lucassen and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/4, 2012. URL https://doi.org/10.1594/PANGAEA.802809.

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

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

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

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

G. M. McFarquhar, C. S. Bretherton, R. Marchand, A. Protat, P. J. DeMott, S. P. Alexander, G. C. Roberts, C. H. Twohy, D. Toohey, S. Siems, Y. Huang, R. Wood, R. M. Rauber, S. Lasher-Trapp, J. Jensen, J. L. Stith, J. Mace, J. Um, E. Järvinen, M. Schnaiter, A. Gettelman, K. J. Sanchez, C. S. McCluskey, L. M. Russell, I. L. McCoy, R. L. Atlas, C. G. Bardeen, K. A. Moore, T. C. J. Hill, R. S. Humphries, M. D. Keywood, Z. Ristovski, L. Cravigan, R. Schofield, C. Fairall, M. D. Mallet, S. M. Kreidenweis, B. Rainwater, J. D'Alessandro, Y. Wang, W. Wu, G. Saliba, E. J. T. Levin, S. Ding, F. Lang, S. C. H. Truong, C. Wolff, J. Haggerty, M. J. Harvey, A. R. Klekociuk, and A. McDonald. 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, 2021. doi:10.1175/BAMS-D-20-0132.1.

Met Office. *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon, 2010. URL `https://scitools.org.uk/cartopy`.

B. Meyer and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/7, 2013. URL `https://doi.org/10.1594/PANGAEA.823259`.

B. Meyer and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS112 (ANT-XXXIII/3), 2018. URL `https://doi.org/10.1594/PANGAEA.895580`.

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

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

U. Neu, M. G. Akperov, N. Bellenbaum, R. Benestad, R. Blender, R. Caballero, A. Cocozza, H. F. Dacre, Y. Feng, K. Fraedrich, J. Grieger, S. Gulev, J. Hanley, T. Hewson, M. Inatsu, K. Keay, S. F. Kew, I. Kindem, G. C. Leckebusch, M. L. R. Liberato, P. Lionello, I. I. Mokhov, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, M. Schuster, I. Simmonds, M. Sinclair, M. Sprenger, N. D. Tilinina, I. F. Trigo, S. Ulbrich, U. Ulbrich, X. L. Wang, and H. Wernli. IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, 94(4):529–547, 2013. doi:10.1175/BAMS-D-11-00154.1.

nextGEMS authors team. nextGEMS Cycle 3, 2023. URL `https://easy.gems.dkrz.de/DYAMOND/NextGEMS/cycle3.html`. Accessed on 12 June 2024.

nextGEMS authors team. nextGEMS website, 2024. URL `https://nextgems-h2020.eu`. Accessed on 17 June 2024.

Q. Niu, G. M. McFarquhar, R. Marchand, A. Theisen, S. M. Cavallo, C. Flynn, P. J. DeMott, C. S. McCluskey, R. S. Humphries, and T. C. J. Hill. 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, 2024. doi:10.1029/2023JD040396.

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

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

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

A. Pérez-Alarcón, P. Coll-Hidalgo, R. M. Trigo, R. Nieto, and L. Gimeno. CyTRACK: An open-source and user-friendly python toolbox for detecting and tracking cyclones. *Environmental Modelling & Software*, 176:106027, 2024. ISSN 1364-8152. doi:10.1016/j.envsoft.2024.106027.

V. Ramadoss, K. Pfannkuch, A. Protat, Y. Huang, S. Siems, and A. Possner. 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, 2024. doi:https://doi.org/10.1029/2022JD038251.

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

M. M. Rienecker, M. J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, M. Sienkiewicz, R. D. Koster, R. Gelaro, I. Stajner, and J. E. Nielsen. The GEOS-5 data assimilation system—documentation of versions 5.0.1, 5.1.0, and 5.2.0. Technical report, NASA Center for AeroSpace Information, 7115 Standard Drive, Hanover, MD

21076-1320, USA, 2008. URL https://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf. Technical Report Series on Global Modeling and Data Assimilation, Volume 27.

O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. *arXiv*, 2015. doi:10.48550/ARXIV.1505.04597.

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

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

V. Schlindwein and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXIX/8, 2014. URL https://doi.org/10.1594/PANGAEA.839406.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS111 (ANT-XXXIII/2), 2019a. URL https://doi.org/10.1594/PANGAEA.903862.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS112 (ANT-XXXIII/3), 2019b. URL https://doi.org/10.1594/PANGAEA.903863.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS117, 2019c. URL https://doi.org/10.1594/PANGAEA.903916.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS118, 2019d. URL https://doi.org/10.1594/PANGAEA.903922.

H. Schmithüsen. Meteorological observations during POLARSTERN cruise PS111 (ANT-XXXIII/2), 2020a. URL https://doi.org/10.1594/PANGAEA.913625.

H. Schmithüsen. Meteorological observations during POLARSTERN cruise PS112 (ANT-XXXIII/3), 2020b. URL https://doi.org/10.1594/PANGAEA.913626.

H. Schmithüsen. Meteorological observations during POLARSTERN cruise PS117, 2020c. URL https://doi.org/10.1594/PANGAEA.913632.

H. Schmithüsen. Meteorological observations during POLARSTERN cruise PS118, 2020d. URL https://doi.org/10.1594/PANGAEA.913633.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS104 (ANT-XXXII/3), 2021a. URL https://doi.org/10.1594/PANGAEA.929235.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS111 (ANT-XXXIII/2), 2021b. URL https://doi.org/10.1594/PANGAEA.929241.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS111, links to files, 2021c. URL https://doi.org/10.1594/PANGAEA.929484.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS112, links to files, 2021d. URL https://doi.org/10.1594/PANGAEA.929488.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS112 (ANT-XXXIII/3), 2021e. URL https://doi.org/10.1594/PANGAEA.929242.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS117, 2021f. URL https://doi.org/10.1594/PANGAEA.935214.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS117, links to files, 2021g. URL https://doi.org/10.1594/PANGAEA.929505.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS118, links to files, 2021h. URL https://doi.org/10.1594/PANGAEA.929510.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS118, 2021i. URL https://doi.org/10.1594/PANGAEA.935216.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS123, 2021j. URL https://doi.org/10.1594/PANGAEA.935189.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS123, links to files, 2021k. URL https://doi.org/10.1594/PANGAEA.935273.

H. Schmithüsen. Continuous meteorological surface measurement during POLARSTERN cruise PS123, 2021l. URL https://doi.org/10.1594/PANGAEA.935226.

H. Schmithüsen. Radiosonde measurements during POLARSTERN cruise PS124, 2021m. URL https://doi.org/10.1594/PANGAEA.935190.

H. Schmithüsen. Ceilometer CL51 raw data measured during POLARSTERN cruise PS124, links to files, 2021n. URL https://doi.org/10.1594/PANGAEA.935274.

H. Schmithüsen, H. Jens, and J. Wenzel. Meteorological observations during POLARSTERN cruise PS123, 2021a.

URL https://doi.org/10.1594/PANGAEA.935269.

H. Schmithüsen, C. Rohleder, F. Otte, and S. Schröter. Meteorological observations during POLARSTERN cruise PS124, 2021b. URL https://doi.org/10.1594/PANGAEA.935270.

M. Schröder and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS96 (ANT-XXXI/2), 2017. URL https://doi.org/10.1594/PANGAEA.873151.

M. Schröder and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track PS111 (ANT-XXXIII/2), 2018. URL https://doi.org/10.1594/PANGAEA.895579.

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

SHiELD authors team. SHiELD: System for High-resolution prediction on Earth-to-Local Domains, 2024. URL https://www.gfdl.noaa.gov/shield/. Accessed on 18 June 2024.

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

B. Stevens, M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader, K. Block, R. Brokopf, I. Fast, S. Kinne, L. Kornblueh, U. Lohmann, R. Pincus, T. Reichler, and E. Roeckner. Atmospheric component of the MPI-M Earth system model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5(2): 146–172, 2013. doi:10.1002/jame.20015.

B. Stevens, M. Satoh, L. Auger, J. Biercamp, C. S. Bretherton, X. Chen, P. Düben, F. Judt, M. Khairoutdinov, D. Klocke, C. Kodama, L. Kornblueh, S.-J. Lin, P. Neumann, W. M. Putman, N. Röber, R. Shibuya, B. Vanniere, P. L. Vidale, N. Wedi, and L. Zhou. DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, 6(1):61, 2019. doi:10.1186/s40645-019-0304-z.

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

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

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

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. V.-B. and. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, feb 2020. doi:10.1038/s41592-019-0686-2.

A. Voldoire, B. Decharme, J. Pianezze, C. Lebeaupin Brossier, F. Sevault, L. Seyfried, V. Garnier, S. Bielli, S. Valcke, A. Alias, M. Accensi, F. Ardhuin, M.-N. Bouin, V. Ducrocq, S. Faroux, H. Giordani, F. Léger, P. Marsaleix, R. Rainaud, J.-L. Redelsperger, E. Richard, and S. Riette. 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, 2017. doi:10.5194/gmd-10-4207-2017.

C. J. Wall, D. L. Hartmann, and P.-L. Ma. 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, 2017. doi:10.1175/JCLI-D-17-0156.1.

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

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

doi:10.1007/s003820100157.

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

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

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

M. Wiegner, I. Mattis, M. Pattantyús-Ábrahám, J. A. Bravo-Aranda, Y. Poltera, A. Haefele, M. Hervo, U. Görsdorf, R. Leinweber, J. Gasteiger, M. Haeffelin, F. Wagner, J. Cermak, K. Komínková, M. Brettle, C. Münkel, and K. Pönitz. Aerosol backscatter profiles from ceilometers: validation of water vapor correction in the framework of CeiLinEx2015. *Atmospheric Measurement Techniques*, 12(1):471–490, 2019. doi:10.5194/amt-12-471-2019.

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

D. A. Wolf-Gladrow and G. Rohardt. Continuous thermosalinograph oceanography along POLARSTERN cruise track ANT-XXVIII/3, 2012. URL `https://doi.org/10.1594/PANGAEA.780004`.

Z. Xia and G. M. McFarquhar. 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, 2024. doi:10.1029/2023JD039869.

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

M. H. Zhang, W. Y. Lin, S. A. Klein, J. T. Bacmeister, S. Bony, R. T. Cederwall, A. D. Del Genio, J. J. Hack, N. G. Loeb, U. Lohmann, P. Minnis, I. Musat, R. Pincus, P. Stier, M. J. Suarez, M. J. Webb, J. B. Wu, S. C. Xie, M.-S. Yao, and J. H. Zhang. Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *Journal of Geophysical Research: Atmospheres*, 110(D15):D15S02, 2005. doi:10.1029/2004JD005021.

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

L. Zhao, Y. Wang, C. Zhao, X. Dong, and Y. L. Yung. 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, 2022. doi:10.1007/s00376-022-2036-z.