



Comparing remotely sensed observations of clouds and aerosols in the Southern Ocean with climate model simulations

A thesis submitted in partial fulfilment of the requirements
for the Degree of Doctor of Philosophy in the University of Canterbury
by Peter Kuma

School of Physical and Chemical Sciences
University of Canterbury
Christchurch
Aotearoa/New Zealand

30 May 2020

Please consider the environment before printing this thesis.

Copyright © 2016–2020 Peter Kuma, Adrian J. McDonald, Olaf Morgenstern, Simon P. Alexander, John J. Cassano, Sally Garrett, Jamie Halla, Sean Hartery, Mike J. Harvey, Simon Parsons, Graeme Plank, Vidya Varma, Jonny Williams, Richard Querel, Israel Silber, Connor J. Flynn, Guang Zeng.

This work is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

Southern Ocean (SO) shortwave (SW) radiation biases are a common problem in contemporary general circulation models (GCMs), with most models exhibiting a tendency to absorb too much incoming SW radiation. These biases have been attributed to deficiencies in the representation of clouds during the austral summer months, either due to cloud cover or cloud albedo being too low. They affect simulation of New Zealand (NZ) and global climate in GCMs due to excessive heating of the sea surface and the effect on large-scale circulation. Therefore, improvement of GCMs is necessary for accurate prediction of future NZ and global climate. Currently the New Zealand Earth System Model (NZESM), based on the UK Hadley Centre Coupled Model version 3 (HadGEM3), is developed at the National Institute of Water and Atmospheric Research (NIWA) and the University of Canterbury. We performed ship-based lidar, radar, radiosonde and weather observations on two SO voyages and processed data from multiple past SO voyages. We used the observations and satellite measurements for evaluation of NZESM and contrasting with the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) to better understand the source of the problem. Due to the nature of lidar observations (the laser signal is quickly attenuated by clouds) they cannot be used for 1:1 comparison with a model without using a lidar simulator, which performs atmospheric radiative transfer calculations of the laser signal. We modify an existing satellite lidar simulator present in the Cloud Feedback Model Intercomparison Project (CFMIP) Observational Simulator Package (COSP) for use with the ground-based lidars used in our observations by modifying the geometry of the radiative transfer calculations, Mie and Rayleigh scattering of the laser signal. We document and make the modified lidar simulator available to the scientific community as part of a newly-developed lidar processing tool called the Automatic Lidar and Ceilometer Framework (ALCF), which enables unbiased comparison between lidar observations and models by performing calibration of lidar backscatter, noise removal and consistent cloud detection. We apply the lidar simulator on NZESM model fields. Significant SW radiation errors in the SO of up to 21 Wm^{-2} are shown to be present in the model. Using the lidar observations, we find that the model underestimates overall cloud cover by about 9% and strongly underestimates boundary layer low-level stratocumulus (Sc) cloud below 1 km and fog. By using radiosonde observations, we find that the observed cloud was strongly linked to the boundary layer stability and sea surface temperature, while this relationship is weaker in the model. We identify that these errors are not due to misrepresentation of large-scale circulation, which is prescribed in our model based on global satellite observations by nudging. We conclude that the problem is likely in the subgrid-scale parametrisation schemes of the boundary layer, convection and large-scale cloud. In order to address the deficiencies identified we perform experimental simulations of NZESM with modifications of the parametrisation schemes. We find that a three-layer cloud pro-

files were common in the Ross Sea region, consisting of cumulus (Cu) below Sc, and corresponding to local thermodynamic levels: lifting condensation level, dry and moist neutral buoyancy levels of parcels lifted from the surface. We find that not enough moisture is transported to the top of the boundary layer to form Sc clouds. By increasing surface moisture flux and convective mass flux in the model we improve the Sc cloud simulation, but we show that a lack of vertical moisture transport across the lifting condensation level from the surface layer to the zone of convective mass flux is a likely limiting factor. We show that the modifications had a positive impact on the Southern Ocean and global radiation balance of up to 5 Wm^{-2} in zonal average over this limited time period. We suggest that further research should focus on the weak vertical coupling between the boundary layer turbulence and boundary layer convection parametrisation in the model, and that the lidar simulator framework is used as a cloud evaluation tool in further studies due to its benefits over more statistical approaches, which tend to hide compensating biases.

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 2: Kuma, P., McDonald, A. J., Morgenstern, O., Alexander, S. P., Cassano, J. J., Garrett, S., Halla, J., Hartery, S., Harvey, M. J., Parsons, S., Plank, G., Varma, V., and Williams, J.: *Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations. Atmospheric Chemistry and Physics (accepted)*, <https://doi.org/10.5194/acp-2019-201>, 2020.

Chapter 3: Kuma, P., McDonald, A. J., Morgenstern, O., Querel, R., Silber, I., Flynn, C.: *Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0), Geoscientific Model Development (submitted)*, <https://doi.org/10.5281/zenodo.3785715>. 2020.

Chapter 4: Kuma, P., McDonald, A. J., Morgenstern, O., Hartery, S., Williams, J., Varma, V., Zeng, G., Harvey, M., Parsons, S., Graeme, P: *Improving Southern Ocean boundary layer cloud parametrisation in the general circulation model HadGEM3-GA7.1/UM11.4, manuscript in preparation*, 2020.

Please detail the nature and extent (%) of contribution by the candidate:

Chapter 2: Peter Kuma participated on methodology development, voyage observations, data analysis, writing and reviewing of the manuscript. Extent: approx. 70%.

Chapter 3: Peter Kuma wrote the code of the framework, performed the data analysis of the case studies and wrote the text of the manuscript. Extent: approx. 80%.

Chapter 4: Peter Kuma participated on the TAN1802 field measurements, performed the model runs, analysis and wrote the manuscript. Extent: approx. 80%.

Please see also "Author contributions" at the end of Chapter 2, 3 and 4 for details.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all. The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Adrian McDonald*

Signature:

A handwritten signature in black ink that reads "A. J. McDonald". The signature is written in a cursive style with a large, sweeping flourish at the end.

Date:

6/5/2020

Acknowledgements

The author would like to thank Adrian McDonald and Olaf Morgenstern for supervising the PhD; University of Canterbury for providing the UC Doctoral Scholarship; the New Zealand Deep South National Science Challenge for providing funding; NIWA, NeSI and the UK Met Office for providing access to their supercomputing resources and the Unified Model; NIWA for making it possible to conduct Southern Ocean observations on R/V Tangaroa; the crew and the science team of the TAN1702 and TAN1802 voyages; everyone who co-authored the papers which form Chapter 2, 3, and 4: Adrian McDonald, Olaf Morgenstern, Simon Alexander, John Cassano, Sally Garrett, Jamie Halla, Sean Hartery, Mike Harvey, Simon Parsons, Graeme Plank, Vidya Varma, Jonny Williams, Richard Querel, Israel Silber, Connor Flynn, Guang Zeng and everyone acknowledged in the chapters.

Contents

Abstract	iii
Acknowledgements	vii
1 Introduction	1
1.1 Objectives	7
1.2 Methods	7
1.2.1 New Zealand Earth System Model	8
1.2.2 CFMIP Observation Simulator Package	8
1.2.3 In situ observations in the Southern Ocean	9
1.2.4 Satellite radiation budget observations	12
1.2.5 Auxiliary software	13
1.3 Outline of the thesis and author’s contributions	14
2 Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 re-analysis using ship-based observations	15
2.1 Introduction	16
2.2 Datasets	18
2.2.1 Ship observations	18
2.2.2 HadGEM3	20
2.2.3 MERRA-2	21
2.2.4 CERES	21
2.2.5 NSIDC sea ice concentration	22
2.3 Methods	22
2.3.1 Lidar simulator	22
2.3.2 Lidar data processing	23
2.4 Spatiotemporal subsets investigated	25
2.5 Results	25
2.5.1 Shortwave radiation balance	25
2.5.2 Cloud occurrence in model and observations	28
2.5.3 Radiosonde observations	30
2.5.4 Zonal plane comparison of GA7.1N and MERRA-2	32
2.6 Discussion	32
2.7 Conclusions	38
3 Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0)	41
3.1 Introduction	42
3.2 Overview of operation of the Automatic Lidar and Ceilometer Framework (ALCF 1.0)	44
3.3 Supported input data: instruments, reanalyses and models	46
3.3.1 Instruments	46

3.3.2	Reanalyses and models	47
3.4	Lidar simulator	49
3.4.1	Rayleigh and Mie scattering	50
3.4.2	Cloud overlap and cloud fraction	53
3.4.3	Multiple scattering	54
3.5	Lidar data processing	54
3.5.1	Noise and subsampling	54
3.5.2	Backscatter calibration	55
3.5.3	Cloud detection	57
3.5.4	Water vapour absorption	57
3.6	Description of case studies	58
3.7	Case study results	59
3.8	Discussion and conclusions	71
4	Improving Southern Ocean boundary layer cloud parametrisation in the HadGEM3-GA7.1/UM11.4 general circulation model	75
4.1	Introduction	76
4.2	Methods	77
4.2.1	The TAN1802 voyage	77
4.2.2	HadGEM3-GA7.1/UM11.4	78
4.2.3	ALCF	79
4.3	Parametrisations	79
4.3.1	Large-scale cloud scheme (PC2)	79
4.3.2	Boundary layer scheme	80
4.3.3	Convection scheme	80
4.3.4	JULES surface flux parametrisation	81
4.3.5	Southern Ocean boundary layer	81
4.3.6	Experimental run	82
4.4	Results	83
4.4.1	Cloud observations	83
4.4.2	Cloud representation	85
4.4.3	Boundary layer mass flux and relative humidity	87
4.4.4	Cloud occurrence statistics	89
4.4.5	Shortwave radiation bias	90
4.4.6	Boundary layer types	91
4.5	Discussion and conclusions	91
5	Conclusions and further work	95
5.1	Further work	97

List of figures

1.1	Biases in the TOA net radiation down	2
1.2	Spaceborne vs. surface lidar scattering ratio	5
1.3	Instruments deployed on Southern Ocean voyages and stations.	10
1.4	Lufft CHM 15k ceilometer volume backscattering coefficient profile plot	11
1.5	Cloud base height distribution	12
1.6	Statistics calculated from observations collected on the TAN1802 voyage	13
2.1	Map showing tracks of voyages	19
2.2	Schematic of the processing pipeline	22
2.3	Geographical distribution of the TOA outgoing SW radiation in CERES, GA7.1N and MERRA-2	26
2.4	Zonal means of the TOA outgoing SW radiation in CERES, GA7.1N and MERRA-2	27
2.5	Scatter plot of SW radiation bias in GA7.1N and MERRA-2	28
2.6	Cloud occurrence frequency as a function of height derived from ceilometer observations and model fields	29
2.7	Cloud cover bias in models relative to observations	29
2.8	Scatter plots of radiosonde measurements on the TAN1802 and NBP1704 voyages	31
2.9	Histogram of $\min\{SLL, LCL\}$	33
2.10	Zonal plane plot of cloud liquid and ice mixing ratios in GA7.1N and MERRA-2 at $60^\circ S$	34
3.1	Scheme showing the operation of the ALCF and the processing commands	45
3.2	Theoretical distributions of cloud droplet radius	53
3.3	Map showing the location of sites	58
3.4	Cloud occurrence histogram as a function of height	60
3.5	Examples of observed and simulated backscatter profiles	61
3.6	The same as Fig. 3.5 but for the Lauder site.	62
3.7	The same as Fig. 3.5 but for the Christchurch site.	63
3.8	Scatter plot showing the average cloud fraction and the average inverse lidar ratio	67
3.9	Backscatter histograms as a function of height	68
3.10	The same as Fig. 3.9 but calculated from clear sky profiles only.	69
3.11	Backscatter noise standard deviation histogram	70
4.1	A map showing the track of the TAN1802 voyage	78
4.2	A scheme showing the operation of the boundary layer and convection schemes	82
4.3	Selected daily backscatter profiles collected on the TAN1802 voyage and radiosonde profiles	84
4.4	Selected daily observed and simulated backscatter profile plots	86
4.5	Figure 4.4 continued.	87
4.6	Mass flux and relative humidity profiles	88
4.7	Figure 4.6 continued.	89
4.8	Cloud occurrence histogram as a function of height	90

4.9	Top of atmosphere outgoing shortwave radiation	91
4.10	Zonal average of top of atmosphere outgoing shortwave radiation	92
4.11	Boundary layer type histograms	92

List of tables

2.1	Table of voyages	18
2.2	Table of deployments	20
2.3	A table showing a "back-of-the-envelope" calculation	38
3.1	Table of ALCs and their technical parameters	46
3.2	Reanalyses and models used in the case studies and some of their main properties	47
3.3	Table of physical quantities.	49
3.4	Table of sensitivity tests	52
3.5	Theoretical molecular backscatter and the calibration coefficient	55
3.6	Location of sites and instruments	59
4.1	Summary of the cloud-related subgrid-scale parametrisations in the UM11.4.	79
4.2	Diagnostic model levels and layers	83

Chapter 1

Introduction

Clouds represent one of the largest sources of uncertainty in estimating global climate sensitivity (Williams and Bodas-Salcedo, 2017). Clouds over the ocean are especially important for determining the radiation budget, due to the low albedo of the sea surface compared to land. Over the Southern Ocean (SO), cloud cover exceeds 80%, with predominantly boundary layer clouds (Mace et al., 2009). Due to its large influence on circulation and atmospheric transports in the Southern Hemisphere, the SO is important for global climate. Unlike most other places on the globe, it is largely unaffected by sources of continental and anthropogenic aerosols, is dominated by a strong circumpolar vortex, and its southern boundary is a permanently ice-covered continent, which could mean that global parametrisations do not apply very well in this region. SO south of 30°S accounts for about 43% of anthropogenic CO₂ and 75% of excess heat uptake (Frölicher et al., 2015). Observations in the SO are sparse, which limits the accuracy of simulations by numerical weather prediction (NWP) models and general circulation models (GCMs). Globally, clouds have a predominantly cooling effect on the climate due to reflection of sunlight, which exceeds the warming effect due to absorption of thermal radiation from the surface, estimates identify 18 Wm⁻² of cooling relative to a cloud-free atmosphere (Zelinka et al., 2017). This effect is about 5 times as large as heating from a doubling of CO₂, which highlights the importance of cloud cover in modulating global climate. Nearly all climate models predict cloud feedback to be positive, i.e. amplification of warming with increasing CO₂ concentration.

Shortwave (SW) radiation bias over the SO of up to 30 Wm⁻² is a well-documented problem in current NWP models and GCMs (Trenberth and Fasullo, 2010) (Fig. 1.1), and it has been the subject of many studies. It manifests both as a bias in SW radiation reaching the surface and as a SW reflectivity bias at the top of the atmosphere (TOA). Bodas-Salcedo et al. (2014) evaluated SW bias in a number of GCMs and found a strong SW bias is a very common problem, leading to overestimated sea surface temperature (SST) in the SO. Trenberth and Fasullo (2010) noted that a poor representation of clouds might lead to unrealistic projections for the Southern Hemisphere. This bias is linked to large-scale model errors such as a double-intertropical convergence zone (Hwang and Frierson, 2013), position of the midlatitude jet and meridional energy transport (Mason et al., 2014). The reasons for the observed SW radiation bias can be numerous, concurrent and compensating. As noted by Kelleher and Grise (2019), cloud biases in the SO can arise either as a result of biases in large-scale dynamics, or cloud parametrisation. To summarise, they can range from microphysical to large-scale dynamics and be due to misrepresentation of: cloud fraction, cloud optical depth, frequency of cloud regimes or types, cloud vertical distribution and overlap, cloud horizontal distribution and homogeneity, cloud phase and supercooled liquid con-

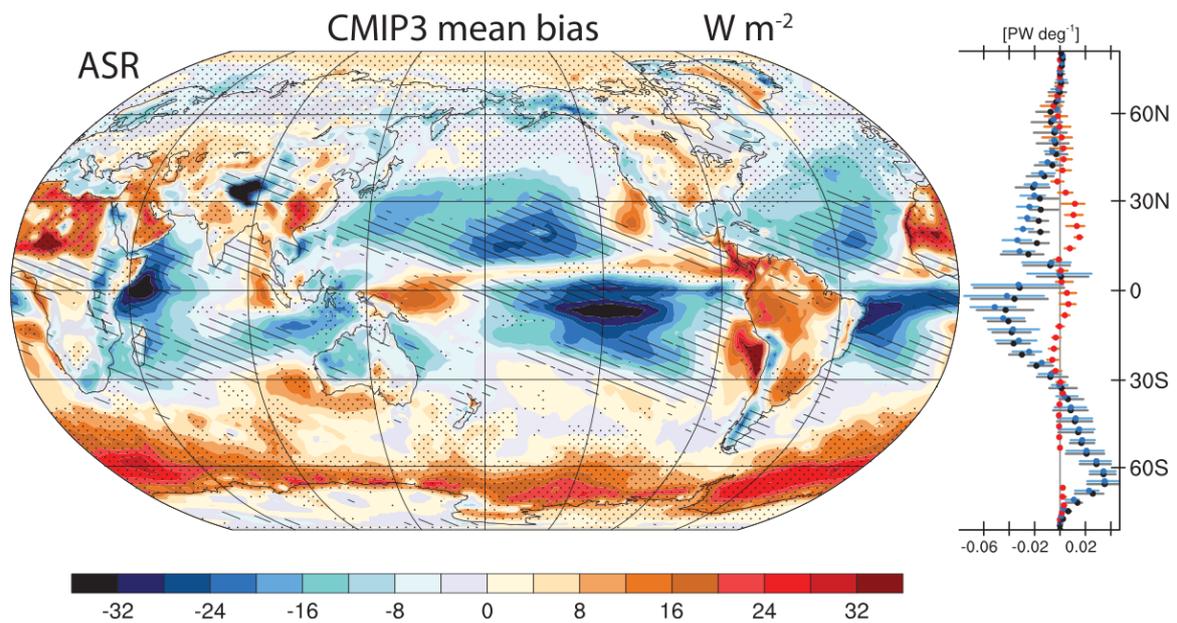


Figure 1.1 | Biases in the TOA net radiation down relative to observations regionally for 1990–99 in W m^{-2} , where stippled (hatched) regions correspond to regions in which at least three quarters of the models share a common positive (negative) bias. (right) The model zonal mean is given (dots) with the 25th to 75th percentile range (lines) over land (red), ocean (blue), and all (black) surfaces. *Adopted from Trenberth and Fasullo (2010).*

tent, surface albedo (sea ice vs. water), moisture fluxes, large-scale circulation, extratropical and polar cyclones, weather regimes, direct and indirect aerosol effects, radiative transfer parametrisation, boundary layer turbulence and convection, among others.

The New Zealand Earth System Model (NZESM) is a relatively new earth system model based on the UK climate model HadGEM/UK Earth System Model (UKESM) (Williams et al., 2016), whose aim is to improve climate predictions for Aotearoa/New Zealand. Reducing SO model biases is essential for achieving this aim. Walters et al. (2017) showed that a clear and extensive SW radiation bias over the SO is present in the atmospheric component of the model Global Atmosphere version 6.0 (GA6.0) compared to satellite radiation budget observations by the Clouds and the Earth’s Radiant Energy System (CERES) (Wielicki et al., 1996). The bias in the context of the UK Met Office models has been studied by Bodas-Salcedo et al. (2012) by assessing cloud regimes in cyclones. Using observations by the International Satellite Cloud Climatology Project (ISCCP) (Rossow and Schiffer, 1999), Multi-angle Imaging SpectroRadiometer (MISR) (Diner et al., 1998), Moderate Resolution Imaging Spectroradiometer (MODIS) (Salomonson et al., 2002), Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) (Winker et al., 2010) and CloudSat (Stephens et al., 2002) satellites, they found that the model underestimates optical depth of stratocumulus and mid-topped clouds. Recently, Davies et al. (2017) studied boundary layer clouds in the SO compared to the Northern Hemisphere, with a focus on supercooled liquid in clouds and cloud homogeneity. They noted that boundary layer clouds are a likely explanation for the bias due to their large fractional coverage over the SO. Examination of cloud cover in the NZESM against passive satellite instruments was performed by Schuddeboom et al. (2017, 2019) and is an ongoing

effort.

While some authors focused on cloud distribution, others emphasised the role of microphysics, especially supercooled liquid content. Because supercooled liquid has a higher SW reflectivity than the equivalent amount of ice particles (in terms of mixing ratio), it has a positive effect on cloud albedo. Morrison et al. (2011) studied the occurrence of supercooled liquid in clouds over the SO using observations by MODIS and found that it is present year-round in low clouds at temperature as low as -40°C . Lawson and Gettelman (2014) noted that supercooled liquid is often underestimated in the Antarctic in GCMs, and mixed-phase clouds can occur at -32°C . They showed that increasing supercooled liquid in the Community Earth System Model (CESM) leads to a cloud radiative effect (CRE) increase of 7.4 Wm^{-2} over Antarctica. More recently, Kay et al. (2016) managed to fix the SW radiation bias in the CESM by increasing supercooled liquid in shallow convective clouds, and notably they also needed to reduce a compensating tropical SW radiation bias to maintain global radiation balance in the model. Bodas-Salcedo et al. (2016) found supercooled liquid to be abundant in the SO in summer and contribute about 30% to the reflected SW radiation. Noh et al. (2019) developed an algorithm for detecting mixed-phase cloud with liquid top in the SO using Himawari geosynchronous (GEO) satellite data, which are thought to be common in the region, but difficult to detect with passive satellite instruments. They focused on liquid-top mixed-phase clouds. In their case studies, they found both supercooled liquid and mixed-phase clouds with liquid top over the SO south of Australia and New Zealand and noted that their algorithm may complement active instruments in detecting mixed-phase cloud.

Several field campaigns were performed in the SO in recent years: Clouds, Aerosols, Precipitation, Radiation, and Atmospheric Composition over the Southern Ocean (CAPRICORN) (Mace and Protat, 2018a,b), Measurements of Aerosols, Radiation and Clouds over the Southern Ocean (MARCUS) (McFarquhar, 2016), the Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES) (McFarquhar et al., 2014) and the Macquarie Island Cloud and Radiation Experiment (MICRE) (DeMott et al., 2018). The SOCRATES campaign consisted of 15 flights of GV HIAPER and a voyage of R/V Investigator from Hobart, Tasmania in January–February 2018, organised by the National Science Foundation (NSF) and the National Center for Atmospheric Research (NCAR). Gettelman et al. (2020) analysed cloud microphysical observations from these flights compared to nudged CAM6 simulations, and found that CAM6 represents cloud properties relatively well, and observed supercooled liquid clouds extensively in cold sectors of cyclones. They found the representation of supercooled liquid better than in CAM5 due to a scheme dependence on the available ice nuclei. They found 50% differences in ice water path (IWP) and liquid water path (LWP) between CAM5 and CAM6. Their model simulates cloud droplet size distribution prognostically and they found a satisfactory agreement with the in situ airborne observations. Mace and Protat (2018a) analysed cloud observations collected on the second CAPRICORN voyage of R/V Investigator from Hobart, Tasmania to 53°S in March–April 2016. In their radar and lidar observations, low cloud below 2 km was predominant during the voyage and the total lidar cloud fraction was 76%, compared to 87% in CloudSat–CALIPSO in the region and time of year in 2007–11, with more high clouds identified by CloudSat–CALIPSO. They also found that about 30% of cloud is detected by a ship-based lidar but not a radar due to low sensitivity of the radar. In terms of cloud phase, they found that ice-phase processes occur 20–40% more often than implied by

CALIPSO due to attenuation of the signal at the cloud top. Thus, one should perhaps be cautious when using active satellite products as a reference for supercooled liquid cloud evaluation in GCMs in this region. They performed 1–2 daily soundings and found MERRA-2 about 1.2 K warmer and 8% drier. They note that unlike the lidar, the radar is unable to reliably detect cloud base height due to frequent precipitation, which cannot be distinguished from cloud. Mace and Protat (2018b) studied stratocumulus clouds occurring during the CAPRICORN voyage. They characterise them as tenuous, supercooled, rarely drizzling and present in cold air advection. They quantify their water path at 15–25 gm^{-2} , effective radius at 8 μm , number concentration at 20 cm^{-2} and optical depth 3–4. It is probably notable, however, that these values can be different in the high-latitude SO, not reached by the CAPRICORN voyage. They hypothesise that these non-precipitating stratocumulus clouds are responsible for the majority of the SO shortwave radiation biases identified in GCMs. The MARCUS field campaign was conducted between November 2017 and March 2018 on *Aurora Australis*. It was focused on collecting biogenic INP concentrations in the region, but a range of ARM instruments were deployed on this ship. Zheng and Li (2019) analysed warm air advection events on the voyages and found that they induce highly-stratified cloud-topped marine boundary layer with stratiform clouds.

Multiple observational datasets are available for assessing the SO biases, largely consisting of satellite datasets, and a relatively few ship- and land-based datasets due to the very large costs and operational difficulties of field campaigns in this remote and extreme-weather region. Satellite observations provide the most complete record both spatially and temporally, although they do not provide historical records prior to 1960s and past observations are limited by instrument capabilities and the availability of derived products. They have been utilised by most studies of clouds in the SO and globally. Satellite instruments are very diverse, though only a few datasets are readily available for studying clouds. Operational GEO satellites provide near-continuous temporal coverage, which makes them ideal for studying clouds, but they have a limited use in high-latitude regions such as the SO. In combination with operational polar-orbiting low Earth orbit (LEO) satellites such as the NASA Polar Operational Environmental Satellites (POES), they have been used to produce a very long-term (1983–present) cloud-oriented dataset ISCCP (Schiffer and Rossow, 1983). However, this dataset is limited by a small number of spectral channels of the Advanced Very High Resolution Radiometer (AVHRR). Other extensive cloud datasets include MODIS on board of the NASA Afternoon Train (A-Train) satellites Aqua and Terra and the Extended AVHRR Polar Pathfinder (APP-x) (Meier et al., 1997). Other notable instruments available for studying clouds include MISR and passive microwave sensors, due to their ability to observe cloud liquid water, total column water vapour, vertically-resolved temperature profile, and ability to see through clouds, even though their relatively low spatial resolution makes passive microwave instruments less popular than passive visible (VIS) and infrared (IR) instruments. Passive VIS and IR satellite observations of clouds are ideal due to their high spatial and temporal resolution, but have a number problems globally and some specifically in polar latitudes (Bromwich et al., 2012):

- Passive instruments can only observe the highest layer of clouds, unless the layer is semi-transparent, meaning that cloud vertical structure is poorly measured with passive instruments.

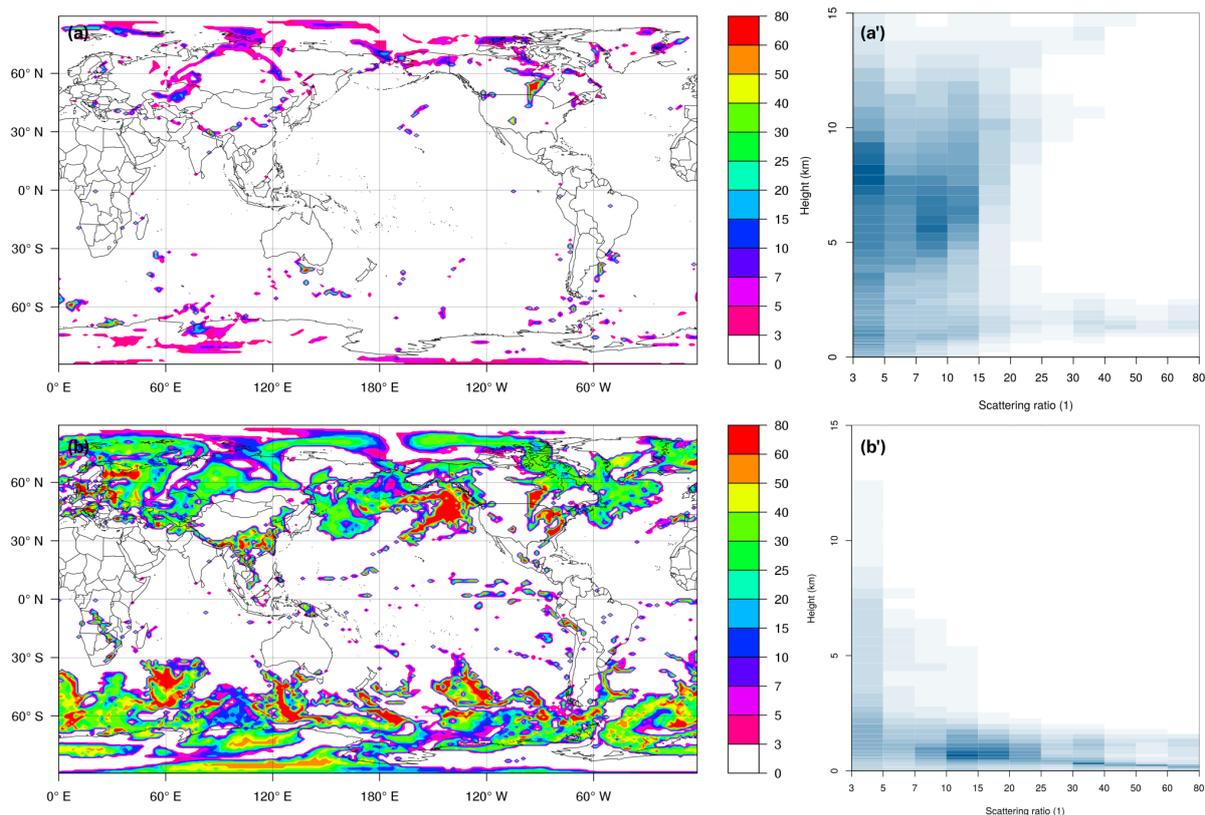


Figure 1.2 | (a), (a') Spaceborne vs. (b), (b') surface lidar scattering ratio (SR) simulated based on the New Zealand Earth System Model (NZESM) output in year 2017 at a lidar wavelength of 532 nm. (a), (b) show SR at a model level corresponding to approximately 300 m above sea level (ASL) over the sea, and (a'), (b') show an SR histogram by height.

- It is difficult to discern clouds from surface ice and snow in the VIS spectrum due to similar albedo and in the IR spectrum due to similar temperature and frequent inversions.
- Poor detection of semi-transparent high clouds, falsely classified as mid-level clouds (Haynes et al., 2011).
- Lack of sunlight limits polar wintertime SW measurements.

Active satellite instruments are affected by signal attenuation by optically thick clouds (lidar), ground clutter (radar), and compared to passive instruments they have smaller spatial coverage, shorter historical record and a small number of spectral bands, limiting their ability to determine cloud microphysical properties (Noh et al., 2017; Mace and Protat, 2018a,b; Gettelman et al., 2020). In addition to spaceborne observations, ground-based and in situ observations can provide an important complementary view of clouds from below. Figure 1.2 shows that scattering ratio (SR) (the ratio of total backscatter to molecular backscatter) in the boundary layer measured by a lidar is much higher when measured by a ground-based lidar than a spaceborne lidar due to obscuration by higher-level cloud. Ground-based and in situ instruments include radars, ceilometers, lidars, pyranometers, sky cameras, radiosondes, dropsondes, in situ aerosol measurements (cloud condensation nuclei and ice nuclei) and airborne observations from drones, weather balloons, kites

and aircraft. These observations are logistically difficult and expensive, and are generally sparse in the SO, with limited time periods and limited historical records. The use of ground-based and in situ observations alone for assessment of GCMs is difficult due to their small representativeness of climatic conditions, and therefore there is a risk of tuning the model to the specific conditions occurring during a case study (Jakob, 2003). Deployments on ships of opportunity can make these types of observations more cost-efficient and common.

Different processing of observations from the same instrument can lead to different results, for example the GCM-Oriented CALIPSO Cloud Product (GOCCP) relative to the standard CloudSat–CALIPSO products (Chepfer et al., 2010). Different thresholds can be applied which define what “cloud” is, and cloud detection is affected by targeting a particular false alarm ratio, such as 5% as in the CloudSat–CALIPSO dataset (Hagihara et al., 2010). Probability of detection (sensitivity) then depends on the receiver operating characteristic (ROC) curve, which in turn depends on instrument noise. Instrument noise and bias can vary over lifetime of the instrument, or between instruments of multi-instrument datasets such as ISCCP. A problem with different processing algorithms was noted by Martucci et al. (2010), who compared manufacturer-supplied cloud base height (CBH) determination between co-located Vaisala CL31 and Jenoptik CHM 15k ceilometers and found a poor agreement, and developed a new algorithm for determining CBH which leads to consistent height between the two instruments.

Due to the reasons outlined above, a combination of multiple satellite passive, active, ground-based and in situ observations are needed to comprehensively assess cloud climatology and biases. This has been also noted by other authors: Williams and Bodas-Salcedo (2017) evaluated cloud representation in the UK Met Office Unified Model (UM) using a multi-dataset and multi-diagnostic approach, and highlighted the importance of using multiple instruments due to compensating errors in GCMs. While use of single or combined satellite observations to assess model performance is common in many studies, combination of ground-based and spaceborne instruments is less common. For example, Muhlbauer et al. (2015) studied cirrus clouds using A-Train observations (CloudSat, CALIPSO, MODIS, CERES), ground-based Atmospheric Radiation Measurement (ARM) radar and aircraft observations. Zhang et al. (2017) performed a comparison of satellite and ground-based cloud observations at an ARM site.

Comparison between models and observations cannot always be performed directly, especially if observations do not produce fields equivalent to model quantities. In such cases observations can be mapped to model fields by inversion algorithms, but this may be unreliable due to a large number of factors involved and a limited view of the instrument (parts of the atmosphere obscured by clouds). Conversely, model fields can be mapped to observations by instrument simulators, and this approach has been used extensively in a number of studies. Satellite simulators such as the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011) solve the problem by transforming model fields to observed fields, which can then be compared directly or statistically.

We provide a further literature review in Chapter 2.

1.1 Objectives

Our objectives are aligned with the New Zealand Deep South National Science Challenge (DSC), whose mission is to “enable New Zealanders to adapt, manage risk, and thrive in a changing climate”, and is broadly in line with the current international research in the area such as SOCRATES (McFarquhar et al., 2014). Here, we focus on complementing other studies evaluating representation of clouds, aerosols and cloud–aerosol interaction in the SO, but also taking into consideration the Southern Hemisphere and global processes, with a particular focus on utilising in situ measurements available from intensive observation periods (IOPs), complemented by land-based stations. For this purpose the COSP simulator needs to be extended to support these instruments. Ground-based observations need to be complemented by satellite observations, especially the global radiation budget measurements by CERES. Other diagnostic means include case studies, by which we can ensure that any improvements are due to the right physical reasons rather than just improving statistics by mutually compensating model errors. Our particular focus is therefore on linking observed biases to model processes. We shall try to evaluate specific deficiencies in the NZESM subgrid-scale parametrisations affecting clouds and radiative transfer, in order to determine the relative importance of cloud macrophysical and microphysical characteristics in the observed biases. This has been explored to some extent by other authors, but not always in the context of the UM or the NZESM, where the causes can be different. While our focus is on evaluation of the NZESM, contrasting with other models, such as atmospheric reanalyses is useful. We shall focus on biases in the SO and the Antarctic, but pay attention to any processes relevant to the Southern Hemisphere and globally. Adjacent to our study will be development of a publicly-available dataset of in situ observations in the SO based on previous and new SO voyages and permanent stations collected by the University of Canterbury and our collaborators. Our main objectives are outlined below:

1. Participate on SO IOPs with the aim of collecting atmospheric observations for model evaluation.
2. Collate and post-process the existing and new SO in situ datasets.
3. Extend the COSP lidar simulator with a ceilometer and ground-based lidar simulator for instruments deployed on the SO voyages.
4. Use in situ and satellite observations in conjunction with the lidar simulator to evaluate SO cloud biases in the NZESM.
5. Perform experimental simulations of the NZESM with the aim of improving the simulation of SO clouds relative to the observations.

1.2 Methods

Achieving our objectives will require a number of modelling and observational resources. As outlined here and discussed in a greater detail in Chapter 2, 3 and 4, these include access to the model output and code of the NZESM, the COSP simulator, publicly-available reanalyses, in situ SO observations and publicly-available satellite datasets.

1.2.1 New Zealand Earth System Model

The NZESM is an actively developed branch of HadGEM/UKESM, with the aim of improving the atmosphere, ocean and cryosphere simulation affecting Aotearoa/New Zealand (Williams et al., 2016). Development is done by the National Institute of Water and Atmospheric Research (NIWA) in Wellington and the University of Canterbury. The NZESM is a fully coupled atmosphere–ocean model, including land surface and sea ice. The parent model UKESM (Walters et al., 2017) is planned to participate in the 6th Climate Model Intercomparison Project (CMIP6) (Eyring et al., 2016; Meehl et al., 2014), which shall eventually contribute to the upcoming Intergovernmental Panel on Climate Change (IPCC) 6th Assessment Report (AR6). Participation of the NZESM in the CMIP6 is currently not envisioned, although any improvements achieved will be continuously contributed to the parent model.

Apart from a standard free-running mode, it is possible to run the NZESM in a nudged mode, continuously modulated by observed meteorological conditions using the ERA-Interim reanalysis (Dee et al., 2011) and prescribed SST and sea ice by the HadISST dataset (Rayner et al., 2003). A nudged run can be useful for comparison with instantaneous values of observational data taken during the simulated period, as opposed to long-term statistics. The model fields can be exported at arbitrary intervals down to the model time step of 20 minutes.

1.2.2 CFMIP Observation Simulator Package

COSP (Bodas-Salcedo et al., 2011) is a satellite instrument simulator package for atmospheric model evaluation developed as part of the CFMIP (Bony et al., 2011), whose purpose is to generate pseudo-measurements and statistics from model fields, which can then be compared to real measurements. A direct comparison without a simulator is often not possible due to a limited field of view (FOV) of the instrument and attenuation by atmospheric constituents (clouds, aerosols, gases), which is a wavelength-dependent process. COSP was utilised in evaluation of GCMs in the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al., 2012). COSP contains multiple simulators of different instruments: ISCCP, MODIS, MISR, CloudSat, CALIPSO, a MilliMeter-wavelength Cloud Radar (MMCR)/Ka-Band ARM Zenith Radar (KAZR) ground-based radar and the Radiative Transfer for Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOSV) (RTTOV). Notably, radar observations are simulated by the QuickBeam simulator (Haynes et al., 2007) and lidar (CALIPSO) observations are simulated by the Active Remote Sensing Simulator (ACTSIM) (Chepfer et al., 2008). In general, these may need to be tuned for any particular instrument being simulated due to different wavelengths, signal modulation, view and error characteristics. COSP allows for comparison of instrument quantities (backscatter, radar reflectivity), or derived products (cloud top/base, cloud phase, ...) between the model and observations. An exact co-located comparison is limited by the relatively low spatial and temporal resolution of GCMs, and pseudo-observations need to be made on subcolumns generated by a cloud generator. Algorithms for calculating derived products are generally not available, and datasets such as CALIPSO-GOCCP were developed for the purpose of comparison of equivalent quantities from observations and the simulator (Chepfer et al., 2010). COSP can be run either online (inside the model) or offline, when fields from a model are provided to COSP after completing the sim-

ulation. Running the simulator offline allows for rapid modification and testing of code. Cloud overlap in COSP is treated by the Subgrid Cloud Overlap Profile Sampler (SCOPS) (Webb et al., 2001), which generates subcolumns based on the grid cell cloud fraction and precipitation fluxes. Either random or maximum–random cloud overlap (Geleyn and Hollingsworth, 1979; Ritter and Geleyn, 1992) is assumed, whereby cloud in the adjacent layers overlaps maximally, and cloud separated by clear layers overlaps randomly.

ACTSIM is a lidar simulator integrated in COSP (Chepfer et al., 2008; Chiriaco et al., 2006). In the current implementation it simulates a spaceborne lidar with a wavelength of 532 nm, aimed at simulating the CALIOP instrument on CALIPSO. The simulator produces attenuated volume backscatter coefficient, which can be compared directly with measurements from a lidar. Support for a ground-based ceilometer such as Lufft CHM 15k or Vaisala CL51 will require modification of ACTSIM. Firstly, the viewpoint from the ground means that the lidar signal passes through atmospheric layers in a reversed order relative to what is assumed for a spaceborne lidar. Secondly, wavelength of our instruments is different from CALIOP (1064 nm and 910 nm), which requires re-calculation of the Mie and Rayleigh scattering coefficients.

1.2.3 In situ observations in the Southern Ocean

In situ SO observations are essential for improving the model SO biases. A set of SO datasets have been collected by the University of Canterbury and partner organisations by deploying our instruments on a number of voyages of opportunity as well as conducting IOPs:

- *Aurora Australis* voyages to Casey, Davis and Mawson, Antarctica (2015–2016).
- Macquarie Island station (2016–2018).
- HMNZS *Wellington* voyages to the Ross Sea (2016).
- RV *Nathaniel B. Palmer* voyage NBP1704 to the Ross Sea (2017).
- RV *Tangaroa* voyages TAN1502, TAN1503 (2015), TAN1702 (2017) and TAN1802 (2018) to the Ross Sea, Chatham Islands, the Campbell Plateau and the Ross Sea, respectively.

The author participated on field measurements on the TAN7102 and TAN1802 voyages and the deployments on the HMNZS *Wellington* and the NBP1704. In addition to the datasets outlined above we have access to a set of ceilometer and lidar observations from the following land-based locations in Aotearoa/New Zealand:

- **Cass**, a deployment of a Vaisala CL51 ceilometer at a station in the Southern Alps, Aotearoa/New Zealand.
- **Lauder**, a dataset of a Sigma Space MiniMPL lidar and a Vaisala CL31 ceilometer observations at a station in Aotearoa/New Zealand made available by NIWA.
- **Christchurch**, a deployment of a Lufft CHM 15k and a Sigma Space MiniMPL on the Ernest Rutherford building of the University of Canterbury, Aotearoa/New Zealand.

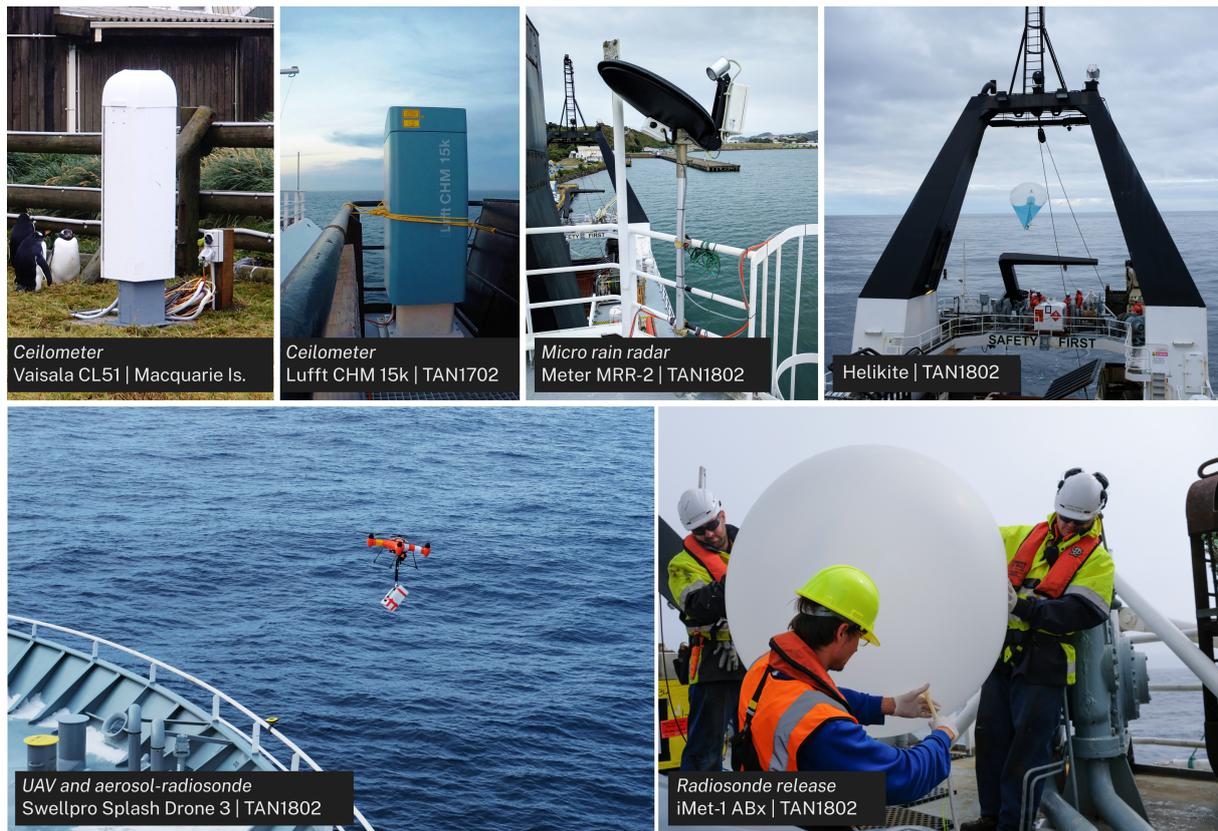


Figure 1.3 | Instruments deployed on Southern Ocean voyages and stations.

Observations collected on the voyages include remote sensing with ceilometers and lidars, a micro rain radar, radiosonde profiles, automatic weather station (AWS) data (temperature, relative humidity, wind, SST, radiometer, ...), unmanned aerial vehicle (UAV) and tethered balloon soundings and aerosol concentration (Fig. 1.3). Below, we briefly describe the instruments.

Lufft CHM 15k is an IR ceilometer operating at a single wavelength of 1064 nm, which makes it suitable for observation of cloud droplets and ice particles of similar size and boundary layer aerosol. The primary purpose of a ceilometer is observation of CBH, although other atmospheric features such as boundary layer height, visibility, precipitation and multiple cloud layers can be detected as well using a suitable algorithm. The primary measured quantity is attenuated volume backscattering coefficient β ($\text{km}^{-1}\text{sr}^{-1}$), which can also be used for a direct comparison with a ceilometer simulator. The instrument allows for an easy deployment in adverse conditions, such as on ships. The ceilometer records data in the NetCDF format (Rew et al., 2006), which makes it easy for processing by various data analysis tools. The averaging period of the instrument is 2 s. This provides spatiotemporal resolution much greater than a GCM. Averaging over longer time periods can be applied to improve the signal-to-noise ratio (SNR).

Vaisala CL51 is an IR ceilometer operating a wavelength of 910 nm. Similar to *Lufft CHM 15k*, it is suitable for observation of cloud droplets and ice particles. The averaging period is 6 s. The firmware contains onboard detection of multiple cloud layers and visibility by a standard algorithm and a Sky Condition Algorithm (SCA). Two-dimensional backscatter profiles are recorded in ASCII-encoded data files.

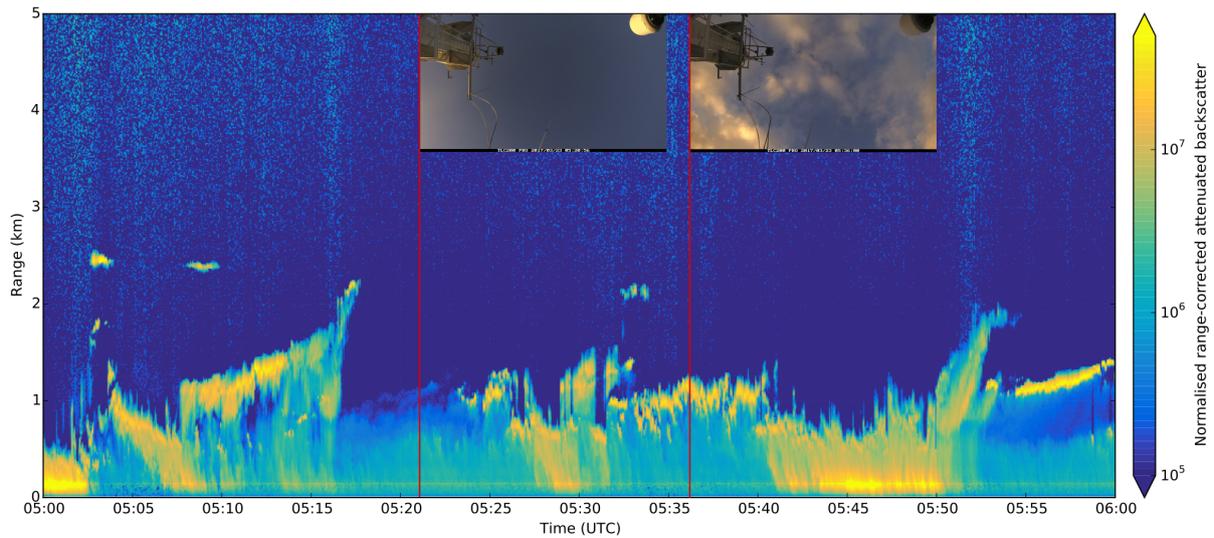


Figure 1.4 | A Lufft CHM 15k ceilometer volume backscattering coefficient profile plot and corresponding sky camera images collected on the TAN1702 voyage on 23 March 2017.

Metek MRR-2 is a micro rain radar, operating at a microwave frequency of 24.230 GHz, i.e. wavelength of 12.38 mm. The wavelength makes it suitable for observation of liquid and ice precipitation. This instrument can be used alongside the Vaisala CL51 instrument to detect period of time with precipitation and rain rate. A Metek MRR-2 was deployed on the TAN1702, TAN1802 and HMNZS *Wellington* voyages.

A sky camera was used as an ancillary instrument providing a visual perspective of the atmospheric conditions (type of cloud, fog, precipitation), as well as a primary instrument for determining cloud fraction. Our deployments included an off-the-shelf time lapse camera Brinno BCC200 (as a low-cost but satisfactory solution) and a fisheye-lens camera. The sampling period can be chosen from a wide range of values; we have determined that a 5-minute interval is sufficient. Figure 1.4 shows an example backscatter profile plot from the TAN1702 voyage, combined with sky camera images. In addition to CBH, these measurements provide a wealth of information about the cloud type, the vertical extent, optical thickness, precipitation, fog and boundary layer aerosol. Some of these, as well as satellite cloud observations in the Ross Sea region, were analysed by co-authored studies: Klekociuk et al. (2019a); Jolly et al. (2018); Hartery et al. (2020b,a).

Preliminary analysis of multiple voyage datasets indicates that low cloud below 2 km constitutes the majority of cloud in the summertime in the SO (Fig. 1.5). Preliminary results from the TAN1802 voyage also show a very high cloud fraction of 94% and a strong peak of boundary layer cloud below 1 km above sea level (ASL) (Fig. 1.6a), the predominance of stratus (52%) and stratocumulus (30%) clouds (Fig. 1.6b), predominantly near-zero SST (Fig. 1.6c) and near-surface air temperature below SST (Fig. 1.6d, e). These results suggest a cold boundary layer destabilised by relatively warm SST and subsequent formation of low stratus and stratocumulus cloud.

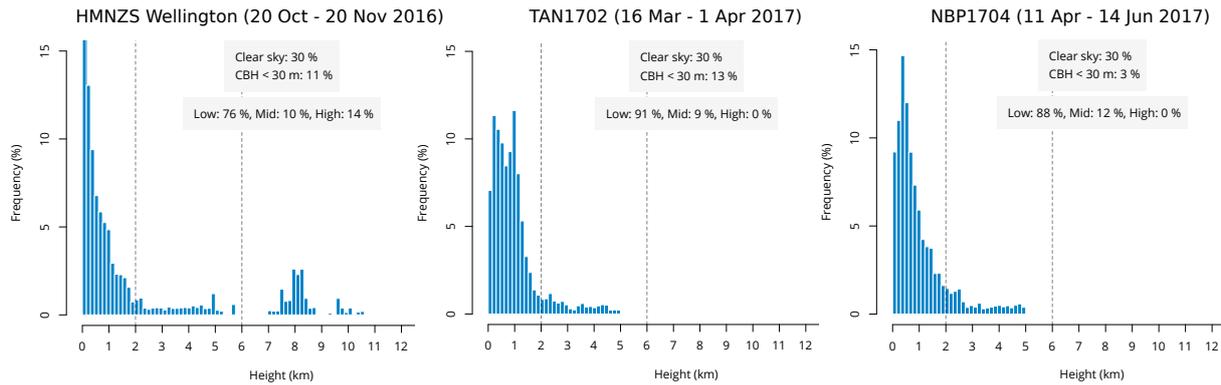


Figure 1.5 | Cloud base height distribution on the HMNZS *Wellington* 2016, TAN702 and NBP1704 voyages derived with a Lufft CHM 15k ceilometer observations (as determined by the instrument’s firmware).

1.2.4 Satellite radiation budget observations

Earth radiation budget (ERB) observations are central for assessment and development of GCMs. LEO satellite observations of TOA SW and longwave (LW) fluxes have been performed starting with the Nimbus satellite series in 1970s (Smith et al., 1977), followed by the Earth Radiation Budget Satellite (ERBS) and NOAA satellites in 1980s (Barkstrom, 1984), the Scanner for Radiation Budget (ScaRaB) project on Meteor-3 and Resurs-01/4 LEO satellites in 1990s (Kandel et al., 1994) and the CERES instruments on a number of LEO satellites from late 1990s to the present day (Wielicki et al., 1996). Geosynchronous satellite measurements have the advantage of continuous temporal sampling, but cannot provide a good angular resolution and observations at high polar latitudes. They have, however, been utilised as part of the Geostationary Earth Radiation Budget (GERB) project. The National Institute of Standards and Technology Advanced Radiometer (NISTAR) instrument on the Deep Space Climate Observatory (DSCOVR) satellite in L1 Lagrangian point provides continuous measurements of the sunlit part of the Earth (Khlopenkov et al., 2017). It has, however, not been used as extensively as earlier satellite observations. Global radiation balance is one of the most commonly adjusted properties of GCMs (Hourdin et al., 2017; Schmidt et al., 2017). It is vital for GCMs to simulate accurate spatiotemporal variability of the radiation budget, as deviations can cause shifts in circulation patterns such as the polar fronts and the inter-tropical convergence zone (ITCZ).

CERES are instruments measuring the ERB, deployed on multiple satellites: TRMM (1997–2015), Terra (2000–present), Aqua (2002–present), Suomi NPP (2011–present) and JPSS-1 (2017–present) (Damadeo and Heather, 2017). They are considered to provide the most reliable measurements of radiation budget, although they are limited by the necessity of temporal (diurnal) and angular interpolation (Smith et al., 2011). Recent version of the CERES Energy Balanced and Filled (EBAF) dataset (Edition 4.0) has been found to decrease clear-sky TOA SW flux in the SO region in January by up to 15 W m^{-2} in summer compared to the previous version (Loeb et al., 2017), which may affect previous results and should be taken into consideration in future analysis.

The Geostationary Earth Radiation Budget (GERB) project involves ERB instruments on Meteosat Second Generation (MSG) GEO satellites (Harries et al., 2005). Measurements began in 2002 on MSG-1 and continue to the present day. Both SW and LW bands are available. The

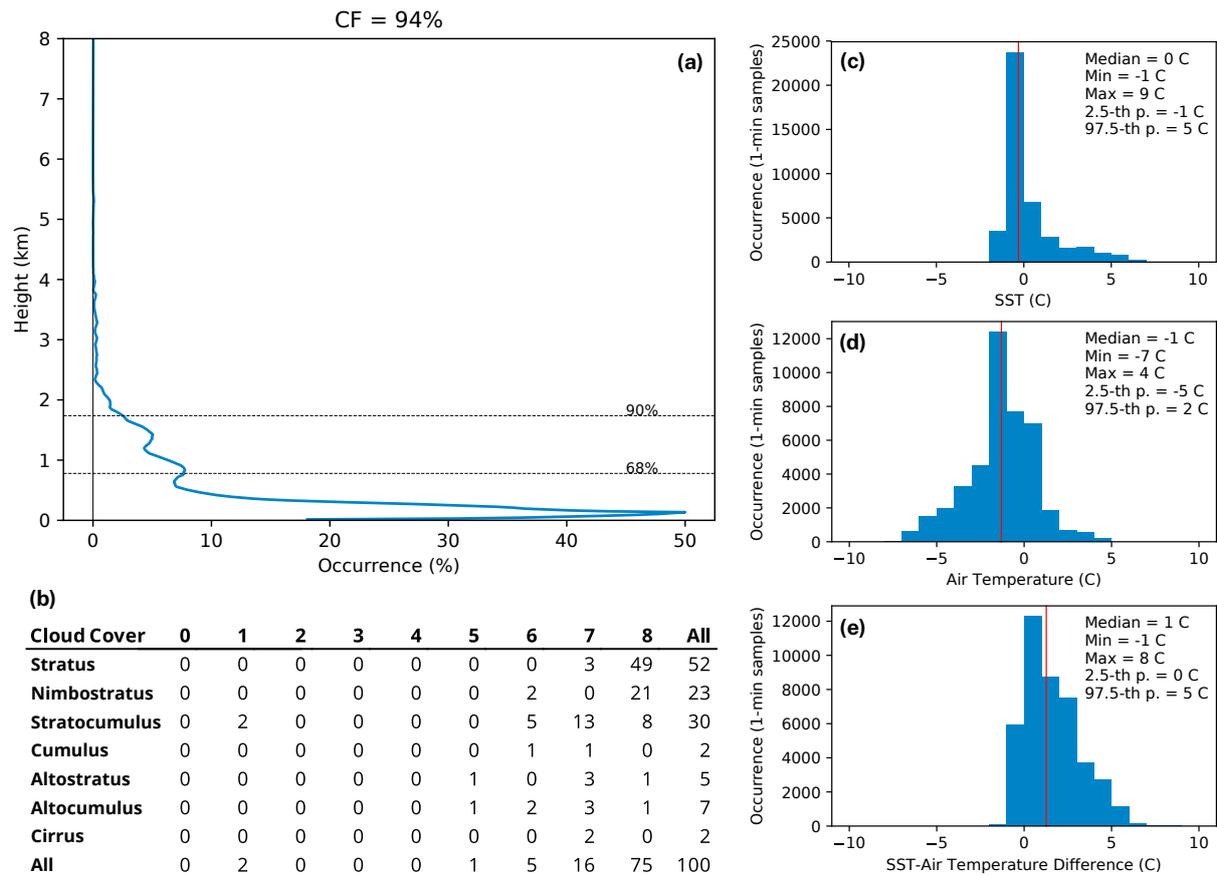


Figure 1.6 | Statistics calculated from observations collected on the TAN1802 voyage. **(a)** cloud occurrence as a function of height, the 68-th and 90-th percentiles and the total cloud fraction (CF) calculated from a Luftt CHM 15k ceilometer observations. **(b)** cloud type and cloud cover (octas) occurrence in % calculated from human observations. Histograms of **(c)** sea surface temperature (SST), **(d)** air temperature and **(e)** SST - air temperature calculated from the automatic weather station (AWS) data.

advantage of GERB over CERES is its continuous temporal and spatial coverage in its FOV (Sandford et al., 2003). However, it does not provide full spatial coverage (polar latitudes and longitudes outside of its FOV). GERB has been used for correction of CERES temporal interpolation (CERES geostationary method), resulting in difference in excess of 25 Wm^{-2} over marine stratus and land convection relative to uncorrected data (CERES-only) (Doelling et al., 2013). It is therefore important to consider the effect of temporal interpolation when comparing regional ERB with a GCM. Because high latitudes are not observed by GEO satellites, this correction cannot be done for latitudes over 60° . This is compensated by the high-revisit frequency of the CERES-carrying satellites at the poles.

1.2.5 Auxiliary software

As part of the observational data processing work we developed open source tools for transforming the native instrument data formats to the more commonly used NetCDF and HDF formats:

- `cl2nc`¹, a tool for converting Vaisala CL51 ceilometer data to NetCDF4.
- `mrr2c`², a tool for converting Metek MRR-2 radar data to HDF5.
- `mpl2nc`³, a tool for converting Sigma Space MiniMPL lidar data to NetCDF4 and applying dead time, overlap and afterpulse calibration.

These tools were made publicly available on the code collaboration network GitHub.

1.3 Outline of the thesis and author's contributions

This thesis consists of the Introduction (Chapter 1), three research chapters (Chapter 2, 3 and 4) and Conclusions and further work (Chapter 5). The three research chapters are accepted for publication (Chapter 2), in review (Chapter 3) and a manuscript in preparation (Chapter 4). The author of this thesis is the primary author of the three manuscripts. Chapter 2 provides a comprehensive literature review on the central topic of this thesis: model cloud biases in the SO. In Chapter 2 we evaluate SO cloud in a nudged run of the GA7.1 and MERRA-2 in comparison with a collection of SO voyage observations. In Chapter 3 we describe a new ground based lidar processing and simulator framework. In Chapter 4 we describe and evaluate an experimental run of the UM11.4 aimed at improving representation of boundary layer cloud in the SO. As of May 2020, Chapter 2 was accepted for publication in the Atmospheric Chemistry and Physics (ACP) (Kuma et al., 2020a), and Chapter 3 is in review in the Geoscientific Model Development Discussions (GMDD) (Kuma et al., 2020b). Co-authored published studies related to this thesis are: Jolly et al. (2018) (published in the ACP), Klekociuk et al. (2019a) (published in the Deep Sea Research Part II: Topical Studies in Oceanography), Hartery et al. (2020b) (published in the Journal of Geophysical Research: Atmospheres) and Hartery et al. (2020a) (submitted to the Geophysical Research Letters).

¹<https://github.com/peterkuma/cl2nc>.

²<https://github.com/peterkuma/mrr2c>.

³<https://github.com/peterkuma/mpl2nc>.

Chapter 2

Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations

Peter Kuma¹, Adrian J. McDonald¹, Olaf Morgenstern², Simon P. Alexander³, John J. Cassano⁴, Sally Garrett⁵, Jamie Halla⁵, Sean Hartery¹, Mike J. Harvey², Simon Parsons¹, Graeme Plank¹, Vidya Varma², and Jonny Williams²

¹School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand

²National Institute of Water and Atmospheric Research, Wellington, New Zealand

³Australian Antarctic Division, Kingston, Australia

⁴Cooperative Institute for Research in Environmental Sciences and Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado, US

⁵New Zealand Defence Force, Wellington, New Zealand

Abstract

Southern Ocean (SO) shortwave (SW) radiation biases are a common problem in contemporary general circulation models (GCMs), with most models exhibiting a tendency to absorb too much incoming SW radiation. These biases have been attributed to deficiencies in the representation of clouds during the austral summer months, either due to cloud cover or cloud albedo being too low. The problem has been the focus of many studies, most of which utilised satellite datasets for model evaluation. We use multi-year ship based observations and the CERES spaceborne radiation budget measurements to contrast cloud representation and SW radiation in the atmospheric component Global Atmosphere (GA) version 7.1 of the HadGEM3 GCM and the MERRA-2 reanalysis. We find that the prevailing bias is negative in GA7.1 and positive in MERRA-2. GA7.1 performs better than MERRA-2 in terms of absolute SW bias. Significant errors of up to 21 Wm^{-2} (GA7.1) and 39 Wm^{-2} (MERRA-2) are present in both models in the austral summer. Using ship-based ceilometer observations, we find low cloud below 2 km to be predominant in the Ross Sea and the Indian Ocean sectors of the SO. Utilising a novel surface lidar simulator developed for this study, derived from an existing COSP-ACTSIM spaceborne lidar simulator, we find that GA7.1 and MERRA-2 both underestimate low cloud and fog occurrence relative to the ship observations on average by 4–9% (GA7.1) and 18% (MERRA-2). Based on radiosonde observations, we also find the low cloud to be strongly linked to boundary-layer atmospheric stability and the sea surface temperature. GA7.1 and MERRA-2 do not represent the observed relationship between boundary layer stability and clouds well. We find that MERRA-2 has a much greater proportion of cloud

liquid water in the SO in austral summer than GA7.1, a likely key contributor to the difference in the SW radiation bias. Our results suggest that subgrid-scale processes (cloud and boundary layer parametrisations) are responsible for the bias, and that in GA7.1 a major part of the SW radiation bias can be explained by cloud cover underestimation, relative to underestimation of cloud albedo.

2.1 Introduction

Clouds are considered one of the largest sources of uncertainty in estimating global climate sensitivity (Boucher et al., 2013; Flato et al., 2014; Bony et al., 2015). Clouds over oceans are especially important for determining the radiation budget due to the low albedo of the sea surface compared to land. Over the Southern Ocean (SO), cloud cover is very high at over 80%, with boundary-layer clouds being particularly common (Mace et al., 2009). Excess downward shortwave (SW) radiation in general circulation models (GCMs), with a bias over the SO of up to 30 Wm^{-2} , is a problem well-documented by Trenberth and Fasullo (2010) and Hyder et al. (2018), and has been the subject of many studies. Bodas-Salcedo et al. (2014) evaluated the SW bias in a number of GCMs and found that a strong SW bias is a very common feature, leading to increased sea surface temperature (SST) in the SO and corresponding biases in the storm track position. Trenberth and Fasullo (2010) note that a poor representation of clouds might lead to unrealistic climate change projections in the Southern Hemisphere. The SW bias has also been linked to large-scale model problems such as the double-Intertropical Convergence Zone (Hwang and Frierson, 2013), biases in the position of the midlatitude jet (Ceppi et al., 2012) and errors in the meridional energy transport (Mason et al., 2014). Bodas-Salcedo et al. (2012) studied the SO SW bias in the context of the Global Atmosphere (GA) 2.0 and 3.0 models and found that mid-topped and stratocumulus clouds are the dominant contributors to the bias.

Due to its extent and magnitude, the SW radiation bias is believed to limit accuracy of the models, especially for modelling the Southern Hemisphere climate. A model based on the Hadley Centre Global Environmental Model version 3 (HadGEM3) is currently used in New Zealand for assessing future climate (Williams et al., 2016). In this paper we evaluate the atmospheric component of HadGEM3, GA7.1 (Walters et al., 2017) and the reanalysis Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) using observations collected in the SO on a number of voyages. Ship-based atmospheric observations in the SO provide a unique view of the atmosphere not available via any other means. Boundary layer observations by satellite instruments are limited by the presence of an almost continuous cloud cover, potentially obscuring the view of low level clouds. The frequently used active instruments CloudSat (Stephens et al., 2002) and Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) (Winker et al., 2010) are both of limited use when observing low level, thick or multi-layer cloud: CloudSat is affected by surface clutter below approximately 1.2 km (Marchand et al., 2008) and the CALIPSO lidar signal cannot pass through thick cloud. Likewise, passive instruments and datasets such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Salomonson et al., 2002) and the International Satellite Cloud Climatology Project (ISCCP) (Rossow and Schiffer, 1999) can only observe radiation scattered or emitted from the cloud top of optically thick clouds. Therefore, one can accurately identify the cloud top height or cloud top pressure with satellite instruments, but

not always the cloud base height (CBH) or the vertical profile of cloud, although there has been some recent progress on deriving CBH statistically from CALIPSO measurements (Mülmenstädt et al., 2018). Ship-based measurements therefore provide valuable extra information.

Multiple explanations of the SW radiation bias have been proposed: cloud underestimation in the cold sectors of cyclones (Bodas-Salcedo et al., 2014), cloud–aerosol interaction (Vergara-Temprado et al., 2018), cloud homogeneity representation (Loveridge and Davies, 2019a), lack of supercooled liquid (cloud liquid at air temperature below 0°C) (Kay et al., 2016; Bodas-Salcedo et al., 2016) and the “too few, too bright” problem (Nam et al., 2012; Klein et al., 2013; Wall et al., 2017). Each model can exhibit the bias for a different set of reasons, and results from one model evaluation therefore do not necessarily explain biases in all other models (Mason et al., 2015). The use of SO voyage data for atmospheric model evaluation is not new, and has recently been used by Sato et al. (2018) to evaluate the impact of SO radiosonde observations on the accuracy of weather forecasting models. Klekociuk et al. (2019a) contrasted SO cloud observations with the ECMWF Interim reanalysis (ERA-Interim) and the Antarctic Mesoscale Prediction System–Weather Research and Forecasting Model (AMPS-WRF) (Powers et al., 2012), and found that these models underestimate the coverage of the predominantly low cloud. Protat et al. (2017) compared ship-based 95 GHz cloud radar measurements at 43–48°S in March 2015 with the Australian Community Climate and Earth-System Simulator (ACCESS) NWP model, a model related to HadGEM3, and found low cloud peaking at 80% cloud cover, which was underestimated in the model. The clouds were also more spread out vertically (especially due to “multilayer” situations defined as co-occurrence of cloud below and above 3 km) and more likely to have intermediate cloud fraction rather than very low or very high cloud fraction. Previous studies have documented that supercooled liquid is often present in the SO cloud in the austral summer months (Morrison et al., 2011; Huang et al., 2012; Chubb et al., 2013; Huang et al., 2016; Bodas-Salcedo et al., 2016; Jolly et al., 2018; Listowski et al., 2019) and is linked to SO SW radiation biases in GCMs, which underestimate the amount of supercooled liquid in clouds in favour of ice. Warm clouds generally reflect more SW radiation than cold clouds containing the same amount of water (Vergara-Temprado et al., 2018). In particular, Kay et al. (2016) reported a successful reduction of SO absorbed SW radiation in the Community Atmosphere Model version 5 (CAM5) by decreasing the shallow convection ice detrainment temperature and thereby increasing the amount of supercooled liquid cloud.

Two common techniques used for model cloud evaluation have been cloud regimes (Williams and Webb, 2009; Haynes et al., 2011; Mason et al., 2014, 2015; McDonald et al., 2016; Jin et al., 2017; McDonald and Parsons, 2018; Schuddeboom et al., 2018, 2019) and cyclone compositing (Bodas-Salcedo et al., 2012; Williams et al., 2013; Bodas-Salcedo et al., 2014, 2016; Williams and Bodas-Salcedo, 2017), both of which link the SW radiation bias to specific cloud regimes and cyclone sectors. We use simple statistical techniques, rather than sophisticated classification or machine learning algorithms, the advantage of which is easier interpretation for the purpose of model development.

We first assess the magnitude of the top of the atmosphere (TOA) SO SW radiation bias in a nudged run of GA7.1 (“GA7.1N”) and MERRA-2 with respect to the Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) and CERES Synoptic (SYN) products (Section 2.5.1). This allows us to identify the underlying magnitude of the SW bias and how this

Table 2.1 | Table of voyages. The table lists voyages analysed in this study. Listed is the voyage name (Voyage), which is the official name of the voyage or an abbreviation for the purpose of this study, ship name (Ship), organisation (Org.), start and end dates of the voyage (Start, End), number of days spent at sea (Days), target region of the SO (Region), maximum and minimum geographical coordinates of the voyage track (Lat., Lon.).

Voyage	Ship	Org.	Start	End	Days	Region	Lat.	Lon.
TAN1502	RV <i>Tangaroa</i>	NIWA	2015-01-20	2015-03-12	51	Ross Sea	41°S–75°S	162°E–174°W
TAN1802	RV <i>Tangaroa</i>	NIWA	2018-02-08	2018-03-21	41	Ross Sea	41°S–74°S	170°E–175°W
HMNZSW16	HMNZS <i>Wellington</i>	RNZN	2016-11-20	2016-12-20	20	Ross Sea	36°S–68°S	166°E–180°E
NBP1704	RV <i>Nathaniel B. Palmer</i>	NSF	2017-04-11	2017-06-13	63	Ross Sea	53°S–78°S	163°E–174°W
AA15 (AA V1–V3)	<i>Aurora Australis</i>	AAD	2015-10-22	2016-02-22	123	Indian O. sector	42°S–69°S	62°E–160°E

might change based on the ship track sampling pattern. We then evaluate cloud occurrence in GA7.1N and MERRA-2 relative to the SO ceilometer observations and compare SO radiosonde observations with pseudo-radiosonde profiles derived from the models (Sections 2.5.2 and 2.5.3). Lastly, we look at zonal plots of potential temperature, humidity, cloud liquid and ice content in GA7.1N and MERRA-2 to show how these models differ in their atmospheric stability and representation of clouds (Section 2.5.4). Our aim is to identify how differences between GA7.1N and MERRA-2 can explain the TOA outgoing SW radiation bias, assuming misrepresentation of clouds is the major contributor to the bias.

2.2 Datasets

We used an observational dataset of ceilometer and radiosonde data comprising multiple SO voyages (Section 2.2.1), GA7.1N atmospheric model simulations (Section 2.2.2) and the MERRA-2 reanalysis (Section 2.2.3). Later in the text, we will refer to GA7.1N and MERRA-2 together as “the models”, even though MERRA-2 is more specifically a reanalysis. CERES satellite observations (Wielicki et al., 1996) were also used as a reference for TOA outgoing SW radiation and an National Snow and Ice Data Center (NSIDC) satellite-based dataset (Maslanik and Stroeve, 1999) was used as an auxiliary dataset for identifying sea ice.

2.2.1 Ship observations

We use ship-based ceilometer and radiosonde observations made in the SO on 5 voyages between 2015 and 2018 (Table 2.1 and Fig. 2.1):¹

- 2015 TAN1502 voyage of the NIWA ship RV *Tangaroa* from Wellington, New Zealand to the Ross Sea.
- 2015–2016 voyages (V1–V3) of the Australian Antarctic Division (AAD) icebreaker *Aurora Australis* from Hobart, Australia to Mawson, Davis, Casey and Macquarie Island (“AA15”).
- 2016 Royal New Zealand Navy (RNZN) ship HMNZS *Wellington* voyages (“HMNZSW16”).
- 2017 NBP1704 voyage of the NSF icebreaker RV *Nathaniel B. Palmer* from Lyttelton, New Zealand to the Ross Sea.

¹The voyage name pattern is a 2–6 character ship name followed by a 2 digit year and a 2 digit sequence number. TANxxxx and NBPxxxx are official voyage names, while HMNZSW16 and AA15 are names made for the purpose of this study.

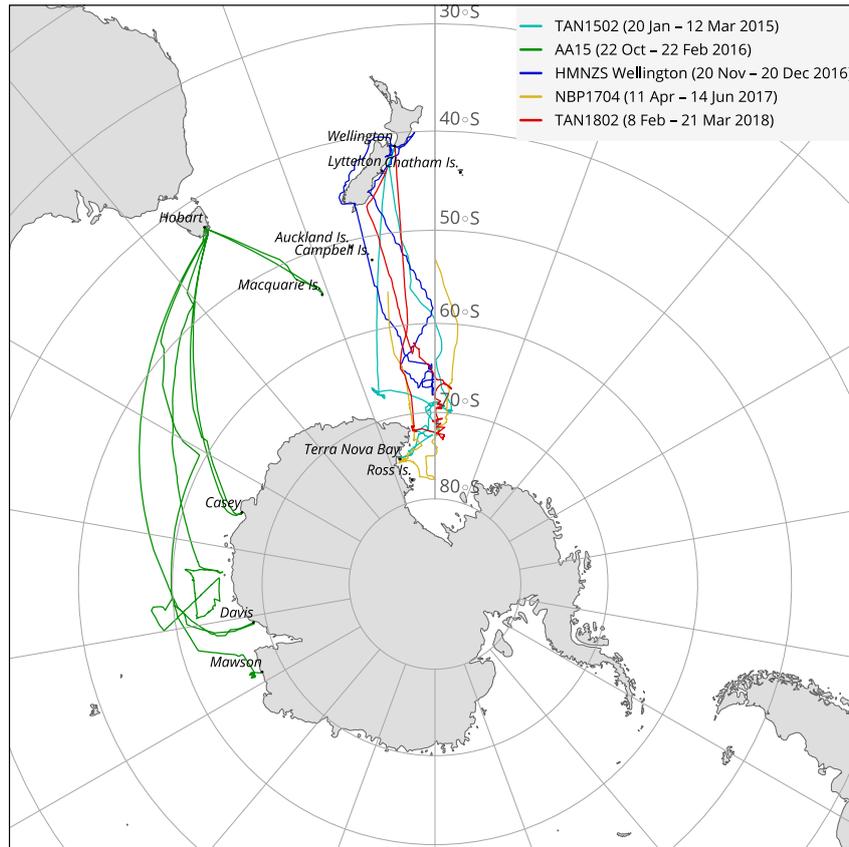


Figure 2.1 | Map showing tracks of voyages used in this study. The ship observational dataset comprises 5 voyages between 2015 and 2018, spanning months from November to June and latitudes between 40°S and 78°S , of which data between 50°S and 70°S are used in this study.

- 2018 TAN1802 voyage of RV *Tangaroa* from Wellington to the Ross Sea (Hartery et al., 2019).

Together, these voyages cover latitudes between 41 and 78°S and the months of November to June inclusive. A total of 298 days of observations were collected. Geographically, the voyages mostly cover the Ross Sea sector of the SO, with only AA15 covering the Indian Ocean sector (Fig. 2.1). This sampling emphasises the Ross Sea sector over other parts of the SO, although the SO SW radiation bias is present at all longitudes in the SO (Section 2.5.1), affected by the atmospheric circulation (Jones and Simmonds, 1993; Sinclair, 1994, 1995; Simmonds and Keay, 2000; Simmonds et al., 2003; Simmonds, 2003; Hoskins and Hodges, 2005; Hodges et al., 2011). The voyage observations were performed using a range of instruments (described below). Table 2.2 details which instruments were deployed on each voyage.

The primary instruments were the Lufft CHM 15k and Vaisala CL51 ceilometers. A ceilometer is an instrument which typically uses a single-wavelength laser to emit pulses vertically into the atmosphere and measures subsequent backscatter resolved on a large number of vertical levels based on the timing of the retrieved signal (Emeis, 2010). Depending on the wavelength, the emitted signal interacts with cloud droplets, ice crystals and precipitation by Mie scattering, and to a lesser extent with aerosol and atmospheric gases by Rayleigh scattering (Bohren and Huffman,

Table 2.2 | Table of deployments. The table cells indicate if data from a given instrument (row) was available from a voyage (column).

Instrument/Voyage	AA15	TAN1502	HMNZSW16	NBP1704	TAN1802
Lufft CHM 15k			✓	✓	✓
Vaisala CL51	✓	✓			
iMet radiosondes					✓
Radiosondes (other)				✓	

1998). The signal is quickly attenuated in thick cloud and therefore it is normally not possible to observe mid and high level parts of such a cloud, or a multi-layer cloud. The main derived quantity determined from the backscatter is CBH, but it is also possible to apply a cloud detection algorithm to determine cloud occurrence by height. The range-normalised signal is affected by noise which increases with the square of range. A major source of noise is solar radiation which causes a diurnal variation in noise levels (Kotthaus et al., 2016). Due to signal attenuation and noise ceilometers cannot measure clouds obscured by a lower cloud, and therefore cannot be used for 1:1 comparison with model clouds without using a lidar simulator, which accounts for this effect (Chepfer et al., 2008). The Lufft CHM 15k ceilometer operates in the near-infrared spectrum at 1064 nm, measuring lidar backscatter up to a maximum height of 15 km, producing 1024 regularly spaced bins (about 15 m resolution). The sampling rate of the instrument is 2 s. The Vaisala CL51 ceilometer operates in the near-infrared spectrum at 910 nm. The sampling rate of the instrument is 2 s and range is 7.7 km, producing 770 regularly spaced bins (10 m resolution).

Radiosonde observations were performed on the TAN1802 and NBP1704 voyages south of 60°S. Temperature, pressure, relative humidity and Global Navigation Satellite System (GNSS) coordinates (from which wind speed and direction are derived) were retrieved to altitudes of about 10–20 km, terminated by a loss of radio communication or balloon burst.

On the TAN1802 voyage we used iMet-1 ABx radiosondes, measuring pressure, air temperature, relative humidity and GNSS coordinates of the sonde (from which wind speed and direction are derived). The sondes were launched three times per day at about 8:00, 12:00 and 20:00 UTC on 100 g Kaymont weather balloons. They reached a typical altitude of 10–20 km, and then terminated by balloon burst or loss of radio communication. We used 10 s resolution profiles generated by the vendor-supplied iMetOS-II control software for further processing.

Automatic weather station (AWS) data were available on the TAN1502, TAN1802 and NBP1704 voyages. These included variables such as air temperature, pressure, sea surface temperature, wind speed and wind direction. Voyage track coordinates were obtained from the ships' GNSS receivers.

2.2.2 HadGEM3

HadGEM3 (Walters et al., 2017) is a general circulation model developed by the UK Met Office and the Unified Model Partnership. It can be used in a "nudging" (Telford et al., 2008) mode, in which winds and potential temperature are relaxed towards the ERA-Interim reanalysis (Dee et al., 2011). The Met Office Global Atmosphere 7.1 (GA7.1) is the atmospheric component of HadGEM3 (Walters et al., 2017), based on the Unified Model (UM) version 11.0.

The model runs used the HadISST sea surface temperature dataset (Rayner et al., 2003) as lateral boundary conditions. The nudged simulations represent atmospheric dynamics as determined by observations. The model was run on a $1.875^\circ \times 1.25^\circ$ (longitude \times latitude) "N96" resolution grid, which corresponds to a horizontal resolution of about 100×140 km at 60° S and 85 vertical levels. The model output fields were sampled every 6 hours (instantaneous) and daily (mean). In our analysis we used a nudged run of GA7.1 ("GA7.1N") between years 2015 and 2018, corresponding to the ship observations.

2.2.3 MERRA-2

Modern-Era Retrospective analysis for Research and Applications (MERRA-2) is a reanalysis provided by the NASA Global Modelling and Assimilation Office (Gelaro et al., 2017). The reanalysis was chosen for its contrasting results of TOA outgoing SW radiation bias in the SO compared to GA7.1. As shown later (Fig. 2.3), its bias is positive rather than negative, when CERES is used as a reference.

We used the following products (Bosilovich et al., 2015):

- 1-hourly average Radiation Diagnostics (product "M2T1NXRAD.5.12.4")
- 3-hourly instantaneous Assimilated Meteorological Fields (product "M2I3NVASM.5.12.4")
- 1-hourly instantaneous Single-Level Diagnostics (product "M2I1NXASM.5.12.4")
- 3-hourly average Assimilated Meteorological Fields (product "M2T3NVASM.5.12.4")
- 1-hourly average Single Level Diagnostics (product "M2T1NXSLV.5.12.4")

We used the "Radiation Diagnostics" in TOA outgoing SW radiation evaluation (Section 2.5.1), the instantaneous "Assimilate Meteorological Fields" and "Single-Level Diagnostics" products to generate simulated ceilometer profiles and pseudo-radiosonde profiles (Section 2.5.2 and 2.5.3), and the average "Assimilate Meteorological Fields" and "Single-Level Diagnostics" to generate zonal plane plots of thermodynamic and cloud fields (Section 2.5.4). The 4-dimensional MERRA-2 fields were provided on pressure and model levels. For our analysis we chose to use the model-level products (72 levels) due to their higher vertical resolution compared to pressure-level products. The analysed time period of MERRA-2 data was 2015–2018.

2.2.4 CERES

The Clouds and the Earth's Radiant Energy System (CERES) is a set of low Earth orbit (LEO) satellite instruments and a dataset of SW and longwave (LW) radiation observations (Loeb et al., 2018; Doelling et al., 2016). The CERES instruments (called FM1 to FM6) provide a continuous record of observations since the first deployment on the Tropical Rainfall Measuring Mission (TRMM) satellite in 1997 (Simpson et al., 1996), and have been flown on Terra, Aqua (Parkinson, 2003), the Suomi NPOESS Preparatory Project (Suomi NPP) and Joint Polar Satellite System-1 (JPSS-1) (Goldberg et al., 2013) satellites since. Currently CERES is considered the best available global Earth radiation datasets, and is often used as the primary dataset for GCM tuning and validation (Schmidt et al., 2017; Hourdin et al., 2017). We used the following CERES products in our analysis:

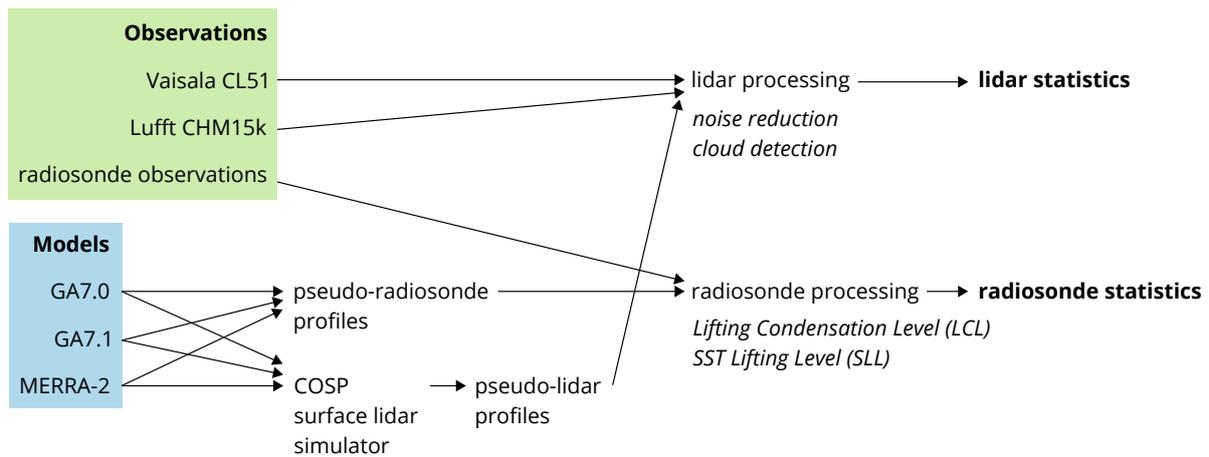


Figure 2.2 | Schematic of the processing pipeline utilised in this study to produce lidar and radiosonde statistics from observations and model data.

- CERES SYN1deg-Day Edition 4A (configuration code 406406 and 407406) product of daily average radiation ("CERES SYN").
- CERES EBAF-TOA Edition 4.1 (CERES_EBAF_Ed4.1) product of monthly energy-balanced average radiation ("CERES EBAF").

Due to the sun-synchronous orbits of the LEO satellite platforms, the Flight Model (FM) instruments of CERES do not capture the full diurnal variation of radiation. The EBAF and SYN1deg products are adjusted for diurnal variation by using 1-hourly geostationary satellite observations between 60°S and 60°N, and use an algorithm to account for changing solar zenith angle and diurnal land heating. The CERES EBAF-TOA Edition 4.1 product is a Level 3B product, which means it has been globally balanced by ocean heat measurements using the Argo network (Roemmich et al., 2009; Roemmich and Team, 2009).

2.2.5 NSIDC sea ice concentration

We used the Near-Real-Time Defense Meteorological Satellite Program (DMPS) Special Sensor Microwave Imager/Sounder (SSMIS) Daily Polar Gridded Sea Ice Concentrations, Version 1 product (NSIDC-0081) (Maslanik and Stroeve, 1999) provided by the National Snow and Ice Data Center (NSIDC) to classify observations into those affected and unaffected by sea ice. The sea ice concentration product has a resolution of 25×25 km. We used a cutoff value of 15% of sea ice concentration for the binary classification of sea ice, in line with previous studies (Comiso and Nishio, 2008).

2.3 Methods

2.3.1 Lidar simulator

CFMIP Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011), a set of instrument simulators developed by the Cloud Feedback Model Intercomparison Project (CFMIP), was ex-

tended with a surface lidar simulator and used to produce virtual lidar measurements from model fields (Kuma et al., 2020b). Resampling, noise reduction and cloud detection were performed on observational and (where applicable) model lidar data in a consistent way to reduce structural uncertainty (see Section 2.3.2). The schematic in Fig. 2.2 shows the processing pipeline utilised in this study.

COSP was originally developed as a satellite simulator package whose aim is to produce virtual satellite (and more recently ground-based) observations from atmospheric model fields in order to improve comparisons of model output with observations (Bodas-Salcedo et al., 2011). This approach is required because physical quantities derived from satellite observations generally do not directly correspond to model fields. COSP accounts for the limited view of the satellite instrument by calculating radiative transfer through the atmosphere, i.e. attenuation by hydrometeors and air molecules and backscattering. COSP comprises multiple instrument simulators, such as MODIS, ISCCP, MISR, CALIPSO and CloudSat. It has been used extensively by previous studies of model cloud, for example by Kay et al. (2012), Franklin et al. (2013), Klein et al. (2013), Williams and Bodas-Salcedo (2017), Jin et al. (2017), and Schuddeboom et al. (2018). COSP is planned to be used in the upcoming Coupled Model Intercomparison Project Phase 6 (CMIP6) (Webb et al., 2017).

For our analysis, we have developed a ground-based lidar simulator by modifying the COSP ACTSIM spaceborne lidar simulator (Chiriaco et al., 2006) (see the Code and data availability section at the end of the document). This required reversing of the vertical layers, as the surface lidar looks from the surface up rather than down from space to the surface, and changing the radiation wavelength affecting Mie scattering by cloud droplets and Rayleigh scattering by air molecules. In this paper we present only a brief description of the surface lidar simulator, with a more complete description planned in an upcoming paper. The new simulator is made available as part of the Automatic Lidar and Ceilometer Framework (ALCF) at <https://alcf-lidar.github.io>.

The recently introduced COSP version 2 (Swales et al., 2018) added support for a surface lidar simulator, although we believe our implementation, developed before COSPv2 was available, is more complete in the present context due to its treatment of Mie scattering at wavelengths other than 532 nm (the wavelength of the CALIPSO lidar). Previously, a surface lidar simulator based on COSP has been used by Chiriaco et al. (2018) and Bastin et al. (2018). A ground-based radar simulator in COSP has also recently been implemented (Zhang et al., 2018).

The surface lidar simulator takes model cloud liquid and ice mixing ratios, cloud fraction and thermodynamic profiles as the input, and calculates vertical profiles of attenuated backscatter. This can be done either by running the simulator "online" within the model code or "offline" on the model output. We used the offline approach in our analysis.

2.3.2 Lidar data processing

Lidar data in this study came from two different instruments: Lufft CHM 15k and Vaisala CL51 ceilometers and the lidar simulator. These instruments use different output formats, wavelengths, sampling rates and range bins, as previously noted. Backscatter and derived fields such as CBH are provided in the firmware generated data products, but the backscatter is uncalibrated and the derived fields such as cloud detection are based on instrument-dependent algorithms. Therefore,

we performed consistent subsampling, noise reduction and cloud detection on data from both instruments, and applied the same methods to the lidar simulator output. As part of the processing we developed a publicly available tool called `cl2nc` ("CL to NetCDF") for converting the Vaisala CL51 ceilometer data format to NetCDF (see the Code and data availability section at the end of the document).

2.3.2.1 Calibration

The backscatter profiles produced by the Lufft CHM 15k and Vaisala CL51 ceilometers are not calibrated to physical units, even though they are expressed in $\text{m}^{-1}\text{sr}^{-1}$. To calibrate these backscatter fields we used the method described by O'Connor et al. (2004). This method uses the lidar ratio (LR) to calculate a calibration factor based on a known value of the LR in fully scattering cloudy scenes (18.8 ± 0.8 sr), such as thick stratocumulus clouds, which are common over the SO. We applied this technique by using visually identified scenes and choosing a calibration factor which achieves the known value. Due to the nature of the conditions (LR can be highly variable even in thick cloud scenes), the calibration is likely accurate to only about 50% of the backscatter value. We do not expect this to have a serious impact on the accuracy of cloud detection completed in this study, largely because the predominantly low cloud tends to cause backscatter orders of magnitude greater than clear air, and because of the very large differences in cloud occurrence between the observations and models.

2.3.2.2 Subsampling, noise removal and cloud detection

In order to simplify further processing and increase the signal-to-noise ratio, we subsampled the ceilometer observations at a sampling rate of 5 minutes by averaging multiple profiles, and vertically averaging on regularly spaced 50 m bins. We expect that in most cases cloud was almost constant on this time and vertical scale, and therefore we were not averaging together different cloud types or clear and cloudy profiles. At the same time as subsampling, we performed noise removal by estimating the noise distribution (mean and standard deviation) based on returns in the uppermost range bins (i.e. 300 samples over 5 min when sampling rate was 2 s), and subtracting the range-scaled noise mean from the backscatter. We then used the range-scaled noise standard deviation (σ) for cloud detection: a bin was considered cloudy if the calibrated backscatter minus 3σ exceeded $20 \times 10^{-6} \text{ m}^{-1}\text{sr}^{-1}$. This threshold was chosen subjectively so that cloud was visually well separated from other features, such as boundary-layer aerosol and noise on backscatter profile plots. The same threshold was used on both the observations and output from the COSP surface lidar simulator and thus should cause little bias.

2.3.2.3 Model lidar data processing

We used the same sampling rate (5 min) and model levels as range bins on the surface lidar simulator output. For each vertical profile we used model data at the same location as the ship and the same time relative to the start of the year. Model data were selected using nearest-neighbour interpolation. The model resolution is lower than the distance travelled by the ship in 5 minutes, therefore the same model data were used multiple times to generate consecutive profiles. However,

we also used the SCOPS (Webb et al., 2001) subcolumn generator included in COSP to generate 10 random samples of cloud for each profile based on cloud fraction and the maximum/random cloud overlap assumption (Bodas-Salcedo, 2010). The lidar simulator processes each sample individually. The resulting cloud occurrence is calculated as the average of the 10 samples. The lidar simulator does not generate noise, and therefore we did not perform any noise removal on the simulated profiles, but we used the same threshold of $20 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$ and vertical bins of 50 m for detecting cloud (as used on the observations). For the MERRA-2 cloud occurrence analysis, we applied the lidar simulator on the 3-hourly instantaneous Assimilated Meteorological Fields (M2I3NVASM.5.12.4) product.

2.4 Spatiotemporal subsets investigated

Because our observational dataset does not span the entire geographical area of the SO and all months of the year, and the atmospheric conditions in the SO are geographically variable, we subset the datasets into a number of geographical regions by latitude and time periods by season. The geographical regions investigated are $50\text{--}75^\circ\text{S}$ by 5 degrees of latitude, and the temporal periods investigated are austral summer of December, January, February (DJF) and autumn months of March, April, May (MAM).

We do not use data from $70\text{--}75^\circ\text{S}$ and $50\text{--}55^\circ\text{S}$ in all parts of the analysis. The data from $70\text{--}75^\circ\text{S}$ are likely affected by circulation induced by land near the Ross Sea (Coggins et al., 2014), and therefore may not be representative of the SO in general. This decision builds on the analysis detailed in Jolly et al. (2018) which shows a significant gradient in cloud properties between the Ross Ice Shelf and the Ross Sea and strong influences associated with synoptic conditions. The data from $50\text{--}55^\circ\text{S}$ were relatively sparse (the ships spent relatively little time passing through this latitudes). Radiosonde observations were only available south of 60°S .

There is likely temporal variability present within the DJF and MAM time periods, but we decided to limit the number of temporal subsets to maintain a reasonable quantity of observations in each subset. The magnitude of the SO TOA outgoing SW radiation bias is primarily modulated by incoming solar radiation, which is the highest in DJF. The voyages do not uniformly cover all geographical regions or time periods, with the largest number of observations in the Ross Sea sector south of New Zealand (TAN1802, TAN1502, HMNZSW16, NBP1704), followed by the Indian Ocean sector south of Western Australia (AA15). Temporally, the voyage observations mostly cover summer to autumn months of the year.

2.5 Results

2.5.1 Shortwave radiation balance

Figure 2.3 shows TOA outgoing SW radiation in CERES, GA7.1 and MERRA-2. We present this panel plot in order to evaluate how well GA7.1N and MERRA-2 are performing in terms of SW radiation bias in the SO relative to CERES. This analysis assumes that CERES is a good observational reference, although it is affected by errors of lower order of magnitude (2.5 Wm^{-2} "regional monthly uncertainty" (Loeb et al., 2018, Sect. 4a)). The plots reveal relatively zonally symmetric

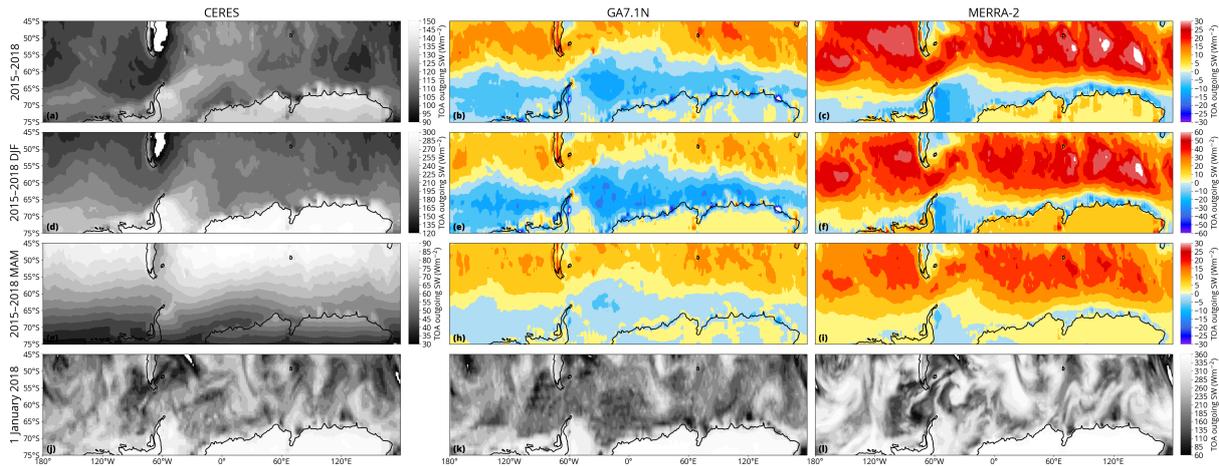


Figure 2.3 | Geographical distribution of the TOA outgoing SW radiation in CERES, GA7.1N and MERRA-2. The plots show global all sky SW radiation as annual (2015–2018; **a–c**), seasonal (2015–2018 DJF, MAM; **d–i**) and daily (1 January 2018; **j–l**) mean. The blue–red colormap shows bias relative to CERES (**b, c, e, f, h, i**), while the grayscale colormap shows absolute values (**a, d, g, j, k, l**).

pattern of negative and positive bias on the annual (Fig. 2.3b, c) and seasonal (Fig. 2.3e, f, h, i) time scales. GA7.1N shows predominantly negative bias, while MERRA-2 shows predominantly positive bias. The annual average is dominated by the bias in DJF due to the relatively strong incoming solar radiation in DJF. The bias displays very similar geographical pattern on the annual scale, DJF and MAM. The bias is much lower in MAM compared to DJF due to lower incoming solar radiation.

We chose 1 January 2018 as a representative day in DJF to show the daily scale. On the daily scale (Fig. 2.3j, k, l), the patterns are closely linked to synoptic features. The region on the eastern side of the Antarctic Peninsula shows the greatest negative bias in the models. The relatively zonally symmetric annual and seasonal means suggest that there is not a significant need for subsetting by longitude, and that latitude averages can be very useful in identifying the key features of the SW radiation biases. The daily synoptic features are generally well-correlated between CERES and the models, which is expected in nudged model runs and reanalyses. MERRA-2 has greater TOA outgoing SW radiation than GA7.1N on all three time periods presented here. Considering that cloud is the dominant factor affecting SW radiation in the SO, this can only be associated with either cloud cover which is too high, or cloud albedo which is too high. GA7.1N reflects too little SW radiation south of 60°S and too much north of 60°S (Fig. 2.3b, e, h). MERRA-2 reflects too much SW radiation in most of the SO except for coastal regions of Antarctica (approx. 65–70°S) and the eastern side of the Antarctic Peninsula. The opposite sign of SW radiation bias in GA7.1N compared to MERRA-2 suggests that contrasting the two models could be useful for uncovering the cause of the bias.

Figure 2.4 shows line plots of zonal mean reflected SW radiation and bias relative to CERES by month in multiple latitude bands between 50 and 70°S, with the southernmost band 65–70°S limited to 180–80°W to avoid covering land areas in Antarctica. The annual cycle follows the expected seasonal pattern modulated by varying incoming solar radiation with maxima of reflected radiation in December and maxima of bias in December and January. The Antarctic sea ice extent,

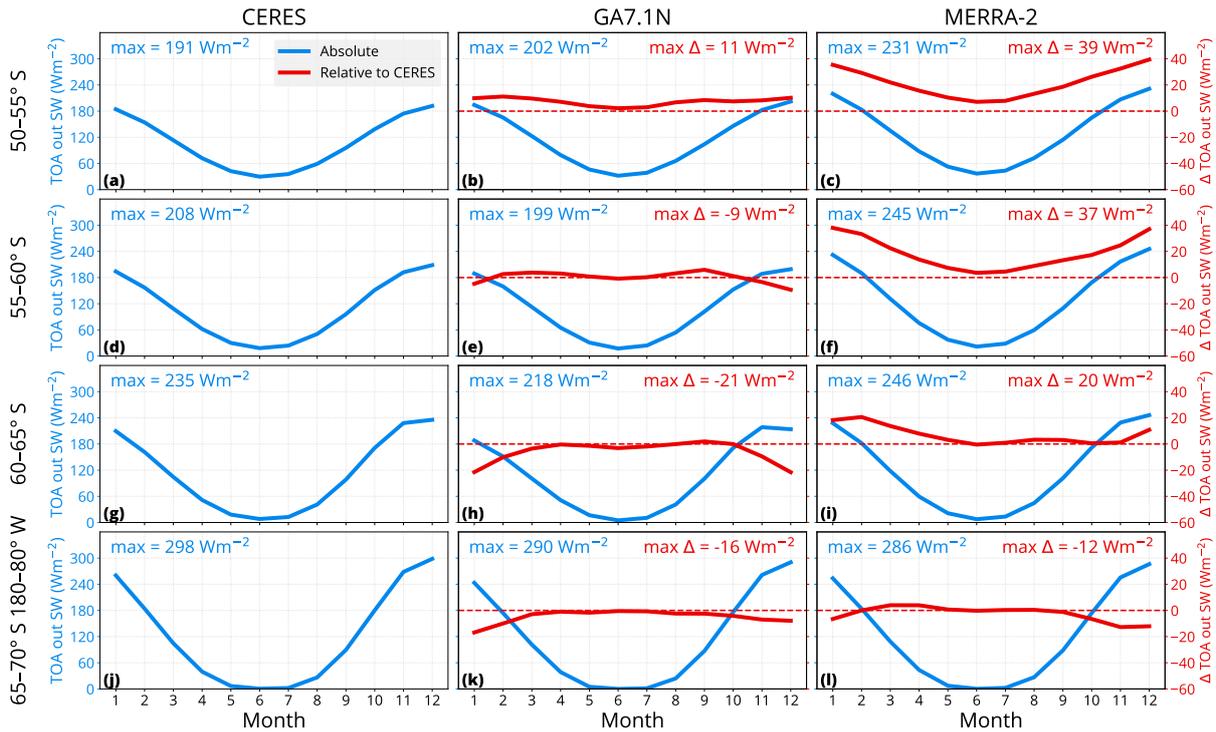


Figure 2.4 | Zonal means of the TOA outgoing SW radiation in CERES, GA7.1N and MERRA-2 during the years 2015–2018 in 5-degree latitude bands between 50 and 70°S. The plots show monthly zonal mean TOA outgoing SW radiation (blue) and its difference relative to CERES (red) as a function of month. Shown are also the maxima ("max") and the difference from CERES ("max Δ ").

at its minimum in February and peaking in September, is also likely a secondary modulating factor of the TOA outgoing SW radiation at higher latitudes. The models represent the seasonal pattern well, but differ substantially during the periods of peak incoming solar radiation. The GA7.1N model (Fig. 2.4b, e, h, k) exhibits bias ranging from -21 to $+11$ Wm^{-2} . The bias is positive north of 55°S and negative south of this latitude, with the greatest absolute bias between 60 and 65°S . MERRA-2 displays a clearly different bias from GA7.1N, ranging from -12 to 39 Wm^{-2} (Fig. 2.4c, f, i, l). The peak SW bias in MERRA-2 is positive for latitudes north of 65°S and negative south of this this latitude. The absolute bias in MERRA-2 is much larger than in GA7.1N north of 60°S and similar to GA7.1N south of this latitude. Therefore, the MERRA-2 results are valuable for contrasting with GA7.1. The strong latitudinal variation of the TOA outgoing SW radiation bias is important to take into consideration. Previous studies of SO clouds often did not discern different latitudes.

Figure 2.5 shows scatter plot of the TOA outgoing SW radiation bias in GA7.1N and MERRA-2 as a function of near-surface air temperature and relative humidity between 55 and 70°S in January 2018. The bias is predominantly negative in GA7.1N and positive MERRA-2. There is a strong cluster of negative bias at temperature around 0°C in GA7.1N and -2°C in MERRA-2, and a cluster of positive bias at higher temperatures. This is consistent with the latitudinal dependence of bias in both models shown above.

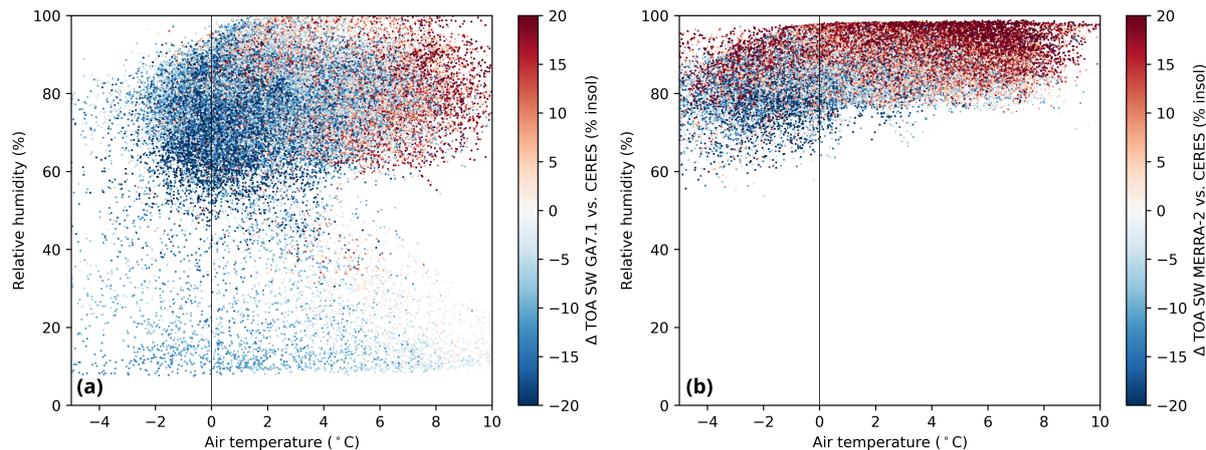


Figure 2.5 | Scatter plot of SW radiation bias in **(a)** GA7.1N and **(b)** MERRA-2 grid cells between 55°S and 70°S in January 2018. Each point represents a daily average of SW radiation bias as a function of near-surface air temperature and near-surface relative humidity. The bias is expressed as a percentage of the incoming solar radiation in the grid cell. The points are a random sample of 100000 points.

2.5.2 Cloud occurrence in model and observations

To understand how clouds contribute to the SW bias, we examine cloud cover and cloud occurrence as a function of height in the models and observations. Figure 2.6 shows cloud occurrence profiles derived from ceilometer observations on different voyages and GA7.1N and MERRA-2 model output derived via the COSP surface lidar simulator, in subsets by latitude and season. Most notably, the observed cloud cover is consistently very high in the observations (80–100%) for all periods and latitude bands examined and greater than 90% in most of the subsets. This finding differs substantially from the modelled cloud cover derived via the surface lidar simulator, which ranges between 69 and 100% in GA7.1N, and is about 4–9% lower than observations across the subsets. The cloud cover in MERRA-2 is also lower than observed and much lower than in GA7.1N, spanning 51–95%. Only in 4 subsets is the cloud cover greater in GA7.1N than observed, and only in 1 subset is the cloud cover greater in MERRA-2 than observed (out of 21 subsets). Our analysis therefore shows that cloud cover is underestimated in both GA7.1N and MERRA-2 in the evaluated geographical regions and seasons.

Examination of the vertical distributions in Fig. 2.6 shows that observations have a strong predominance of cloud below 2 km and peaking below 500 m in most subsets, including a substantial amount of surface-level fog in some subsets. In contrast, GA7.1N and MERRA-2 simulate clouds at a higher altitude, peaking at about 500 m and generally the peak is higher than in observed clouds. Especially, clouds below 500 m and fog appear to be lacking in the models.

The subsets in Fig. 2.6 are derived from uneven length of ship observations (1.0–28.9 days) due to the limited availability of data. The longer subsets (Fig. 2.6a4, b4, c2, c4, f1) appear marginally more consistent between the models and observations in terms of the cloud occurrence profile, but the cloud cover is still markedly underestimated. Figure 2.7 shows the model subsets of Fig. 2.6 as points by their cloud cover bias relative to observations. It can be seen that GA7.1N underestimates cloud cover by about 4% and MERRA-2 by 16% when non-weighted averages are considered, and

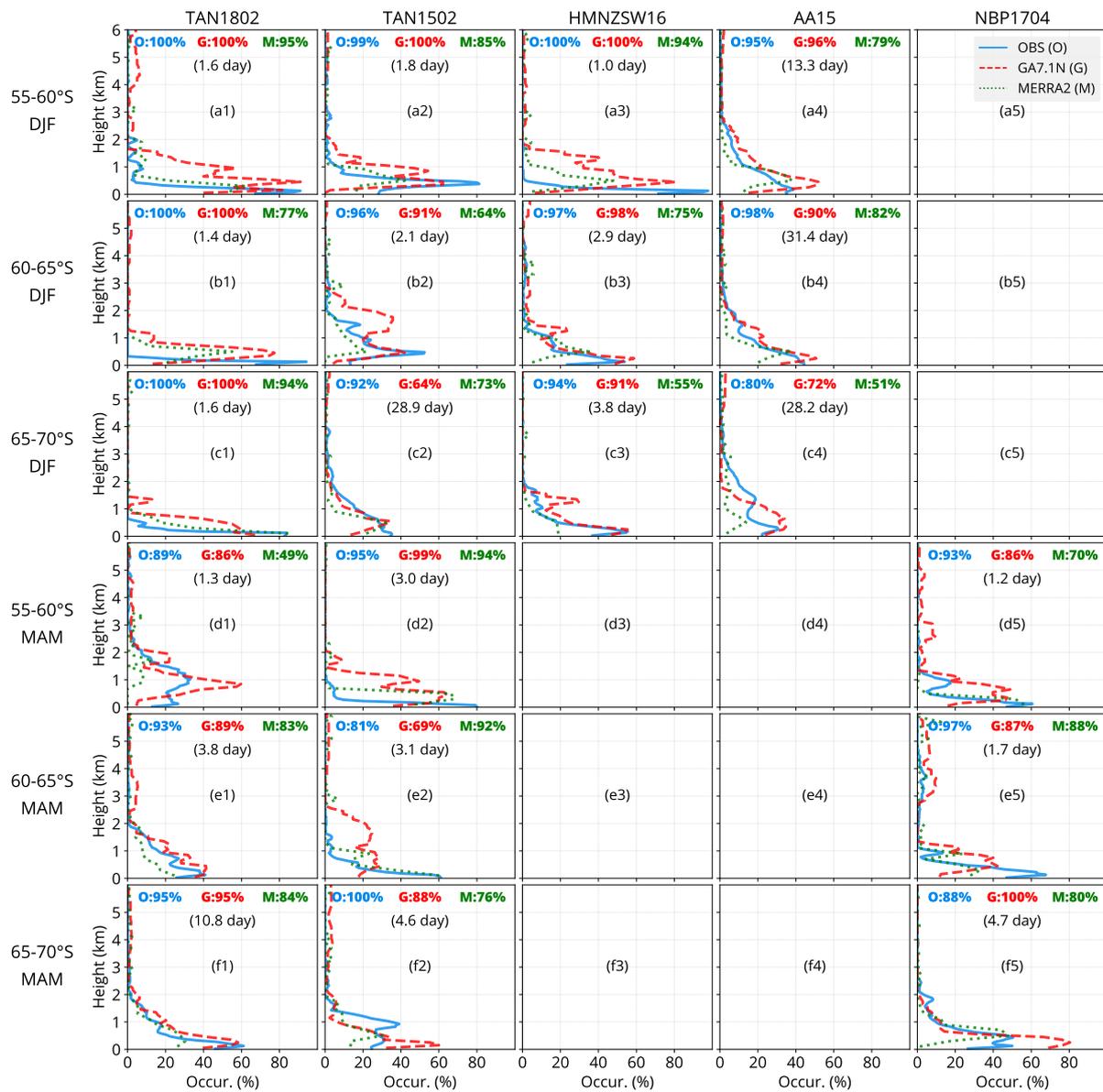


Figure 2.6 | Cloud occurrence frequency as a function of height derived from ceilometer observations (OBS) and model fields (GA7.1N and MERRA-2). The observational and model data were subsetted by latitude and season (DJF, MAM) along the voyage track. The numbers at the top of each panel show total (vertically integrated) cloud cover and the number of days the ship spent passing through the spatiotemporal subset. The height in the plots is limited to 6 km. There was no significant amount of cloud detected above this level.

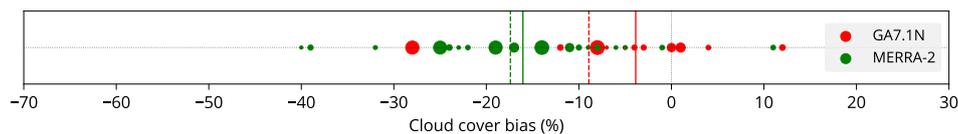


Figure 2.7 | Cloud cover bias in models relative to observations. The points represent subsets as in Fig. 2.6. The size of the circles is proportional to the number of days of observations in the subset. The solid lines are averages, and dashed lines are averages weighted by the number of days the ship spent passing through the spatiotemporal subset.

by 9% (GA7.1N) and 18% (MERRA-2) when weighted averages are considered. Due to the nature of the lidar measurements, middle to high clouds may be obscured by low clouds, as the laser signal is quickly attenuated by thick cloud. Therefore, the lack of clouds above 2 km in the plots does not imply that no clouds are present. The lidar simulator, however, ensures unbiased 1:1 comparison with observations by accounting for the signal attenuation. The results demonstrate the value of surface cloud measurements in the SO relative to satellite measurements such as CloudSat and CALIPSO, which would likely provide a biased sample of these clouds because of "ground clutter" and obscuration by higher-level clouds, respectively (Alexander and Protat, 2018).

2.5.3 Radiosonde observations

We use radiosonde measurements performed on TAN1802 and NBP1704 to evaluate boundary layer properties and correlate them with clouds observed by a ceilometer. We compare the observations with "pseudo-radiosonde" profiles extracted from model fields at the same location and time. The location is based on the GNSS coordinates of the ship at the time of the balloon launch (the balloon trajectory length was generally not long enough to span multiple model grid cells in the lower troposphere).

We define a new quantity "SST lifting level" (SLL) derived from SST and boundary layer atmospheric potential temperature, defined as the level to which an air parcel with the same temperature as SST, rising from the sea surface, would rise adiabatically by buoyancy. That is, it is the level closest to the surface at which potential temperature is equal to SST, provided the air parcel is permitted to rise to this level by buoyancy (otherwise the air parcel does not rise and SLL is 0 m). This quantity is applicable in sea ice-free conditions in the SO, when cold Antarctic air is warmed by the open sea surface and is lifted by buoyancy until it reaches a limit imposed by the atmospheric stability of the atmosphere. Alongside the lifting condensation level (LCL) we found SLL to be a useful quantity for evaluation of CBH. The authors are not aware of any previous use of SLL, but this definition is supported by observations (see below).

Apart from SLL and LCL, we also use the lower tropospheric stability (LTS) (Klein and Hartmann, 1993). LTS is defined as the difference between potential temperature at 700 hPa and sea level pressure (Klein and Hartmann, 1993). It has been used in multiple previous studies (Williams et al., 2006; Franklin et al., 2013; Williams et al., 2013; Naud et al., 2014).

Figure 2.8 shows the observed and modelled relationship between CBH and the minimum of SLL and LCL (" $\min\{SLL, LCL\}$ "), LTS, SLL and LCL. A large fraction of the observed points (OBS) in Fig. 2.8a lie close to the origin (40% in the first 100 m in observations, vs. 26% and 17% in GA7.1N and MERRA-2, respectively), which suggests that near zero $\min\{SLL, LCL\}$ is a good indicator of fog or very low cloud, a relationship not well-represented in the models. The remaining observed points show a close equivalence between $\min\{SLL, LCL\}$ and CBH, while the models do not represent this equivalence well. The histogram in Fig. 2.8a reveals that about 42% of observed profiles have CBH within 100 m of $\min\{SLL, LCL\}$, while only about 28% of GA7.1N and 21% of MERRA-2 profiles do.

Using SLL or LCL as a predictor for CBH individually resulted in a weaker relationship than $\min\{SLL, LCL\}$: 25% and 31% of OBS profiles have CBH within 100 m of SLL and LCL, respectively (Fig. 2.8c, d). This suggests that $\min\{SLL, LCL\}$ is more strongly related to CBH than SLL or LCL

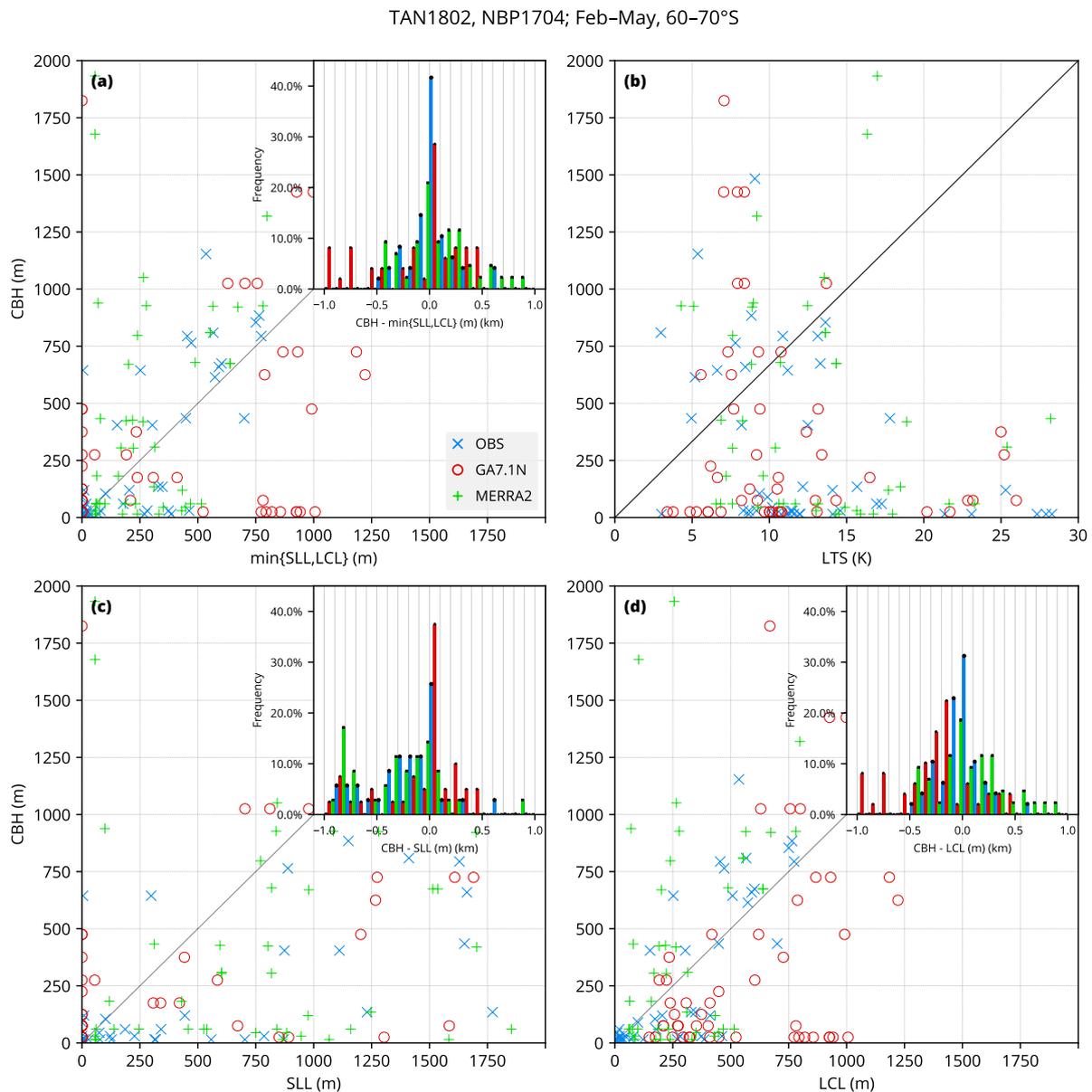


Figure 2.8 | Scatter plots of radiosonde measurements on the TAN1802 and NBP1704 voyages between February and May and 60–70°S latitude. Corresponding profiles from GA7.1N and MERRA-2 are selected, i.e. having the same geographical coordinates and the same time of the year. Each point on the scatter plots represents a radiosonde profile. The plots compare three datasets: observations (OBS), GA7.1N and MERRA-2. The radiosonde observations are matched with ceilometer (OBS) and COSP-based CBH (GA7.1N and MERRA-2). **(a)** shows the points as a function of $\min\{SLL, LCL\}$ and CBH. The inset histogram shows distribution of the difference of CBH and $\min\{SLL, LCL\}$ in bins of 100 m, where each bin contains three bars for the three datasets. **(b, c, d)** show the points as a function of LTS, SLL and LCL, respectively.

individually. Figure 2.8b shows CBH as a function of LTS. LTS does not display a good predictive ability for CBH in this dataset, with the exception of very stable profiles (LTS > 15 K), when observed CBH was below 250 m in all but one case.

Figure 2.9 shows the distribution of $\min\{\text{SLL}, \text{LCL}\}$ derived from radiosonde observations and model fields. In observations, the quantity almost consistently peaks near the ground and reaches up to 1.5 km in ice-free cases (Fig. 2.9a1–a5, b4). GA7.1N represents this distribution relatively well. This is not the case with MERRA-2, which is less likely to peak near the ground (Fig. 2.9a3, a5, c4). The sea-ice cases (Fig. 2.9b5, b6) show markedly different observed distribution of the quantity, with peak at about 300 m. GA7.1N and MERRA-2 represent the distribution over sea ice relatively poorly.

2.5.4 Zonal plane comparison of GA7.1N and MERRA-2

In order to better understand the differences in the SW radiation bias between GA7.1N and MERRA-2, we inspect zonal plane plots of cloud occurrence and thermodynamic fields of the models in DJF 2017/18 and 1 January 2018 (Fig. 2.10). The figure shows seasonal and daily average cloud liquid and ice mixing ratio contours plotted over two different backgrounds – potential temperature and relative humidity (RH). The daily average plots (Fig. 2.10c, d) show a very pronounced difference between the cloud liquid amount between the two models, with MERRA-2 simulating a much greater amount of cloud liquid. In contrast, GA7.1N simulates cloud with ice, which are nearly absent in MERRA-2 at the chosen contour levels. The liquid content is generally concentrated near SLL in MERRA-2, but much less so in GA7.1N, where SLL is often at 0 m. The cloud ice in GA7.1N generally has significantly greater vertical extent than the cloud liquid. These differences are also present on the seasonal scale (Fig. 2.10a, b). The difference in potential temperature between the models is relatively small. GA7.1N, however, shows a slightly higher potential temperature. The RH field is very different between GA7.1N and MERRA-2, with MERRA-2 simulating higher RH by about 10%.

Perhaps most interestingly, the vertically integrated liquid and ice content (Fig. 2.10i, j) is very different between the models. Both models simulate almost the same liquid + ice total, but the phase composition of cloud in GA7.1N is majority ice, while in MERRA-2 it is almost entirely liquid.

2.6 Discussion

The TOA outgoing SW radiation assessment shows that the models exhibit monthly average biases of up to 39 Wm^{-2} (MERRA-2, 50–55°S in December), and that these biases have a significant latitudinal dependency, with the opposite sign of bias between different latitude bands. In GA7.1N the bias is predominantly negative, while in MERRA-2 the bias is predominantly positive. Similar pattern of bias is present in both models. The bias is positive north of 55°S (65°S) in GA7.1N (MERRA-2) and negative south of this latitude. This finding is consistent with Schuddeboom et al. (2019), who observed opposite sign of SW cloud radiative effect south and north of 55°S in GA7.1.

A very similar geographical pattern of bias is present in DJF and MAM, suggesting that similar

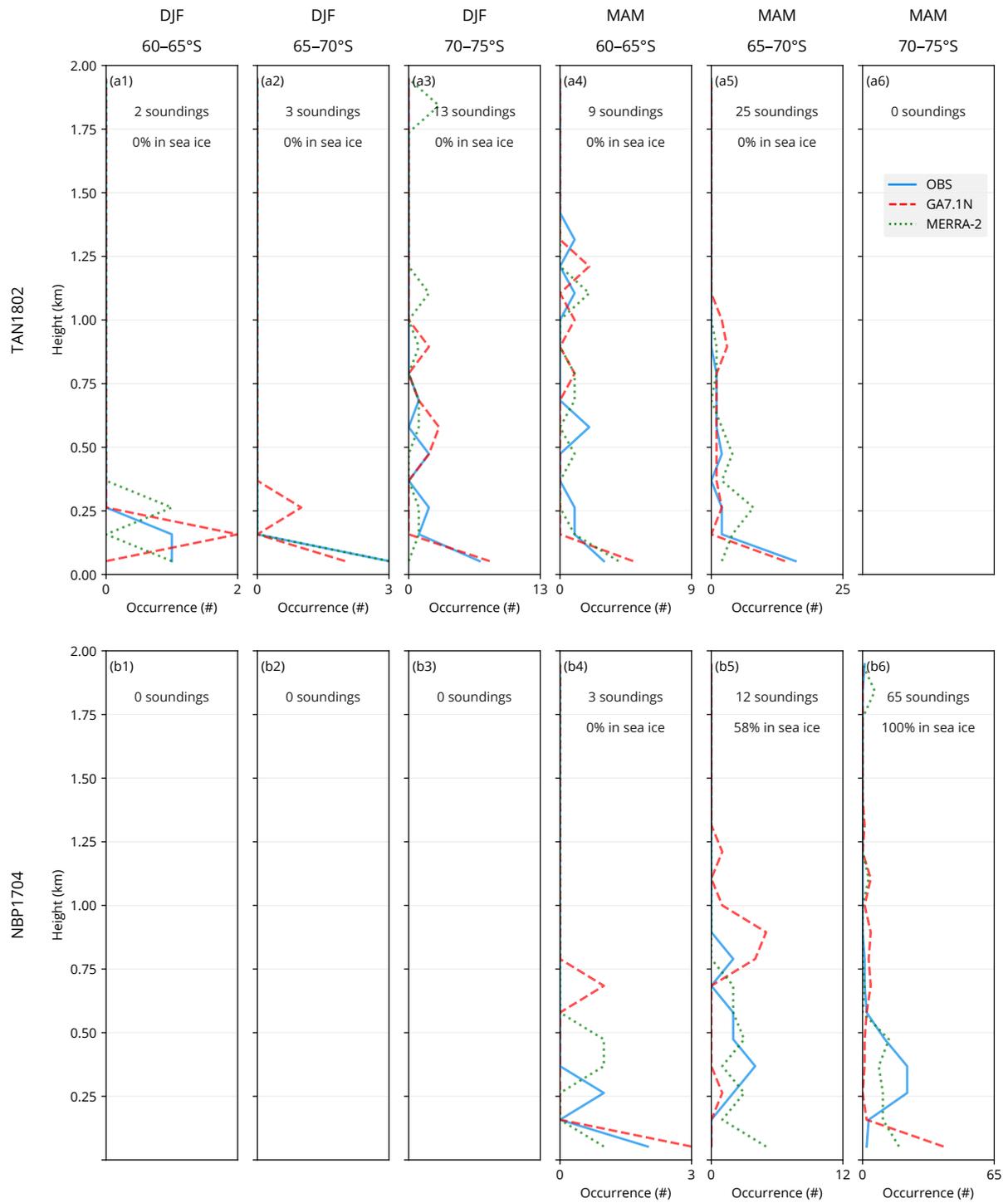


Figure 2.9 | Histogram of $\min\{SLL, LCL\}$ derived from radiosonde observations (OBS) on TAN1802 and NBP1704, and the equivalent profiles in GA7.1N and MERRA-2. Shown are subsets by latitude between 60 and 75°S and seasons DJF and MAM. The numbers at the top of each panel indicate the number of profiles which make up the histogram and the percentage of sea ice cases determined from NSIDC satellite-derived sea ice concentration.

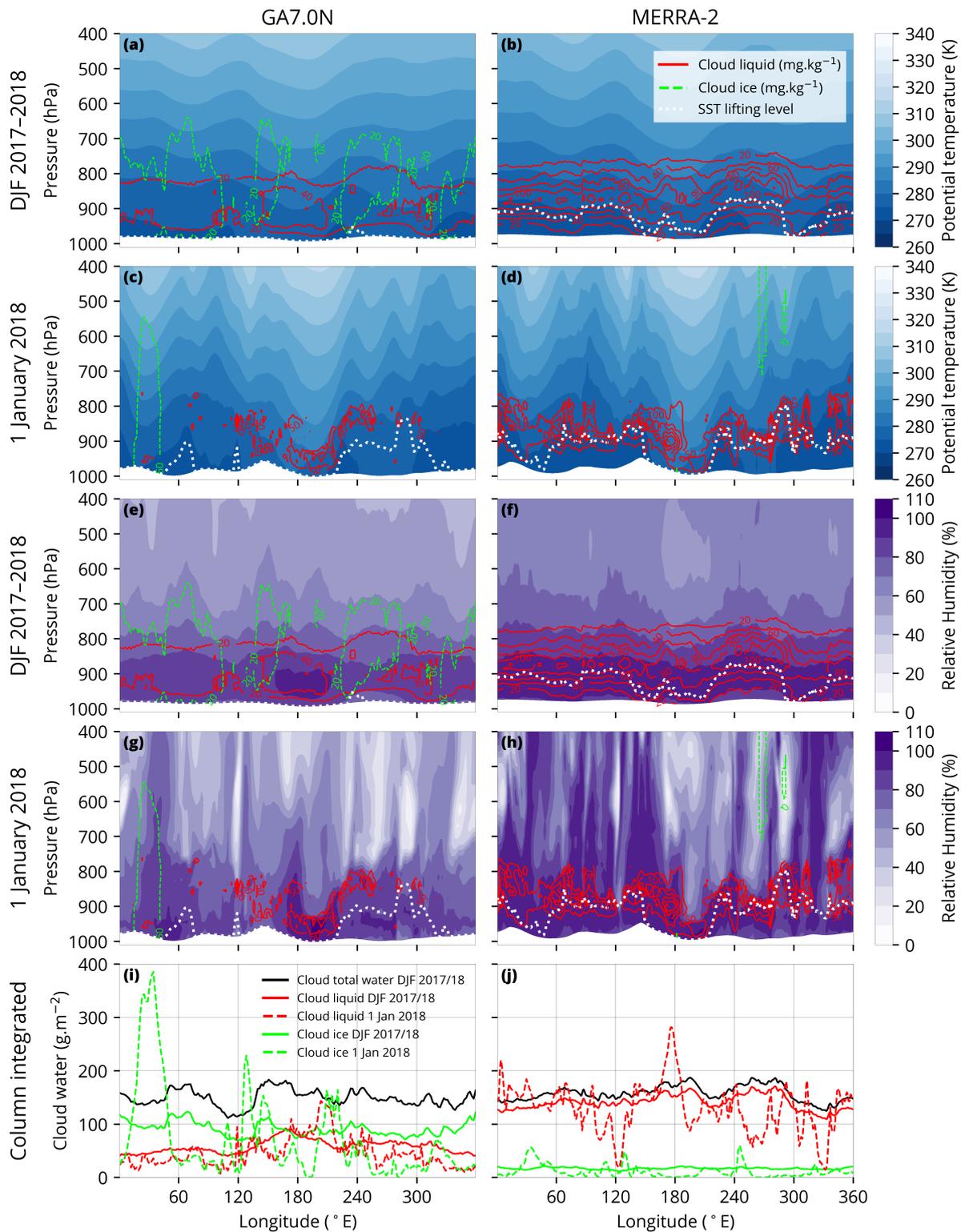


Figure 2.10 | Zonal plane plot of cloud liquid and ice mixing ratios in GA7.1N and MERRA-2 at 60°S. The cloud liquid and ice mixing ratios are plotted as contours on top of the potential temperature fields (a–d) and relative humidity fields (e–h). SLL is indicated by a white line. (a, b, e, f) show a seasonal average in DJF 2017/2018 and (c, d, g, h) show a daily average on 1 January 2018. (i, j) show the column-integrated values of cloud liquid and ice water as a function of longitude corresponding to the plots above. All liquid shown in the plots is supercooled (air temperature is less than 0°C everywhere).

cloud biases are present in both seasons. This is also supported by Fig. 2.6, which does not display a significant difference in observed cloud occurrence and bias in the models between DJF and MAM. Consistent with the maximum of incoming solar radiation, December and January were found to be the months with the greatest absolute bias in the models. Therefore, fixing the representation of clouds in the SO in these months is relatively more important than in other months.

Figure 2.5 suggests that the bias correlates not only with latitude, but also with near-surface air temperature. The negative bias is strongly clustered around 0°C in GA7.1N, and -2°C in MERRA-2, and positive bias is predominantly correlated with higher temperature.

The ship-based lidar cloud occurrence revealed close to 100% cloud cover in multiple subsets. Subsetting allowed us to identify whether the cloud cover is substantially different by latitude and season, and also sample independent weather situations (it is expected that cloud occurrence profiles are highly correlated over several days due to persistence of synoptic situations). The subsets show a relatively consistent cloud occurrence profile peaking below 500 m, and almost zero above 2 km (possibly also due to obscuration of lidar signal by lower clouds). The models generally do not reproduce this profile well. Apart from underestimating the total cloud cover, the peak of cloud occurrence in the models is higher than observed. Improving the cloud profile representation in the models is likely key for improving the SW radiation bias.

The effect of clouds on SW radiation is the product of cloud cover (the fraction of the sky containing clouds) and cloud albedo (the fraction of SW radiation reflected by the clouds). With our ship-based lidar observations we measured cloud cover (total, and cloud cover as a function of height), while we did not measure cloud albedo. The cloud cover was almost consistently underestimated in both GA7.1N and MERRA-2 across all latitudes. At the same time, the satellite observations show that MERRA-2 reflects too much all-sky SW radiation. Therefore, the cloud albedo in MERRA-2 must be too high in order to cause too much all-sky SW radiation reflection despite the lack of cloud cover. This effect is visible on the daily scale in Fig. 2.3j–l, where the individual clouds in MERRA-2 appear significantly brighter than on satellite observations.

Remarkably, the observed cloud occurrence profiles appear to be similar between the DJF and MAM seasons and latitude bands between 55 and 70°S (Fig. 2.6): if we focus on the subsets with more than 10 days (Fig. 2.6a4, b4, c2, c4, f1), i.e. not heavily skewed toward a single weather situation, we find that they are all characterised by a peak below 500 m of 25–60% and falling to near-zero above 2–3 km, sometimes with a minor secondary peak between 1 and 2 km. The simulated profiles show a slightly higher altitude of the primary peak between 0 and 1 km, underestimated in MERRA-2 by up about two thirds, falling to near-zero between 2 and 3 km, without any substantial secondary peak. The total cloud fraction appears to be more strongly underestimated at high latitudes in GA7.1N in DJF, by 8–28% (Fig. 2.6c2, c4) vs. 8% (Fig. 2.6b4). This is an important consideration in connection with the SW radiation bias, which shows a strong latitudinal gradient of the TOA outgoing SW radiation bias in the models (Fig. 2.3, 2.4). Based on the the presented results a plausible explanation for the SW radiation bias could be overestimation of cloud albedo north of about 55°S (65°S) in GA7.1N (MERRA-2) causing positive TOA outgoing SW radiation bias north of this latitude and underestimation of cloud cover over the whole SO causing negative TOA outgoing SW radiation bias south of this latitude.

In the ship observations we found a notable correspondence between CBH, SLL and LCL.

Boundary layer thermodynamics, determining the lifting levels, is a plausible driver of cloud formation in the absence of other forcing. We examined SLL in models and radiosonde observations, and found differences which are likely too small to explain the cloud occurrence differences between the models and ceilometer observations. Bodas-Salcedo et al. (2012), in their analysis of an earlier version of the GA model (GA3.0) using cyclone composites also noted that biases in thermodynamics are not likely to explain the SW radiation bias, but may still play a significant role. The presence of positive TOA outgoing SW radiation bias in the SO between 50 and 55°S in GA7.1, which contrasts with the negative bias south of the latitude, is important because it places a limit on the applicability of other studies which used SO observational data from regions north of 55°S (Lang et al., 2018).

In Section 2.5.3 we show that $\min\{\text{SLL}, \text{LCL}\}$ has a stronger equivalence to CBH than SLL, LCL individually or LTS. This relationship becomes quite notable when examining the individual voyage radiosonde profiles (not presented here). We hypothesise that the theoretical reason for this relationship is the following. When SLL is higher than LCL, an air parcel warmed by the sea surface to temperature close to SST rises by buoyancy past LCL to a level with the equivalent potential temperature. The water vapour starts to condensate at LCL (assuming enough cloud condensation nuclei are present at 100% saturation), forming cloud with CBH equal to LCL. If SLL is lower than LCL, the air parcel rises to the level of equivalent potential temperature, where air lifted from the sea surface eventually accumulates, potentially forming cloud if enough moisture is transported from the sea surface. The models do not represent the observed relationship well, and improving this relationship may be one way of improving the cloud simulation.

Considering the strong observed relationship between $\min\{\text{SLL}, \text{LCL}\}$ and CBH (CBH tends to occur at the same level as $\min\{\text{SLL}, \text{LCL}\}$), we evaluated the distribution of $\min\{\text{SLL}, \text{LCL}\}$ in the models in comparison with radiosonde observations (Fig. 2.9). We found that GA7.1N represents this distribution relatively well in sea-ice-free cases, while MERRA-2 underestimates cases when $\min\{\text{SLL}, \text{LCL}\}$ was near the surface. This may be the reason for the underestimation of very low cloud and fog in this model identified in the comparison with lidar observations. Therefore, improving the distribution of the quantity in MERRA-2 may lead to improvement of low cloud simulation.

It is interesting to contrast our results with previous studies which used cyclone compositing for the TOA SW radiation bias evaluation in GCMs. We cannot make substantial conclusions from our results on how much of the model bias is attributable to cyclones. It appears, however, that the cloud cover and cloud liquid and ice mixing ratio bias in GA7.1N is systematic rather than isolated to cyclonic activities due to its relative consistency across spatiotemporal subsets in the high latitude SO. This does not rule out even greater biases related to cyclonic sectors. Specifically, Bodas-Salcedo et al. (2014) evaluated a large set of models, including HadGEM2-A, a predecessor model to HadGEM3, likely affected by similar biases, and found that about 80% of grid cells south of 55°S could be classified as affected by a cyclone, and that these grid cells were responsible for the majority of the total SW radiation bias. Moreover, their cyclone compositing showed that the bias in HadGEM2-A was largely negative in the cold quadrants, and near zero in the warm quadrants. Their results also indicate a strong contrast in SW bias south and north of 55°S, similar to the result we found in GA7.1N. We think these results can be reconciled with our study by

assuming that the model has a particular difficulty in representing cloud in situations when near-surface air temperature is lower than the SST. In these regions the heat flux is from the ocean to the atmosphere is positive, which in the austral summer predominantly occur south of 55°S and in the cold sectors of cyclones. The cloud representation when near-surface air temperature is greater than SST is relatively more accurate, this case occurring predominantly north of 55°S and in the warm sector of cyclones. As shown in Fig. 2.5, the negative TOA outgoing SW radiation bias in the models is clustered at zero and sub-zero temperatures. This suggests a possible explanation that sub-zero air mass advecting from Antarctica or from sea ice covered areas over warm water (cold-air outbreaks) could be inducing low cloud and fog, and this process is not well represented in the models (Bodas-Salcedo et al., 2012).

Previous studies have documented that supercooled liquid is often present in the SO cloud in summer months (Morrison et al., 2011; Huang et al., 2012; Chubb et al., 2013; Huang et al., 2016; Bodas-Salcedo et al., 2016; Jolly et al., 2018; Listowski et al., 2019). We cannot substantially add to these findings with our observations, although preliminary analysis of a polarising lidar Sigma Space MiniMPL profiles from the TAN1802 voyage suggests supercooled liquid was commonly present in the ubiquitous stratocumulus cloud. The side-by-side comparison of cloud liquid and ice mixing ratios on the zonal plane (Fig. 2.10) suggests that models can differ significantly in their representation of cloud phase, with GA7.1N simulating markedly less supercooled liquid than MERRA-2. This is the most likely the explanation for the overestimation of TOA outgoing SW radiation in MERRA-2, despite the underestimated cloud cover in this model. If cloud cover is increased in MERRA-2 to better match with the lidar observations, the cloud albedo would have to be lowered to obtain a reasonable match of TOA outgoing SW radiation with CERES.

The 2016–2018 voyages may have been affected by the unusually low sea ice extent (discussed below), which can have a significant effect on cloud (Frey et al., 2018; Taylor et al., 2015). The modulating effect of sea ice on cloud in the SO has previously been shown by Listowski et al. (2019) and there is an apparent difference in cloud between the Ross Sea and Ross Ice Shelf as shown by Jolly et al. (2018), with cloud over the ice shelf having smaller cloud cover, a greater amount of altostratus cloud and a smaller amount of deep convective cloud. The sea ice and ice shelves block transport of heat and moisture to the atmosphere. Their low thermal conductivity and high albedo mean the surface can cool to very low temperature and thus have an effect on the radiation balance of the atmosphere. We did not focus on sea ice conditions, since one can expect the effect of cloud biases on the SW radiation bias over sea ice to be small – the ice surface is already highly reflective in the SW, and the presence of cloud has little impact on the grid cell SW reflectivity (the SW albedo of cloud is similar to sea ice, depending on the sea ice concentration).

The Antarctic sea ice extent has undergone a rapid decrease starting in the spring of 2016 after about a decade of slightly increasing extent (Turner et al., 2017; Stuecker et al., 2017; Doddridge and Marshall, 2017; Kushahara et al., 2018; Schlosser et al., 2018; Ludescher et al., 2018). The sea ice extent due to this decrease was found to be the lowest on observational record since 1979, and the Ross Sea was particularly affected by this anomaly. The unusually low sea ice extent likely affected atmospheric observations made on the voyages presented in this study, e.g. the TAN1802 voyage in February and March 2018 to the Ross Sea experienced no sea ice during the entire voyage. Because sea ice is an important factor influencing the atmospheric boundary-layer stability and radiation

Table 2.3 | A table showing a “back-of-the-envelope” calculation how the GA7.1N peak TOA outgoing SW radiation bias (Fig. 2.4) would change if the cloud cover were increased by 5% (Fig. 2.7), assuming the cloud albedo does not change. The “corrected” TOA outgoing SW radiation is calculated by multiplying the original value by 1.05.

Latitude	TOA out. SW at max. Δ (Wm^{-2})	Max. Δ TOA out. SW (Wm^{-2})	Corrected Max. Δ TOA out. SW (Wm^{-2})	Explained error
55–60°S	199	-9	0.95	111%
60–65°S	214	-21	-10.3	51%
65–70°S	243	-16	3.85	76%

balance, a significant secondary effect on cloud cover, cloud phase and opacity is expected. Sea ice is, however, not expected to be responsible for the SO SW radiation bias described here, because the bias is present even when sea ice concentration is prescribed from satellite observations, as is the case in the nudged run GA7.1 and the MERRA-2 reanalysis. Given that few of the ship-based observations were collected before 2016, we cannot reliably estimate how the anomalous sea ice extent affected our results.

In our results we found that even when model atmospheric dynamics is prescribed based on past observations, the TOA outgoing SW radiation bias is large and cloud occurrence, especially of low cloud and fog, is underestimated. CBH is found to be strongly linked to the boundary layer thermodynamics, and this link does not seem to be well represented in GA7.1N and MERRA-2. We therefore expect that cloud and boundary layer parametrisations (as part of subgrid scale processes in the models) are responsible for this bias. We have identified parts of the GA7.1N model most likely responsible: the large-scale cloud scheme, the PC2 scheme (Wilson et al., 2008a,b) and the boundary layer scheme. A future study should focus on these schemes to identify the parts responsible for the bias. In particular, the model should improve simulation of very low cloud and fog and achieve a closer match between the lifting levels and CBH (Fig. 2.8a).

In Table 2.3 we present a simple calculation how the GA7.1N peak TOA outgoing SW radiation bias would change if the cloud cover were increased by 5% (as suggested by Fig. 2.7), assuming the cloud albedo does not change. This correction would explain 51–111% of the bias depending on the latitude. The remaining part of the bias must be attributed to cloud albedo. One way this could be improved is by increasing the supercooled liquid fraction, or by increasing the total cloud water (liquid + ice) path. Therefore, our results suggest that in GA7.1N underestimation of cloud cover is responsible for the majority of the negative TOA outgoing SW radiation bias, relative to underestimation of cloud albedo.

2.7 Conclusions

We analysed 4 years of observational SO ship data, and contrasted them with a nudged run of the GA7.1 GCM, and MERRA-2 reanalysis. We used satellite observations of the Earth radiation budget to assess the TOA outgoing SW radiation bias in the SO in the models. We examined the total cloud cover and vertical distribution of cloud as measured by ceilometers and simulated by a ceilometer simulator based on the model data. We also compared SO radiosonde observations from two voyages with pseudo-radiosonde profiles from the models in order to assess boundary layer stability and the correlation between cloud base and atmospheric lifting levels. We also com-

pared model fields of cloud liquid and ice content, potential temperature and relative humidity in a zonal plane analysis across the SO to contrast cloud and thermodynamics simulated by GA7.1N and MERRA-2.

The SO SW radiation bias is significant in GA7.1N and MERRA-2, and tends to be positive in the northern parts of the SO and negative in the southern parts of the SO in both models. MERRA-2 shows greater absolute bias than GA7.1N. SO ship-based lidar and radiosonde observations are a valuable tool for model cloud evaluation, considering the amount of low cloud in this region which is likely poorly sampled by satellite instruments due to possible obscuration by higher overlapping cloud. The main findings of this study are that multi-year ship-based observations:

- corroborate satellite-based evidence of underestimated cloud cover, with both GA7.1N and MERRA-2 underestimating cloud cover on average by about 4–9% (GA7.1N) and 18% (MERRA-2),
- show that low cloud below 2 km is almost continuous in the SO in summer months in sea ice-free conditions, and not well represented in the models,
- indicate that boundary layer thermodynamics is a strong driver of cloud in the SO, and this relationship is not well represented in the models,
- suggest that subgrid-scale processes in situations when near-surface atmospheric temperature is lower or close to SST are responsible for the cloud misrepresentation.

Here, we introduced a new quantity (a thermodynamic level) called SST lifting level (SLL), which is the level of neutral buoyancy of an adiabatically lifted parcel with temperature equal to SST. The motivation for introducing this level was the frequently observed occurrence of cloud base at this height, together with LCL. We think that this is explained by the strongly thermodynamically-driven cloud in the Southern Ocean boundary layer and is linked to the particular conditions of the summertime Southern Ocean: sub-zero temperature of the near-surface atmosphere, destabilised by the relatively warmer (near-zero) sea surface.

Future studies of SO cloud representation in the GA model could focus on specific details of the model subgrid-scale cloud processes (such as the large scale cloud, boundary layer and convection schemes), and how their tuning impacts cloud occurrence distributions compared to the ship observations. The stark difference between GA7.1N and MERRA-2 cloud liquid and ice content also remains to be explained, and could provide valuable insight for improving the SO SW radiation bias in the model and the reanalysis.

Code and data availability

The original COSP version 1 simulator is open source and available publicly at <https://github.com/CFMIP/COSPv1>. The modified COSP version 1 simulator including the ground-based lidar simulator used in this study is open source and available at <https://alcf-lidar.github.io>. The cl2nc software for converting Vaisala CL51 data to NetCDF is available at <https://github.com/peterkuma/cl2nc>. The CERES EBAF and SYN1deg products are available publicly from the CERES website: <https://ceres.larc.nasa.gov/>. The Neal-Real-Time DMPS SSMIS Daily Polar Gridded Sea Ice Concentrations product is available publicly from the NSIDC website: <https://nsidc.org/data/nsidc-0081>. The Hadley Centre Sea Ice and Sea Surface Temperature data

set (HadISST) is available publicly from the Met Office website: <https://www.metoffice.gov.uk/hadobs/hadisst/>. The MERRA-2 data are available publicly from the MERRA-2 website: <https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/>. The ship-based observations dataset as well as all processing code is available on request from the authors.

Author contributions

Peter Kuma participated on methodology development, voyage observations, data analysis, writing and reviewing of the manuscript. Adrian McDonald participated on conceptualisation, funding acquisition, methodology development, voyage observations, data analysis, writing and reviewing of the manuscript. Olaf Morgenstern participated on model development, methodology development, data analysis, writing and reviewing of the manuscript. Simon Alexander, John Cassano, Jamie Halla, Sean Hartery, Sally Garrett, Mike Harvey, Simon Parsons, and Graeme Plank participated on voyage observations and reviewing of the manuscript. Vidya Varma and Jonny Williams participated on model development and reviewing of the manuscript.

Acknowledgements

We would like to thank everyone who participated on obtaining the Southern Ocean voyage observations, especially Kelly Schick and Peter Guest for performing ceilometer and radiosonde measurements on RV *Nathaniel B. Palmer*; the Royal New Zealand Navy for ceilometer and radar measurements on HMNZS *Wellington*; Alex Schuddeboom for deployment of instruments on RV *Nathaniel B. Palmer*, the crew of the TAN1502, *Aurora Australis* V1–V3 2015/16, HMNZS *Wellington*, NBP1704 and TAN1802 voyages. Logistical and technical support for the ceilometer observations made aboard *Aurora Australis* during the summer of 2015/16 were provided as part of the Australian Antarctic Science project 4292. We acknowledge the Met Office for use of the MetUM, and for providing the HadGEM3 model. We acknowledge NASA-GMAO and ECMWF for the MERRA-2 and ERA-Interim reanalyses, respectively. In this analysis we used publicly available satellite datasets provided by NASA and NSIDC. The CERES data were obtained from the NASA Langley Research Center CERES ordering tool (<https://ceres.larc.nasa.gov>). We wish to acknowledge the contribution of the NeSI high-performance computing facilities to the results of this research. New Zealand’s national facilities are provided by the NZ eScience Infrastructure and funded jointly by NeSI’s collaborator institutions and through the Ministry of Business, Innovation & Employment’s Research Infrastructure programme (<https://www.nesi.org.nz>). We would like to acknowledge the financial support that made this work possible provided by the Deep South National Science Challenge via the “Clouds and Aerosols” project. We acknowledge the software tools Python, R (R Core Team, 2018), numpy (Oliphant, 2006), scipy (Jones et al., 2001–), matplotlib (Hunter, 2007), Climate Data Operators (CDO) (Schulzweida, 2018) and parallel (Tange et al., 2011), which we used in our data analysis.

Chapter 3

Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0)

Peter Kuma¹, Adrian J. McDonald¹, Olaf Morgenstern², Richard Querel³, Israel Silber⁴, and Connor J. Flynn⁵

¹University of Canterbury, Christchurch, New Zealand

²National Institute of Water & Atmospheric Research (NIWA), Wellington, New Zealand

³National Institute of Water & Atmospheric Research (NIWA), Lauder, New Zealand

⁴Department of Meteorology and Atmospheric Science, Pennsylvania State University, PA, USA

⁵School of Meteorology, University of Oklahoma, Norman, OK, USA

Abstract

Automatic lidars and ceilometers provide valuable information on cloud and aerosols, but have not been used systematically in the evaluation of GCMs and NWP models. Obstacles associated with the diversity of instruments, a lack of standardisation of data products and open processing tools mean that the value of the large ALC networks worldwide is not being realised. We discuss a tool, called the Automatic Lidar and Ceilometer Framework (ALCF), that overcomes these problems and also includes a ground-based lidar simulator, which calculates the radiative transfer of laser radiation, and allows one-to-one comparison with models. Our ground-based lidar simulator is based on the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) which has been used extensively for spaceborne lidar intercomparisons. The ALCF implements all steps needed to transform and calibrate raw ALC data and create simulated backscatter profiles for one-to-one comparison and complete statistical analysis of cloud. The framework supports multiple common commercial ALCs (Vaisala CL31, CL51, Lufft CHM 15k and Sigma Space MiniMPL), reanalyses (JRA-55, ERA5 and MERRA-2) and models (AMPS and the Unified Model). To demonstrate its capabilities, we present case studies evaluating cloud in the supported reanalyses and models using CL31, CL51, CHM 15k and MiniMPL observations at three sites in New Zealand. We show that the reanalyses and models generally underestimate cloud fraction and overestimate cloud albedo, the common "too few too bright" problem. If sufficiently high temporal resolution model output is available (better than 6 hourly), a direct comparison of individual clouds is also possible. We demonstrate that the ALCF can be used as a generic evaluation tool to examine cloud occurrence and cloud properties in reanalyses, NWP models and GCMs, potentially utilising the large amounts of ALC data already available. This tool is likely to be particularly useful for the analysis and improvement of low-level cloud simulations which are not well monitored from

space. This has previously been identified as a critical deficiency in contemporary models, limiting the accuracy of weather forecasts and future climate projections.

3.1 Introduction

Automatic lidars and ceilometers (ALC) are active ground-based instruments which emit laser pulses in the ultraviolet, visible or infrared (IR) part of the electromagnetic spectrum and measure radiation backscattered from atmospheric constituents such as cloud and fog liquid droplets and ice crystals, haze, aerosol (particulate matter, ash) and atmospheric gases (Emeis, 2010). Vertical profiles of attenuated backscattered radiation can be produced by measuring received power as a function of time elapsed between emitting the pulse and receiving the backscattered radiation. Quantities such as cloud base height (CBH) and a cloud mask (Pal et al., 1992; Wang and Sassen, 2001; Martucci et al., 2010; Costa-Surós et al., 2013; Van Tricht et al., 2014; Liu et al., 2015a,b; Lewis et al., 2016; Cromwell and Flynn, 2019; Silber et al., 2018), aerosol optical depth (Marenco et al., 1997; Welton et al., 2000, 2002; Wiegner and Geiß, 2012; Wiegner et al., 2014; Jin et al., 2015; Dionisi et al., 2018) and boundary layer height (Eresmaa et al., 2006; Münkler et al., 2007; Emeis et al., 2009; Tsaknakis et al., 2011; Milroy et al., 2012; Knepp et al., 2017) can be derived from the backscatter profile. Lidars equipped with polarisation or multiple wavelengths can also provide depolarisation ratio or colour ratio, respectively, which can be used to infer cloud phase or particle types. Doppler lidars can measure wind speed in the direction of the lidar orientation. ALCs are commonly deployed at airports, where they provide CBH, fog and aerosol observations needed for air traffic control. Large networks of up to thousands of ACLs have been deployed worldwide: Cloudnet (Illingworth et al., 2007), E-PROFILE (Illingworth et al., 2018), EARLINET (Pappalardo et al., 2014), ICENET (Cazorla et al., 2017), MPLNET (Welton et al., 2006) and ARM (Stokes and Schwartz, 1994; Campbell et al., 2002). The purpose of these networks is to observe cloud, fog, aerosol, air quality, visibility and volcanic ash, provide input to numerical weather prediction (NWP) model evaluation and assimilation (Illingworth et al., 2015b, 2018) and for climate studies. These networks are usually composed of multiple types of ACLs, with Vaisala CL31, CL51, Lufft (formerly Jenoptik) CHM 15k and Sigma Space MiniMPL the most common. Complex lidar data processing has been set up on some of these networks. Notably, at the SIRTA site in France, lidar ratio (LR) comparable with a lidar simulator (Chiriaco et al., 2018) is calculated as part of their "ReOBS" processing method. Intercomparison and calibration campaigns such as CeiLinEx2015 (Mattis et al., 2016) and INTERACT-I(-II) (Rosoldi et al., 2018; Madonna et al., 2018) have been performed. Lidar data processing involves a number of tasks such as resampling, calibration, noise removal and cloud detection. Some of these are implemented in the instrument firmware of ALCs. This, however, means that lidar backscatter and detected cloud and cloud base are not comparable between different instruments. In most cases the algorithms are not publicly documented, making it impossible to compare the data with values from a model or a lidar simulator without a systematic bias.

Atmospheric model evaluation is an ongoing task, and a critical part of the model improvement process (Eyring et al., 2019; Hourdin et al., 2017; Schmidt et al., 2017). Traditionally, various types of observational and model datasets have been utilised – weather and climate station data, up-

per air soundings, ground-based and satellite remote sensing datasets, and high-resolution model simulations, amongst others. Clouds are one of the most problematic phenomena in atmospheric models due to their transient nature, high spatial and temporal variability, and sensitivity to a complex combination of conditions such as relative humidity, aerosols (presence of cloud condensation nuclei and ice nuclei), thermodynamic and dynamic conditions. At the same time, clouds have a very substantial effect on the atmospheric shortwave and longwave radiation balance, and any cloud misrepresentation has a strong effect on other components of the model, limiting the ability to accurately represent past and present climate and predict future climate (Zadra et al., 2018). An improved understanding of clouds and cloud feedbacks is one of the focuses of the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016), and comparison of model cloud with observations is one of the key points of the Cloud Feedback Model Intercomparison Project (CFMIP) (Webb et al., 2017). Satellite observations make up the majority of the data used to evaluate model clouds. These include passive visible and IR low earth orbit and geostationary radiometers measuring, among others, features such as cloud cover, cloud top height (CTH), cloud top temperature; passive microwave instruments measuring total column water; active radars and lidars measuring cloud vertical profiles. Ground-based remote sensing instruments include radars, lidars, ceilometers, radiometers and sky cameras. As pointed out by Williams and Bodas-Salcedo (2017), using a wide range of different observational datasets including satellite and ground-based for general circulation model (GCM) evaluation is important due to limitations of each dataset.

Model cloud is commonly represented by the mixing ratio of liquid and ice and cloud fraction (CF) on every model grid cell and vertical level. In addition some models provide the cloud droplet effective radius used in radiative transfer calculations. Remote sensing observations do not match the representation of the atmospheric model fields directly because of their different resolutions, limited field of view (FOV) and attenuation by atmospheric constituents before reaching the instrument's receiver. Instrument simulators bridge this gap by converting the model fields to quantities which emulate those measured by the instrument, which can then be compared directly with observations. One such collection of instrument simulators is the CFMIP Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011; Swales et al., 2018), which has been used for more than a decade for evaluation of models using satellite, and more recently ground-based, observations. The simulators in COSP include active instruments such as spaceborne and ground-based radars: Cloud Profiling Radar (CPR) on CloudSat (Stephens et al., 2002), Ka-band ARM Zenith Radar (KAZR); lidars: Cloud-Aerosol Lidar Orthogonal Polarization (CALIOP) on CALIPSO (Winker et al., 2009), Cloud-Aerosol Transport System (CATS) on ISS (McGill et al., 2015), the Atmospheric Lidar (ATLID) on EarthCARE (Illingworth et al., 2015a); and spaceborne passive instruments: ISCCP (Rossow and Schiffer, 1991), MODIS (Parkinson, 2003) and MISR (Diner et al., 1998). The more recent addition of ground-based radar (Zhang et al., 2018) and lidar (Chiriaco et al., 2018; Bastin et al., 2018) opens up new possibilities to use the large amount of remote sensing data obtained from ground-based active remote sensing instruments. In practice, ground-based observational remote sensing data are not straightforward to use without a substantial amount of additional processing. Some previous studies have also compared models and ground-based radar and lidar observations without the use of an instrument simulator (Bouniol et al., 2010; Hansen et al., 2018), though for the reasons identified above this is not advisable.

In this study we introduce a software package called the Automatic Lidar and Ceilometer Framework (ALCF) for evaluating model cloud using ALC observations. It extends and integrates the COSP lidar simulator (Chiriaco et al., 2006; Chepfer et al., 2007, 2008) with pre- and post-processing steps, and allows the simulator to be run offline on model output, instead of having to be integrated inside the model. This makes it possible to compare ALC data at any location without having to run the model with a specific configuration. Multiple ALCs, reanalyses and model output formats are supported. The original COSP lidar simulator was extended with Rayleigh and Mie scattering at multiple lidar wavelengths. Observational ALC data from a number of common instruments can be processed by re-sampling to a common resolution, removing noise, detecting cloud and calculating statistics. The same steps can be performed on the simulated lidar data from the model (the output of running COSP on the model data), allowing for one-to-one comparison of model and observations. A particular focus of our work was on applying the same processing steps on the observed and simulated backscatter in order to avoid biases. The ALCF is made available under an open source license (MIT) at <https://alcf-lidar.github.io>, and as a permanent archive of code and technical documentation on Zenodo at <https://doi.org/10.5281/zenodo.3779518>.

A relatively small amount of other open source code is available for ALC data processing. A lidar simulator has been developed as part of the Goddard Satellite Data Simulator Unit (G-SDSU) (G-S, 2019), a package based on the instrument simulator package SDSU (Masunaga et al., 2010). The Community Intercomparison Suite (CIS) (Watson-Parris et al., 2016) allows for subsetting, aggregation, co-location and plotting of mostly satellite data with a focus on model–observations intercomparison. The STRAT lidar data processing tools are a collection of tools for conversion of raw ALC data, visualisation and feature classification (Morille et al., 2007).

Here, we provide an overview (Sect. 3.2) and describe the supported ALCs, reanalyses and models (Sect. 3.3). We present a set of case studies at three sites in New Zealand (NZ) (Sect. 3.6) to demonstrate the value of this new tool. Furthermore, we describe the lidar simulator (Sect. 3.4), observed and simulated lidar data processing steps (Sect. 3.5) and the results of the case studies (Sect. 3.7).

3.2 Overview of operation of the Automatic Lidar and Ceilometer Framework (ALCF 1.0)

The ALCF performs the necessary steps to simulate ALC backscatter based on 4-dimensional atmospheric fields from reanalyses, NWP models and GCMs, and to transform the observed raw ALC backscatter profiles to profiles comparable with the simulated profiles. It does so by extracting 2-dimensional (time×height) profiles from the model data, performing radiative transfer calculations based on a modified COSP lidar simulator (Sect. 3.4), absolute calibration and resampling of observed backscatter to common resolution and performing comparable cloud detection on the simulated and observed backscatter. The framework supports multiple common ALCs (Sect. 3.3.1), reanalyses and models (Sect. 3.3.2). The schematic in Fig. 3.1 illustrates this process as well as the ALCF commands which perform the individual steps. The following commands are implemented: **model**, **simulate**, **lidar**, **stats** and **plot**. The commands are normally executed in a sequence, which is also implemented by a meta-command **auto**, which is

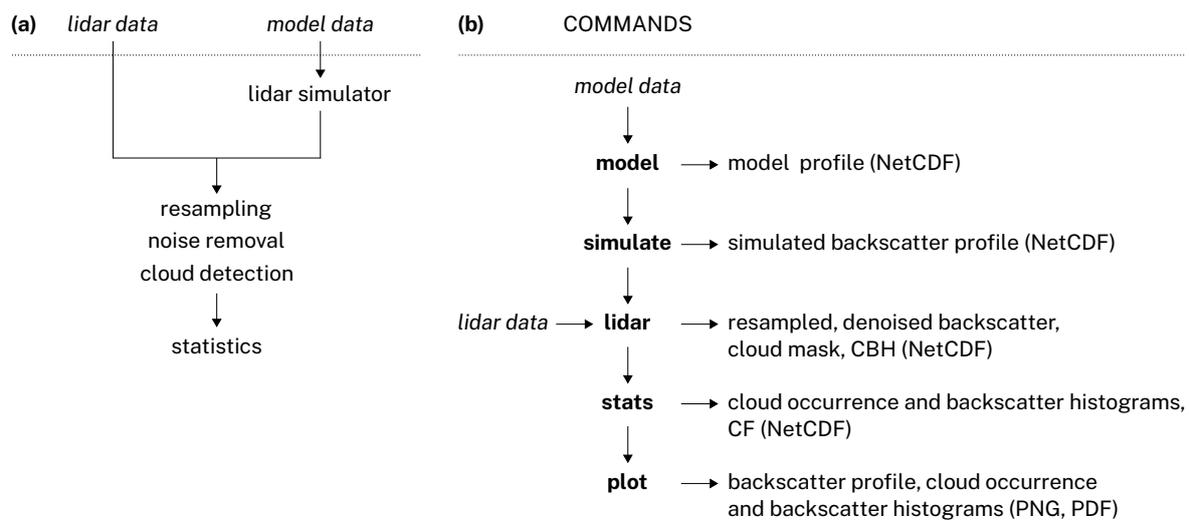


Figure 3.1 | (a) Scheme showing the operation of the ALCF and (b) the processing commands.

equivalent to executing a sequence of commands. The commands are described in detail in the technical documentation available online at <https://alcf-lidar.github.io>, on Zenodo at <https://doi.org/10.5281/zenodo.3779518> and in the Supplementary information. The physical basis is described here.

The **model** command extracts 2-dimensional profiles of cloud liquid and ice content (and other thermodynamic fields) from the supported NWP model, GCM and reanalysis data (*model data* in Fig. 3.1) at a geographical point, along a ship track or a flight path. The resulting profiles are recorded as NetCDF files. Sect. 3.3.2 describes the supported reanalyses and models. The *model data* can be either in one of the supported model output formats, or a new module for reading arbitrary model output can be written, providing that the required atmospheric fields are present in the model output. The required model fields are: per-level specific cloud liquid water content, specific cloud ice water content, cloud fraction, geopotential height, temperature, surface-level pressure and orography. No physical calculations are performed by this command. The atmospheric profiles are extracted by a nearest-neighbour selection.

The **simulate** command runs the lidar simulator described in Sect. 3.4 on the extracted model data (the output of the **model** command) and produces simulated backscatter profiles. This command runs the COSP-derived lidar simulator, which performs radiative transfer calculations of the laser radiation through the atmosphere. The resulting simulated backscatter profiles are the output of this command.

The **lidar** command applies various processing algorithms on either the simulated backscatter data (the output of the **simulate** command) or the observed ALC data (*lidar data* in Fig. 3.1) (Sect. 3.5). The data are resampled to increase the signal-to-noise ratio (SNR), noise is subtracted, LR is calculated, a cloud mask is calculated by applying a cloud detection algorithm and CBH is determined from the cloud mask. Absolute calibration (Sect. 3.5.2) can also be applied in this step by multiplying the observed backscatter by a calibration coefficient. This is important in order to obtain unbiased backscatter profiles comparable with the simulated backscatter profiles. Sect. 3.3.1

Table 3.1 | Table of ALCs and their technical parameters. Power is calculated as Pulse×Pulse Repetition Frequency (PRF).

Instrument	λ (nm)	Laser	Rate ¹ (s)	Res. ² (m)	Depol. ³	Pulse ⁴ (μ J)	Range ⁵ (km)	PRF (kHz)	Overlap ⁶ (m)	Power (mW)
CHM 15k	1064	Nd:YAG	2–600	5	no	7–9	15.4	5–7	1000 ⁷	48
CL31	910	InGaAs	2–120	10	no	1.2	7.7	10	70 ⁷	12
CL51	910	InGaAs	6–120	10	no	3	15.4	6.5	230 ⁸	20
MiniMPL	532	Nd:YAG	1–900	5–75	yes	3–4	30.0	2.5	2000 ⁸	9

¹Sampling rate. ²Vertical (range) resolution. ³Depolarisation. ⁴Pulse energy. ⁵Maximum range. ⁶Range of full overlap. ⁷Hopkin et al. (2019). ⁸Madonna et al. (2018).

describes the supported instruments. The *lidar data* can be in one of the supported instrument formats. If the native instrument format is not NetCDF, it has to be converted from the native format with the auxiliary command `convert` or one of the conversion programs: `cl2nc` (Vaisala CL31, CL51), `mpl2nc` or `SigmaMPL` (Sigma Space MiniMPL).

The `stats` step calculates summary statistics from the output of the `lidar` command. These include CF, cloud occurrence by height, backscatter histograms and the averages of LR and backscatter.

The `plot` command plots the backscatter profiles produced by the `lidar` command (Fig. 3.5, 3.6, 3.7), and the statistics produced by the `stats` command: cloud occurrence (Fig. 3.4), backscatter histograms (Fig. 3.9) and backscatter noise standard deviation histograms (Fig. 3.11).

3.3 Supported input data: instruments, reanalyses and models

3.3.1 Instruments

The primary focus of the framework is to support common commercial ALCs. Ceilometers are considered the most basic type of lidar (Emeis, 2010; Kotthaus et al., 2016) intended as commercial products designed for unattended operation. They are used routinely to measure CBH, but most instruments also provide the full vertical profiles of backscatter. Therefore, they are suitable for model evaluation by comparing not only CBH, but also cloud occurrence as a function of height. Their compact size and low cost make it possible to deploy a large number of these instruments in different locations, or use them in unusual settings such as mounted on ships (Klekociuk et al., 2019b; Kuma et al., 2020a). Common off-the-shelf ceilometers are the Lufft CHM 15k, Vaisala CL31 and CL51. Some lidars offer higher power and therefore higher SNR, and capabilities not present in ceilometers such as dual polarisation, multiple wavelengths, Doppler shift measurement and Raman scattering. Below we describe ALCs supported by the framework and used in our case studies: Lufft CHM 15k, Vaisala CL31 and CL51 and Sigma Space MiniMPL. Table 3.1 lists selected parameters of the supported ALCs.

Lufft CHM 15k (previously Jenoptik CHM 15k) is a ceilometer operating at a wavelength of 1064 nm in the near IR spectrum. The maximum range of the instrument is 15.4 km, vertical sampling resolution 5 m and sampling rate up to 2 s. The wavelength in the near IR spectrum ensures low molecular backscatter. The instrument produces NetCDF files containing uncalibrated attenuated backscatter profiles and various derived variables.

Vaisala CL31 and CL51 are ceilometers operating at a wavelength of 910 nm in the near IR spec-

Table 3.2 | Reanalyses and models used in the case studies and some of their main properties. The temporal and horizontal grid resolution and vertical levels listed is the resolution of the model output available. The horizontal grid resolution is determined at 45°S. The internal resolution of the model may be different (see Sect. 3.3.2 for details). The reanalyses and the UM use regular longitude-latitude grids, while the AMPS horizontal grid is regular in the South Pole stereographic projection.

Model/Grid	Type	Time resolution	Horizontal grid resolution	Vertical levels
AMPS/D01	NWP	3 h	$0.27^\circ \times 0.19^\circ$ (21×21 km)	60
ERA5	Reanalysis	1 h	$0.25^\circ \times 0.25^\circ$ (20×28 km)	37
JRA-55	Reanalysis	6 h	$1.25^\circ \times 1.25^\circ$ (98×139 km)	37
MERRA-2	Reanalysis	3 h	$0.625^\circ \times 0.50^\circ$ (49×56 km)	72
UM (GA7.1)/N96	GCM	20 min.	$1.875^\circ \times 1.25^\circ$ (147×139 km)	85

trum. The maximum range of CL31 and CL51 is 7.7 km and 15.4 km and the sampling rate is 2 and 6 s, respectively. The vertical resolution is 10 m. The wavelength is characterised by relatively low molecular backscatter (but higher than 1064 nm) and is affected by water vapour absorption (Wiegner and Gasteiger, 2015; Wiegner et al., 2019), which can cause additional absorption of about 20% in the mid-latitudes and 50% in the tropics (see also Sect. 3.5.4). The instruments produce data files containing uncalibrated attenuated backscatter which can be converted to NetCDF (see `cl2nc` in the Code and data availability section). The firmware configuration option "noise_h2 off" results in backscatter range correction to be selectively applied under a certain critical range and above this range only if cloud is present (Kotthaus et al., 2016, Sect. 3.2). This was the case with our case study dataset (Sect. 3.6). We apply range correction on the uncorrected range gates during lidar data processing. The critical range in CL51 is not documented, but was determined as 6000 m based on an observed discontinuity.

Sigma Space (part of Hexagon) Mini Micro Pulse Lidar (MiniMPL) (Spinhirne, 1993; Campbell et al., 2002; Flynn et al., 2007) is a dual-polarisation micro pulse lidar (meaning that it uses a high pulse repetition rate (PRF) and low pulse power) operating at a wavelength of 532 nm (green) in the visible spectrum. The maximum range of the instrument is 30 km. The vertical resolution is up to 5 m and sampling rate up to 1 s. The shorter wavelength is affected by stronger molecular backscatter than 910 nm and 1064 nm. The instrument can be housed in a protective enclosure, which makes it suitable for deployment in a relatively harsh environment. A scanning head can be mounted on the enclosure which provides configurable scanning by elevation angle and azimuth. The instrument produces data files containing raw backscatter which can be converted to NetCDF containing normalised relative backscatter (NRB) with a vendor-provided tool `SigmaMPL` (see also `mpl2nc` in the Code and data availability section).

3.3.2 Reanalyses and models

Below we briefly describe reanalyses and models¹ used in the case studies presented here (Sect. 3.6). We used publicly available output from three reanalyses and one NWP model. In addition, we performed nudged GCM simulations with high-temporal resolution output with the Unified

¹We use the term "reanalysis" when referring to ERA5, JRA-55 and MERRA-2 even though the reanalyses are based on atmospheric models. We use the term "model" when referring to AMPS and the UM, which are atmospheric models.

Model (UM). Table 3.2 lists some of the main properties of the reanalyses and models.

The Antarctic Mesoscale Prediction System (AMPS) (Powers et al., 2003) is a limited-area NWP model based on the polar fifth-generation Pennsylvania State University-National Center for Atmospheric Research Mesoscale Model (Polar MM5), now known as the Polar Weather Research and Forecasting (WRF) model (Hines and Bromwich, 2008). The model serves operational and scientific needs in Antarctica, but its largest grid also covers the South Island of NZ. AMPS forecasts are publicly available on the Earth System Grid (Williams et al., 2009). The forecasts are produced on several domains. The largest domain D01 used in the presented analysis covers NZ and has horizontal grid spacing of approximately 21 km over NZ. The model uses 60 vertical levels. The model output is available in 3-hourly intervals and initialised at 00:00 and 12:00 UTC. The initial and boundary conditions are based on the Global Forecasting System (GFS) global NWP model. AMPS assimilates local Antarctic observations from human-operated stations, automatic weather stations (AWS), upper-air stations and satellites.

ERA5 (ECMWF, 2019) is a reanalysis produced by the European Centre For Medium-Range Weather Forecasts (ECMWF) currently available for the time period 1979 to present, with a plan to extend the time period to 1950. The reanalysis is based on the global NWP model Integrated Forecast System (IFS) version CY41R2. It uses a 4D-Var assimilation of station, satellite, radiosonde, radar, aircraft, ship-based and buoy data. The model has 137 vertical levels. Atmospheric fields are interpolated from horizontal resolution equivalent of 31 km and 137 model levels on regular longitude-latitude grid of 0.25° and 37 pressure levels, and made available to the end-users. In this analysis we use the hourly data on pressure and single levels.

Japanese 55-year reanalysis (JRA-55) (Ebita et al., 2011; Kobayashi et al., 2015; Harada et al., 2016) is a global reanalysis produced by the Japan Meteorological Agency (JMA) and the Central Research Institute of Electric Power Industry (CRIEPI) based on the JMA Global Spectral Model (GSM). The reanalysis is available from 1958 onward. The reanalysis is based on the JMA operational assimilation system. JRA-55 uses a 4D-Var assimilation of surface, upper-air, satellite, ship-based and aircraft observations. The model uses 60 vertical levels and a horizontal grid with resolution approximately 60 km. In this analysis we use the 1.25° isobaric analysis and forecast fields interpolated to 37 pressure levels.

Modern-Era Retrospective analysis for Research and Applications (MERRA-2) (Gelaro et al., 2017) is a reanalysis produced by the NASA Global Modeling and Assimilation Office (GMAO). The reanalysis is based on the Goddard Earth Observing System (GEOS) atmospheric model. The model has approximately $0.5^\circ \times 0.65^\circ$ horizontal resolution and 72 vertical levels. It performs 3D-Var assimilation of station, upper-air, satellite, ship-based and aircraft data in 6-hourly cycles. In this analysis, we use the MERRA-2 3-hourly instantaneous model-level assimilated meteorological fields (M2I3NVASM) version 5.12.4 product.

The UK Met Office Unified Model (UM) (Walters et al., 2019) is an atmospheric model for weather forecasting and climate projection developed by the UK Met Office and the Unified Model Partnership. The UM is the atmospheric component, called Global Atmosphere (GA), of the HadGEM3-GC3.1 GCM and the UKESM1 earth system model (ESM). In this analysis we performed custom nudged runs of the UM (Telford et al., 2008) in the GA7.1 configuration with 20 min. time step and output temporal resolution on a New Zealand eScience Infrastructure (NeSI)/National Insti-

Table 3.3 | Table of physical quantities.

Symbol	Name	Units	Expression
α_s (α_e)	Volume scattering (extinction) coefficient	m^{-1}	
Ω	Solid angle	sr	
$P_\pi(\theta)$	Scattering phase function at angle θ	1	$\int_{4\pi} P_\pi(\theta) d\Omega = 4\pi$
β	Volume backscattering coefficient	$\text{m}^{-1}\text{sr}^{-1}$	$\beta = \alpha_s P_\pi(\pi)/(4\pi)$
N	Particle number concentration	m^{-3}	
$n(r)$	Number distribution of particle size	m^{-4}	$N = \int_0^\infty n(r) dr$
Q_s (Q_e)	Scattering (extinction) efficiency of spherical particles	1	$\alpha_s = Q_s \pi r^2 N$, $\alpha_e = Q_e \pi r^2 N$
Q_b	Backscattering efficiency of spherical particles	sr^{-1}	$\beta = Q_b \pi r^2 N$
S	Lidar ratio (extinction-to-backscatter ratio)	sr	$S = \alpha_e / \beta$
k	Backscatter-to-extinction ratio	sr^{-1}	$k = 1/S$
r_{eff}	Effective radius	m	$r_{\text{eff}} = \int_0^\infty r^3 n(r) dr / \int_0^\infty r^2 n(r) dr$
σ_{eff}	Effective standard deviation	m	$\sigma_{\text{eff}} = (\int_0^\infty (r - r_{\text{eff}})^2 r^2 n(r) dr) / (\int_0^\infty r^2 n(r) dr)$
k_B	Boltzmann constant	JK^{-1}	$k_B \approx 1.38 \times 10^{-23} \text{JK}^{-1}$
p	Atmospheric pressure	Pa	
T	Atmospheric temperature	K	
ρ	Liquid (or ice) density	$\text{kg}\cdot\text{m}^{-3}$	
ρ_{air}	Air density	$\text{kg}\cdot\text{m}^{-3}$	
q	Cloud liquid (or ice) mass mixing ratio	1	

tute of Water & Atmospheric Research (NIWA) supercomputer (Williams et al., 2016). The model was nudged to the ERA-Interim (Dee et al., 2011) atmospheric fields of horizontal wind speed and potential temperature and the HadISST sea surface temperature (SST) and sea ice dataset (Rayner et al., 2003). The model uses 85 vertical levels and a horizontal grid resolution of $1.875^\circ \times 1.25^\circ$.

3.4 Lidar simulator

The COSP lidar simulator Active Remote Sensing Simulator (ACTSIM) was introduced by Chiriaco et al. (2006) for the purpose of deriving simulated CALIOP measurements (Chepfer et al., 2007, 2008). The simulation is implemented by applying the lidar equation on model levels. Scattering by cloud particles and air molecules is calculated using the Mie and Rayleigh theory, respectively. CALIOP operates at a wavelength of 532 nm, and calculations in the original COSP simulator use this wavelength. We implemented a small set of changes to the lidar simulator to support a number of ALCs with different operating wavelengths.

The lidar equation (Emeis, 2010) is based on the radiative transfer equation (Goody and Yung, 1995; Liou, 2002; Petty, 2006; Zdunkowski et al., 2007), which relates transmission of radiation to scattering, emission and absorption in media such as the atmosphere. The lidar equation assumes laser radiation passes through a horizontally homogeneous atmosphere where it is absorbed and scattered. A fraction of laser radiation is scattered back to the instrument and reaches the receiver. Scattering and absorption in the atmosphere is determined by its constituents – gases, liquid droplets, ice crystals and aerosol particles. Atmospheric model output typically contains 4-dimensional fields of mass mixing ratios of liquid and ice and CF. The lidar equation can be applied on these output fields to simulate the backscattered radiation received by the instrument. Table 3.3 lists physical quantities used in the following sections. Here, we use radiative transfer notation similar to Petty (2006) and the notation of the original lidar simulator (Chiriaco et al., 2006).

3.4.1 Rayleigh and Mie scattering

The Rayleigh volume backscattering coefficient β_{mol} ($\text{m}^{-1}\text{sr}^{-1}$) in ACTSIM is parametrised by the following equation (Eq. (8) in Chiriaco et al. (2006)):

$$\beta_{\text{mol}} = \frac{p}{k_B T} (5.45 \times 10^{-32}) \left(\frac{\lambda}{550\text{nm}} \right)^{-4.09} = \frac{p}{k_B T} C_{\text{mol}}, \quad (3.1)$$

where for lidar wavelength $\lambda = 532$ nm, $C_{\text{mol}} = 6.2446 \times 10^{-32}$; k_B is the Boltzmann constant $k_B \approx 1.38 \times 10^{-23}$ JK⁻¹, p is the atmospheric pressure and T is the atmospheric temperature. We multiply this equation by $\exp(4.09(\log(532) - \log(\lambda)))$ (where the value of λ is in nm) to get molecular backscatter for wavelengths other than 532 nm, which allows us to support multiple commercially available instruments. The strength of molecular backscattering is usually lower than backscattering from clouds for the relevant wavelengths.

The lidar signal at visible or near IR wavelengths is scattered by atmospheric constituents of similar size in the Mie scattering regime (Mie, 1908). In the most simple approximation, one can assume spherical dielectric particles. The scattering from these particles depends on the relative size of the wavelength and the (spherical) particle radius r , expressed by the dimensionless size parameter x :

$$x = \frac{2\pi r}{\lambda}. \quad (3.2)$$

While the wavelength is approximately constant during the operation of the lidar², the particle size comes from a distribution of sizes, typically approximated in NWP models and GCMs by a Gamma or log-normal distribution with a given mean and standard deviation. Some models provide the mean as effective radius r_{eff} . If the effective radius is not provided by the model, the lidar simulator assumes a value $r_{\text{eff}} = 30$ μm by default, which is the value assumed by the original ACTSIM simulator.

In order to support multiple laser wavelengths, it is necessary to calculate backscattering efficiency due to scattering by a distribution of particle sizes. We use the computer code MIEV developed by Warren J Wiscombe (Wiscombe, 1979, 1980) to calculate backscattering efficiency for a range of the size parameter x and integrate for a distribution of particle sizes. The resulting pre-calculated LR (extinction-to-backscatter ratio) as a function of the effective radius is included in the lidar simulator for fast lookup during the simulation.

Cloud droplet and ice crystal size distribution parameters are an important assumption in lidar simulation due to the dependence of Mie scattering on the ratio of wavelength and particle size (the size parameter x). NWP models and GCMs traditionally use the effective radius r_{eff} and effective standard deviation σ_{eff} (or an equivalent parameter such as effective variance ν_{eff}) to parametrise this distribution. Knowledge of the real distribution is likely highly uncertain due to a large variety of clouds occurring globally and the limited ability to predict microphysical cloud properties

²The actual lidar wavelength is not constant and is characterised by a central wavelength and width. The central wavelength may fluctuate with temperature (Wiegner and Gasteiger, 2015).

Table 3.4 | Table of sensitivity tests of theoretical distribution assumption, effective radius r_{eff} and effective standard deviation σ_{eff} of the cloud droplet/ice crystal size distribution. μ and σ are the mean and standard deviation of a normal distribution corresponding to the log-normal distribution, calculated numerically from r_{eff} and σ_{eff} . μ_* and σ_* are the actual mean and standard deviation of the distribution (calculated numerically).

Distribution	r_{eff} (μm)	σ_{eff} (μm)	μ	σ	μ_* (μm)	σ_* (μm)
log-normal	20	10	2.44	0.47	12.76	6.26
log-normal	20	5	2.84	0.25	17.72	4.43
log-normal	10	5	1.74	0.47	6.40	3.20
Gamma	20	10			9.98	7.00
Gamma	20	5			17.50	4.68
Gamma	10	5			5.00	3.54

in models. In this section we introduce theoretical assumptions used in the lidar simulator based on established definitions of the effective radius and effective standard deviation and two common distributions. Edwards and Slingo (1996) discuss the effective radius in the context of model radiation schemes, and we will primarily follow the definitions detailed in Chang and Li (2001) and Petty and Huang (2011). The practical result of this section (and the corresponding offline code) is pre-calculated backscatter-to-extinction ratios as a function of the effective radius in the form of a lookup table included in the lidar simulator, and used in the online calculations. The offline code is provided and can be re-used for calculation of the necessary lookup tables for different lidar wavelengths, should the user of the code want to support another instrument.

The effective radius r_{eff} and effective standard deviation σ_{eff} are defined by:

$$r_{\text{eff}} = \frac{\int_0^\infty r^3 n(r) dr}{\int_0^\infty r^2 n(r) dr}, \quad \sigma_{\text{eff}}^2 = \frac{\int_0^\infty (r - r_{\text{eff}})^2 r^2 n(r) dr}{\int_0^\infty r^2 n(r) dr}, \quad (3.3)$$

where $n(r)$ is the probability density function (PDF) of the distribution. Here, we follow Petty and Huang (2011), who define the effective variance ν_{eff} which relates to σ_{eff} by $\nu_{\text{eff}} = \sigma_{\text{eff}}^2 / r_{\text{eff}}^2$. Due to lack of knowledge about the real distribution of particle radii, it has to be modelled by a theoretical distribution, such as a log-normal or Gamma distribution. The original ACTSIM simulator assumes a log-normal distribution (Chiriaco et al., 2006) with the PDF:

$$n(r) \propto \frac{1}{r} \exp\left(-\frac{(\log r - \mu)^2}{2\sigma^2}\right), \quad (3.4)$$

where μ and σ are the mean and the standard deviation of the corresponding normal distribution, respectively. Chiriaco et al. (2006) use the value of $\sigma = \log(1.2) = 0.18$ "for ice clouds" (the value for liquid cloud does not appear to be documented). In our parametrisation we used a combination of r_{eff} and σ_{eff} to constrain the theoretical distribution, where the effective standard deviation σ_{eff} was assumed to be one fourth of the effective radius r_{eff} . This choice is approximately consistent with $\sigma = \log(1.2) = 0.18$ at $r_{\text{eff}} = 20 \mu\text{m}$ (see Table 3.4, described below). In future updates, the values could be based on in situ studies of size distribution or taken from the atmospheric model output if available.

From the expression for the n-th moment of the log-normal distribution $E[X^n] = \exp(n\mu +$

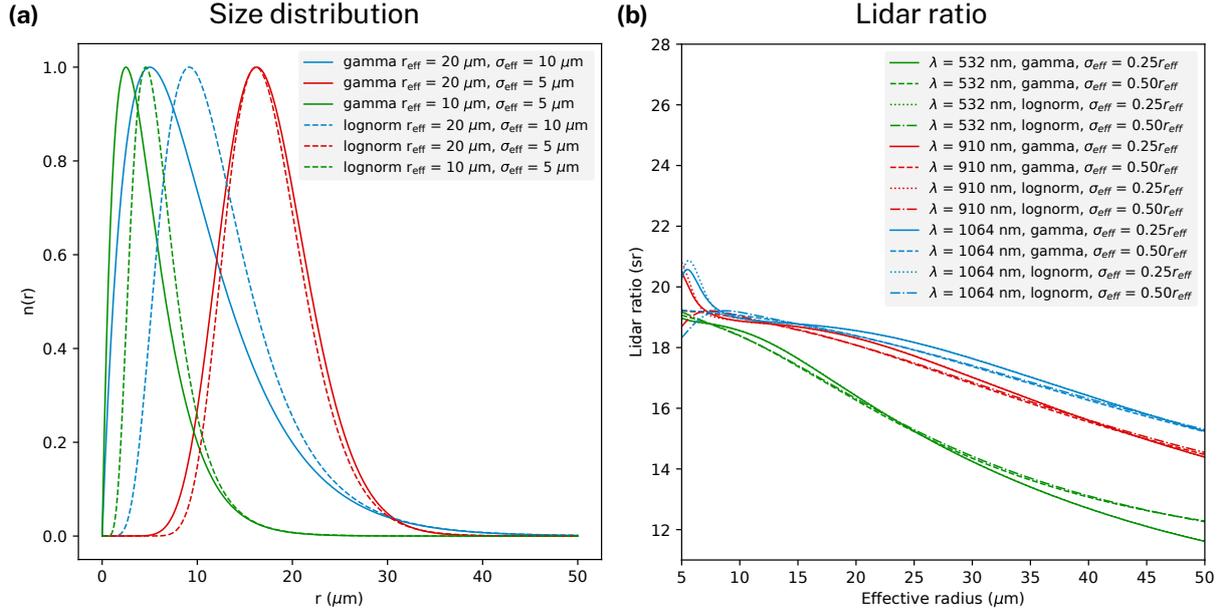


Figure 3.2 | (a) Theoretical distributions of cloud droplet radius based on the log-normal and Gamma distributions parametrised by multiple choices of the effective radius r_{eff} and effective standard deviation σ_{eff} . **(b)** Lidar ratio (LR) as a function of effective radius calculated for different theoretical cloud droplet size distributions, laser wavelengths and effective standard deviation ratios.

$n^2 \frac{\sigma^2}{2}$) and Equation (3.3) we calculate r_{eff} and σ_{eff} of the log-normal distribution:

$$r_{\text{eff}} = \frac{E[r^3]}{E[r^2]} = \exp\left(\mu + \frac{5}{2}\sigma^2\right), \quad (3.5)$$

$$\begin{aligned} \sigma_{\text{eff}}^2 &= \frac{E[(r - r_{\text{eff}})^2 r^2]}{E[r^2]} = \frac{E[r^4] - 2E[r^3]r_{\text{eff}} + r_{\text{eff}}^2 E[r^2]}{E[r^2]} = \frac{\exp(4\mu + 8\sigma^2) - \exp(4\mu + 7\sigma^2)}{\exp(2\mu + 2\sigma^2)} = \\ &= \exp(2\mu + 6\sigma^2) - \exp(2\mu + 5\sigma^2). \end{aligned} \quad (3.6)$$

We find μ and σ for given r_{eff} and σ_{eff} numerically by root-finding using the equations above. In practice, we find that the root-finding converges well for r_{eff} between 5 and 50 μm , which is the range mostly likely to be applicable in practice.

The Gamma distribution follows the PDF:

$$n(r) \propto r^{(1-3\nu_{\text{eff}})/\nu_{\text{eff}}} \exp\left(-\frac{r}{r_{\text{eff}}\nu_{\text{eff}}}\right) \quad (3.7)$$

(see e.g. Eq. 13 in Petty and Huang (2011) or Eq. 1 in Bréon and Doutriaux-Boucher (2005)). In this case, the distribution depends explicitly on r_{eff} and σ_{eff} , and as such does not require numerical root-finding.

Figure 3.2a shows the log-normal and Gamma distributions calculated for a number of r_{eff} and σ_{eff} values, and Table 3.4 summarises properties of these distributions. The actual mean and standard deviation of the distributions do not necessarily correspond well with the effective radius and effective standard deviation.

In ACTSIM, the volume extinction coefficient α_e is calculated by integrating the extinction by individual particles over the particle size distribution:

$$\alpha_e = \int_0^\infty Q_e \pi r^2 n(r) dr \approx Q_e \pi \int_0^\infty r^2 n(r) dr = Q_e \frac{3q\rho_{\text{air}}}{4\rho r_{\text{eff}}}, \quad (3.8)$$

assuming approximately constant extinction efficiency $Q_e \approx 2$ (which is approximately true for the interesting range of r_{eff} and laser wavelengths), and using the relationship between the cloud liquid (or ice) mass mixing ratio q and $\int_0^\infty r^2 n(r) dr$:

$$q\rho_{\text{air}} = \int_0^\infty \frac{4}{3}\pi r^3 \rho n(r) dr = \frac{4}{3}\pi\rho \int_0^\infty r^3 n(r) dr = \frac{4}{3}\pi\rho r_{\text{eff}} \int_0^\infty r^2 n(r) dr, \quad (3.9)$$

where ρ and ρ_{air} are the densities of liquid water (or ice) and air, respectively.

Likewise, the volume backscatter coefficient from particles β_p is calculated by integrating backscattering by individual particles over the particle size distribution:

$$\beta_p = \int_0^\infty Q_s \pi r^2 \frac{P_\pi(\pi)}{4\pi} n(r) dr. \quad (3.10)$$

where Q_s is scattering efficiency and $P_\pi(\pi)$ is scattering phase function at 180° . Since the normalisation of $n(r)$ is not known until the online phase of calculation, the backscatter-to-extinction ratio from particles $k_p = \beta/\alpha_e$ can be calculated offline instead (the requirement for normalisation of $n(r)$ is avoided by appearing in both the numerator and denominator):

$$k_p = \beta_p/\alpha_e = \frac{\int_0^\infty Q_s r^2 P_\pi(\pi)/(4\pi) n(r) dr}{\int_0^\infty Q_e r^2 n(r) dr}. \quad (3.11)$$

We pre-calculate this integral numerically for a permissible interval of r_{eff} (5–50 μm) at 500 evenly spaced wavelengths, and store the result as a lookup table for the online phase. The integral in the numerator is numerically hard to calculate due to strong dependency of $P_\pi(\pi)$ on r . Figure 3.2b shows LR as a function of r_{eff} , calculated for log-normal and Gamma particle size distributions with $\sigma_{\text{eff}} = 0.25r_{\text{eff}}$ and $\sigma_{\text{eff}} = 0.5r_{\text{eff}}$. This corresponds to the lookup table we use in the online phase of the lidar simulator. As can be seen in Fig. 3.2, LR depends only weakly on the choice of the distribution type and the effective standard deviation ratio.

3.4.2 Cloud overlap and cloud fraction

Model cloud is defined by the liquid and ice mass mixing ratio and cloud fraction in each atmospheric layer. The lidar simulator simulates radiation passing vertically at a random location within the grid cell. Therefore, it is necessary to generate a random vertical cloud overlap based on the cloud fraction in each layer, as the overlap is not defined explicitly in the model output. Two common methods of generating overlap are the random and maximum–random overlap (Geleyn and Hollingsworth, 1979). In the random overlap method, each layer is either cloudy or clear with a probability given by CF, independent of other layers. The maximum–random overlap assumes that adjacent layers with non-zero CF are maximally overlapped, whereas layers separated by zero CF layers are randomly overlapped. COSP implements cloud overlap generation in the Subgrid

Cloud Overlap Profile Sampler (SCOPS) (Klein and Jakob, 1999; Webb et al., 2001; Chepfer et al., 2008). The ALC lidar simulator uses SCOPS to generate 10 random subcolumns for each profile, using the maximum–random overlap assumption as the default setting of a user-configurable option. The backscatter profile and cloud occurrence can be plotted for any subcolumn. Due to the random nature of the overlap, the backscatter profile may differ from the observed profile even if the model is correct in its cloud simulation. The random overlap generation should, however, result in unbiased cloud statistics.

3.4.3 Multiple scattering

Due to a finite FOV of the lidar receiver, a fraction of the laser radiation scattered forward will remain in FOV. Therefore, the effective attenuation is smaller than calculated with the assumption that all but the backscattered radiation is removed from FOV and cannot reach the receiver. The forward scattering can be repeated multiple times before a fraction of the radiation is backscattered, eventually reaching the receiver. To account for this multiple scattering effect, the COSP lidar simulator uses a multiple scattering correction coefficient η , by which the volume scattering coefficient is multiplied before calculating the layer optical thickness (Chiriaco et al., 2006; Chepfer et al., 2007, 2008). The theoretical value of η is between 0 and 1 and depends on the receiver FOV and optical properties of the cloud. For CALIOP at $\lambda = 532$ nm a value of 0.7 is used in the COSP lidar simulator. Hogan (2006) implemented fast approximate multiple scattering code. This code has recently been used by Hopkin et al. (2019) in their ceilometer calibration method. They noted that η is usually between 0.7 and 0.85 for wavelengths between 905 and 1064 nm. The ALC simulator presented here does not use an explicit calculation of η , but retains the value of $\eta = 0.7$. This value is also used when calculating LR from the vertically integrated backscatter (Sect. 3.5.2). The code of Hogan (2006) "Multiscatter" is publicly available (<http://www.met.reading.ac.uk/clouds/multiscatter/>) and could be used in a later version of the framework to improve the accuracy of simulated attenuation and calibration.

3.5 Lidar data processing

Scheme in Fig. 3.1 outlines the processing done in the framework. The individual processing steps are described below.

3.5.1 Noise and subsampling

ALC signal reception is affected by a number of sources of noise such as sunlight and electronic noise (Kotthaus et al., 2016). Range-independent noise can be removed by assuming that the attenuated volume backscatter coefficient at the highest range gate is dominated by noise. This is true if the highest range is not affected by clouds, aerosol, and if contributions from molecular scattering are negligible. The supported instruments have a range of approximately 8 (CL31), 15 (CL51, CHM 15k) and 30 km (MiniMPL). By assuming the distribution of noise at the highest level is approximately normal, the mean and standard deviation can be calculated from a sample over a period of time such as 5 minutes, which is short enough to assume the noise is constant over this period, and long enough to achieve accurate estimates of the standard deviation. The mean and standard

Table 3.5 | Theoretical molecular backscatter calculated at pressure 1000 hPa and temperature 20°C and the calibration coefficient, relative to the instrument native units, determined for the instrument based on the molecular backscatter and stratocumulus lidar ratio calibration methods.

Instrument	Wavelength (nm)	Molecular backscatter ($\times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$)	Calibration coefficient
CHM 15k	1064	0.0906	0.34
CL31	910	0.172	1.45×10^{-3}
CL51	910	0.172	1.2×10^{-3}
MiniMPL	532	1.54	3.75×10^{-6}

deviation can then be scaled by the square of the range to estimate the distribution of range independent noise at each range bin. By subtracting the noise mean from the measured backscatter we get the expected backscatter. The result of the noise removal algorithm is the expected backscatter and its standard deviation at each range bin.

3.5.2 Backscatter calibration

ALCs often report backscatter in arbitrary units (a.u.) or as NRB (MiniMPL). If they report it in units of $\text{m}^{-1} \text{sr}^{-1}$, these values are often not calibrated to represent the true absolute backscatter. Assuming that range-dependent corrections (overlap, dead time and afterpulse) have been applied on the backscatter in a. u., the reported backscatter is proportional to the true attenuated backscatter (inclusive of noise backscatter). In order to have a comparable quantity to the lidar simulator and consistent input to the subsequent processing (e.g. cloud detection), calibration is required. Several methods of calibration have been described previously: calibration based on LR in fully attenuating liquid stratocumulus clouds (O'Connor et al., 2004; Hopkin et al., 2019), calibration based on molecular backscatter (Wiegner et al., 2014) and calibration based on a high spectral resolution lidar reference (Heese et al., 2010; Jin et al., 2015). In addition, calibration can be assisted by sunphotometer or radiosonde measurements (Wiegner et al., 2014).

A relatively large variability of the calibration coefficient has been determined for instruments of the same model (Hopkin et al., 2019). However, past studies can be useful for determining an approximate value of the coefficient before applying one of the calibration methods. For the CL51, Jin et al. (2015) reported a value of 1.2 ± 0.1 based on a multi-wavelength lidar reference. Hopkin et al. (2019) reported mean values 1.4–1.5 for a number of CL31 instruments (software version 202). For CHM 15k, Hopkin et al. (2019) reported mean values between 0.3 and 0.8 for a majority of the instruments examined. The ALCF provides per-instrument default values of the calibration coefficient (Table 3.5), but a unit-specific coefficient should be determined for an analysed instrument during the lidar data processing step.

Calibration based on LR in fully opaque liquid stratocumulus clouds has been applied successfully on large networks of ALCs. It utilises the fact that given suitable conditions vertically integrated backscatter is proportional to LR of the cloud, which can be theoretically derived if the cloud droplet effective radius can be assumed. The theoretically derived value is about 18.8 sr for common ALC wavelengths and a relatively large range of effective radii (O'Connor et al., 2004). Another factor which needs to be known or assumed is the multiple scattering coefficient, which

tends to be about 0.7-1.0 in common ALCs. Due to its relatively simple requirements, this method is possibly the easiest ALC calibration method. The ALCF implements this calibration method by letting the user identify time periods with fully opaque liquid stratocumulus cloud, for which the mean LR is calculated. The ratio of the observed LR and the theoretical LR is equivalent to the calibration coefficient. This implementation, while very easy to perform, has multiple limitations, some of which are highlighted by Hopkin et al. (2019):

1. Aerosol can cause additional attenuation and scattering, which results in LR which is different from the theoretical value by an unknown factor. Therefore, a frequent re-calibration may be necessary.
2. The multiple scattering coefficient assumption may not be accurate for the given instrument.
3. The 910 nm wavelength of CL31 and CL51 is affected by water vapour absorption which causes additional attenuation, which is currently not taken into account in the calculation of LR.
4. Near-range backscatter retrieval is affected by receiver saturation and incomplete overlap. Therefore, using stratocumulus clouds above approximately 2 km for this calibration method is recommended. This range is instrument dependent.
5. The composition of the stratocumulus cloud may be uncertain. At temperature between 0 and -30°C these clouds may contain both liquid and ice which results in a different LR than expected.

These limitations could be addressed in the future by (1) using sunphotometer observations as an optional input to determine the aerosol optical depth (AOD), (2) calculating the multiple scattering coefficient more accurately (such as with the Multiscatter package of Hogan (2006)), (3) calculating the water vapour absorption explicitly based on water vapour, temperature and pressure fields from a reanalysis or radiosonde profile data, (4) correcting the near-range backscatter based on the integrated backscatter distribution as a function of height of the maximum backscatter (Hopkin et al., 2019, Sect. 5.1), (5) combining the backscatter profile with temperature field from a reanalysis to exclude cold clouds.

Molecular (Rayleigh) backscattering can be accurately calculated if temperature and pressure of the atmospheric profile is known (Sect. 3.4.1). This can be employed for absolute calibration of ALCs. Given the low SNR of low-power ALCs, several hours of integration are required to identify the molecular backscatter (Wiegner et al., 2014). The molecular backscatter is attenuated by an unknown amount of aerosol with unknown LR, and the near-range backscatter is affected by a potentially inaccurate overlap correction. Therefore, this method alone produces calibration coefficient which depend on the atmospheric conditions. We found that all studied ALCs except for the CL31 are capable of observing the molecular backscatter (Section 3.7). Therefore, this method may be used in addition to the liquid stratocumulus LR method for cross-validation of the calibration.

3.5.3 Cloud detection

Cloud is the most strongly attenuating feature in ALC backscatter measurements. Due to this attenuation, the lidar signal is quickly attenuated in thick cloud and can fall below the noise level before reaching the top of the cloud. This means that the first cloud base can be detected reliably (unless the cloud is too thin or too high and obscured by backscatter noise), while the cloud top or multi-layer cloud cannot be observed reliably under all conditions. The opposite is true for spaceborne lidars, which can detect the cloud top reliably but cannot always detect the cloud base. Therefore, ALC observations can be regarded as complementary to spaceborne lidar observations. By applying a suitable algorithm, one can detect CBH, CTH and identify cloud layers. Instrument firmware often determines CBH and sometimes cloud layers as part of its internal processing, often using an undisclosed algorithm which is not comparable between different instruments and potentially not even different versions of the instrument firmware (Kotthaus et al., 2016). Mattis et al. (2016) compared a large number of ALCs and found differences of up to 70 m between the reported CBH, and others found relatively large differences as well (Liu et al., 2015b; Silber et al., 2018). Alternatively to the instrument reported CBH and cloud layers, it is possible to detect cloud based on the backscatter profile. A relatively large number of cloud detection algorithms have been proposed (Wang and Sassen, 2001; Morille et al., 2007; Martucci et al., 2010; Van Tricht et al., 2014; Silber et al., 2018; Cromwell and Flynn, 2019). We use a simple algorithm based on a backscatter threshold applied on the denoised attenuated backscatter, assuming that the noise can be represented by a normal distribution at the highest range, which is unlikely to contain cloud or aerosol if the instrument is pointing vertically (this may not be true, however, for CL31 which has a maximum range of just 7.7 km). This assumption neglects the range-dependent molecular backscatter, which is relatively small at the ceilometer wavelengths examined (910 nm and 1064 nm). A cloud mask is determined positive where the backscatter is greater than a chosen threshold plus three standard deviations of noise at the given range. A threshold of $2 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$ was found to be a good compromise between false detection and misses. This value is above the maximum molecular backscatter, which is approximately $1.54 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$ at the surface in the case of the MiniMPL (wavelength 532 nm). Noise is not simulated by the lidar simulator, but the cloud detection algorithm allows for coupling of simulated and observed profiles, whereby the noise standard deviation is taken from the corresponding location in the observed profile. With 5 minute averaging, when the standard deviation of noise is relatively low, we found that the coupling does not make substantial differences to the detected cloud (not shown). While the threshold-based algorithm is less sophisticated than other methods of cloud detection, the vertical resolution of the simulated backscatter is likely too low and the vertical derivatives of the simulated backscatter too crudely represented (as can be seen in Fig. 3.5b–f, 3.6b–f, 3.7b–e) to apply any algorithm based on the vertical derivatives of backscatter. Using the same cloud detection algorithm on observed and simulated backscatter is essential for an unbiased one-to-one comparison of cloud.

3.5.4 Water vapour absorption

Previous studies have noted that ceilometers which utilise the wavelength of 910 nm such as the Vaisala CL31 and CL51 are affected by additional absorption of laser radiation by water vapour



Figure 3.3 | Map showing the location of sites. Data at three sites in New Zealand were analysed: Cass, Lauder and Christchurch.

(Wiegner and Gasteiger, 2015; Wiegner et al., 2019; Hopkin et al., 2019). The wavelength coincides with water vapour absorption bands between 900 and 930 nm, while the other common ceilometer wavelength of 1064 nm is not affected. Wiegner and Gasteiger (2015) reported that it can cause absorption of the order of 20% in the extratropics and 50% in the tropics. The lidar simulator does not currently account for this. However, as the water vapour concentration is available from the reanalyses and models, it should be possible to use a line-by-line model to calculate the water vapour volume absorption coefficient for each vertical layer during the integration process. Water vapour also affects calibration of observed backscatter. In order to use the liquid stratocumulus LR calibration method, the backscatter has to be corrected for water vapour absorption to achieve high accuracy of calibration. Hopkin et al. (2019) used a simplified approach based on a parametrised curve and reported a difference from explicit radiative transfer calculations of 2% in the United Kingdom atmosphere (Middle Wallop). In the future either approach should be used to include water vapour absorption in the simulator, or remove the effect of water vapour absorption from the observed lidar backscatter to achieve an improved one-to-one comparison between the observations, reanalyses and models.

3.6 Description of case studies

The case studies analysed here were selected to include all instruments supported by the framework. We compare four different instruments (CHM 15k, CL31, CL51, MiniMPL) deployed at three locations in NZ (Lauder, Christchurch, Cass) with three reanalyses (MERRA-2, ERA5, JRA-55),

Table 3.6 | Location of sites and instruments. The time periods are inclusive.

Site	Coordinates	Surface altitude (m)	Instruments	Time period	Missing period	Days
Cass	43.0346°S 171.7594°E	577	CL51	19 Sep–1 Oct 2014		13
Lauder	45.0379°S 169.6831°E	370	MiniMPL, CL31	12–24 Jan 2018		13
Christchurch	43.5225°S 172.5841°E	45	MiniMPL, CHM 15k	17 July–19 August 2019	22–31 July	24

one NWP model (AMPS) and one GCM (UM). These case studies aim to demonstrate capability rather than to comprehensively evaluate cloud simulation in the models and reanalyses. The work detailed in Kuma et al. (2020a) provides a detailed evaluation of the UM and MERRA-2 relative to shipborne ceilometer observations. Figure 3.3 shows the location of the sites and Table 3.6 summarises the case studies, which are also described in greater detail below. The sites were chosen from available datasets to demonstrate the use of the framework with all supported instruments. Two of the sites also had co-located instruments: CL31 and MiniMPL in Lauder, and CHM 15k and MiniMPL in Christchurch. The MiniMPL in Lauder and Christchurch were two different units.

Cass is a field station of the University of Canterbury located at an altitude of 577 m in the Southern Alps of the South Island of NZ. The station is located far from any settlements and likely affected little by anthropogenic aerosol relative to the other sites. We have analysed 13 days of observations with a CL51 at this station performed in September and October 2014.

Lauder is a field station of NIWA located inland in the Central Otago region on the South Island of NZ. The station is situated in a rural area relatively far from large human settlements. We have analysed 13 days of co-located MiniMPL and CL31 observations made in January 2018. The MiniMPL was operated in a fixed vertical scanning mode during this period (elevation angle 90°).

Observations at the Christchurch site were performed at the University of Canterbury campus on the Ernest Rutherford building rooftop. Christchurch is located on the east coast of the South Island of NZ. Its climate is affected by the ocean, its proximity to the hilly area of the Banks Peninsula, the Canterbury Plains and föhn-type winds (Canterbury northwester) resulting from its position on the lee side of the Southern Alps. The city is affected by significant wintertime air pollution from domestic wood burning and transport. The orography of the city and the adjacent Canterbury Plains is very flat, making it prone to inversions. The Ernest Rutherford building is a 5 floor building situated in an urban area, surrounded by multiple buildings of similar height. We have analysed 24 days of co-located MiniMPL and CHM 15k observations performed in July and August 2019. The MiniMPL was operated in a fixed vertical scanning mode (elevation angle 90°). The nudged run of the UM was only available up to year 2018. Therefore, it was not analysed for this site.

3.7 Case study results

To demonstrate the ways that the ALCF can be used we compared a total of 50 days of ALC observations with simulated lidar backscatter at three sites in NZ (Sect. 3.6). The observed backscatter was normalised to calibrated absolute range-corrected total volume backscattering coefficient. The noise mean as determined at the furthest range was removed from the backscatter. Cloud detection based on an attenuated absolute backscatter coefficient threshold of $2 \times 10^{-6} \text{m}^{-1} \text{sr}^{-2}$ and three

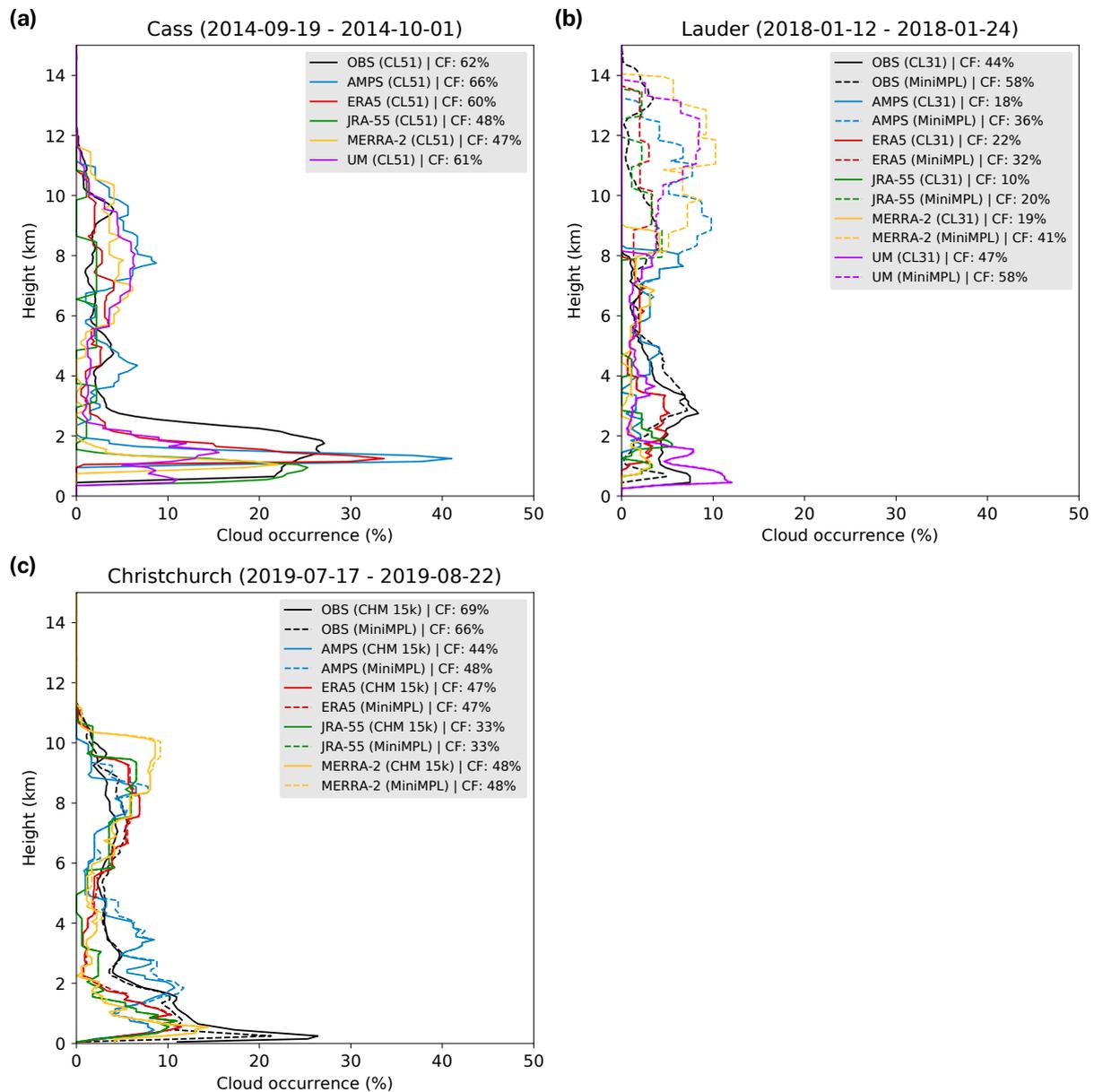


Figure 3.4 | Cloud occurrence histogram as a function of height above the mean sea level observed at three sites and simulated by the lidar simulator based on atmospheric fields for five reanalyses and models. Shown is also the total cloud fraction (CF). The histogram is calculated from the cloud mask as determined by the cloud detection algorithm.

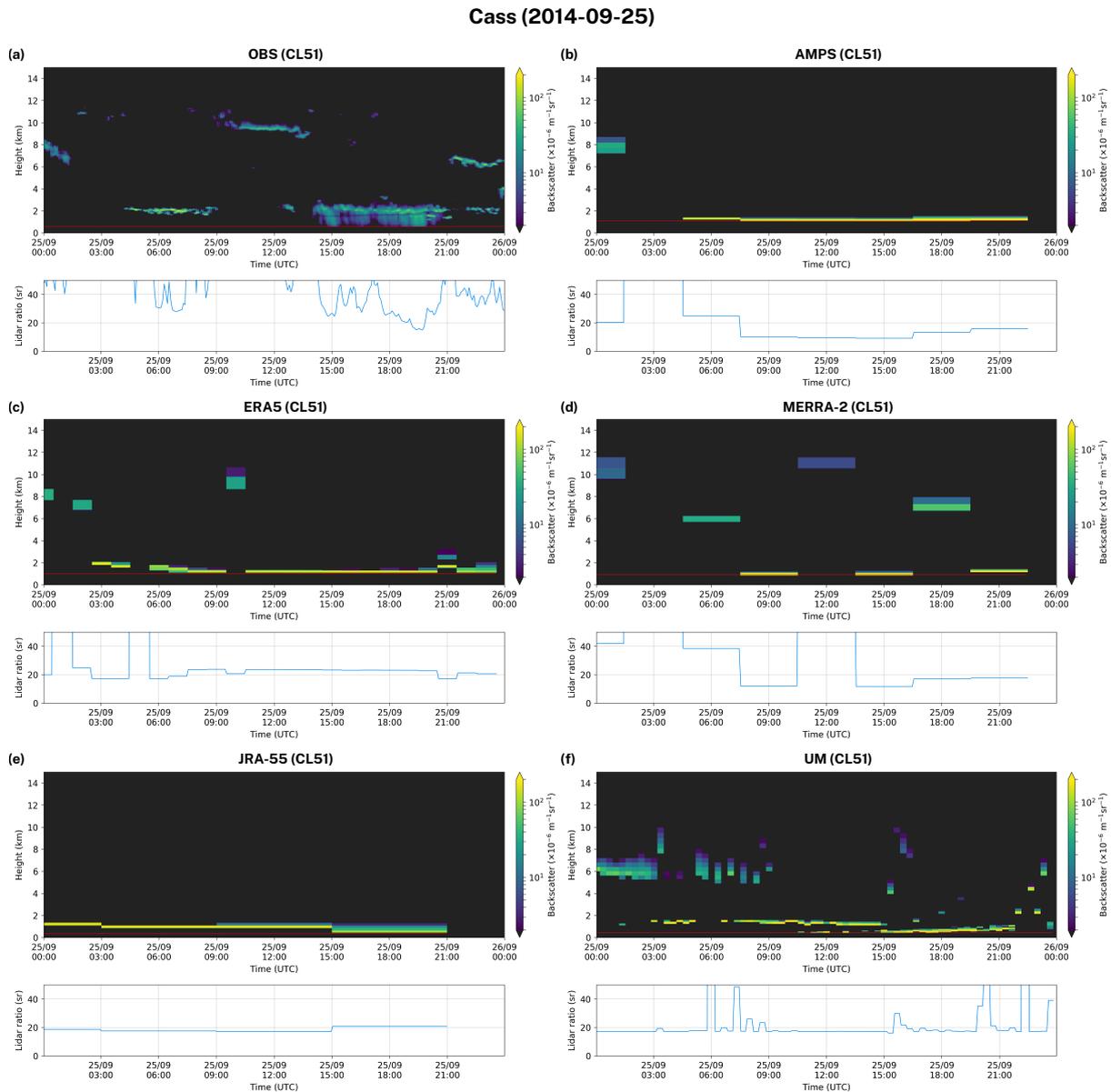


Figure 3.5 | Examples of observed (OBS) and simulated backscatter profiles during 24 hours taken from the three case studies presented here. The observed backscatter was normalised to absolute units and denoised. The vertical levels were interpolated based on the nearest neighbour assumption. The cloud detection threshold of $2 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$ corresponds the low end of the colour scale, i.e. all visible backscatter in the plots was determined as cloud, unless the difference from the threshold was smaller than three noise standard deviations. The first subcolumn of 10 columns generated by the Subgrid Cloud Overlap Profile Sampler (SCOPS) was selected to make the plots. The red line signifies the station altitude.

Lauder (2018-01-16)

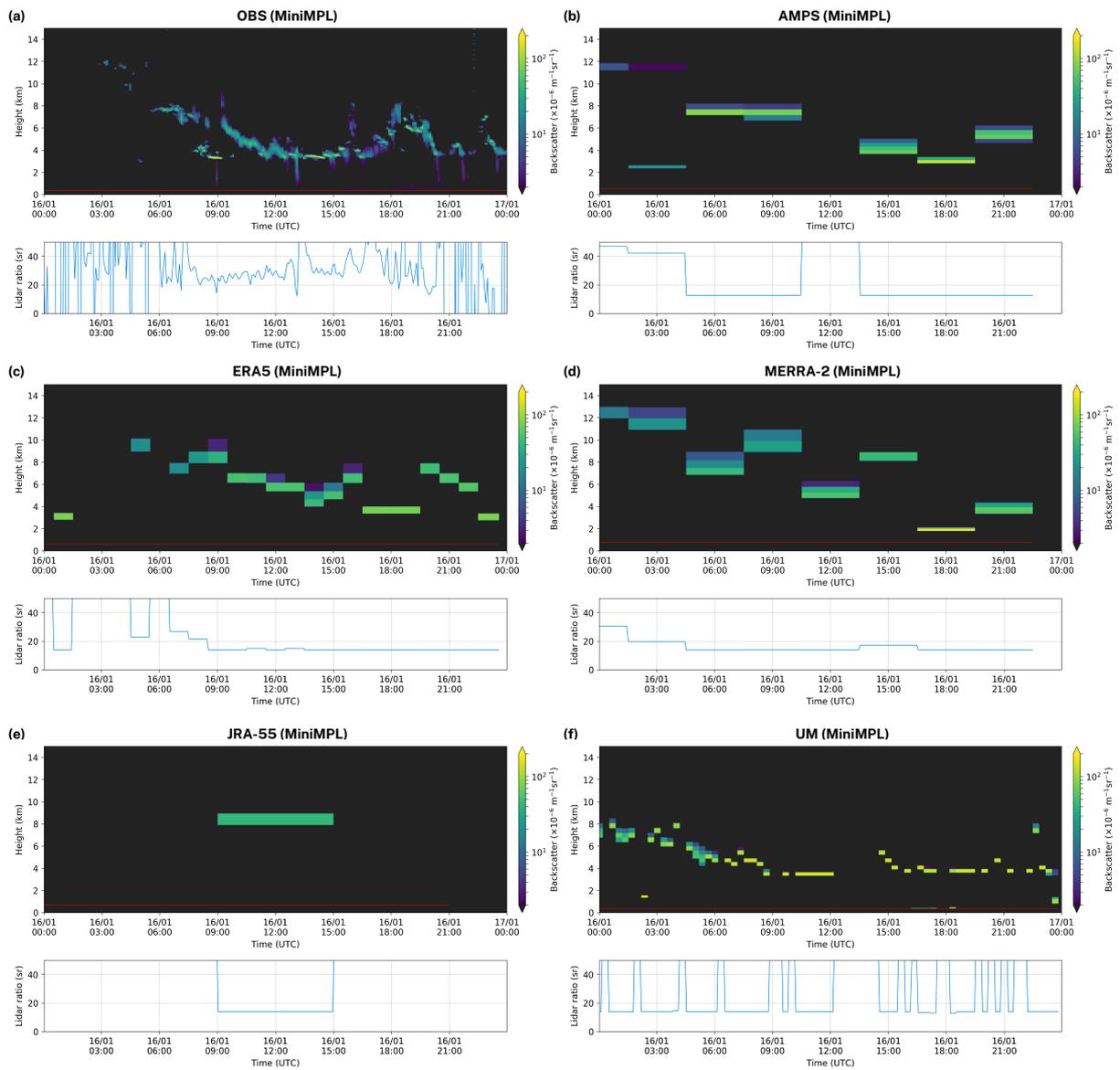


Figure 3.6 | The same as Fig. 3.5 but for the Lauder site.

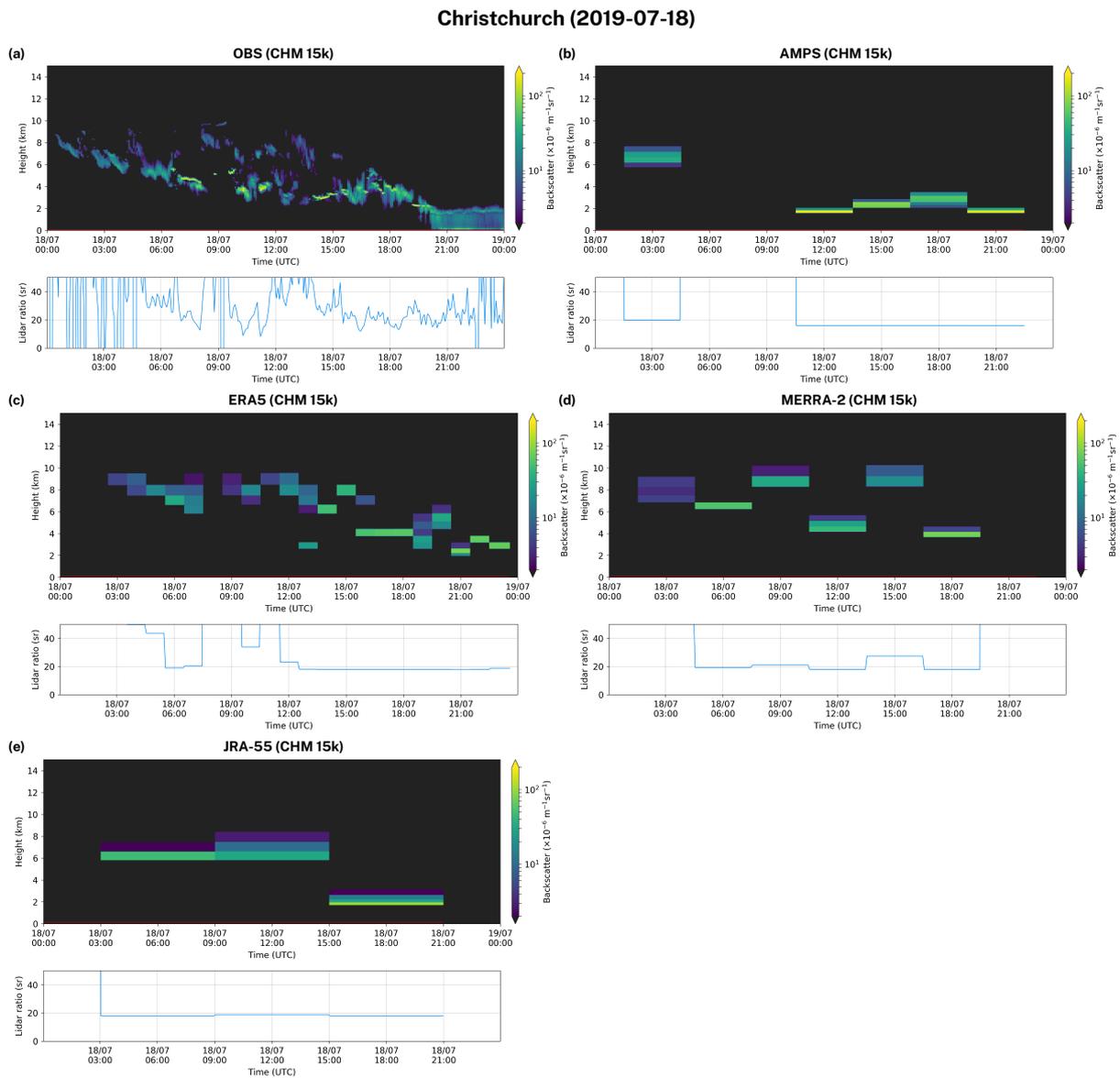


Figure 3.7 | The same as Fig. 3.5 but for the Christchurch site.

noise standard deviations was applied to derive a cloud mask and CBH. We compare the statistical cloud occurrence as a function of height above the mean sea level (ASL) (Fig. 3.4) and individual backscatter profiles (selected profiles are shown in Fig. 3.5, 3.6 and 3.7) in this section.

We analysed 13 days of CL51 observations at the Cass field station in late winter. Due to the location of the station at a relatively high altitude in a varied terrain of the Southern Alps, the models with their relatively coarse horizontal grid resolution do not represent the terrain and position accurately. The orography representation of the models meant that the virtual altitude of the station was 1115 m (AMPS), 1051 m (ERA5), 401 m (JRA-55), 914 m (MERRA-2) and 428 m (UM). The virtual position, which is the centre of the nearest model grid cell to the site location, ranged from relatively close in the Southern Alps (AMPS, ERA5, MERRA-2, UM) to relatively far on the West Coast of NZ (JRA-55) depending on the horizontal resolution of the grid. The time period examined was characterised by diverse cloud occurrence with periods of low cloud and precipitation, mid-level cloud, fog, high cloud and clear skies. Precipitation, currently not simulated by the lidar simulator, was present in about 18% of the observed backscatter profiles, as determined by visual inspection. Figure 3.4a shows that predominantly low cloud and precipitation between the ground and 2 km ASL in 25% of profiles was observed. Cloud between 3 and 12 km ASL was observed about evenly in 2% of profiles. While the reanalyses and models were able to partially reproduce the peak of cloud occurrence near 1 km ASL, the peak they displayed is less vertically broad than observed, and in the UM the peak was much weaker than observed. The lack of precipitation simulation might also have contributed to this apparent difference between observed and simulated cloud. Above 3 km ASL, the reanalyses and models tended to overestimate cloud, with only ERA5 and JRA-55 simulating close to the observed cloud occurrence. The observed total CF was 62%. AMPS overestimated this value by 4 percentage points (pp), ERA5 and the UM reproduced almost the exact value (within 2 pp), while the other reanalyses (JRA-55 and MERRA-2) underestimated CF by about 15 pp.

We also analysed 13 days of CL31 and MiniMPL observations at the Lauder station in summer. During the time period relatively diverse cloud was observed, with periods of low, mid- and high cloud, clear sky and a small fraction of profiles with precipitation (about 3%). The altitude of the station of 370 m ASL generally had a much higher equivalent in the reanalyses and models: 565 m (AMPS), 642 m (ERA5), 681 m (JRA-55) and 786 m (MERRA-2) due to the presence of hills in the surrounding region (the station is in a high valley), with the exception of the UM where the altitude was 385 m. The virtual station position in the reanalyses and models ranged from relatively close to the station in the same geographical region (AMPS, ERA5), a nearby location in a more hilly region (JRA-55), a relatively distant location in the adjacent Dunstan Mountains (MERRA-2) and a relatively distant location in Central Otago (UM). Figure 3.4b shows that the CL31 observed relatively even cloud occurrence between the ground and 3 km ASL at 8%, falling off to about 3% between 4 and 8 km ASL (the maximum lidar range of CL31 is 7.7 km). The MiniMPL observed much weaker backscatter than CL31 below 3 km ASL, which was identified as an overlap calibration issue in the MiniMPL. The MiniMPL observed substantial amounts of cloud above 8 km, not present in the CL31 observations due to its range limitation. Overall, the observed cloud occurrence had two peaks at ground to 3 km ASL and at about 9 km ASL. The simulated cloud occurrence was generally underestimated between the ground and 5 km ASL, with

the exception of the UM which reproduced the lower half of the peak accurately, and ERA5 which reproduced the upper half of the peak accurately. Above 5 km ASL, the cloud occurrence was well reproduced in ERA5 and JRA-55, and strongly overestimated in AMPS, MERRA-2 and the UM. The reanalyses and models also tended to have two peaks at about 2 km ASL and 11 km ASL, but these were quite different from the observed peaks, with the lower peak underestimated by about 5 pp in the reanalyses and models and the higher peak overestimated by about 5–10 pp. The total CF was observed as 44% and 58% by CL31 and MiniMPL, respectively. CF observed by the MiniMPL was likely higher due to its higher maximum lidar range (CL31 missed a substantial amounts of high cloud due to this limitation). The total CF was strongly underestimated by the reanalyses and models by up to 34 pp (CL31) and 38 pp (MiniMPL), with the exception of the UM which simulated the correct CF within 3 pp.

The Christchurch observations were taken during a total of 24 days in mid- to late winter. The cloud situations were characterised by the frequent occurrence of low cloud and fog, with relatively diverse mid- and high level cloud and periods of clear sky also present (not shown). Precipitation was present in about 9% of profiles and fog in about 11% of profiles. As the site location is relatively flat (Canterbury Plains), the models didn't have any difficulty in reproducing the altitude of the site, which was 32 m (AMPS), 72 m (ERA5), 143 m (JRA-55) and 76 m (MERRA-2). The virtual location was within the boundaries of the city (AMPS), on the Canterbury Plains close to the city boundaries (ERA5, MERRA-2), and over Lake Ellesmere about 20 km from the city (JRA-55). Figure 3.4c shows that the co-located CHM 15k and MiniMPL observed a strong peak of cloud occurrence of 26% (CHM 15k) and 20% (MiniMPL) at about 500 m ASL. This was likely due to the combined precipitation and fog as well as false detection of aerosol as cloud. The observed cloud occurrence had a local minimum of 2% at about 5 km ASL, a secondary peak of 5% at 7 km ASL, and fell off 0% at 11 km ASL. The CHM 15k and MiniMPL observations showed inconsistencies of up to 5 pp. The reanalyses and models strongly underestimated low cloud by about 10–20 pp. With the exception of AMPS, they underestimated mid-level cloud by about 5 pp and represented high cloud relatively accurately. The total CF observed was 69% and 66% for the CHM 15k and MiniMPL, respectively, while the reanalyses and models strongly underestimated CF by up to 36 pp (JRA-55), with underestimates around 20 pp common.

Figures 3.5, 3.6, 3.7 show images of backscatter for three separate days taken from the three case studies. The selected days represent some of the best-matching profiles and demonstrate how well the reanalyses and models can simulate cloud under favourable conditions. As can be seen in the figures, the UM performs the best in terms of temporal and height accuracy of the simulated cloud (Fig. 3.5f, 3.6f), followed by ERA5 (Fig. 3.5c, 3.6c, 3.7c). This is likely due to the high output temporal resolution of the UM and ERA5 of 20 min. and 1 h, respectively. The UM and ERA5 were able to represent the relatively fine structure of cloud and to a lesser extent the optical thickness (inferred from the strength of backscattering) of the cloud. Deficiencies, however, are readily identifiable. The low cloud in the UM (Fig. 3.5f) appears to be shifted by several hours relative to observations (Fig. 3.5a) and the high cloud has a greater vertical extent in the UM. Likewise, the altocumulus cloud observed in Fig. 3.6a is shifted by several hours in the UM (Fig. 3.6f). The stratocumulus and nimbostratus cloud, identified visually based on the backscatter profiles, in ERA5 (Fig. 3.5c) is markedly lower than observed (Fig. 3.5a), as well as optically thicker than in reality.

The mid-level cloud in ERA5 (Fig. 3.6c) was located about 2 km higher than observed (Fig. 3.6a). Precipitation observed in Fig. 3.7a towards the end of the analysed period was not present in the ERA5 simulated profile (Fig. 3.7c) due to lack of precipitation simulation in the current lidar simulator (even though rain and snow specific content is available from the reanalysis). MERRA-2 was the third-best performing reanalysis in terms of cloud representation accuracy. The reanalysis managed to capture the overall structure of clouds (Fig. 3.5c, g, k), but substantial discrepancies were present, some of which were likely due to the relatively low temporal resolution of the reanalysis (3 h). AMPS was identified as the fourth-best in terms of cloud representation accuracy despite the highest spatial resolution of the model. Fig. 3.5h shows that while the overall structure of the cloud is represented in the reanalysis, it was not able to represent any of the finer details. This is likely partially due to the temporal resolution of the model output of just 3 h (AMPS has, however, relatively high horizontal grid resolution of 21 km). A visual comparison of AMPS and MERRA-2 with observations suggests that the MERRA-2 simulations are more accurate than AMPS despite the fact that the horizontal grid resolution of AMPS is much finer than MERRA-2. This demonstrates that other factors in the model than resolution have stronger influence on the quality of cloud simulation. JRA-55 was identified as the last in terms of cloud representation accuracy. JRA-55 has the lowest temporal resolution of the studied reanalyses and models of just 6 h, as well as the lowest horizontal grid resolution of 139 km. Therefore, it cannot be expected to capture any fine details of cloud. In the presented profiles (Fig. 3.5d, l) one can see that the cloud is only crudely represented. JRA-55 was able to represent the stratocumulus cloud of Fig. 3.5a, although its temporal extent and optical thickness were overestimated. The mid-level cloud of Fig. 3.5i was relatively well-represented in terms of height and optical thickness (Fig. 3.5l), given the low temporal resolution of the reanalysis. We stress that a direct backscatter profile intercomparison is highly dependent on the temporal resolution of the model output. The statistical intercomparison, however, should still give unbiased results if the cloud physics is accurately simulated by the atmospheric model.

The comparison between observations, reanalyses and models can be summarised by considering that the radiative effect of clouds is the product of cloud occurrence/fraction and their albedo. CF can be derived from ALC observations by determining the cloud mask and identifying cloudy profiles which contain detected cloud. Cloud albedo cannot be directly observed with ALCs due to the unknown backscatter-to-extinction ratio of the cloud layers, which depends on the cloud phase, the effective radius and the shape of ice crystals. However, we can use LR, calculated from the vertically integrated backscatter, as a proxy for cloud albedo. Clouds which are not fully opaque result in LR which is greater than the expected theoretical LR based on the Mie theory (Fig. 3.2). The inverse of LR (extinction-to-backscatter ratio) is therefore approximately proportional to the albedo of the cloud. Figure 3.8a, b shows an absolute and relative (respectively) scatter plot of the average cloud fraction and the average of the inverse of LR calculated from cloudy profiles. As already shown in Fig. 3.4, the reanalyses and models generally underestimate CF. However, Fig. 3.8 shows that this is compensated by overestimated albedo (higher LR^{-1} indicates higher cloud albedo), which is higher in all reanalyses and models than the corresponding ALC observations. Further examination shows that the UM and ERA5 show the closest proximity to the corresponding observations in terms of the average CF and LR^{-1} . The Lauder CL31 and MiniMPL also show

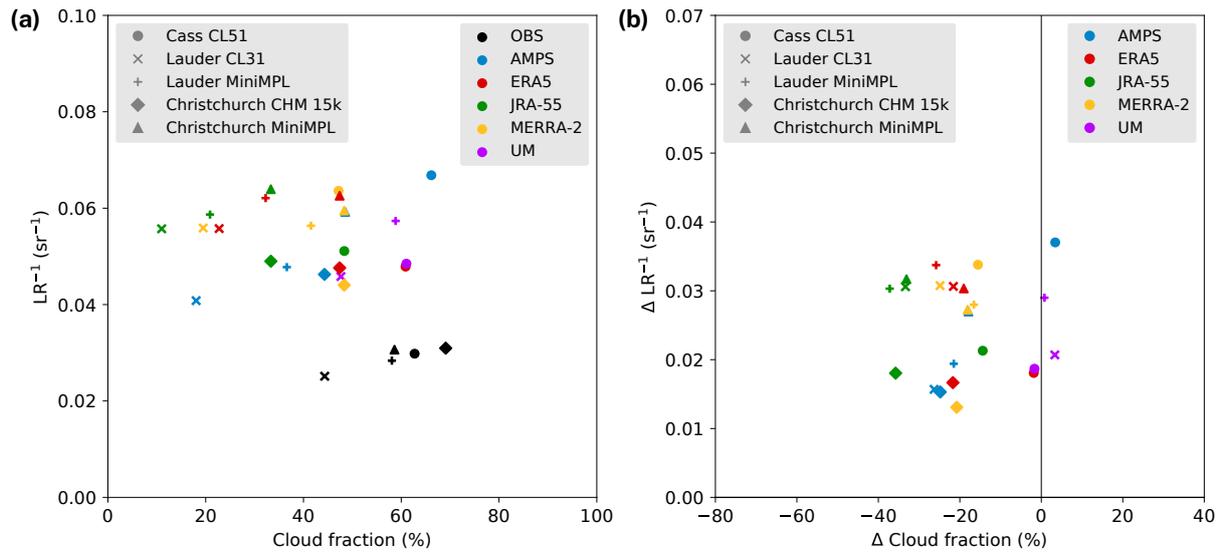


Figure 3.8 | (a) Scatter plot showing the average cloud fraction calculated from all profiles and the average inverse lidar ratio (LR^{-1}) calculated from cloudy profiles for the case studies. The site and instrument are coded by the shape of the marker and the observation (legend in the top left corner), reanalysis or model is coded by the colour of the marker (legend in the top right corner). (b) The same as (a), but showing relative differences between the reanalyses and models relative to the corresponding observations.

a stark difference in CF, with the CL31 detecting much smaller CF due to its maximum vertical range limitation of 7.7 km, which causes a significant fraction of high cloud to be omitted. The results shown in Fig. 3.8 suggest that the “too few too bright” model cloud problem identified in previous studies (Nam et al., 2012; Klein et al., 2013; Wall et al., 2017; Kuma et al., 2020a) is also observed in the current case studies, whereby the simulated CF is too low, but its exaggerated albedo means that the top of the atmosphere (TOA) shortwave radiative balance can be approximately correct.

Figure 3.9 shows backscatter histograms as a function of height for small values of backscatter (up to $2 \times 10^{-6} m^{-1} sr^{-1}$) observed and simulated at the sites of the case studies, calculated for the entire time period of each case study. The scale of values is below cloud backscatter, and therefore shows backscatter which results from molecular and aerosol scattering and noise. Molecular backscattering depends on the atmospheric pressure and temperature as well as the lidar wavelength. It causes the main “streak” (a local maximum) visible in each of the histograms. The observed molecular backscattering value at the surface approximately corresponds to the theoretically calculated value at each wavelength: $0.0906 \times 10^{-6} m^{-1} sr^{-1}$ ($\lambda = 1064$ nm), $0.172 \times 10^{-6} m^{-1} sr^{-1}$ ($\lambda = 910$ nm) and $1.54 \times 10^{-6} m^{-1} sr^{-1}$ ($\lambda = 532$ nm) at 1000 hPa and 20°C (Table 3.5). The molecular backscattering in the boundary layer is, however, superimposed on backscattering by aerosol and cloud. In the case of the MiniMPL observations at the Christchurch site (Fig. 3.9i), the molecular backscatter streak has multiple secondary streaks. These are caused by different levels of attenuation by cloud and aerosol during the period of the observations. These secondary streaks were also partially reproduced by the simulator (Fig. 3.9j). A smaller portion of the width of the streak is also caused by fluctuations of atmospheric temperature and pressure. Under suitable conditions,

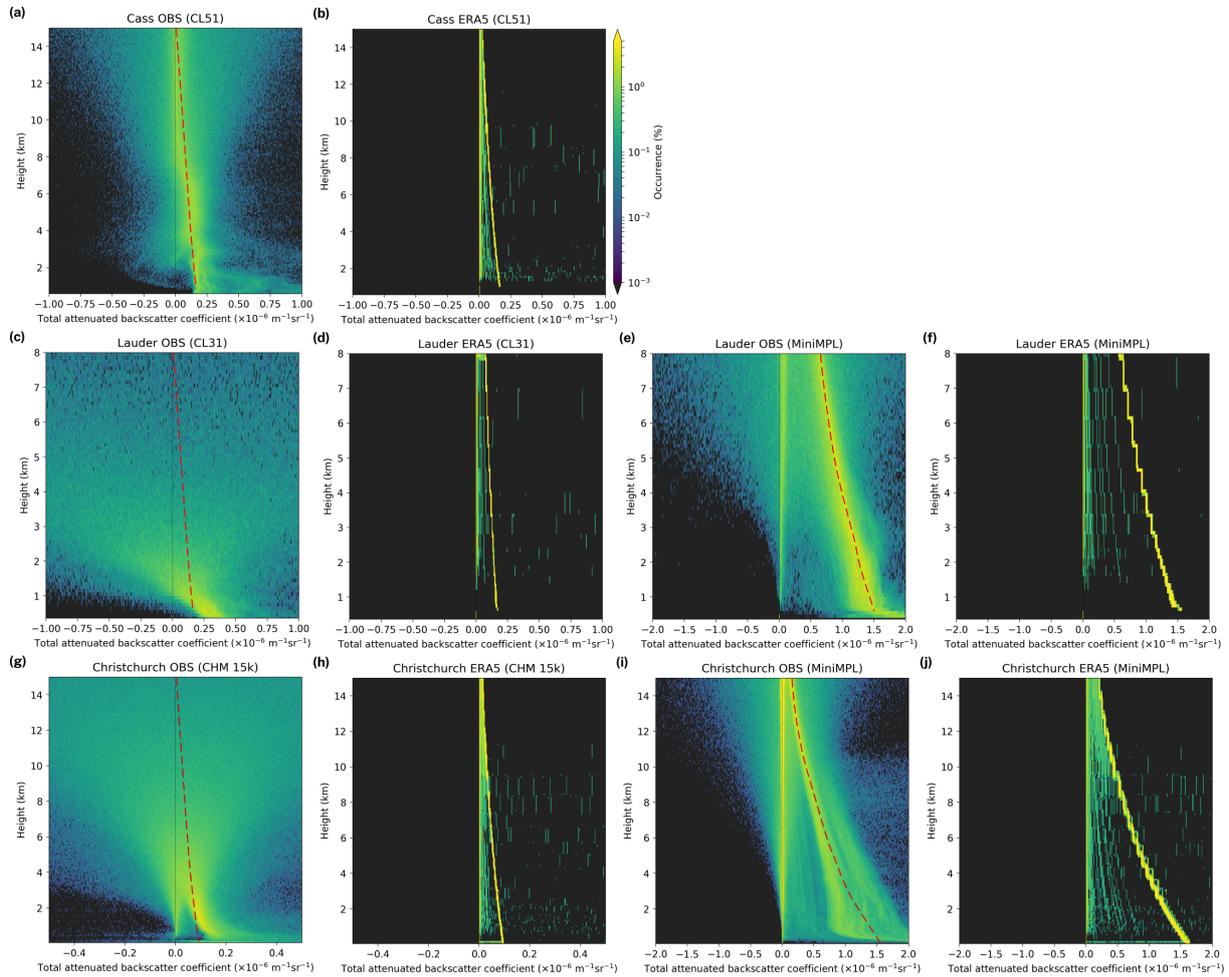


Figure 3.9 | Backscatter histograms as a function of height observed and simulated at three different sites of the case studies calculated from all profiles. The plots show the distribution of backscatter for values which are on the scale of noise, molecular and aerosol backscatter ($[-0.5, 0.5]$ for CHM 15k, $[-1, 1]$ for CL31 and CL51 and $[-2, 2] \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$ for MiniMPL). The simulated backscatter is based on the ERA5 atmospheric fields. Visible in the plots is backscatter caused by molecular backscatter (the main “streak”), noise when signal is fully attenuated by cloud (the zero-centred “streak”), and the range-dependent noise (the zero-centred “cone”). The molecular backscatter is marked by a red dashed line on the observed backscatter plots, the shape of which is taken from the simulated molecular backscatter for the corresponding instrument and site.

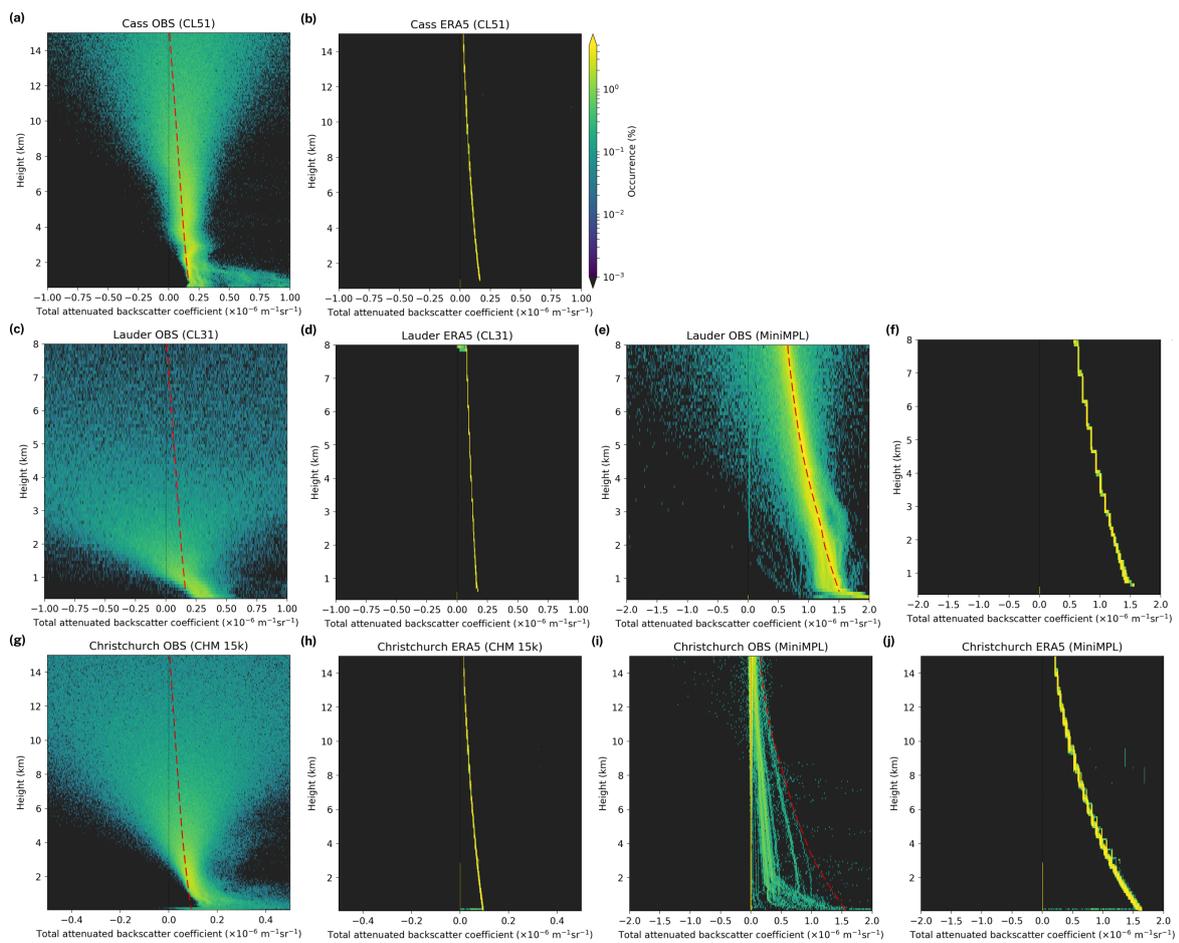


Figure 3.10 | The same as Fig. 3.9 but calculated from clear sky profiles only.

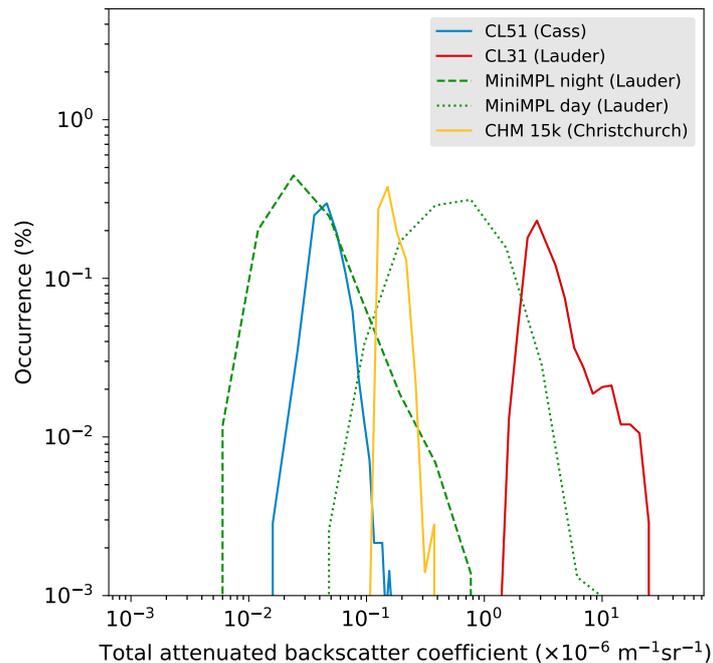


Figure 3.11 | Backscatter noise standard deviation histogram calculated for each instrument at sites of the case studies from clear sky profiles over the whole time period. The noise distribution is calculated at the furthest range. Shown is the range-scaled noise distribution at a range of 7 km. The MiniMPL nighttime and daytime distributions are calculated from nighttime and daytime profiles only, determined by the position of the Sun below and above the horizon, respectively, at the geographical position of the site and time when the profile was observed.

the molecular backscatter can be used for absolute calibration of an instrument. With the exception of CL31 (Fig. 3.9c), the molecular backscatter can be identified in the observed backscatter in each case. Therefore, it is possible to choose a calibration coefficient such that the observed and simulated molecular backscatter overlap. This can be considered a viable alternative to the liquid stratocumulus LR calibration method, or as a means of cross-validating the instrument calibration. However, it should be noted that the accuracy of this method is affected by an unknown amount of aerosol attenuation. Cloudy profiles can be filtered when calculating the histogram, and therefore the effect of cloud attenuation can be minimised. In addition to the molecular backscatter streak, there is a zero-centred streak visible in the histograms. This is caused by noise when the signal is fully attenuated by cloud. Lastly, a zero-centred "cone" of noise is visible in the observed backscatter, increasing with the square of range. The size of this cone is particularly large in the case of the CL31 (Fig. 3.9c), which is most likely the result of its low receiver sensitivity and low power compared to the other instruments. The standard deviation of the cone at the furthest range is used to determine the noise standard deviation used by the cloud detection algorithm (Sect. 3.5.3).

Figure 3.10 shows the same information as Fig. 3.9, but for clear sky profiles only. Here, it can be seen that the zero-centred peak caused by the complete attenuation by cloud is no longer present. There is a clear overlap between the centre of the noise cone with the simulated molecular backscatter; i.e. the noise cone is centred at the observed molecular backscatter. This is visible

with all instruments including CL31 (Fig. 3.10c), where the overlap between the observed and simulated molecular backscatter is most clearly visible at about 1 km ASL. Below 1 km ASL, the effect of boundary layer aerosol distorts the molecular backscatter by an unknown quantity. The clear sky histograms as shown in Fig. 3.9 may therefore be preferable to the all-sky histograms of Fig. 3.9 for calibration by fitting the molecular backscatter. Note that Figure 3.10i does not show a good match with the simulated molecular backscatter (Fig. 3.10j) likely due to miscalibration of the MiniMPL in the lower troposphere, which causes spurious cloud to be detected. The majority of profiles were therefore filtered out as cloudy. The dead time, afterpulse and overlap MiniMPL calibration supplied by the vendor appears to be deficient and causes range-dependent bias in the backscatter profile.

We now examine the noise in each instrument using the ALCF. Figure 3.11 shows the distribution of standard deviation of backscatter noise determined at the highest observable range of each instrument and range-scaled to 7 km. It can be seen that the CL31 is affected by the greatest amount of noise, peaking at about $2 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$. This is at the threshold of cloud detection of $2 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$. Therefore, thin cloud may be obscured by noise at higher ranges with this instrument. The MiniMPL, operating in the visible spectral range, shows a bimodal distribution of backscatter noise depending on sunlight. During daytime, it peaks at about $0.7 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$, which is the second highest of the analysed instruments. During nighttime, it peaks at about $0.02 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$, which is the lowest of the analysed instruments. The CHM 15k histogram peaks between the nighttime and daytime MiniMPL at $0.2 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$. CL51 observed the second lowest amount of noise after the nighttime MiniMPL at about $0.04 \times 10^{-6} \text{m}^{-1} \text{sr}^{-1}$. The difference between nighttime and daytime backscatter noise in the MiniMPL has been previously analysed by Silber et al. (2018) (Fig. S3) and these results confirm their findings.

3.8 Discussion and conclusions

We presented the Automatic Lidar and Ceilometer Framework, which combines lidar processing and lidar simulation for the purpose of model evaluation. The lidar simulation is based on the COSP spaceborne lidar simulator by accounting for the different geometry and lidar wavelength. New lookup tables for Mie scattering were calculated for a number of ALC wavelengths and noise removal and cloud detection algorithms were implemented. The framework supports the most common ALCs and reanalyses. We demonstrated the use of the framework on ALC observations at three different sites in New Zealand, and applied the lidar simulator to three reanalyses and two models. We found that while some reanalyses and models such as the UM and ERA5 show relatively good correspondence with observed cloud, others performed relatively poorly. All reanalyses and models underestimated the total CF by up to 38 pp, with underestimation by 20 pp common. In some cases, the observed and simulated backscatter profiles matched relatively closely in terms of time and altitude, and a better match was observed with reanalyses with high output temporal resolution such as the UM and ERA5, while reanalyses with low temporal resolution didn't allow for reliable direct (non-statistical) comparison of cloud. However, it is clear that more factors than the horizontal and vertical resolution influence the cloud simulation accuracy; especially the cloud, boundary layer and convection schemes employed by the atmospheric model. The

reanalysis and model output temporal, horizontal grid resolution and vertical resolution are not always the same as the internal resolution of the underlying atmospheric model. Both have an impact on the comparison between simulated and observed backscatter and cloud. While the output resolution should not have an impact on the long-term statistics, it can be a limiting factor for direct backscatter profile comparison. We demonstrated that the ALCF could be used to identify substantial differences in cloud backscatter which were present in all reanalyses and models. We showed that all the studied instruments except for the CL31 are capable of detecting molecular backscatter and that this can be used for calibration or for cross-validation of other calibration methods. We found that the nighttime MiniMPL was subject to the least amount of noise of all the instrument examined, followed by the CL51, CHM 15k, daytime MiniMPL and CL31. Noise in the MiniMPL was shown to have a bimodal distribution due to day/nighttime. The ALCF can therefore be useful for testing the quality of collected data.

Currently the framework has several limitations which should be addressed in the future. The water vapour absorption at 910 nm likely affects the instrument calibration of the CL31 and CL51 ceilometers and limits the accuracy of the one-to-one comparison, even though due to the relatively high backscatter caused by cloud, the calculated cloud masks are unlikely to be strongly affected. The lidar simulator currently does not simulate backscattering from precipitation. Observed precipitation is generally detected as "cloud" by the cloud detection algorithm, while the simulated profile contains no backscatter at the location of precipitation (backscattering and attenuation by rain drops and snow should be implemented in the lidar simulator in the future). If desired, the backscatter profiles affected by precipitation can be excluded before the comparison or their fraction determined by visually inspecting the observed backscatter to assess their possible effect on the statistical results. In addition, backscattering from ice crystals is currently treated in the same way as backscattering from liquid droplets, i.e. the same effective radius of spheres is assumed and the particle concentration is calculated from the model ice mass mixing ratio in the atmosphere layer. In reality, the non-spherical shape of ice crystals, the different typical effective radius and refractive index of ice result in different backscattering from ice crystals versus liquid droplets for the same mass mixing ratio. The ice mass mixing ratio reported by models is usually much lower than the liquid mass mixing ratio, therefore the simulated backscatter from cold cloud tends to be much lower than from warm cloud. The ALCs also suffer from various measurement deficiencies. Notably incomplete overlap, dead time and afterpulse corrections tend to give sub-optimal results at the near range. It is possible to use semi-automated methods to correct for these deficiencies, such as by calculating the integrated backscatter distribution by height of the maximum backscatter and correcting for the range-dependent bias (Hopkin et al., 2019, Sect. 5.1). This method could be implemented in the framework to enable range-dependent calibration of the observed backscatter.

The presented framework streamlines lidar data processing and tasks related to lidar simulation and model comparison. The framework was recently used by Kuma et al. (2020a) for Southern Ocean model cloud evaluation in the GA7.1 model and MERRA-2 reanalysis. Considering the existing extensive ALC networks worldwide there is a wealth of global data. We therefore think that ALCs should have a greater role in model evaluation. Satellite observations have long been established in this respect due to their availability, spatial and temporal coverage and their well-developed derived products and tools. ALCs, with their diverse formats and decentralised nature,

have so far lacked derived products and tools which would make them more accessible for model evaluation. We hope that this software will enable more model evaluation studies based on ALC observations. Development of lidar data processing is currently hampered by closed development of code. We note that code has very rarely been made available with past ALC studies. Continued improvement of publicly available code for lidar data processing is needed to achieve faster development of ground-based remote sensing and make it more attractive for GCM, NWP model and reanalysis evaluation.

Code and data availability

The *ALCF* is open source and available at <https://alcf-lidar.github.io> and as a permanent archive of code and technical documentation on Zenodo at <https://doi.org/10.5281/zenodo.3779518>. The technical documentation is also in the Supplementary information. A tool for converting Vaisala CL31 and CL51 data files to NetCDF *cl2nc* is open source and available at <https://github.com/peterkuma/cl2nc>. A tool for converting MiniMPL raw binary data files to NetCDF *mpl2nc* is open source and available at <https://github.com/peterkuma/mp12nc>. The observational data used in the case studies are available upon request. The reanalyses data used in the case studies are publicly available online from the respective projects. The *Unified Model* data used in the case studies are available upon request. The Unified Model is proprietary to the UK Met Office and is made available under a licence. For more information, readers are advised to contact the UK Met Office.

Author contributions

Peter Kuma wrote the code of the framework, performed the data analysis of the case studies and wrote the text of the manuscript. Adrian McDonald and Olaf Morgenstern provided continuous scientific input on the code development, analysis and text of the manuscript. Richard Querel, Israel Silber and Connor J. Flynn provided calibration of the MiniMPL data and substantial discussion of the theoretical concepts. All authors reviewed the manuscript.

Acknowledgements

We would like to acknowledge the New Zealand Deep South National Science Challenge project which provided funding for our work; the New Zealand eScience Infrastructure (NeSI) which provided supercomputing resources to run the Unified Model; Vidya Varma, Jonny Williams, Guang Zeng and Wolfgang Hayek for their contribution to setting up a nudged run of the Unified Model; Graeme Plank and Graeme MacDonald who participated on the installation of the Vaisala CL51 at the Cass field station; the COSP project for the code which we used as the basis for the lidar simulator; the AMPS, JRA-55, ERA5, MERRA-2 models and reanalyses which provided public access to their data; the open source libraries numpy (Van Der Walt et al., 2011), scipy (Virtanen et al., 2019), matplotlib (Hunter, 2007), netCDF4 (Rew and Davis, 1990) and Astropy (Price-Whelan et al., 2018) and the Python programming language (Rossum, 1995) which we used in the implementation of our code; the R programming language (R Core Team, 2017), the Natural Earth dataset (<https://www.naturalearthdata.com>) and the Shuttle Radar Topography Mission (SRTM) Version 3 Global 1 arc second digital elevation model (Werner, 2001; NASA JPL, 2013) which we used to produce a map of sites; GitHub which provided free hosting of our code; and the Linux-based (Torvalds, 1997) operating systems Devuan GNU+Linux and Debian GNU/Linux on which we produced this analysis.

Chapter 4

Improving Southern Ocean boundary layer cloud parametrisation in the HadGEM3-GA7.1/UM11.4 general circulation model

Peter Kuma¹, Adrian J. McDonald¹, Olaf Morgenstern², Sean Hartery¹, Jonny Williams², Vidya Varma², Guang Zeng², Mike Harvey², Simon Parsons¹, Graeme Plank¹

¹University of Canterbury, Christchurch, Aotearoa/New Zealand

²NIWA, Wellington, Aotearoa/New Zealand

Abstract

Southern Ocean cloud biases and the corresponding shortwave and longwave radiation biases are a longstanding problem in general circulation models. We use one month of high-resolution summertime Southern Ocean voyage ceilometer and radiosonde observations collected on the Aotearoa/New Zealand research vessel Tangaroa in February–March 2018 in the Ross Sea region and a nudged run of HadGEM3-GA7.1/UM11.4 to evaluate the impact of modifications in the convection and boundary layer schemes on cloud simulation in a set of representative case studies. We use the recently-developed Automatic Lidar and Ceilometer Framework (ALCF) to assess and improve the representation of stratocumulus (Sc) cloud, currently strongly underestimated in the model. We show that two- and three-layer cloud profiles of cumulus (Cu) below Sc corresponding to local thermodynamic levels were a common occurrence, where the Cu cloud base height coincides with the lifting condensation level (LCL) and the Sc cloud heights coincide with the neutral buoyancy level of dry and moist (respectively) air parcels lifted from the surface. While the thermodynamic state of the atmosphere is simulated well by the model, too little moisture appears to be transported to the top of the boundary layer. By increasing surface moisture flux and convective mass flux in the model we can improve the Sc cloud simulation in case studies, but we demonstrate that a lack of vertical moisture transport across the LCL from the surface layer to the zone of convective mass flux remains a likely limiting factor. We also show that the modifications made have a positive impact on the Southern Ocean and global radiation balance of up to 5 Wm^{-2} based on zonal averages over the time period examined.

4.1 Introduction

Cloud biases in general circulation models (GCMs) are a long-standing problem (Trenberth and Fasullo, 2010; Vignesh et al., 2020). Correct cloud representation in GCMs is critical due to the large impact of clouds on the planetary albedo and longwave emissivity as well as being a latent heat source and the source of precipitation. Cloud biases are a key contributor to the error in the simulation of the radiation balance in contemporary GCMs (Li et al., 2013). Clouds in these models are parametrised by subgrid-scale parametrisation schemes. Cloud typically varies at scales much smaller than the resolution of the models, which is typically on the order of 10–200 km in the horizontal. Therefore, many of the physical processes responsible for cloud formation and removal, such as convection and cloud microphysics, are not resolved at the model resolution and have to be parametrised. The nature of these parametrisations means that they inevitably lead to large uncertainties. An alternative to such schemes are cloud resolving models (CRMs), which operate at a much higher horizontal spatial resolution (on the order of 1 km) (Guichard and Couvreux, 2017; Satoh et al., 2019). However, at present it is not feasible to use these models for long-term climate projections due to their computational demands. Therefore, a continued improvement of existing cloud schemes is needed.

Cloud parametrisation in GCMs is traditionally improved by intercomparing with observations such as the ISCCP satellite-based cloud dataset (Rossow and Schiffer, 1991) or more recently the CloudSat–CALIPSO datasets (Stephens et al., 2002; Winker et al., 2003). By accomplishing a good match with past and present observations of cloud it is assumed that cloud representation in simulations of future climate will be reasonable. However, the physical processes responsible for any cloud bias are not necessarily obvious from a simple intercomparison, which may be due to processes outside of the cloud parametrisation scheme (Morcrette and Petch, 2010). The presence of compensating errors can also make the model perform worse when only one of the errors is corrected (Hourdin et al., 2017; Schuddeboom et al., 2019).

Here, we focus on the Southern Ocean (SO) boundary layer parametrisations in the Global Atmosphere 7.1 (GA7.1)/Unified Model 11.4 (UM11.4) component of the Hadley Centre Global Environment Model 3 (HadGEM3) (Walters et al., 2019), with the aim of improving the simulation of stratocumulus clouds. This follows the findings of Kuma et al. (2020a) of underestimated cloud occurrence below 2 km above sea level (ASL) in GA7.1 relative to a comprehensive set of ship-based observations. The GA7.1 model is based the UK Met Office Unified Model (UM), which is also used for operational numerical weather prediction (NWP). HadGEM3 (GC3.1) has participated in the 6th Coupled Model Intercomparison Project (CMIP6). As shown by Kuma et al. (2020a), cloud fraction in a previous version of the model is underestimated in the SO by about 6–8%, with low cloud below 1 km and fog particularly underestimated compared to ship-based ceilometer observations. Schuddeboom et al. (2019) and Kuma et al. (2020a) noted that there is a clear division at a latitude of 55°S between positive and negative shortwave (SW) radiation biases in HadGEM3-GA7.1 and that the positive vs. negative bias appears to be linked to near-surface air temperature, with negative bias strongly associated with close to zero near-surface air temperature. Therefore, we expect that near-surface temperature and sea surface temperature (SST) have a significant role in the cloud bias, potentially by destabilisation of the relatively cold Antarctic air in the surface

layer by near-zero SST. Loveridge and Davies (2019b) also implicated quiescent conditions in the SO boundary layer (BL) cloud bias in GA7.0. Considering these past findings, here we focus on BL cloud south of 60°S, which we will identify as "high-latitude SO", as opposed to "low-latitude SO" north of 60°S.

The UM parametrises clouds using the prognostic PC2 scheme (Wilson et al., 2008a,b), in which the cloud condensate (liquid and ice) and cloud fraction are prognostic variables defined on every grid cell. The prognostic variables can be advected with the large-scale flow, which is in contrast to earlier schemes which were purely diagnostic (Smith, 1990) and based on the assumption that any supersaturation is turned into cloud condensate (Jakob, 2000). Cloud condensate and cloud fraction originating in the convection scheme are added to the prognostic fields. injected into the condensate and cloud fraction of the cloud scheme and thus become prognostic. Schemes which used prognostic condensate and diagnostic cloud fraction also exist (Sundqvist, 1978; Sundqvist et al., 1989).

An improvement in the SO cloud representation in the MRI-ESM2 GCM has recently been reported by Kawai et al. (2019). This model uses a cloud scheme based on Tiedtke (1993). This is also the basis of the PC2 scheme, although both schemes contain many modifications from the original scheme. Therefore, it may be possible to gain insight from this study on improvements in the PC2 scheme which would enhance the simulation of clouds in this region. Kawai et al. (2019) reported that their updated stratocumulus scheme resulted in about a 10% increase in SO cloud cover and a 10 Wm^{-2} increase in top of the atmosphere (TOA) outgoing SW radiation over the SO, reducing the model bias.

This paper is structured as follows. In Sect. 4.2 we briefly describe the TAN1802 voyage dataset, the HadGEM3-GA7.1/UM11.4, and the lidar processing and simulator framework ALCF. In Sect. 4.3 we describe the relevant characteristics of the cloud-related parametrisation schemes in the UM and an experimental run of the HadGEM3-GA7.1/UM11.4. Finally, in Sect. 4.4 we contrast the results of the experimental run with the control run and observations.

4.2 Methods

4.2.1 The TAN1802 voyage

In our analysis we use ceilometer, radiosonde, and AWS data collected on the TAN1802 voyage of R/V Tangaroa in the Ross Sea. The data were collected between 16 February and 15 March 2018 UTC (inclusive), when the ship was south of 60°S. The deployed ceilometer was a Lufft CHM 15k operating at an infrared wavelength of 1064 nm and directed vertically between the surface and 15 km above sea level (ASL). We process the raw backscatter profiles with the ALCF (Sect. 4.2.3). In addition, we use radiosonde profiles sampled during the voyage at between 1–3 times daily intervals using the iMet-1 ABx radiosondes. Near-surface air temperature, relative humidity (RH) and SST were sampled continuously during the voyage. The TAN8102 voyage atmospheric measurements are also described in greater detail by Kuma et al. (2020a) and Hartery et al. (2020b). The track of the voyage is shown in Fig. 4.1.

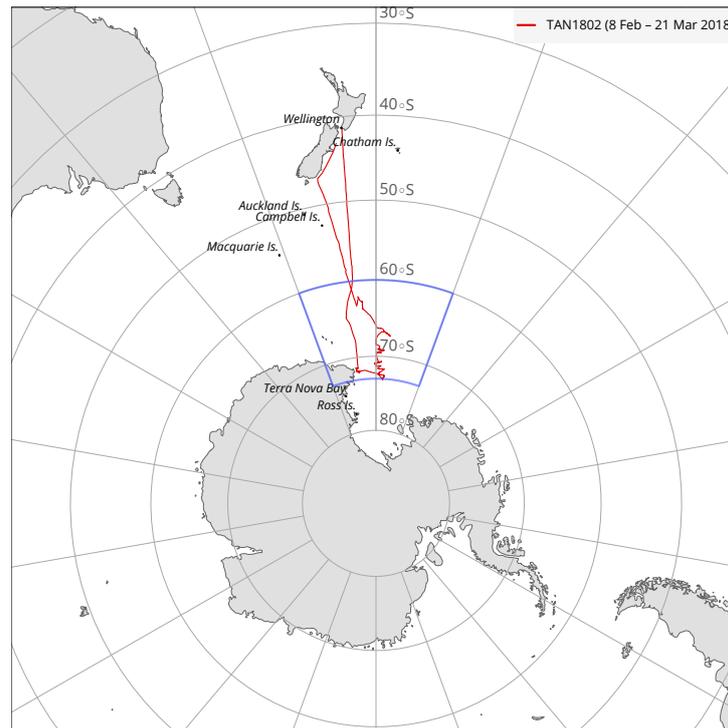


Figure 4.1 | A map showing the track of the TAN1802 voyage and the region between 60°S and 73°S of the TAN1802 ceilometer, radiosonde and AWS data analysed here.

4.2.2 HadGEM3-GA7.1/UM11.4

The GA7.1 model is based on the UK Met Office Unified Model (UM), which is an atmospheric model used for both NWP and climate-modelling purposes in global and regional configurations (Walters et al., 2019). HadGEM3 is a proprietary model developed by the UK Met Office and partner organisations in a number of countries. We analyse data from a relatively recent version UM11.4 released on 21 Jun 2019. The experiments were performed on the NeSI supercomputer in Wellington, Aotearoa/New Zealand (Williams et al., 2016). The horizontal grid scale of the model is N96, corresponding to about 100×140 km horizontal grid cells at 60°S. The time step of the model is 20 minutes to aid comparison with observations. We use instantaneous 3-dimensional atmospheric fields exported from the model at each time step and process these with the ALCF (Sect. 4.2.3) to obtain simulated lidar data corresponding to the TAN1802 voyage. We use UM11.4 in a “nudged” configuration in which SST and sea ice concentration are prescribed from the HadISST dataset (Rayner et al., 2003) and atmospheric fields are nudged towards the ERA-Interim reanalysis (Dee et al., 2011). In this configuration the large-scale dynamics is effectively prescribed while parametrised fields such as cloud are the result of the model’s physics. Here, we analyse two simulations produced by the UM11.4: (1) an unmodified control run “UM11.4cnt” and (2) an experimental run “UM11.4ext” (Sect. 4.3.6). The cloud- and BL-related parametrisations in the HadGEM3-GA7.1/UM11.4 are described in Sect. 4.3.

Table 4.1 | Summary of the cloud-related subgrid-scale parametrisations in the UM11.4.

Parametrisation	Principle	References
Convection scheme	Mass flux with CAPE or surface buoyancy flux closure.	Gregory and Rowntree (1990) Grant and Brown (1999) Grant (2001)
Boundary layer scheme	First-order closure turbulence.	Lock et al. (2000) Martin et al. (2000)
Large-scale cloud scheme (PC2)	Prognostic cloud liquid and ice.	Wilson et al. (2008a,b)
JULES surface flux parametrisation	The COARE algorithm.	Fairall et al. (2003)

4.2.3 ALCF

The Automatic Lidar and Ceilometer Framework (ALCF) (Kuma et al., 2020b) is a lidar simulator and lidar processing tool, based on the Cloud Feedback and Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (Bodas-Salcedo et al., 2011). The ALCF enables consistent comparison between ground-based lidar observations and atmospheric models. We use the ALCF package to process the ceilometer observations collected on TAN1802 and to simulate backscatter profiles from the UM11.4 model output. Using this tool, we derive a cloud mask from the observations, calculate simulated backscatter from 20-min resolution model data and derive a cloud mask from the simulated backscatter. The resulting backscatter can be compared statistically as well as on co-located diagrams thanks to the model nudging, which means it represents real weather situations (unlike a free-running model).

4.3 Parametrisations

The UM11.4 relies on the parametrisation of subgrid-scale processes, which cannot be represented by the atmospheric dynamics on the relatively coarse grid of the model. The parametrisations directly affecting cloud are the convection, boundary layer, and large-scale cloud (PC2) schemes and the related parametrisation of surface fluxes in the surface component of HadGEM3 called the Joint UK Land Environment Simulator (JULES) (Best et al., 2011; Clark et al., 2011). The formation and destruction of cloud is therefore an interplay between a number of parametrised processes. This interplay means that it is typically not obvious from model evaluation studies which process is responsible for a bias in cloud occurrence or TOA radiation balance. The parametrisations outlined above operate independently on each individual horizontal grid cell of the model. Table 4.1 summarises the cloud-related parametrisation schemes, described in a greater detail below with a specific focus on aspects important to BL cloud.

4.3.1 Large-scale cloud scheme (PC2)

The PC2 scheme is based on Tiedtke (1993). The motivation to develop a new scheme was to allow for detrainment of convection into cloud fraction and break the diagnostic link between cloud fraction and condensate (Wilson et al., 2015). The scheme contains a separate treatment of large-scale (stratiform) and convective cloud, both of which contribute to the cloud condensate and cloud fraction by detraining moisture and condensate. Moisture in the PC2 scheme within a grid cell is assumed to have a probability distribution function around the grid mean, parametrised

by a critical RH (Gregory et al., 2002). Therefore, supersaturation within the grid cell can occur before mean RH reaches 100%. This probability density function (PDF) is defined explicitly in PC2, in contrast with previous schemes (Tiedtke, 1993), which assumed an implicit PDF (Wilson et al., 2008a). The PC2 scheme calculates the evolution of condensate and cloud fraction by summing the contributions of a number of sources and sinks in each model time step: advection, boundary layer (Sect. 4.3.2), convection (Sect. 4.3.3) and precipitation. The convection scheme can be a source of condensate, cloud fraction, water vapour and heat via "inhomogeneous forcing" in PC2. Condensate and cloud fraction are injected from the convective plumes into the prognostic fields. Water vapour and heat changes can generate additional condensate if supersaturation occurs.

4.3.2 Boundary layer scheme

The BL scheme in UM11.4 is a first-order closure turbulence scheme described by Lock et al. (2000) and Martin et al. (2000). We use the standard version 9C in our analysis. At an initial stage, the BL scheme diagnoses the BL type into one of six types depending on the liquid–frozen virtual potential temperature (θ_{vl}) profile (Lock et al., 2000, Fig. 1). Turbulence and convection are then applied selectively in the relevant layers, while turbulence is applied in the unstable surface mixed layer (SML), and convection is applied in the layer between the LCL and the capping inversion, and cloud-top driven turbulence just below the capping inversion in the Sc cloud layer. Our particular focus is on the BL Type V (i.e. decoupled Cu below Sc), which is the type most commonly associated with Sc biases in UM11.4 compared to our observational dataset, and also the most prevalent type in the high-latitude SO (Sect. 4.4.6). The turbulent locally-mixed layers are characterised by diffusion coefficients K_h and K_m . Depending on conditions, these are based on the surface wind shear, surface buoyancy gradient, Richardson number (Ri), cloud-top radiative cooling and cloud-top buoyancy reversal. This is in contrast with non-local mixing performed by the mass flux parametrisation in the convection scheme.

4.3.3 Convection scheme

UM11.4 offers several versions of the convection scheme. We use the standard version 6, which is based on a mass flux parametrisation described by Gregory and Rowntree (1990). The convection scheme parametrises convective plumes covering a fraction of the grid cell, consisting of updrafts and downdrafts which transport scalar quantities, such as heat and moisture, and vector quantities, such as momentum, vertically across model levels. The updrafts and downdrafts exchange air with the environment (i.e. the remaining part of the grid cell) by entrainment and detrainment. Air ascending in the updraft can undergo condensation due to moist adiabatic cooling, forming cumulus (Cu) cloud, and also detraining cloud condensate into the environment. An updraft generally terminates at the capping inversion, which is diagnosed by the z_{par} level at the beginning of the model time step. At this level, the updraft is detrained completely into the environment by a so-called "forced detrainment". Mass flux, in the units of $\text{Pa}\cdot\text{s}^{-1}$, is determined by conservation of mass at each model level (Gregory and Rowntree, 1990). This, however, requires a closure at the bottom level of the convective region. The scheme implements two types of closure: a closure based on the convective available potential energy (CAPE) and on surface buoyancy flux. The choice of closure is based on the diagnosed type of convection, which can be shallow, mid-level or deep. In

our analysis we focus on shallow convection, which was predominant in our case study. Shallow convection in the scheme is based on the surface buoyancy flux closure. As described by Grant and Brown (1999) and Grant (2001), this closure is based on a similarity theory linking surface buoyancy flux to the initial mass flux at the cloud base, as determined computationally by large eddy simulations (LES) of profiles observed during the Barbados Oceanographic and Meteorological Experiment (BOMEX) (Davidson, 1968), a North Sea field campaign (Smith and Jonas, 1995) and the Atlantic Trade-wind Experiment (ATEX). Currently a new convection scheme “CoMorph” is in development by the UK Met Office with the intention to eventually replace the existing convection scheme.

4.3.4 JULES surface flux parametrisation

JULES is the surface parametrisation model used in conjunction with the UM (Best et al., 2011; Clark et al., 2011). Among other processes, JULES parametrises fluxes of momentum, heat and moisture from the sea surface to the the surface layer of the atmosphere. The fluxes are determined by surface transfer coefficients for momentum and scalars C_D and C_M (respectively), which in turn are determined by the sea surface roughness for momentum and scalars z_{0m} and z_{0h} (respectively). Multiple options of surface roughness calculation are available (Lock et al., 2019), two of which we will briefly describe here. The standard option is based on the Coupled Ocean–Atmosphere Response Experiment (COARE) algorithm (Fairall et al., 2003). The momentum roughness length is based on a generalised Charnock’s formula which accounts for low wind speeds. Charnock’s coefficient a is parametrised by a linear relationship with 10-m wind speed, subject to minimum and maximum bounds. The scalar roughness length is based on an empirical dependence on the roughness Reynolds number R_r determined by a set of field campaigns (COARE-plus). R_r , in turn, is derived from the momentum roughness length, friction velocity and kinematic viscosity (Fairall et al., 2003). A simpler second option of surface roughness length calculation is available which uses a fixed Charnock’s coefficient and a fixed scalar roughness length.

4.3.5 Southern Ocean boundary layer

Figure 4.2 shows a schematic of operation of the UM11.4 boundary layer and convection parametrisation most commonly occurring during the TAN1802 voyage in situations with Sc cloud, and Table 4.2 summarises the BL diagnostic levels of the model. This situation corresponds to the BL Type V in Lock et al. (2000) (decoupled Sc over Cu). In the UM11.4 this is parametrised by the BL scheme first-order turbulence scheme in the surface and Sc layers and the convection scheme mass flux parametrisation between these two layers. The level separating the surface turbulence and mass flux is the LCL, above which moist convection is expected to occur. As noted by Lock et al. (2000), this vertical separation of the parametrisation schemes is undesirable, but also unavoidable due to the BL scheme partially accounting for the vertical transport by convection. As we will show later, this vertical separation of the two parametrisation schemes may be responsible for the lack of Sc in the model compared to observations (Sect. 4.4.3). For the purpose of BL cloud simulation, we are chiefly concerned with the vertical moisture and air mass transport from the sea surface to the Sc layer. This process is simulated by:

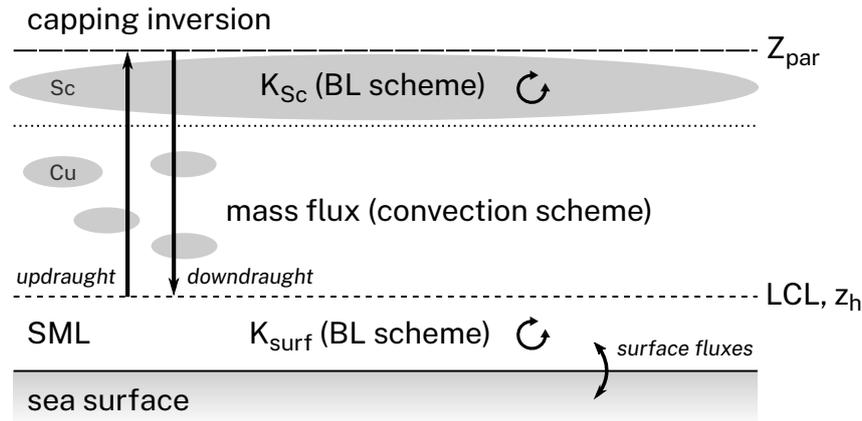


Figure 4.2 | A scheme showing the operation of the boundary layer (BL) and convection schemes in a cumulus (Cu) below stratocumulus (Sc) profile commonly observed in the summertime Southern Ocean. Surface mixed layer (SML), lifting condensation level (LCL).

1. A flux of moisture from the sea surface to the surface layer, determined by the sea surface roughness in the surface scheme (JULES).
2. Mixing of moisture and heat in the SML by turbulence up to the LCL.
3. Flux of mass and moisture by convection from the LCL to the capping inversion.
4. Forced detrainment of air mass from the convective updraft into the environment below the capping inversion (corresponding to the z_{par} level), identified in the model as the capping level for a moist lifted parcel.

If these processes are sufficient, enough moisture (saturated air) is detrained below the capping inversion to increase RH above the critical RH and condense the excess water vapour into Sc cloud. Deficiencies in any of above processes or their interconnections can result in a deficiency in RH and a lack of Sc compared to observations. Our aim is therefore to identify which of these processes or interconnections are underestimated.

4.3.6 Experimental run

We prepared an experimental run of the UM11.4 ("UM11.4exp") in order to evaluate the effect of tuning of the processes outlined in Sect. 4.3.5 on the simulation of Sc in the Ross Sea to compare with the TAN1802 voyage observations. In order to increase the moisture transport from the surface to the sub-capping-inversion layer, we applied the configuration and code modifications described below:

- We used the fixed sea surface roughness length option (Sect. 4.3.4) with a fixed scalar surface roughness length of 10^{-4} m. This value is approximately equal to the maximum in the empirical fit of COARE data (Fairall et al., 2003). The motivation for this change is to increase the surface flux of moisture to the greatest physically meaningful value.

Table 4.2 | Diagnostic model levels and layers relevant to shallow boundary layer convection, turbulence and cloud, listed from the lowest to the highest.

Level	Description
SL (surface layer)	Layer between the surface and the lower of $0.1z_h$ and level where θ_{el} starts to increase with height.
SML (surface mixed layer)	Turbulently well-mixed layer between the surface and the LCL.
LCL (lifting condensation level)	Traditional definition (e.g. Wallace and Hobbs (2006)).
z_h	Level equal to the LCL if cumulus-capped, or z_{par} if not.
z_{par}	Top of a diagnostic moist parcel ascent.

- We increased a coefficient relating sub-cloud convective velocity scale to cumulus mass flux for shallow convection (c_{mass}) by 50%. c_{mass} determines the initial mass flux in the surface buoyancy flux closure (Sect. 4.3.3). This has the effect of increasing the initial mass flux by 50%. The motivation for this change is to increase the speed of convective updrafts, by which the forced detrainment of saturated air below the capping inversion is increased.

4.4 Results

In this section we compare the experimental run UM11.4exp described in Sec. 4.3.6 with the control run UM11.4cnt, in-situ observations on TAN1802 and satellite observations from the CERES instruments' synoptic (SYN) product (Loeb et al., 2018) in the time period between 16 February 2018 and 15 March 2018.

4.4.1 Cloud observations

As shown previously by Kuma et al. (2020a), the BL cloud base in the SO commonly corresponds to either the LCL or the SST lifting level (SLL). Here, we show that in the TAN1802 voyage observations, three clouds layers were often observed and corresponded to three different thermodynamic levels. Figure 4.3 shows several days of co-located ceilometer and radiosonde observations. Based on the radiosonde profiles, we calculated the following thermodynamic levels: the LCL, SLL and saturated SLL (SLL_s), which uses the same assumption as SSL, but allows the parcel to ascend by moist adiabatic processes above the LCL. On a majority of days with Sc cloud, we observed that the LCL corresponded with the cloud base of relatively thin Cu fractus below 1 km ASL (visible on all days in Fig. 4.3). The much thicker layer of Sc corresponded with SLL between 1 and 2 km ASL, i.e. due to parcels lifted by dry convection. In several cases, SLL_s corresponded with a third layer of cloud above Sc (Fig. 4.3a, d, e), i.e. due to parcels lifted by moist convection. The third layer was, however, intermittent and much less significant than the Sc cloud based identified near SLL. Examination of radiosonde profiles, shows that SLL and SLL_s were both characterised by an inversion (most clearly visible in Fig. 4.3e'). Physically, this inversion would act to prevent further ascent of a parcel and cause a forced detrainment of convective updrafts, thus leading to accumulation of moisture below the inversion and the gradual formation of Sc cloud. As discussed later, we think that this process is underestimated in the UM11.4, which explains the absence of Sc layers in the model compared to observations.

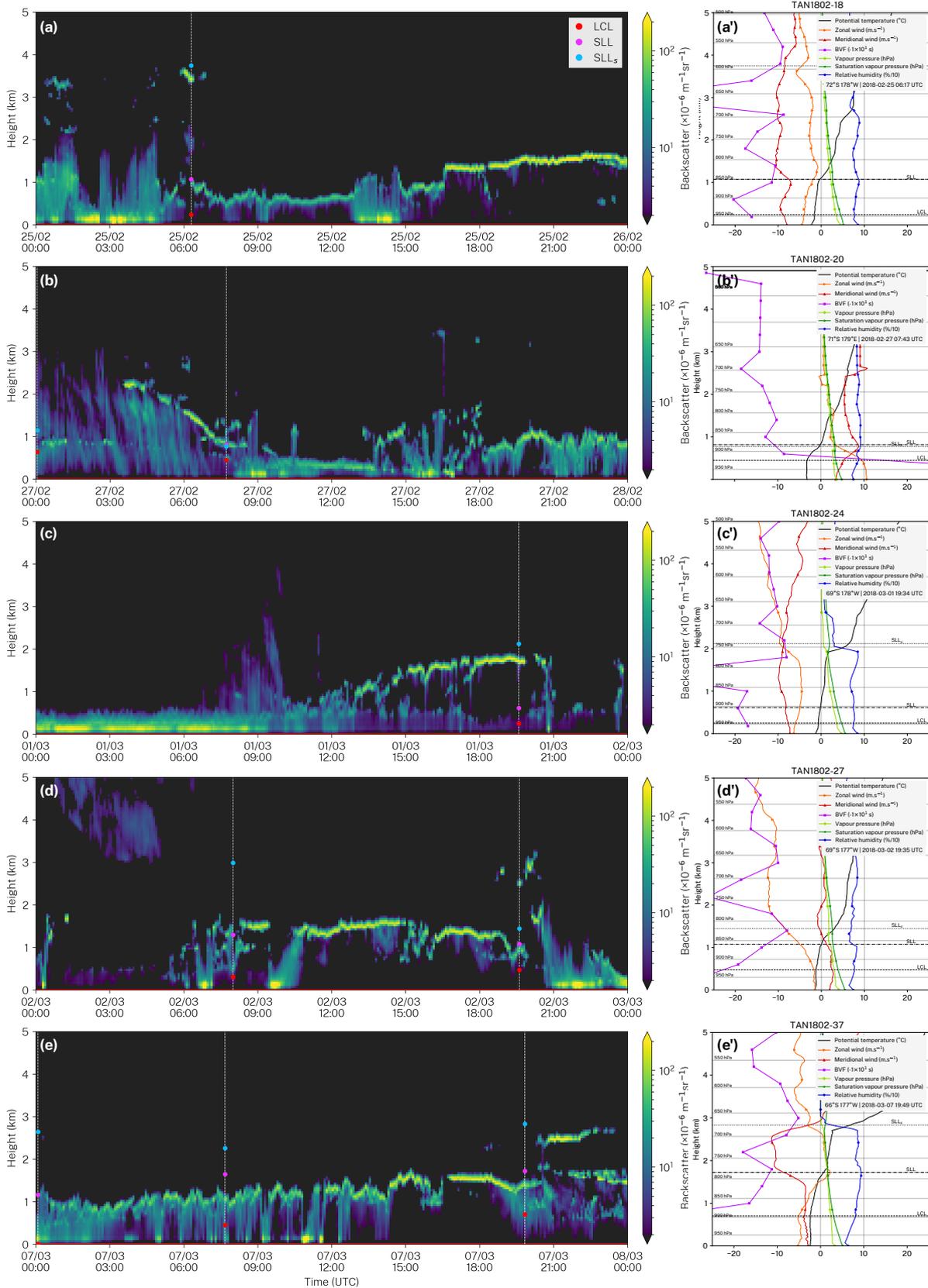


Figure 4.3 | Selected daily backscatter profiles collected on the TAN1802 voyage and radiosonde profiles. Shown are the lifting condensation level (LCL), sea surface temperature (SST) lifting level (SLL) and SST lifting level for a saturated parcel (SLL_s). Radiosonde launch times are indicated by a vertical line on the backscatter plots and the height of the LCL, SLL and SLL_s is indicated by a coloured dot.

4.4.2 Cloud representation

A characteristic feature of the SO cloud observed on the TAN1802 voyage by a ceilometer were layers of relatively optically thick, but geometrically thin Sc at between 1 and 3 km ASL (Fig. 4.3, 4.4 and 4.5; in particular Fig. 4.4a1 17:00–0:00, b1 0:00–12:00, c1 4:00–9:00, e1 12:00–20:00, Fig. 4.5a1 0:00–4:00). These were commonly accompanied by Cu fractus below the Sc layer. Apart from these two types of cloud, a significant amount of fog was observed on the voyage (Sect. 4.4.4). In Fig. 4.4 and 4.5 we compare the observed and simulated backscatter in the UM11.4cnt and UM11.4exp on multiple days of the voyage, as well as the thermodynamic levels. The selected days contain substantial amount of Sc cloud layers, and we omitted days which were dominated by clear sky, fog and precipitation. We note that temporal and cloud height correspondence between the observations and simulations is generally very good, partly due to the high temporal output resolution of the model of 20 min (Kuma et al., 2020b). Both are characterised by boundary layer cloud below 2 km ASL, and this altitude corresponds very well with the z_{par} level diagnosed by the model, i.e. the highest level reached by convection. Likewise, the LCL corresponds relatively well between the radiosonde observations and the model (red dot in OBS vs. red line in the UM in Fig. 4.4 and 4.5; see also Kuma et al. (2020a)). However, clouds simulated by the UM11.4cnt are clearly different from observations in a number of ways. While the periods of fog are relatively well simulated, the Cu fractus cloud layers forming at and above the LCL are clearly overestimated in the UM11.4cnt. Most seriously, the very well-defined Sc cloud layers are completely absent in the UM11.4cnt, and are only represented by intermittent vertical streaks of cloud instead of a coherent horizontal "stratiform" development. We can expect this factor to have a strong impact on the SW radiation balance due to overestimated Cu reflectivity and underestimated Sc reflectivity. Our aim with the experimental run is therefore to improve the simulation of Sc cloud as a major deficiency of the control run.

By increasing surface moisture flux in the UM11.4exp we increase the amount of moisture in the surface layer, and by increasing mass flux we in turn increase the amount of air mass and moisture transported from the SML to z_{par} . The impact on the simulated cloud backscatter is an increase in the amount of cloud just below z_{par} (Fig. 4.4 and 4.5), which is more consistent with observations. Some of the more prominent examples of an improved Sc are 4.4b3 3:00–9:00, c3 17:00–0:00, e3 12:00–21:00 and Fig. 4.5b3. A secondary effect of the increased mass flux is a decrease of the simulated Cu fractus as compensating convective downdrafts detrain warmer and drier air at the bottom of the plumes near the LCL (for example Fig. 4.4b3 3:00–9:00, e3 12:00–20:00 and Fig. 4.5a3, b3). This is potentially a positive development considering the overestimated Cu fractus in the UM11.4cnt compared to TAN1802 observations.

Figures 4.4 and 4.5 also show contour lines of the convective detrainment rate (DTRU) as diagnosed by the model's convection scheme. These largely signify the forced detrainment from convective updrafts of the mass flux scheme, whereby saturated air is removed from the updraft and detrained into the environment, thus increasing RH. Importantly, in the UM11.4cnt DTRU is locally concentrated in the vertical layer where we expect Sc cloud. However, this is apparently not enough to raise RH beyond the critical RH to initiate sufficient Sc cloud formation (also discussed later in Sect. 4.4.3). This indicates that the model is qualitatively correct in its representation of the BL, but that the magnitude of the change is too small. In the UM11.4exp DTRU was expanded

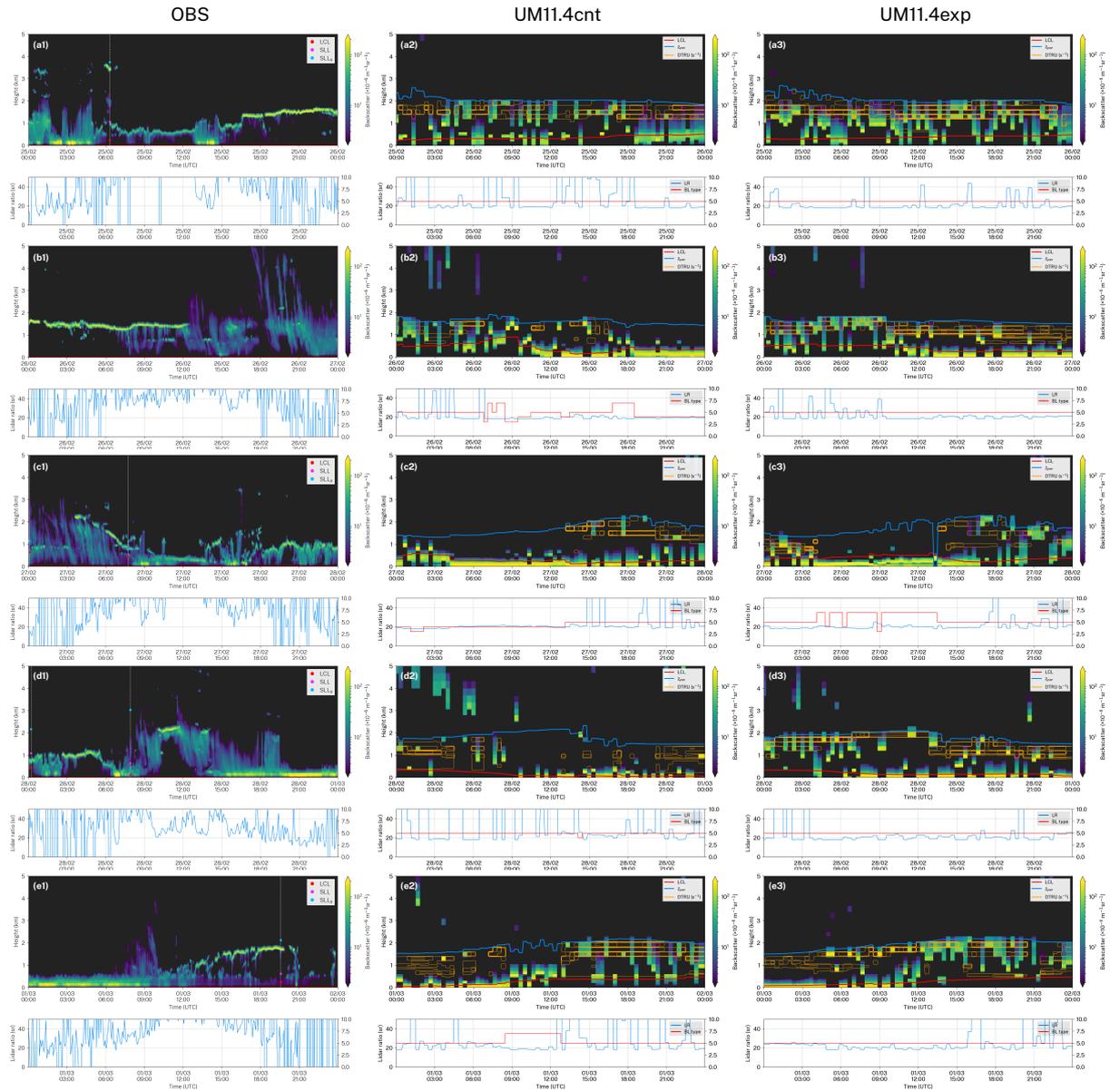


Figure 4.4 | Selected daily observed and simulated backscatter profile plots. The observations (OBS) were collected on the TAN1802 voyage and the simulations are based on the UM11.4 control (UM11.4cnt) and experimental (UM11.4exp) runs at the same time and geographical location as corresponding the observations. The plots also show the radiosonde lifting condensation level (LCL), sea surface temperature (SST) lifting level (SLL), saturated SST lifting level (SSL_s), the model LCL, diagnostic parcel height (z_{par}) and updraft detrainment rate (DTRU). The vertical extent is limited to 5 km ASL.

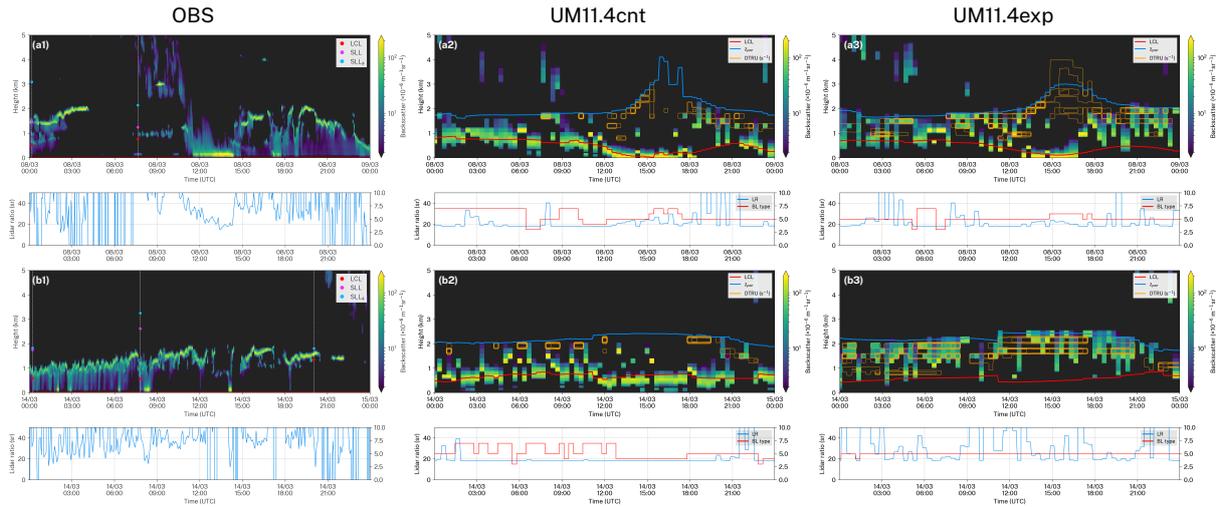


Figure 4.5 | Figure 4.4 continued.

and intensified. This was likely the key contributor to the increased Sc cloud, although a deficiency of this type of cloud persists in the UM11.4exp.

4.4.3 Boundary layer mass flux and relative humidity

Model BL cloud in the UM11.4 is driven by the underlying process of vertical moisture and air mass transport due to turbulence and convection (which can in turn be driven by large-scale dynamics). The large-scale cloud (PC2) scheme ensures that if enough RH is present in a particular model layer, liquid or ice cloud condensation happens. We examined the BL mass flux and RH in the UM11.4cnt and UM11.4exp on a number of days during the TAN1802 voyage (Fig. 4.6 and 4.7), corresponding to days analysed in Sect. 4.4.2. As explained in Sect. 4.3.5, the mass flux parametrisation extends vertically between the LCL and z_{par} . In this convective layer mass flux is positive and transports air from the LCL to z_{par} in updrafts and from z_{par} to the LCL in downdrafts. This corresponds with positive DTRU just below z_{par} . In some instances, intermittent mass flux was simulated by the UM11.4cnt (e.g. Fig. 4.7c1). This is most likely related to the stabilisation of the layer by compensating downdrafts, effectively shutting down convection until sufficient warm and moist air is replenished from the surface layer. We hypothesise that this may be an indication that the mass flux parametrisation operates more quickly than the surface turbulence with respect to moisture transport and therefore these two processes are not currently well-tuned to operate together. This appears to be the case in the unmodified model, but is intensified further in the experimental run.

RH in the UM11.4cnt was characterised by two local peaks at the LCL and z_{par} during periods when Sc cloud were observed. This is partially consistent with the expectation of Sc occurring preferentially at z_{par} . This was, however, apparently not enough for sufficient cloud formation in this layer since RH is typically below 85%. In the UM11.4exp run, the positive mass flux extent was increased and intensified, which resulted in a greater extent and intensity of DTRU. The problem of mass flux intermittency was still present in the experimental run. The increased mass flux had an obvious impact on RH, which separated the LCL and z_{par} levels more clearly as two local peaks of

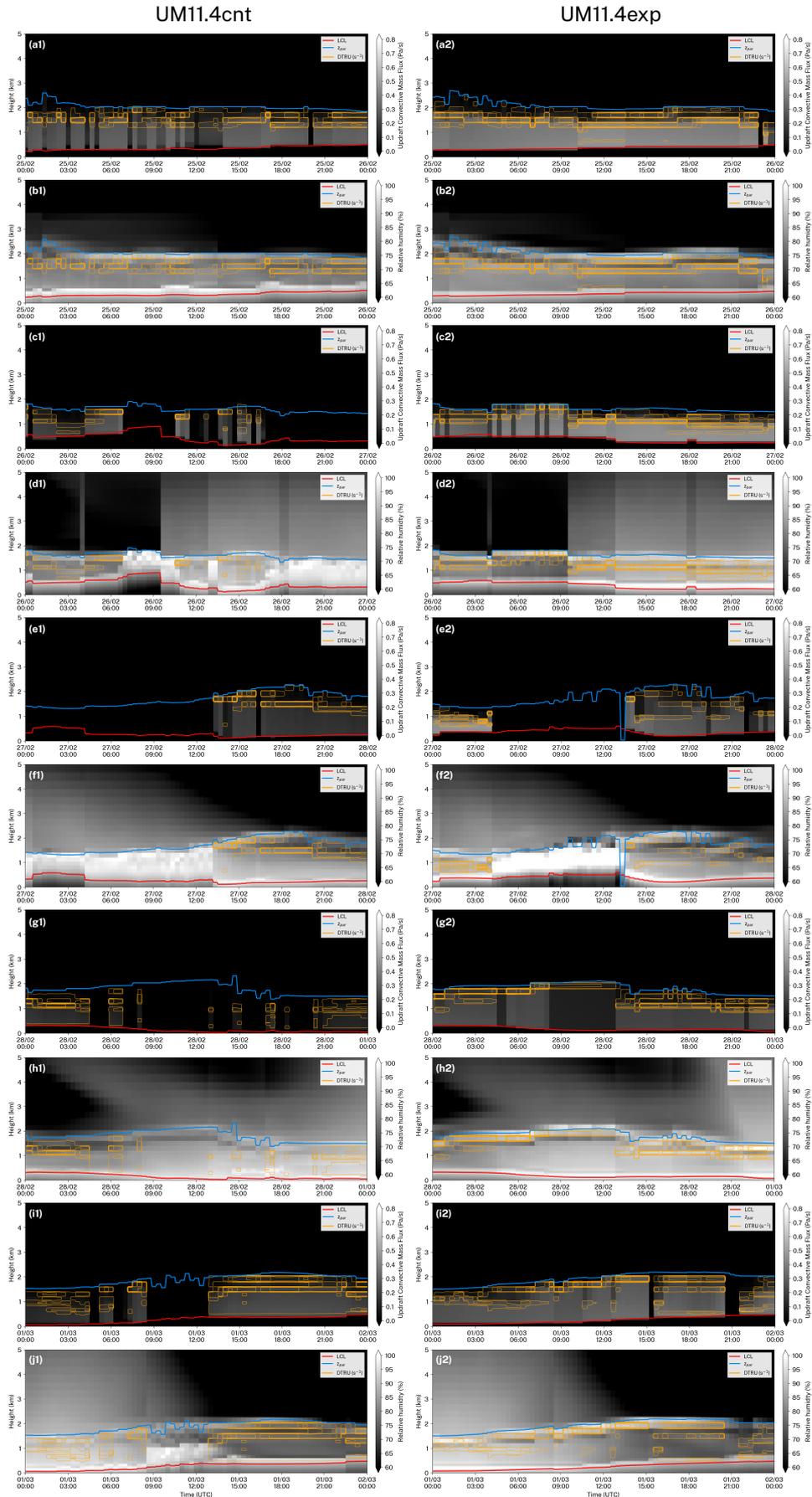


Figure 4.6 | Mass flux (odd rows) and relative humidity (even rows) profiles in the UM11.4 control (UM11.4cnt) and experimental (UM11.4exp) runs, corresponding to the plots in Fig. 4.4.

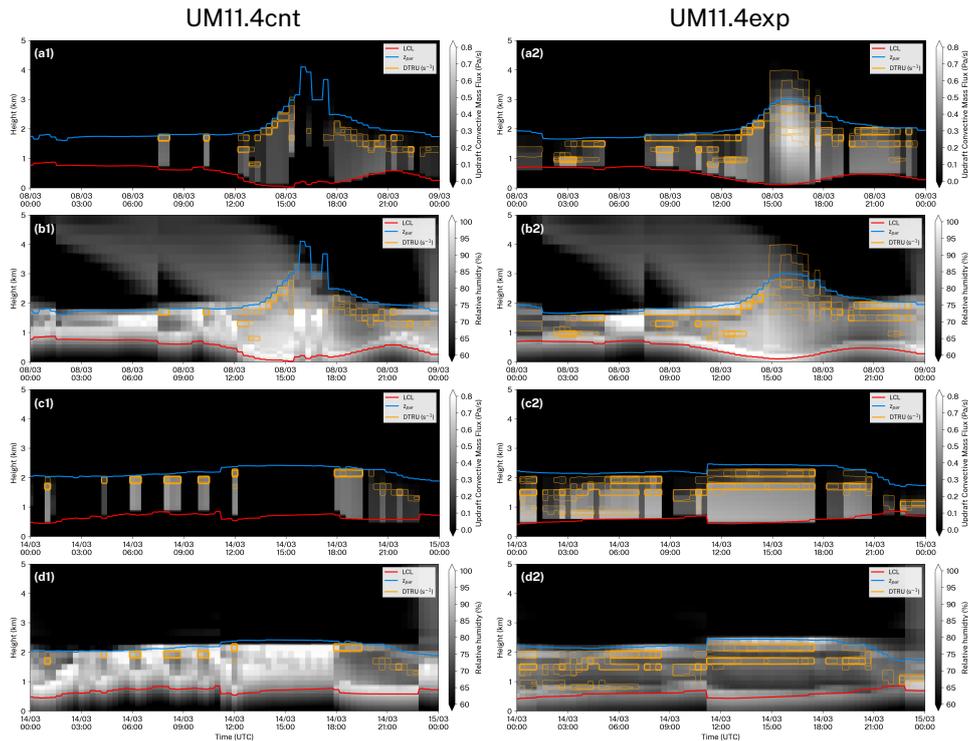


Figure 4.7 | Figure 4.6 continued.

RH, with a relatively dry convective layer between these two levels. RH at z_{par} was also increased, and this lead to a greater formation of Sc cloud (Sect. 4.4.2). However, the modifications in the UM11.4exp did not appear to be sufficient to fully address the problem of underestimated Sc cloud. RH at z_{par} was still too low to enhance cloud formation in many cases, and our experiments with increasing mass flux further (not shown) indicate that the mass flux parametrisation cannot by itself solve the problem further. By a more extreme tuning we also risk increasing mass flux to unphysical values. Instead, the problem appears to be with the coupling of the surface layer turbulence with the mass flux parametrisation, whereby not enough air mass and moisture is transported across the LCL from the SML to the convection layer. In other words, the BL scheme mixes moisture within the surface layer (up to the LCL) by turbulence and the convection scheme transports air between the LCL and z_{par} by updrafts and downdrafts, but there is too little flux across the LCL, effectively decoupling the two schemes. Compared to our radiosonde observations, the peak of RH in the SML seems unphysical and produced artificially by the separation of the BL and convection schemes into two distinct vertical sections.

4.4.4 Cloud occurrence statistics

While the case-based comparison of observed and simulated backscatter shows an improvement of Sc cloud simulation, it is important that the impact on the long-term cloud occurrence statistics is positive. We analysed cloud occurrence by height as determined by a cloud masking algorithm applied to the backscatter profiles (Kuma et al., 2020b). Figure 4.8 compares cloud occurrence in observations, the UM11.4cnt and UM11.4exp runs. The observed cloud peaked near the surface

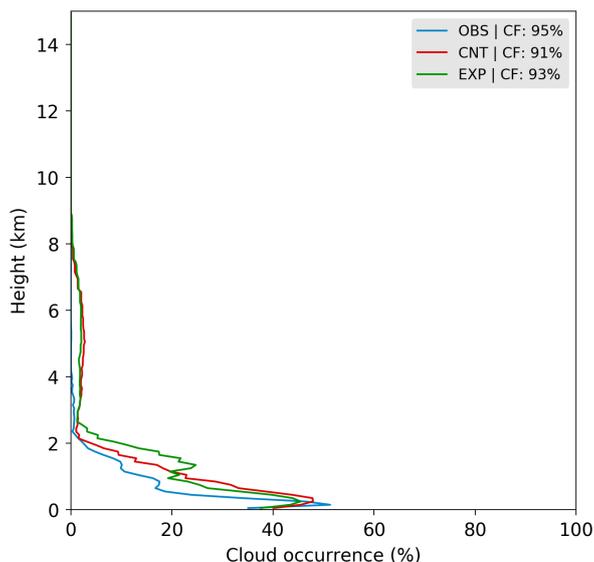


Figure 4.8 | Cloud occurrence histogram as a function of height calculated from 28 days of the TAN1802 voyage calculated from ceilometer observations (OBS), simulated lidar data in the control run of the UM11.4 (CNT) and the experimental run of the UM11.4 (EXP). Shown is also the total cloud fraction.

due to the frequent fog occurrence previously discussed in Kuma et al. (2020a), with two smaller peaks at about 500 m and 1 km ASL due to Sc cloud. Mid-level and high cloud above 2 km ASL was insignificant, probably due to observational constraints (lidar signal fully attenuated by fog and low cloud). The total cloud (+fog) fraction was observed to be 95%. The UM11.4cnt represented the statistical cloud occurrence remarkably well, especially of fog, but possibly overestimated cloud below 2 km ASL. Considering the results of the case study approach (Sect. 4.4.2), we think this is related to the overestimation of Cu cloud in the model. The peaks related to Sc cloud were not present in the UM11.4cnt, as one would expect from the lack of Sc cloud visible on the daily backscatter plots. The UM11.4exp showed a very similar cloud occurrence as the UM11.4cnt, with the exception of a much stronger peak which we believe is associated with the Sc cloud at about 1.5 km ASL. Therefore, the modifications in the UM11.4ext had the desirable effect of increasing Sc cloud statistically, but the overall pattern of vertical cloud occurrence is not substantially better than in the UM11.4cnt simulation. The total cloud fraction in the UM11.4exp was marginally improved at 93% (relative to observed 95% and control of 91%).

4.4.5 Shortwave radiation bias

While the modifications in the UM11.4exp were aimed at achieving a better match with lidar observations of BL cloud, these are also expected to result in an improved TOA radiation balance due to the strong effect of clouds on the planetary albedo. Figure 4.9 shows the absolute and relative reflected SW radiation in CERES, the UM11.4cnt and UM11.4exp during the time period of in-situ observations. Similar to previous versions of the UM (Kuma et al., 2020a; Schuddeboom et al., 2019), the bias in the UM11.4cnt in the SO is characterised by a bipolar zonally-symmetric pattern of negative biases in the high-latitude SO and positive bias in the low-latitude SO (Fig. 4.9e). As shown in our previous multi-voyage ceilometer evaluation (Kuma et al., 2020a), this is likely

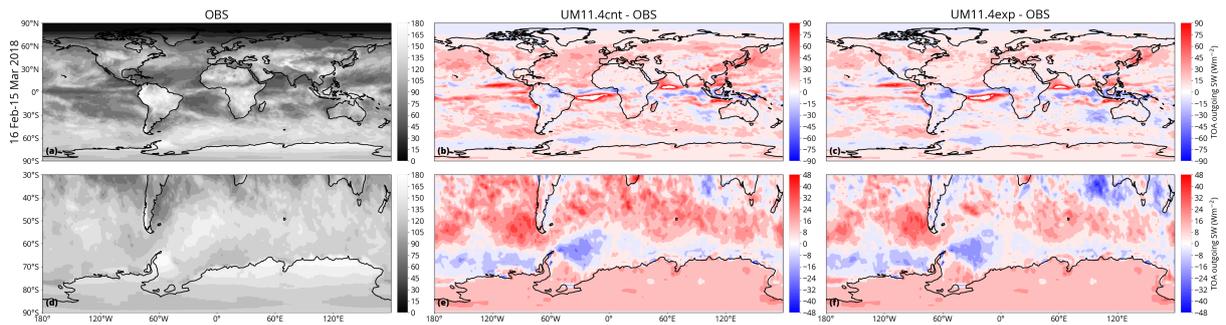


Figure 4.9 | Top of atmosphere outgoing shortwave radiation in CERES SYN (OBS) and bias in the UM11.4 control (UM11.4cnt) and experimental (UM11.4exp) runs relative to CERES SYN during the period of TAN1802 observations 16 February–15 March 2018. **(a, d)** show absolute values and **(b, c, e, f)** show values relative to CERES SYN. The top row **(a–c)** shows the whole globe, the bottom row **(d–f)** shows the Southern Ocean specifically.

caused by the “too few, too bright” cloud problem. The experimental run displays an improved SO SW radiation bias, especially by unexpectedly reducing the positive bias in the low-latitude SO (Fig. 4.9f). Globally, the bias in the UM11.4cnt was positive in most regions (Fig. 4.9b). This has been reduced in the UM11.4exp, without a significant deleterious impact on the existing regions of negative bias. We stress, however, the limited time period of the comparison (16 February–15 March 2018).

The zonal average of the TOA SW bias over the time period of the in-situ observations shows mostly positive bias in the UM11.4cnt, peaking at about 24 W m^{-2} near the equator (Fig. 4.10). In the UM11.4exp the bias is reduced by up to 5 W m^{-2} , with most latitudes experiencing a reduction of bias. Surprisingly, the southern part of the SO was largely unchanged, despite the improvement in the representation of the Sc cloud relative to the in-situ observations.

4.4.6 Boundary layer types

The BL scheme classifies a grid cell as one of six grid cell BL types (Sect. 4.3.2). In the UM11.4cnt in the time period of the in-situ observations, the most frequent BL types in the SO were Type I (stable BL) and V (decoupled Sc over Cu), peaking at about 80 and 90% in some regions of the SO, respectively (Fig. 4.11a, e). Type I was prevalent mostly in the low-latitude SO, while Type V was prevalent primarily in the high-latitude SO. This difference might partially explain the different SW radiation bias of these two zones (Sect. 4.4.5). In the UM11.4exp, the distribution of BL types has significantly changed globally relative to the control run (Fig. 4.11g–l). The most significant change in the SO was the increase in the occurrence of Type V in both the low- and high-latitude SO (Fig. 4.11k), while Type I had a minor increase in the SO (Fig. 4.11g). This suggests that the increase of Type V in the low-latitude SO might be associated with the improved SW radiation bias in this zone.

4.5 Discussion and conclusions

We analysed 28 days of voyage data in the Ross Sea and identified a common three-layer cloud profiles composed of Cu fractus, Sc and occasional Ac, associated with the thermodynamic levels LCL,

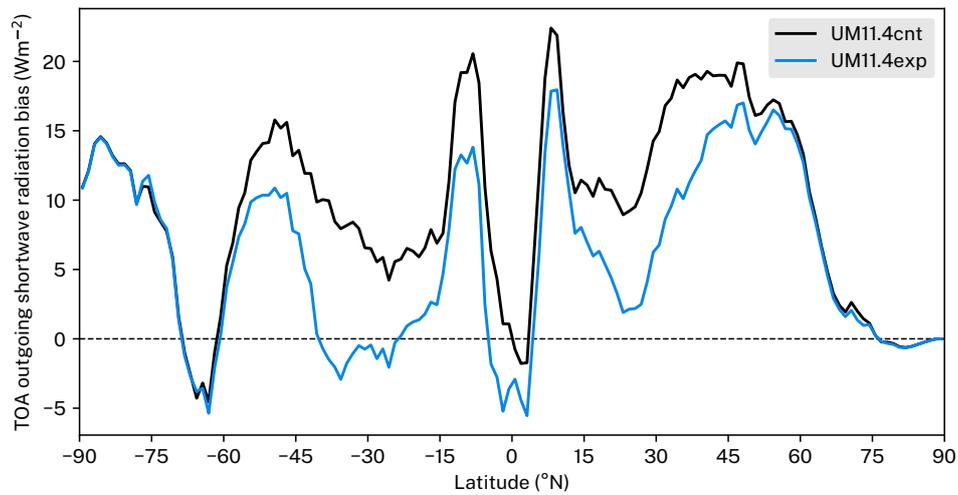


Figure 4.10 | Zonal average of top of atmosphere outgoing shortwave radiation in the UM11.4 control (UM11.4cnt) and experimental (UM11.4exp) runs relative to CERES SYN during the period of TAN1802 observations 16 February–15 March 2018.

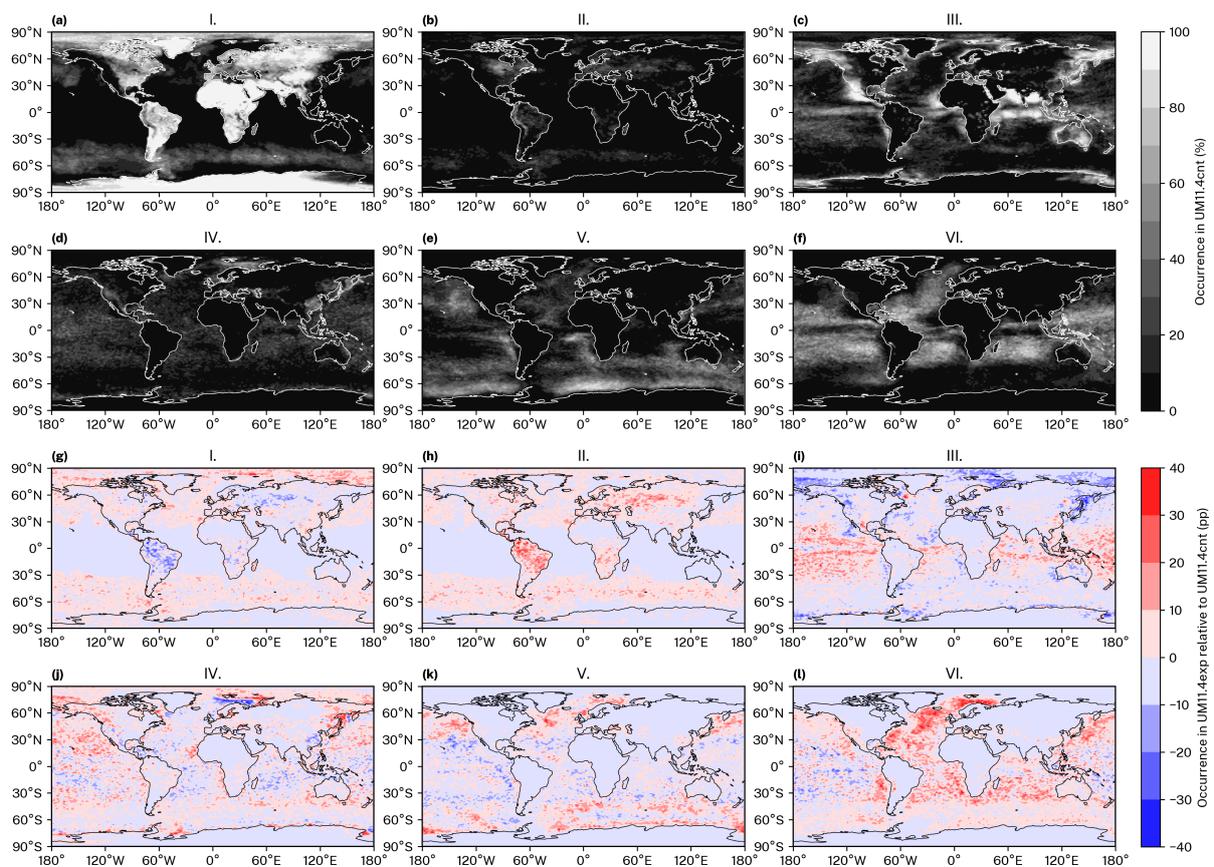


Figure 4.11 | Boundary layer type histograms of Type I–VI (Lock et al., 2000) in the UM11.4cnt (a–f) and UM11.4exp relative to the UM11.4cnt (g–l) expressed as percentage point (pp) difference.

SLL and SLL_s, respectively. This suggests a strong role of thermodynamics in the SO BL cloud formation. We analysed a control run of the UM11.4 in comparison with ceilometer observations using a lidar simulator framework. We found Sc cloud grossly underrepresented in the model, indicating that the current BL and convection schemes are not able to simulate this type of cloud. Considering these results, we prepared an experimental run which increased the amount of moisture flux from the sea surface and increased the convective mass flux, in order to generate more Sc by increasing RH at the top of the BL. We showed that this experimental run was more successful in simulating Sc, but other modifications are likely needed to achieve a satisfactory correspondence with the observations. The experimental run showed a greater ability to couple the surface with the top of the BL, but the connection between the BL and convection scheme appears to be too weak to allow sufficient transport of air mass and moisture across the LCL. Therefore, the artificial vertical separation of operation of the BL and convection schemes to surface–LCL and LCL– z_{par} regions appears to hinder the transport of moisture across the LCL, and to the top of the BL.

We showed that the experimental run not only improves Sc cloud in the Ross Sea, but also improves the SW radiation balance in the SO, especially in the low-latitude SO north of 60°S. In other parts of the globe the effect was also positive, with a decrease of zonal SW radiation bias of up to 5 Wm⁻² by reducing the common positive bias of the model. The effect on the BL types are primarily switching from the stable BL Type I to the convective Type V (Sc over Cu) across the whole of the SO. Currently we do not suggest that the tuning in the experimental model is integrated into the main model due to the modifications being relatively extreme. Our results, however, suggest that if the coupling between the BL and convection schemes across the LCL is improved, either more minor or no tuning would be required to obtain sufficient moisture transport from the surface to the top of the BL for Sc cloud formation. The SW radiation bias results indicate that the effect on the rest of the globe might be positive rather than negative, which would be otherwise expected if only one region (the Ross Sea) is taken into consideration in any tuning of the model. We also showed that a ground-based lidar simulator applied on high temporal resolution model output can be a useful tool for an analysis and improvement of BL cloud using a case study approach.

Author contributions

Peter Kuma participated on the TAN1802 field measurements, performed the model runs, analysis and wrote the manuscript; Adrian McDonald organised the deployment of instruments on the TAN1802 voyage and participated on continuous discussion of the analysis and manuscript; Olaf Morgenstern participated on continuous discussion of the analysis and manuscript; Sean Hartery participated on the TAN1802 field measurements; Jonny Williams, Vidya Varma and Guang Zeng contributed substantially to the nudged UM setup; Mike Harvey organised the deployment of instruments on the TAN1802 voyage; Simon Parsons and Graeme Plank participated on the deployment of instruments on the TAN1802 voyage. All authors reviewed the manuscript.

Acknowledgements

We would like to acknowledge the contributions of those who participated on the TAN1802 voyage and deployment, especially the crew of the voyage; Alex Schuddeboom for reviewing the manuscript; the UK Met Office for providing the HadGEM3-GA7.1/UM11.4 model; the CERES project for providing publicly available data via the NASA Langley Research Center CERES ordering tool (<https://ceres.larc.nasa.gov>); the NeSI high-performance computing facilities provided by the NZ eScience Infrastructure and funded jointly by NeSI's col-

laborator institutions and the Ministry of Business, Innovation & Employment's Research Infrastructure programme (<https://www.nesi.org.nz>); the financial support of the Deep South National Science Challenge via the Clouds and Aerosols project; the open source software we used to produce this analysis: Python, R, numpy, scipy, matplotlib, netCDF4, Linux and Devuan GNU/Linux.

Chapter 5

Conclusions and further work

We conducted ship-based observations on the 2-week-long voyage TAN1702 of RV *Tangaroa* to the Campbell Plateau in March 2017 and the 6-week-long voyage TAN1802 of RV *Tangaroa* to the Ross Sea in February–March 2018. These included ceilometer, lidar, radar, radiosonde, all-sky camera, automatic and human weather observations and aerosol measurements. We deployed ceilometers, radar and a sky cameras on additional voyages: the December 2016 voyages of HMNZS *Wellington* and an April–May voyage NBP1704 of RV *Nathaniel B. Palmer* to the Ross Sea. In addition, we collated and processed observations from previous voyages of *Aurora Australis*, TAN1502, TAN1503 and land-based ceilometer observations on Macquarie Island. We set up and performed nudged simulations of the Global Atmosphere version 7.1 (GA7.1) on a supercomputer of the New Zealand eScience Infrastructure (NeSI) and the National Institute of Water & Atmospheric Research (NIWA). We compared the model output with our collection of ship-based observations, focusing on geometrical properties of clouds observed by ceilometers and their links to thermodynamical profiles of the atmosphere observed with radiosondes. We contrasted these with the MERRA-2 reanalysis. Both the nudged GA7.1 and MERRA-2 showed significant cloud occurrence biases of 4–18% less cloud than in ceilometer observations. Despite of this, both showed a bipolar shortwave (SW) radiation bias by latitude, with outgoing SW radiation underestimated in high-latitude SO and overestimated in low-latitude SO. This suggests that compensating biases of cloud occurrence vs. cloud optical thickness are present in the models. This bipolar SW radiation bias was strongly correlated with near-surface air temperature, with relatively low air temperature regions exhibiting a negative SW radiation bias and relatively high temperature regions exhibiting a positive SW radiation bias. By analysing lifting levels, lifting condensation level (LCL) and the level of neutral buoyancy of an air parcel with potential temperature equal to sea surface temperature (SST), we showed that cloud base in the region is strongly linked to these levels, and these levels were a better predictor for cloud base than the previously utilised lower tropospheric stability (LTS). This finding suggests that local thermodynamics is a strong driver of the summertime SO stratocumulus (Sc) clouds, a finding also supported by Hartery et al. (2020a). Interestingly, we found very large differences in cloud phase simulated by the nudged GA7.1 and MERRA-2. MERRA-2 simulated much greater amount of liquid cloud (the majority of the water path), while the GA7.1 simulated slightly more ice than liquid. Even though we did not use an observational reference for cloud phase, we can say that the cloud phase was most likely behind the difference in cloud biases between the GA7.1 and MERRA-2 (in addition to the identified cloud occurrence biases), which were much more positive in MERRA-2 (liquid cloud reflects more SW radiation for the same amount of water path). We conclude that cloud geometry (vertical distribution of

cloud fraction) biases play at least an equally significant role in radiation biases as cloud phase and these two can be compensating biases, a fact which was not given enough recognition in previous studies. The source of the bipolar biases and its strong relationship with near-surface temperature, however, is still relatively unclear.

In order to compare ceilometer and lidar observations with models, we developed the Automatic Lidar and Ceilometer Framework (ALCF), a tool which extends the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) lidar simulator with support for a range of ground-based instruments and a processing pipeline which allows for an unbiased comparison of simulated and observed backscatter and a cloud mask. This tool was the subject of Chapter 3. While the COSP lidar simulator provided a basis for the physical simulation of laser radiation transfer through the atmosphere, more work needed to be done to support the ground-based instruments with their variety of field of view (FOV), wavelengths and calibration issues. By developing and documenting this tool, we enabled future studies utilising off-the-shelf ground-based lidars for model evaluation. The potential of ground-based lidars has so far been underutilised compared to similar active satellite observations (CALIPSO) due to the lack of software processing tools and the difficulty of accurate calibration of these instruments. The problems addressed by our work included Mie scattering at different laser wavelengths depending on droplet size distribution parameters, absolute calibration by utilising Sc clouds and molecular backscattering, backscatter noise removal, cloud mask/cloud base determination and evaluation of noise characteristics of various common lidars. We demonstrated the usefulness of this new tool on several case studies. Several common atmospheric models and reanalyses are supported by this framework, which allowed us to compare a range of publicly-available model output with ceilometer and lidar observations at several stations in Aotearoa/New Zealand. We showed how this tool can be used for comparing vertical cloud occurrence, backscatter "curtain" plots and cloud fraction vs. optical thickness. More processing algorithms can be incorporated in this modular tool in the future to allow for calculation of derived products.

Lastly, we focused on evaluation of SO boundary layer cloud in the GA7.1 and how the BL turbulence and convection parametrisations affect BL cloud. Previously, we found that Sc cloud below 2 km above sea level (ASL) and fog were predominant in the SO, but underestimated in the model. We used the ceilometer observations collected on TAN1802 to evaluate the model cloud in a case study based approach, where we compared "curtain" backscatter plots between the model and observations using the ALCF. This allowed us to identify in detail that the model is missing the extensive layers of Sc cloud found in observations, while overestimating cumulus (Cu) clouds occurring below the Sc layers. In observations, these layers were found to correspond to LCL (Cu clouds), and the level of dry and moist neutral buoyancy of a parcel with potential temperature equal to SST lifted from the sea surface (Sc clouds). Therefore, BL thermodynamics was identified as being key to resolving these biases. By running a model experiment with an increased surface flux and mass flux, we tested a hypothesis that not enough moisture transport is simulated to accumulate moisture at the top of the BL for cloud condensation. When compared with observations, this experiment led to an improved occurrence of Sc clouds, but due to a limiting factor we were unable to attain a full correspondence with observations. We identified the coupling between the turbulence and convection parametrisations across LCL as the most likely bottleneck. Therefore, it

seems that the separation of parametrisation into turbulence and convection driven vertical regions appears to limit the formation of Sc clouds in the BL in the summertime SO. As noted by Lock et al. (2000), "although this use of two discrete mixing schemes is to some extent undesirable, it has been found to be essential, as allowing the eddy-viscosity-based scheme to act in steady cumulus layers can rapidly result in them erroneously evolving into well-mixed stratocumulus, ...". Our experimental run showed promising results in terms of SW radiation biases globally (evaluated on a short monthly time period in February–March 2018), where the zonal means were reduced by up to 5 W m^{-2} . We conclude that further modifications to the BL schemes are needed to improve Sc cloud simulation in the SO. The tuning in the experimental run also needs to be evaluated globally over a long time period to determine if it can be used operationally. This approach when individual clouds can be compared between a nudged model and observations means that compensating biases are avoided. Secondary to this approach should be evaluation of cloud optical thickness, cloud phase and the cloud–aerosol effects, which all contribute to the radiation biases. In this sense, our work is complementary to that of Revell et al. (2019), Hartery et al. (2020b) and Hartery et al. (2020a).

We can summarise our contributions as follows. The relatively large SO cloud biases and the resulting radiation biases are still common in GCMs and reanalyses today (Gettelman et al., 2020). Due to the relatively large extent of the SO, its role in the uptake of excess heat and CO_2 and being a buffer zone for warming of the Antarctic ice sheets, and the Southern Hemisphere meridional circulation, it has a crucial role in the Earth's climate system. Clouds are the major factor in Earth's albedo variability. Therefore, any cloud biases can result in large biases in a model's large scale atmospheric and oceanic circulation and the cryosphere. Unless these biases are minimised in GCM simulations of the past and present climate, the accuracy of future climate simulations is limited, especially when it comes to predicting the change of cloud cover, the Earth's albedo, circulation and the ice sheets. Using ship-based observations, we narrowed down the cloud climatology in the SO, its main drivers and factors involved in the cloud biases. We identified BL parametrisations of turbulence and convection as a major deficiency in the model, and suggested concrete improvements. This work can translate into eventually improving these parametrisations. The improved understanding of the drivers can be applied in the development of other GCMs and reanalyses. Development of the ALCF as a tool for evaluation of models using ground-based lidar observations can streamline the process of using these type of observations globally, and provide a unique opportunity to evaluate BL clouds, some of which are not visible with spaceborne instruments. This new tool substantially lowers the bar for utilising these instruments, and therefore it can enable future studies of BL clouds.

5.1 Further work

Even though we analysed a large part of the ground-based SO observations available to us (Chapter 2), they are still underutilised. For example, we have not analysed the dual-polarisation lidar data collected on TAN1802 by a MiniMPL. If properly calibrated, these could provide crucial information about supercooled liquid cloud. The micro rain radar (MRR-2) deployments have so far been underutilised. They allow for detection of precipitation, rain rate and vertical velocity in precipitation. With analysis of the raw data it is also possible to get measures of snowfall rate and their

related vertical motion which is more variable than rainfall. Model evaluation of precipitation is needed to complement cloud evaluation in order to make sure the right amount of moisture is being removed from the BL by precipitation and the precipitation has the correct phase (liquid or ice). Similar to clouds, precipitation has an effect on shortwave (SW) and longwave (LW) radiation. It is therefore important for the radiation balance, considering its frequent occurrence in the SO. A dual camera setup was installed on the TAN1802 voyage in order to allow for stereoscopic determination of cloud base height. This imagery has not been utilised so far, but a co-authored study has utilised sky camera images collected on the *Aurora Australis* voyages in comparison with ceilometer observations (Klekociuk et al., 2019a). Co-location with the ceilometer observations provides an opportunity for algorithm development and cross-validation of results. Likewise, cloud type and cloud fraction can be determined from sky camera imagery using a suitable algorithm. A pair of sky cameras could be utilised as a low-cost instrument for determining cloud base height, cloud fraction and cloud type on ships of opportunity (Klekociuk et al., 2019a).

As we identified, ground-based observations have a large additional value to satellite observations when it comes to BL cloud evaluation. More ground-based observations in the SO could narrow the spatial and temporal gaps in our understanding of the region. Therefore, making ship-based atmospheric observations more accessible is an important task. Currently, deployment of instruments such as lidar and radars on ships is logistically difficult and expensive. Smaller and less expensive instruments could mean that they can be deployed more widely and installed permanently on some ships. Smaller lidars typically have lower power and inferior noise characteristics, but a wide coverage could outweigh these deficiencies. Progress in lidar development could also mean smaller instruments may become equally powerful. The commercial nature of common instruments means that some instruments suffer from a number of technical problems and vendors typically show little interest in resolving these. Open hardware and open software design of instruments could substantially accelerate progress and collaboration. Off-the-shelf software defined radio (SDR) receivers and transceivers¹ have recently become widely available. They could provide a basis for development of improved radars which can utilise many frequencies. Open software for processing instrument data could mean that implementation of standard techniques and algorithms for calibration, resampling, noise removal and calculation of derived products become more available to the scientific community. Likewise, public sharing of ground-based observational data and building of dataset collections, similar to the common practice of releasing satellite datasets, has a potential to accelerate atmospheric research in the SO. Utilising unusual platforms such as the Argo floats (Roemmich et al., 2009; Roemmich and Team, 2009) as platforms for atmospheric measurements could provide a vast amount of data and a relatively dense coverage in the SO. Air–sea surface fluxes appear to be implicated in the GA7.1/UM11.4 cloud biases (Chapter 4). Yet, the Coupled Ocean–Atmosphere Response Experiment (COARE) formulas are potentially not well-tuned for high latitudes and atmospheric reanalyses exhibit large air–sea flux biases in the SO (Cerovečki et al., 2011). Evaluation of model air–sea flux biases relative to reliable in situ flux measurements is needed to make sure they are not a major cause of the BL cloud biases. Other related biases are in the representation of marine aerosols, which act as a source of cloud condensation nuclei (CCN) and ice nucleating particles (INPs) (Hartery et al., 2020b).

¹<https://limemicro.com>.

A reliable comparison of models and observations using a lidar simulator requires accurate absolute calibration of the observed backscatter (Chapter 3). We have identified various deficiencies in the vendor-supplied calibration such as the dead time, afterpulse and overlap calibration in the MiniMPL. These often result in range-dependent bias of the volume backscattering coefficient, which may be impossible to reliably correct for without dedicated calibration measurements. Currently, the ALCF only supports height-independent calibration utilising observations of liquid Sc clouds and comparison with a theoretically expected molecular backscatter (Chapter 4). An automated approach should be developed to calculate dead time, afterpulse and overlap corrections as an alternative to the vendor corrections, which appear to be unreliable. This could lower the difficulty of utilising ceilometer and lidar data by the wider scientific community, which is a significant barrier for a wide utilisation of ceilometers and lidars for model evaluation. Currently, the ALCF does not support calculation of cross-polarised backscatter. Adding support for simulation of cross-polarised backscatter would mean that both channels of the Sigma Space MiniMPL can be utilised in a comparison with a model and therefore evaluate the cloud phase. Likewise, aerosol and precipitation is currently not taken into account by the lidar simulator. This not only causes a systematic bias in comparison with observations, but also prevents the simulator being used for model aerosol and precipitation evaluation.

A large amount of automatic lidar and ceilometer (ALC) data have been collected by regional and global networks such as Cloudnet (Illingworth et al., 2007), E-PROFILE (Illingworth et al., 2018), EARLINET (Pappalardo et al., 2014), ICENET (Cazorla et al., 2017) and MPLNET (Weldon et al., 2006). Most of these networks utilise ceilometers and lidars already supported by the ALCF such as Vaisala CL31, CL51, Lufft CHM 15k and Sigma Space MiniMPL. These observations haven't been used extensively for model evaluation. There is a potential to process and compare these observations with models using the ALCF. Such comparison would complement past comparisons with satellite observations, and could rival satellite observations in terms of global spatial and temporal coverage.

Here, we utilised the Clouds and the Earth's Radiant Energy System (CERES) satellite observations of top of the atmosphere (TOA) radiation. Even though the errors of these observations may be relatively low compared to current model cloud biases, systematic errors in the dataset exist due to unequal diurnal sampling and the need for temporal and angle interpolation of the raw measurements to calculate daily averages. The relatively new instrument National Institute of Standards and Technology Advanced Radiometer (NISTAR) on the L1-stationed Deep Space Climate Observatory (DSCOVR) provides a unique viewpoint for Earth radiation observations, from which the sunlit part of the Earth, including polar latitudes, is visible continuously. It would be valuable to compare model radiation fluxes with these new observations to rule out the effect of systematic biases of CERES on the model TOA radiation evaluation.

An unresolved question about the origin of the bipolar SW radiation bias in the high- and low-latitude SO in GA7.1 remains. We know that the model underestimates low cloud, and at the same time, we identified some evidence that the optical thickness of individual clouds is overestimated. This may be the underlying reason for the overestimated reflected TOA SW radiation in the low-latitude SO.

We identified that BL parametrisation of mass flux can be tuned to enable Sc cloud simulation in

the SO (Chapter 4). However, a good observational reference for mass flux and the similarity relationships is lacking in the SO. The BL turbulence and convection parametrisation is based on large eddy simulations (LES) initialised from field experiments in the tropical and midlatitude ocean. These may be inappropriate for the SO. Therefore, new LES initialised from SO field experiments are needed to make sure these parametrisations are correct in this region.

We have shown that nudged simulations of GA7.1 provide a very good basis for identifying model biases compared to ground-based observations. As opposed to a free-running model, it can be reasonably assumed that any biases are not the result of a different weather situation simulated by the model than in reality, and that they are largely due to errors in the subgrid-scale parametrisation processes not assimilated in the ERA-Interim reanalysis driving the model or not an input to the model nudging algorithm. As shown in Chapter 4, a side-by-side comparison of modelled and observed cloud is feasible, and this can be utilised in future studies of model clouds.

In Chapter 2 we presented a dataset of ground-based observations in the SO. This dataset provides a unique and comprehensive information on SO atmospheric conditions and clouds. Work is underway to make this dataset documented and publicly available. To this end, derived products should be developed in order to make it easy for the scientific community to use this dataset. In general, a more concentrated effort is needed to streamline public sharing of atmospheric observations, especially considering the global and accelerating effect of climate change, which has been called "the defining challenge of our time" (WMO, 2020) by the United Nations Secretary-General A. Guterres. It is the author's opinion that the seriousness and urgency of the situation is vastly underestimated even by the atmospheric science community, which continues to hinder international cooperation by not sharing data and model code, and by publishing scientific results in paywalled journals, and therefore putting the well-being of future generations in jeopardy. A question remains whether the parametrisations can be formulated in a more physical manner without relying on the similarity theory relations, which may not be universally applicable globally.

In Chapter 4 we identified that increasing convective mass flux and surface heat flux can improve Sc cloud simulation. However, it remains to be proven if mass flux or surface fluxes are underestimated relative to a physical reference (either an observational reference or large eddy simulations). More model experiments need to be performed which increase flux between the turbulently-mixed surface mixed layer and the convective layer, as well as longer term climate simulation to make sure the modifications address the actual problem without introducing compensating biases, and they do not have a negative impact on the global radiation balance throughout the year. A new BL scheme "CoMorph" is currently in development at the UK Met Office. Our findings could contribute to the development of this new scheme, but more experiments need to be performed with this new scheme and comparison of this new UK Met Office scheme with the ground-based SO observations should be performed.

Bibliography

- G-SDSU, URL <https://cloud.gsfc.nasa.gov/index.php?section=14>, 2019.
- Alexander, S. and Protat, A.: Cloud properties observed from the surface and by satellite at the northern edge of the Southern Ocean, *Journal of Geophysical Research: Atmospheres*, 123, 443–456, <https://doi.org/10.1002/2017jd026552>, 2018.
- Barkstrom, B. R.: The earth radiation budget experiment (ERBE), *Bulletin of the American Meteorological Society*, 65, 1170–1185, <https://doi.org/10.1002/0471743984.vse2742>, 1984.
- Bastin, S., Chiriaco, M., and Drobinski, P.: Control of radiation and evaporation on temperature variability in a WRF regional climate simulation: comparison with colocated long term ground based observations near Paris, *Climate dynamics*, 51, 985–1003, <https://doi.org/10.1007/s00382-016-2974-1>, 2018.
- Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., Gedney, N., et al.: The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes, *Geoscientific Model Development*, 4, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>, 2011.
- Bodas-Salcedo, A.: COSP user’s manual: Version 1.3.1, 2010.
- Bodas-Salcedo, A., Webb, M., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S., Zhang, Y., Marchand, R., Haynes, J., Pincus, R., et al.: COSP: Satellite simulation software for model assessment, *Bulletin of the American Meteorological Society*, 92, 1023–1043, <https://doi.org/10.1175/2011BAMS2856.1>, 2011.
- Bodas-Salcedo, A., Williams, K., Field, P., and Lock, A.: The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: The role of midlatitude cyclone clouds, *Journal of Climate*, 25, 7467–7486, <https://doi.org/10.1175/JCLI-D-11-00702.1>, 2012.
- Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N., Dufresne, J.-L., Koshiro, T., Stevens, B., Wang, Z., and Yokohata, T.: Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models, *Journal of Climate*, 27, 41–56, <https://doi.org/10.1175/JCLI-D-13-00169.1>, 2014.
- Bodas-Salcedo, A., Hill, P., Furtado, K., Williams, K., Field, P., Manners, J., Hyder, P., and Kato, S.: Large contribution of supercooled liquid clouds to the solar radiation budget of the Southern Ocean, *Journal of Climate*, 29, 4213–4228, <https://doi.org/10.1175/jcli-d-15-0564.1>, 2016.
- Bohren, C. F. and Huffman, D. R.: *Absorption and scattering of light by small particles*, John Wiley & Sons, 1998.
- Bony, S., Webb, M., Bretherton, C., Klein, S., Siebesma, P., Tselioudis, G., and Zhang, M.: CFMIP: Towards a better evaluation and understanding of clouds and cloud feedbacks in CMIP5 models, *Clivar Exchanges*, 56, 20–22, 2011.
- Bony, S., Stevens, B., Frierson, D. M., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., et al.: Clouds, circulation and climate sensitivity, *Nature Geoscience*, 8, 261, <https://doi.org/10.1038/ngeo2398>, 2015.
- Bosilovich, M., Lucchesi, R., and Suarez, M.: MERRA-2: File specification, 2015.
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., et al.: Clouds and aerosols, in: *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 571–657, Cambridge University Press, 2013.
- Bouniol, D., Protat, A., Delanoë, J., Pelon, J., Piriou, J.-M., Bouysse, F., Tompkins, A. M., Wilson, D. R., Morille, Y., Haeffelin, M., et al.: Using continuous ground-based radar and lidar measurements for evaluating the representation of clouds in four operational models, *Journal of Applied Meteorology and Climatology*, 49, 1971–1991, <https://doi.org/10.1175/2010JAMC2333.1>, 2010.
- Bréon, F.-M. and Doutriaux-Boucher, M.: A comparison of cloud droplet radii measured from space, *IEEE Transactions on Geoscience and Remote Sensing*, 43, 1796–1805, <https://doi.org/10.1109/TGRS.2005.852838>, 2005.
- Bromwich, D. H., Nicolas, J. P., Hines, K. M., Kay, J. E., Key, E. L., Lazzara, M. A., Lubin, D., McFarquhar, G. M., Gorodetskaya, I. V., Grosvenor, D. P., et al.: Tropospheric clouds in Antarctica, *Reviews of Geophysics*, 50, <https://doi.org/10.1029/2011RG000363>, 2012.
- Campbell, J. R., Hlavka, D. L., Welton, E. J., Flynn, C. J., Turner, D. D., Spinhirne, J. D., Scott III, V. S., and Hwang, I.: Full-time, eye-safe cloud and aerosol lidar observation at atmospheric radiation measurement program sites: Instruments and data processing, *Journal of Atmospheric and Oceanic Technology*, 19, 431–442, [https://doi.org/10.1175/1520-0426\(2002\)019<0431:FTESCA>2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019<0431:FTESCA>2.0.CO;2), 2002.
- Cazorla, A., Casquero-Vera, J. A., Román, R., Guerrero-Rascado, J. L., Toledano, C., Cachorro, V. E., Orza, J. A. G., Cancillo, M. L., Serrano, A., Titos, G., Pandolfi, M., Alastuey, A., Hanrieder, N., and Alados-Arboledas, L.: Near-real-time processing of a ceilometer network assisted

- with sun-photometer data: monitoring a dust outbreak over the Iberian Peninsula, *Atmospheric Chemistry and Physics*, 17, 11 861–11 876, <https://doi.org/10.5194/acp-17-11861-2017>, 2017.
- Ceppi, P., Hwang, Y.-T., Frierson, D. M., and Hartmann, D. L.: Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing, *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2012GL053115>, 2012.
- Cerovečki, I., Talley, L. D., and Mazloff, M. R.: A comparison of Southern Ocean air–sea buoyancy flux from an ocean state estimate with five other products, *Journal of Climate*, 24, 6283–6306, <https://doi.org/10.1175/2011jcli3858.1>, 2011.
- Chang, F. and Li, Z.: The effect of droplet size distribution on the determination of cloud droplet effective radius, in: 11th ARM Science Team Meeting, Atlanta, Ga, pp. 19–23, 2001.
- Chepfer, H., Chiriaco, M., Vautard, R., and Spinhirne, J.: Evaluation of MM5 optically thin clouds over Europe in fall using ICESat lidar spaceborne observations, *Monthly weather review*, 135, 2737–2753, <https://doi.org/10.1175/MWR3413.1>, 2007.
- Chepfer, H., Bony, S., Winker, D., Chiriaco, M., Dufresne, J.-L., and Sèze, G.: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model, *Geophysical Research Letters*, 35, <https://doi.org/10.1029/2008GL034207>, 2008.
- Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J., Minnis, P., Stubenrauch, C., and Zeng, S.: The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP), *Journal of Geophysical Research: Atmospheres*, 115, <https://doi.org/10.1029/2009JD012251>, 2010.
- Chiriaco, M., Vautard, R., Chepfer, H., Haeffelin, M., Dudhia, J., Wanherdrick, Y., Morille, Y., and Protat, A.: The ability of MM5 to simulate ice clouds: Systematic comparison between simulated and measured fluxes and lidar/radar profiles at the SIRTa atmospheric observatory, *Monthly weather review*, 134, 897–918, <https://doi.org/10.1175/mwr3102.1>, 2006.
- Chiriaco, M., Dupont, J.-C., Bastin, S., Badosa, J., Lopez, J., Haeffelin, M., Chepfer, H., and Guzman, R.: ReOBS: a new approach to synthesize long-term multi-variable dataset and application to the SIRTa supersite, *Earth System Science Data*, 10, 919, <https://doi.org/10.5194/essd-10-919-2018>, 2018.
- Chubb, T. H., Jensen, J. B., Siems, S. T., and Manton, M. J.: In situ observations of supercooled liquid clouds over the Southern Ocean during the HIAPER Pole-to-Pole Observation campaigns, *Geophysical Research Letters*, 40, 5280–5285, <https://doi.org/10.1002/grl.50986>, 2013.
- Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al.: The Joint UK Land Environment Simulator (JULES), model description—Part 2: carbon fluxes and vegetation dynamics, *Geoscientific Model Development*, 4, 701–722, <https://doi.org/10.5194/gmd-4-701-2011>, 2011.
- Coggins, J. H., McDonald, A. J., and Jolly, B.: Synoptic climatology of the Ross Ice Shelf and Ross Sea region of Antarctica: k-means clustering and validation, *International journal of climatology*, 34, 2330–2348, <https://doi.org/10.1002/joc.3842>, 2014.
- Comiso, J. C. and Nishio, F.: Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data, *Journal of Geophysical Research: Oceans*, 113, <https://doi.org/10.1029/2007JC004257>, 2008.
- Costa-Surós, M., Calbó, J., González, J., and Martín-Vide, J.: Behavior of cloud base height from ceilometer measurements, *Atmospheric Research*, 127, 64–76, <https://doi.org/10.1016/j.atmosres.2013.02.005>, 2013.
- Cromwell, E. and Flynn, D.: Lidar cloud detection with fully convolutional networks, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 619–627, IEEE, 2019.
- Damadeo, K. and Heather, H.: CERES: Clouds and the Earth’s Radiant Energy System, URL <https://eosps0.gsfc.nasa.gov/sites/default/files/publications/CERES%20Mission%20Brochure%20508.pdf>, 2017.
- Davidson, B.: The Barbados oceanographic and meteorological experiment, *Bulletin of the American Meteorological Society*, 49, 928–935, <https://doi.org/10.1175/1520-0477-49.9.928>, 1968.
- Davies, R., Loveridge, J., and Rampal, N.: Observed cloud morphology and inferred microphysics over the South Pacific from MISR and MODIS measurements of shortwave reflectivity, in: AIP Conference Proceedings, vol. 1810, p. 130001, AIP Publishing, <https://doi.org/10.1063/1.4975583>, 2017.
- Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, D. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- DeMott, P., Hill, T., Marchand, R., and Alexander, S.: Macquarie Island Cloud and Radiation Experiment (MICRE) Ice Nucleating Particle Measurements Field Campaign Report, 2018.
- Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A., et al.: Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview, *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1072–1087, <https://doi.org/10.1109/36.700992>, 1998.
- Dionisi, D., Barnaba, F., Diémoz, H., Di Liberto, L., and Gobbi, G. P.: A multiwavelength numerical model in support of quantitative retrievals of aerosol properties from automated lidar ceilometers and test applications for AOT and PM₁₀ estimation, *Atmospheric Measurement Techniques*, 11, 6013, <https://doi.org/10.5194/amt-11-6013-2018>, 2018.
- Doddridge, E. W. and Marshall, J.: Modulation of the seasonal cycle of Antarctic sea ice extent related to the Southern Annular Mode, *Geophysical Research Letters*, 44, 9761–9768, <https://doi.org/10.1002/2017gl074319>, 2017.
- Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., Wielicki, B. A., Young, D. F., and Sun, M.: Geostationary enhanced temporal interpolation for CERES flux products, *Journal of Atmospheric and Oceanic Technology*, 30, 1072–1090, <https://doi.org/10.1175/JTECH-D-12-00136.1>, 2013.

- Doelling, D. R., Haney, C. O., Scarino, B. R., Gopalan, A., and Bhatt, R.: Improvements to the geostationary visible imager ray-matching calibration algorithm for CERES Edition 4, *Journal of Atmospheric and Oceanic Technology*, 33, 2679–2698, <https://doi.org/10.1175/jtech-d-16-0113.1>, 2016.
- Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., et al.: The Japanese 55-year reanalysis “JRA-55”: an interim report, *Sola*, 7, 149–152, <https://doi.org/10.2151/jmsj.2016-015>, 2011.
- 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), <https://doi.org/10.24381/cds.bd0915c6>, URL <https://cds.climate.copernicus.eu/cdsapp#!/home>, 2019.
- Edwards, J. and Slingo, A.: Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model, *Quarterly Journal of the Royal Meteorological Society*, 122, 689–719, <https://doi.org/10.1002/qj.49712253107>, 1996.
- Emeis, S.: Surface-based remote sensing of the atmospheric boundary layer, vol. 40, Springer Science & Business Media, <https://doi.org/10.1007/978-90-481-9340-0>, 2010.
- Emeis, S., Schäfer, K., and Münkel, C.: Observation of the structure of the urban boundary layer with different ceilometers and validation by RASS data, *Meteorologische Zeitschrift*, 18, 149–154, <https://doi.org/10.1127/0941-2948/2009/0365>, 2009.
- Eresmaa, N., Karppinen, A., Joffre, S., Räsänen, J., and Talvitie, H.: Mixing height determination by ceilometer, *Atmospheric Chemistry and Physics*, 6, 1485–1493, <https://doi.org/10.5194/acp-6-1485-2006>, 2006.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., et al.: Taking climate model evaluation to the next level, *Nature Climate Change*, p. 1, <https://doi.org/10.1038/s41558-018-0355-y>, 2019.
- Fairall, C., Bradley, E. F., Hare, J., Grachev, A., and Edson, J.: Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm, *Journal of climate*, 16, 571–591, [https://doi.org/10.1175/1520-0442\(2003\)016<0571:BPOASF>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2), 2003.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models, in: *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 741–866, Cambridge University Press, <https://doi.org/10.1017/CBO9781107415324.020>, 2014.
- Flynn, C. J., Mendozaa, A., Zhengb, Y., and Mathurb, S.: Novel polarization-sensitive micropulse lidar measurement technique, *Optics express*, 15, 2785–2790, <https://doi.org/10.1364/OE.15.002785>, 2007.
- Franklin, C. N., Sun, Z., Bi, D., Dix, M., Yan, H., and Bodas-Salcedo, A.: Evaluation of clouds in ACCESS using the satellite simulator package COSP: Regime-sorted tropical cloud properties, *Journal of Geophysical Research: Atmospheres*, 118, 6663–6679, <https://doi.org/10.1002/jgrd.50496>, 2013.
- Frey, W., Morrison, A., Kay, J., Guzman, R., and Chepfer, H.: The combined influence of observed Southern Ocean clouds and sea ice on top-of-atmosphere albedo, *Journal of Geophysical Research: Atmospheres*, 123, 4461–4475, <https://doi.org/10.1029/2018JD028505>, 2018.
- Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., and Winton, M.: Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models, *Journal of Climate*, 28, 862–886, <https://doi.org/10.1175/JCLI-D-14-00117.1>, 2015.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *Journal of Climate*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- Geleyn, J. and Hollingsworth, A.: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation, *Contributions to Atmospheric Physics*, 52, 1979.
- Gettelman, A., Bardeen, C., McCluskey, C. S., Järvinen, E., Stith, J., and Bretherton, C.: Simulating Observations of Southern Ocean Clouds and Implications for Climate, *Earth and Space Science Open Archive*, <https://doi.org/10.1002/essoar.10502322.1>, 2020.
- Goldberg, M. D., Kilcoyne, H., Cikanek, H., and Mehta, A.: Joint Polar Satellite System: The United States next generation civilian polar-orbiting environmental satellite system, *Journal of Geophysical Research: Atmospheres*, 118, 13–463, <https://doi.org/10.1002/2013jd020389>, 2013.
- Goody, R. M. and Yung, Y. L.: *Atmospheric radiation: theoretical basis*, Oxford University Press, 2 edn., 1995.
- Grant, A.: Cloud-base fluxes in the cumulus-capped boundary layer, *Quarterly Journal of the Royal Meteorological Society*, 127, 407–421, <https://doi.org/10.1002/qj.49712757209>, 2001.
- Grant, A. and Brown, A.: A similarity hypothesis for shallow-cumulus transports, *Quarterly Journal of the Royal Meteorological Society*, 125, 1913–1936, <https://doi.org/10.1002/qj.49712555802>, 1999.
- Gregory, D. and Rowntree, P.: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure, *Monthly Weather Review*, 118, 1483–1506, [https://doi.org/10.1175/1520-0493\(1990\)118<1483:AMFCSW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2), 1990.
- Gregory, D., Wilson, D., and Bushell, A.: Insights into cloud parametrization provided by a prognostic approach, *Quarterly Journal of the Royal Meteorological Society*, 128, 1485–1504, <https://doi.org/10.1002/qj.200212858305>, 2002.

- Guichard, F. and Couvreur, F.: A short review of numerical cloud-resolving models, *Tellus A: Dynamic Meteorology and Oceanography*, 69, 1373–1378, <https://doi.org/10.1080/16000870.2017.1373578>, 2017.
- Hagihara, Y., Okamoto, H., and Yoshida, R.: Development of a combined CloudSat-CALIPSO cloud mask to show global cloud distribution, *Journal of Geophysical Research: Atmospheres*, 115, <https://doi.org/10.1029/2009JD012344>, 2010.
- Hansen, A., Ament, F., Grützun, V., and Lammert, A.: Model evaluation by a cloud classification based on multi-sensor observations, *Geoscientific Model Development Discussions*, 2018, 1–20, <https://doi.org/10.5194/gmd-2018-259>, 2018.
- Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., Onoda, H., Onogi, K., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: Representation of atmospheric circulation and climate variability, *Journal of the Meteorological Society of Japan. Ser. II*, 94, 269–302, <https://doi.org/10.2151/jmsj.2016-015>, 2016.
- Harries, J. E., Russell, J., Hanafin, J., Brindley, H., Futyan, J., Rufus, J., Kellock, S., Matthews, G., Wrigley, R., Last, A., et al.: The geostationary earth radiation budget project, *Bulletin of the American Meteorological Society*, 86, 945–960, <https://doi.org/10.1175/BAMS-86-7-945>, 2005.
- Hartery, S., Kuma, P., McGregor, J., Marriner, A., Sellegri, K., Saint-Macary, A., Law, C., von Hobem, M., Kremser, S., Lennartz, S., Archer, S., DeMott, P., Hill, T., Querel, R., Brailsford, G., Geddes, A., Parsons, S., McDonald, A., and Harvey, M.: Atmospheric Measurements During the Antarctic and Southern Ocean Marine Environment and Ecosystem Study (ASOMEES), manuscript in preparation, 2019.
- Hartery, S., Kuma, P., Harvey, M. J., and McDonald, A. J.: Quantification of Boundary Layer Mixing over the Southern Ocean Using In-Situ and Remotely-Sensed Measurements, *Geophysical Research Letters*, <https://doi.org/10.1002/essoar.10502904.1>, in review, 2020a.
- Hartery, S., Toohey, D., Revell, L., Sellegri, K., Kuma, P., Harvey, M., and McDonald, A. J.: Constraining the surface flux of sea spray particles from the Southern Ocean, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032026, <https://doi.org/10.1029/2019JD032026>, 2020b.
- Haynes, J., Luo, Z., Stephens, G., Marchand, R., and Bodas-Salcedo, A.: A multipurpose radar simulation package: QuickBeam, *Bulletin of the American Meteorological Society*, 88, 1723–1727, <https://doi.org/10.1175/BAMS-88-11-1723>, 2007.
- Haynes, J. M., Jakob, C., Rossow, W. B., Tselioudis, G., and Brown, J.: Major characteristics of Southern Ocean cloud regimes and their effects on the energy budget, *Journal of Climate*, 24, 5061–5080, <https://doi.org/10.1175/2011JCLI4052.1>, 2011.
- Heese, B., Flentje, H., Althausen, D., Ansmann, A., and Frey, S.: Ceilometer lidar comparison: backscatter coefficient retrieval and signal-to-noise ratio determination, *Atmospheric Measurement Techniques*, 3, 1763–1770, <https://doi.org/10.5194/amt-3-1763-2010>, 2010.
- Hines, K. M. and Bromwich, D. H.: Development and testing of Polar Weather Research and Forecasting (WRF) model. Part I: Greenland ice sheet meteorology, *Monthly Weather Review*, 136, 1971–1989, <https://doi.org/10.1175/2007MWR2112.1>, 2008.
- Hodges, K. I., Lee, R. W., and Bengtsson, L.: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25, *Journal of Climate*, 24, 4888–4906, <https://doi.org/10.1175/2011jcli4097.1>, 2011.
- Hogan, R. J.: Fast approximate calculation of multiply scattered lidar returns, *Applied Optics*, 45, 5984–5992, <https://doi.org/10.1364/AO.45.005984>, 2006.
- Hopkin, E., Illingworth, A. J., Charlton-Perez, C., Westbrook, C. D., and Ballard, S.: A robust automated technique for operational calibration of ceilometers using the integrated backscatter from totally attenuating liquid clouds, *Atmospheric Measurement Techniques Discussions*, 2019, 1–28, <https://doi.org/10.5194/amt-2018-427>, 2019.
- Hoskins, B. J. and Hodges, K. I.: A new perspective on Southern Hemisphere storm tracks, *Journal of Climate*, 18, 4108–4129, <https://doi.org/10.1175/jcli3570.1>, 2005.
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., et al.: The art and science of climate model tuning, *Bulletin of the American Meteorological Society*, 98, 589–602, <https://doi.org/10.1175/BAMS-D-15-00135.1>, 2017.
- Huang, Y., Siems, S. T., Manton, M. J., Protat, A., and Delanoë, J.: A study on the low-altitude clouds over the Southern Ocean using the DARDAR-MASK, *Journal of Geophysical Research: Atmospheres*, 117, <https://doi.org/10.1029/2012JD017800>, 2012.
- Huang, Y., Siems, S. T., Manton, M. J., Rosenfeld, D., Marchand, R., McFarquhar, G. M., and Protat, A.: What is the role of sea surface temperature in modulating cloud and precipitation properties over the Southern Ocean?, *Journal of Climate*, 29, 7453–7476, <https://doi.org/10.1175/JCLI-D-15-0768.1>, 2016.
- Hunter, J. D.: Matplotlib: A 2D graphics environment, *Computing in science & engineering*, 9, 90, <https://doi.org/10.1109/MCSE.2007.55>, 2007.
- Hwang, Y.-T. and Frierson, D. M.: Link between the double-Intertropical Convergence Zone problem and cloud biases over the Southern Ocean, *Proceedings of the National Academy of Sciences*, 110, 4935–4940, <https://doi.org/10.1073/pnas.1213302110>, 2013.
- Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., Wood, R. A., Meijers, A. J., Mulcahy, J., Field, P., et al.: Critical Southern Ocean climate model biases traced to atmospheric model cloud errors, *Nature communications*, 9, <https://doi.org/10.1038/s41467-018-05634-2>, 2018.
- Illingworth, A., Hogan, R., O’connor, E., Bouniol, D., Brooks, M., Delanoë, J., Donovan, D., Eastment, J., Gaussiat, N., Goddard, J., et al.: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations, *Bulletin of the American Meteorological Society*, 88, 883–898, <https://doi.org/10.1175/BAMS-88-6-883>, 2007.
- Illingworth, A., Cimini, D., Haeefe, A., Haeffelin, M., Hervo, M., Kotthaus, S., Löhnert, U., Martinet, P., Mattis, I., O’Connor, E., et al.: How can

- Existing Ground-Based Profiling Instruments Improve European Weather Forecasts?, *Bulletin of the American Meteorological Society*, <https://doi.org/10.1175/BAMS-D-17-0231.1>, 2018.
- Illingworth, A. J., Barker, H., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D. P., et al.: The EarthCARE satellite: The next step forward in global measurements of clouds, aerosols, precipitation, and radiation, *Bulletin of the American Meteorological Society*, 96, 1311–1332, <https://doi.org/10.1175/BAMS-D-12-00227.1>, 2015a.
- Illingworth, A. J., Cimini, D., Gaffard, C., Haefelin, M., Lehmann, V., Löhnert, U., O'Connor, E. J., and Ruffieux, D.: Exploiting existing ground-based remote sensing networks to improve high-resolution weather forecasts, *Bulletin of the American Meteorological Society*, 96, 2107–2125, <https://doi.org/10.1175/BAMS-D-13-00283.1>, 2015b.
- Jakob, C.: The representation of cloud cover in atmospheric general circulation models, Ph.D. thesis, Fakultät für Physik der Ludwig-Maximilians-Universität München, 2000.
- Jakob, C.: An improved strategy for the evaluation of cloud parameterizations in GCMs, *Bulletin of the American Meteorological Society*, 84, 1387–1402, <https://doi.org/10.1175/bams-84-10-1387>, 2003.
- Jin, D., Oreopoulos, L., and Lee, D.: Regime-based evaluation of cloudiness in CMIP5 models, *Climate dynamics*, 48, 89–112, <https://doi.org/10.1007/s00382-016-3064-0>, 2017.
- Jin, Y., Kai, K., Kawai, K., Nagai, T., Sakai, T., Yamazaki, A., Uchiyama, A., Batdorj, D., Sugimoto, N., and Nishizawa, T.: Ceilometer calibration for retrieval of aerosol optical properties, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 153, 49–56, <https://doi.org/10.1016/j.jqsrt.2014.10.009>, 2015.
- Jolly, B., Kuma, P., McDonald, A., and Parsons, S.: An analysis of the cloud environment over the Ross Sea and Ross Ice Shelf using Cloud-Sat/CALIPSO satellite observations: the importance of synoptic forcing, *Atmospheric Chemistry and Physics*, 18, 9723–9739, 2018.
- Jones, D. A. and Simmonds, I.: A climatology of Southern Hemisphere extratropical cyclones, *Climate Dynamics*, 9, 131–145, <https://doi.org/10.1007/BF00209750>, 1993.
- Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python, URL <http://www.scipy.org>, accessed 23 November 2018, 2001–.
- Kandel, R., Monge, J.-L., Viollier, M., Pakhomov, L., Adasko, V., Reitenbach, R., Raschke, E., and Stuhlmann, R.: The ScaRaB project: Earth radiation budget observations from the Meteor satellites, *Advances in Space Research*, 14, 47–54, [https://doi.org/10.1016/0273-1177\(94\)90346-8](https://doi.org/10.1016/0273-1177(94)90346-8), 1994.
- Kawai, H., Yukimoto, S., Koshiro, T., Oshima, N., Tanaka, T., Yoshimura, H., and Nagasawa, R.: Significant improvement of cloud representation in the global climate model MRI-ESM2, *Geoscientific Model Development*, 12, 2875–2897, <https://doi.org/10.5194/gmd-12-2875-2019>, 2019.
- Kay, J., Hillman, B., Klein, S., Zhang, Y., Medeiros, B., Pincus, R., Gettelman, A., Eaton, B., Boyle, J., Marchand, R., et al.: Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators, *Journal of Climate*, 25, 5190–5207, <https://doi.org/10.1175/JCLI-D-11-00469.1>, 2012.
- Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., and Bitz, C.: Global climate impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model (CESM), *Journal of Climate*, 29, 4617–4636, <https://doi.org/10.1175/JCLI-D-15-0358.1>, 2016.
- Kelleher, M. K. and Grise, K. M.: Examining Southern Ocean Cloud Controlling Factors on Daily Time Scales and Their Connections to Mid-latitude Weather Systems, *Journal of Climate*, 32, 5145–5160, <https://doi.org/10.1175/JCLI-D-18-0840.1>, 2019.
- Khlopenkov, K., Duda, D., Thieman, M., Minnis, P., Su, W., and Bedka, K.: Development of multi-sensor global cloud and radiance composites for earth radiation budget monitoring from DSCOVR, in: *Remote Sensing of Clouds and the Atmosphere XXII*, vol. 10424, p. 104240K, International Society for Optics and Photonics, <https://doi.org/10.1117/1.2278645>, 2017.
- Klein, S. A. and Hartmann, D. L.: The seasonal cycle of low stratiform clouds, *Journal of Climate*, 6, 1587–1606, [https://doi.org/10.1175/1520-0442\(1993\)006<1587:TSCOLS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2), 1993.
- Klein, S. A. and Jakob, C.: Validation and sensitivities of frontal clouds simulated by the ECMWF model, *Monthly Weather Review*, 127, 2514–2531, [https://doi.org/10.1175/1520-0493\(1999\)127<2514:VASOFC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2514:VASOFC>2.0.CO;2), 1999.
- Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., and Gleckler, P. J.: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, *Journal of Geophysical Research: Atmospheres*, 118, 1329–1342, <https://doi.org/10.1002/jgrd.50141>, 2013.
- Klekociuk, A. R., French, W. J. R., Alexander, S. P., Kuma, P., and McDonald, A. J.: The state of the atmosphere in the 2016 southern Kerguelen Axis campaign region, *Deep Sea Research Part II: Topical Studies in Oceanography*, <https://doi.org/10.1016/j.dsr2.2019.02.001>, 2019a.
- Klekociuk, A. R., French, W. J. R., Alexander, S. P., Kuma, P., and McDonald, A. J.: The state of the atmosphere in the 2016 southern Kerguelen Axis campaign region, *Deep Sea Research Part II: Topical Studies in Oceanography*, <https://doi.org/10.1016/j.dsr2.2019.02.001>, 2019b.
- Knepp, T. N., Szykman, J. J., Long, R., Duvall, R. M., Krug, J., Beaver, M., Cavender, K., Kronmiller, K., Wheeler, M., Delgado, R., Hoff, R., Berkoff, T., Olson, E., Clark, R., Wolfe, D., Van Gilst, D., and Neil, D.: Assessment of mixed-layer height estimation from single-wavelength ceilometer profiles, *Atmospheric Measurement Techniques*, 10, 3963–3983, <https://doi.org/10.5194/amt-10-3963-2017>, 2017.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Mori, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al.: The JRA-55 reanal-

- ysis: General specifications and basic characteristics, *Journal of the Meteorological Society of Japan*. Ser. II, 93, 5–48, <https://doi.org/10.2151/jmsj.2015-001>, 2015.
- Kotthaus, S., O'Connor, E., Münkel, C., Charlton-Perez, C., Haeffelin, M., Gabey, A. M., and Grimmond, C. S. B.: Recommendations for processing atmospheric attenuated backscatter profiles from Vaisala CL31 ceilometers, *Atmospheric Measurement Techniques*, 9, 3769–3791, <https://doi.org/10.5194/amt-9-3769-2016>, 2016.
- Kuma, P., McDonald, A., Morgenstern, O., Alexander, S., Cassano, J., Garrett, S., Halla, J., Hartery, S., Harvey, M., Parsons, S., Plank, G., Varma, V., and Williams, J.: Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations, *Atmospheric Chemistry and Physics*, <https://doi.org/10.5194/acp-2019-201>, accepted, 2020a.
- Kuma, P., McDonald, A., Morgenstern, O., Querel, R., Silber, I., and Flynn, C.: Ground-based lidar simulator framework for comparing models and observations (ALCF 1.0), *Geoscientific Model Development Discussions*, 2020, 1–45, <https://doi.org/10.5194/gmd-2020-25>, in review, 2020b.
- Kusahara, K., Reid, P., Williams, G. D., Massom, R., and Hasumi, H.: An ocean-sea ice model study of the unprecedented Antarctic sea ice minimum in 2016, *Environmental Research Letters*, 13, 084 020, <https://doi.org/10.1088/1748-9326/aad624>, 2018.
- Lang, F., Huang, Y., Siems, S., and Manton, M.: Characteristics of the Marine Atmospheric Boundary Layer Over the Southern Ocean in Response to the Synoptic Forcing, *Journal of Geophysical Research: Atmospheres*, 123, 7799–7820, <https://doi.org/10.1029/2018jd028700>, 2018.
- Lawson, R. P. and Gettelman, A.: Impact of Antarctic mixed-phase clouds on climate, *Proceedings of the National Academy of Sciences*, 111, 18 156–18 161, <https://doi.org/10.1073/pnas.1418197111>, 2014.
- Lewis, J. R., Campbell, J. R., Welton, E. J., Stewart, S. A., and Haftings, P. C.: Overview of MPLNET version 3 cloud detection, *Journal of Atmospheric and Oceanic Technology*, 33, 2113–2134, <https://doi.org/10.1175/JTECH-D-15-0190.1>, 2016.
- Li, J.-L., Waliser, D., Stephens, G., Lee, S., L'Ecuyer, T., Kato, S., Loeb, N., and Ma, H.-Y.: Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis, *Journal of Geophysical Research: Atmospheres*, 118, 8166–8184, <https://doi.org/10.1002/jgrd.50378>, 2013.
- Liou, K.-N.: An introduction to atmospheric radiation, vol. 84, Elsevier, 2 edn., 2002.
- Listowski, C., Delanoë, J., Kirchgaessner, A., Lachlan-Cope, T., and King, J.: Antarctic clouds, supercooled liquid water and mixed-phase investigated with DARDAR: geographical and seasonal variations, *Atmospheric Chemistry and Physics*, 19, 6771–6808, <https://doi.org/10.5194/acp-19-6771-2019>, 2019.
- Liu, J., Li, Z., Zheng, Y., and Cribb, M.: Cloud-base distribution and cirrus properties based on micropulse lidar measurements at a site in southeastern China, *Advances in Atmospheric Sciences*, 32, 991–1004, <https://doi.org/10.1007/s00376-014-4176-2>, 2015a.
- Liu, L., Sun, X.-j., Liu, X.-c., Gao, T.-c., and Zhao, S.-j.: Comparison of cloud base height derived from a ground-based infrared cloud measurement and two ceilometers, *Advances in Meteorology*, 2015, <https://doi.org/10.1155/2015/853861>, 2015b.
- Lock, A., Brown, A., Bush, M., Martin, G., and Smith, R.: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests, *Monthly weather review*, 128, 3187–3199, [https://doi.org/10.1175/1520-0493\(2000\)128<3187:ANBLMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2), 2000.
- Lock, A., Edwards, J., and Boutle, I.: Unified Model Documentation Paper 024: The Parametrization of Boundary Layer Processes, UK Met Office, 2019.
- Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition 4.0 Data Product, *Journal of Climate*, <https://doi.org/10.1175/jcli-d-17-0208.1>, 2017.
- Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product, *Journal of Climate*, 31, 895–918, <https://doi.org/10.1175/JCLI-D-17-0208.1>, 2018.
- Loveridge, J. and Davies, R.: Cloud Heterogeneity in the Marine Midlatitudes: Dependence on Large-Scale Meteorology and Implications for General Circulation Models, *Journal of Geophysical Research: Atmospheres*, 124, 3448–3463, <https://doi.org/10.1029/2018JD029826>, 2019a.
- Loveridge, J. and Davies, R.: Cloud Heterogeneity in the Marine Midlatitudes: Dependence on Large-Scale Meteorology and Implications for General Circulation Models, *Journal of Geophysical Research: Atmospheres*, 124, 3448–3463, <https://doi.org/10.1029/2018JD029826>, 2019b.
- Ludescher, J., Yuan, N., and Bunde, A.: Detecting the statistical significance of the trends in the Antarctic sea ice extent: an indication for a turning point, *Climate Dynamics*, <https://doi.org/10.1007/s00382-018-4579-3>, 2018.
- Mace, G. G. and Protat, A.: Clouds over the Southern Ocean as observed from the R/V Investigator during CAPRICORN. Part I: Cloud occurrence and phase partitioning, *Journal of Applied Meteorology and Climatology*, 57, 1783–1803, <https://doi.org/10.1175/JAMC-D-17-0194.1>, 2018a.
- Mace, G. G. and Protat, A.: Clouds over the Southern Ocean as observed from the R/V Investigator during CAPRICORN. Part II: The properties of nonprecipitating stratocumulus, *Journal of Applied Meteorology and Climatology*, 57, 1805–1823, <https://doi.org/10.1175/JAMC-D-17-0195.1>, 2018b.
- Mace, G. G., Zhang, Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., and Winker, D.: A description of hydrometeor layer occur-

- rence statistics derived from the first year of merged Cloudsat and CALIPSO data, *Journal of Geophysical Research: Atmospheres*, 114, <https://doi.org/10.1029/2007JD009755>, 2009.
- Madonna, F., Rosoldi, M., Lolli, S., Amato, F., Vande Hey, J., Dhillon, R., Zheng, Y., Brettle, M., and Pappalardo, G.: Intercomparison of aerosol measurements performed with multi-wavelength Raman lidars, automatic lidars and ceilometers in the framework of INTERACT-II campaign, *Atmospheric Measurement Techniques*, 11, 2459–2475, <https://doi.org/10.5194/amt-11-2459-2018>, 2018.
- Marchand, R., Mace, G. G., Ackerman, T., and Stephens, G.: Hydrometeor detection using CloudSat—An Earth-orbiting 94-GHz cloud radar, *Journal of Atmospheric and Oceanic Technology*, 25, 519–533, <https://doi.org/10.1175/2007JTECHA1006.1>, 2008.
- Marenco, F., Santacesaria, V., Bais, A. F., Balis, D., di Sarra, A., Papayannis, A., and Zerefos, C.: Optical properties of tropospheric aerosols determined by lidar and spectrophotometric measurement (Photochemical Activity and Solar Ultraviolet Radiation campaign), *Applied Optics*, 36, 6875–6886, <https://doi.org/10.1364/AO.36.006875>, 1997.
- Martin, G., Bush, M., Brown, A., Lock, A., and Smith, R.: A new boundary layer mixing scheme. Part II: Tests in climate and mesoscale models, *Monthly weather review*, 128, 3200–3217, [https://doi.org/10.1175/1520-0493\(2000\)128<3200:ANBLMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3200:ANBLMS>2.0.CO;2), 2000.
- Martucci, G., Milroy, C., and O'Dowd, C. D.: Detection of cloud-base height using Jenoptik CHM15K and Vaisala CL31 ceilometers, *Journal of Atmospheric and Oceanic Technology*, 27, 305–318, <https://doi.org/10.1175/2009JTECHA1326.1>, 2010.
- Maslanik, J. and Stroeve, J.: Near-Real-Time DMSP SSMIS Daily Polar Gridded Sea Ice Concentrations, Version 1, <https://doi.org/10.5067/U8C09DWVX9LM>, 1999.
- Mason, S., Jakob, C., Protat, A., and Delanoë, J.: Characterizing observed midtopped cloud regimes associated with Southern Ocean short-wave radiation biases, *Journal of Climate*, 27, 6189–6203, <https://doi.org/10.1175/JCLI-D-14-00139.1>, 2014.
- Mason, S., Fletcher, J. K., Haynes, J. M., Franklin, C., Protat, A., and Jakob, C.: A hybrid cloud regime methodology used to evaluate Southern Ocean cloud and shortwave radiation errors in ACCESS, *Journal of Climate*, 28, 6001–6018, <https://doi.org/10.1175/jcli-d-14-00846.1>, 2015.
- Masunaga, H., Matsui, T., Tao, W.-k., Hou, A. Y., Kummerow, C. D., Nakajima, T., Bauer, P., Olson, W. S., Sekiguchi, M., and Nakajima, T. Y.: Satellite data simulator unit: A multisensor, multispectral satellite simulator package, *Bulletin of the American Meteorological Society*, 91, 1625–1632, <https://doi.org/10.1175/2010BAMS2809.1>, 2010.
- Mattis, I., Begbie, R., Boyouk, N., Bravo-Aranda, J. A., Brettle, M., Cermak, J., Drouin, M.-A., Geiß, A., Görsdorf, U., Haefele, A., Haeffelin, M., Hervo, M., Komínková, K., Leinweber, R., Müller, G., Münkel, C., Pattantyús-Ábrahám, M., Pönitz, K., Wagner, F., and Wiegner, M.: The ceilometer inter-comparison campaign CeilEx2015, in: EGU General Assembly Conference Abstracts, pp. EPSC2016–9687, 2016.
- McDonald, A. and Parsons, S.: A Comparison of Cloud Classification Methodologies: Differences Between Cloud and Dynamical Regimes, *Journal of Geophysical Research: Atmospheres*, 123, 11–173, <https://doi.org/10.1029/2018jd028595>, 2018.
- McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., and Schuddeboom, A.: An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states, *Journal of Geophysical Research: Atmospheres*, 121, <https://doi.org/10.1002/2016jd025199>, 2016.
- McFarquhar, G.: Measurements of Aerosols, Radiation, and Clouds over the Southern Ocean (MARCUS) Science Plan, 2016.
- McFarquhar, G., Wood, R., Bretherton, C., Alexander, S., Jakob, C., Marchand, R., Protat, A., Quinn, P., Siems, S., and Weller, R.: The Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES): An Observational Campaign for Determining Role of Clouds, Aerosols and Radiation in Climate System, in: AGU Fall Meeting Abstracts, 2014.
- McGill, M. J., Yorks, J. E., Scott, V. S., Kupchock, A. W., and Selmer, P. A.: The Cloud-Aerosol Transport System (CATS): A technology demonstration on the International Space Station, in: Lidar Remote Sensing for Environmental Monitoring XV, vol. 9612, p. 96120A, International Society for Optics and Photonics, 2015.
- Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S., and Stevens, B.: Climate model intercomparisons: preparing for the next phase, *Eos, Transactions American Geophysical Union*, 95, 77–78, <https://doi.org/10.1002/2014EO090001>, 2014.
- Meier, W. N., Maslanik, J. A., Fowler, C. W., and Key, J. R.: Multiparameter AVHRR-derived products for Arctic climate studies, *Earth interactions*, 1, 1–29, [https://doi.org/10.1175/1087-3562\(1997\)001<0001:MADPFA>2.3.CO;2](https://doi.org/10.1175/1087-3562(1997)001<0001:MADPFA>2.3.CO;2), 1997.
- Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Annalen der physik*, 330, 377–445, <https://doi.org/10.1002/andp.19083300302>, 1908.
- Milroy, C., Martucci, G., Lolli, S., Loaec, S., Sauvage, L., Xueref-Remy, I., Lavrič, J. V., Ciais, P., Feist, D. G., Biavati, G., et al.: An assessment of pseudo-operational ground-based light detection and ranging sensors to determine the boundary-layer structure in the coastal atmosphere, *Advances in Meteorology*, 2012, <https://doi.org/10.1155/2012/929080>, 2012.
- Morcrette, C. J. and Petch, J. C.: Analysis of prognostic cloud scheme increments in a climate model, *Quarterly Journal of the Royal Meteorological Society*, 136, 2061–2073, <https://doi.org/10.1002/qj.720>, 2010.
- Morille, Y., Haeffelin, M., Drobinski, P., and Pelon, J.: STRAT: An automated algorithm to retrieve the vertical structure of the atmosphere from single-channel lidar data, *Journal of Atmospheric and Oceanic Technology*, 24, 761–775, <https://doi.org/10.1175/JTECH2008.1>, 2007.
- Morrison, A. E., Siems, S. T., and Manton, M. J.: A three-year climatology of cloud-top phase over the Southern Ocean and North Pacific, *Journal of Climate*, 24, 2405–2418, <https://doi.org/10.1175/2010JCLI3842.1>, 2011.
- Muhlbauer, A., Ackerman, T., Lawson, R., Xie, S., and Zhang, Y.: Evaluation of cloud-resolving model simulations of midlatitude cirrus with

- ARM and A-train observations, *Journal of Geophysical Research: Atmospheres*, 120, 6597–6618, <https://doi.org/10.1002/2014JD022570>, 2015.
- Mülmenstädt, J., Sourdeval, O., Henderson, D. S., L'Ecuyer, T. S., Unglaub, C., Jungandreas, L., Böhm, C., Russell, L. M., and Quaas, J.: Using CALIOP to estimate cloud-field base height and its uncertainty: the Cloud Base Altitude Spatial Extrapolator (CBASE) algorithm and dataset, *Earth System Science Data*, 10, 2279–2293, <https://doi.org/10.5194/essd-10-2279-2018>, URL <https://www.earth-syst-sci-data.net/10/2279/2018/>, 2018.
- Münkel, C., Eresmaa, N., Räsänen, J., and Karppinen, A.: Retrieval of mixing height and dust concentration with lidar ceilometer, *Boundary-layer meteorology*, 124, 117–128, <https://doi.org/10.1007/s10546-006-9103-3>, 2007.
- Nam, C., Bony, S., Dufresne, J.-L., and Chepfer, H.: The 'too few, too bright' tropical low-cloud problem in CMIP5 models, *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2012gl053421>, 2012.
- NASA JPL: NASA Shuttle Radar Topography Mission Global 3 arc second [Data set], NASA EOSDIS Land Processes DAAC, <https://doi.org/https://doi.org/10.5067/MEaSURES/SRTM/SRTMGL3.003>, 2013.
- Naud, C. M., Booth, J. F., and Del Genio, A. D.: Evaluation of ERA-Interim and MERRA cloudiness in the Southern Ocean, *Journal of Climate*, 27, 2109–2124, <https://doi.org/10.1175/JCLI-D-13-00432.1>, 2014.
- Noh, Y.-J., Forsythe, J. M., Miller, S. D., Seaman, C. J., Li, Y., Heidinger, A. K., Lindsey, D. T., Rogers, M. A., and Partain, P. T.: Cloud-base height estimation from VIIRS. Part II: A statistical algorithm based on A-Train satellite data, *Journal of Atmospheric and Oceanic Technology*, 34, 585–598, <https://doi.org/10.1175/JTECH-D-16-0110.1>, 2017.
- Noh, Y.-J., Miller, S. D., Heidinger, A. K., Mace, G. G., Protat, A., and Alexander, S. P.: Satellite-based detection of daytime supercooled liquid-topped mixed-phase clouds over the Southern Ocean using the Advanced Himawari Imager, *Journal of Geophysical Research: Atmospheres*, 124, 2677–2701, <https://doi.org/10.1029/2018JD029524>, 2019.
- O'Connor, E. J., Illingworth, A. J., and Hogan, R. J.: A technique for autocalibration of cloud lidar, *Journal of Atmospheric and Oceanic Technology*, 21, 777–786, [https://doi.org/10.1175/1520-0426\(2004\)021<0777:ATFAOC>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0777:ATFAOC>2.0.CO;2), 2004.
- Oliphant, T. E.: *A guide to NumPy*, vol. 1, Trelgol Publishing USA, 2006.
- Pal, S. R., Steinbrecht, W., and Carswell, A. I.: Automated method for lidar determination of cloud-base height and vertical extent, *Applied optics*, 31, 1488–1494, <https://doi.org/10.1364/AO.31.001488>, 1992.
- Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, *Atmospheric Measurement Techniques*, 7, 2389–2409, <https://doi.org/10.5194/amt-7-2389-2014>, 2014.
- Parkinson, C. L.: Aqua: An Earth-observing satellite mission to examine water and other climate variables, *IEEE Transactions on Geoscience and Remote Sensing*, 41, 173–183, <https://doi.org/10.1109/tgrs.2002.808319>, 2003.
- Petty, G. W.: *A First Course in Atmospheric Radiation*, Sundog Publishing, 2 edn., 2006.
- Petty, G. W. and Huang, W.: The modified gamma size distribution applied to inhomogeneous and nonspherical particles: Key relationships and conversions, *Journal of the Atmospheric Sciences*, 68, 1460–1473, <https://doi.org/10.1175/2011JAS3645.1>, 2011.
- Powers, J. G., Monaghan, A. J., Cayette, A. M., Bromwich, D. H., Kuo, Y.-H., and Manning, K. W.: Real-Time Mesoscale Modeling Over Antarctica: The Antarctic Mesoscale Prediction System, *Bulletin of the American Meteorological Society*, 84, 1533–1546, <https://doi.org/10.1175/BAMS-84-11-1533>, 2003.
- Powers, J. G., Manning, K. W., Bromwich, D. H., Cassano, J. J., and Cayette, A. M.: A decade of Antarctic science support through AMPS, *Bulletin of the American Meteorological Society*, 93, 1699–1712, <https://doi.org/10.1175/bams-d-11-00186.1>, 2012.
- Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., and et al.: The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package, *The Astronomical Journal*, 156, 123, <https://doi.org/10.3847/1538-3881/aabc4f>, 2018.
- Protat, A., Schulz, E., Rikus, L., Sun, Z., Xiao, Y., and Keywood, M.: Shipborne observations of the radiative effect of Southern Ocean clouds, *Journal of Geophysical Research: Atmospheres*, 122, 318–328, <https://doi.org/10.1002/2016jd026061>, 2017.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>, 2017.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>, 2018.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research: Atmospheres*, 108, 4407, <https://doi.org/10.1029/2002JD002670>, 2003.
- Revell, L. E., Kremser, S., Hartery, S., Harvey, M., Mulcahy, J. P., Williams, J., Morgenstern, O., McDonald, A. J., Varma, V., Bird, L., et al.: The sensitivity of Southern Ocean aerosols and cloud microphysics to sea spray and sulfate aerosol production in the HadGEM3-GA7.1 chemistry–climate model, *Atmospheric Chemistry and Physics*, 19, 15 447–15 466, <https://doi.org/10.5194/acp-19-15447-2019>, 2019.
- Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, *IEEE computer graphics and applications*, 10, 76–82, <https://doi.org/10.1109/38.56302>, 1990.

- Rew, R., Hartnett, E., Caron, J., et al.: NetCDF-4: Software implementing an enhanced data model for the geosciences, in: 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology, 2006.
- Ritter, B. and Geleyn, J.-F.: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, *Monthly weather review*, 120, 303–325, [https://doi.org/10.1175/1520-0493\(1992\)120<0303:acrsmf>2.0.co;2](https://doi.org/10.1175/1520-0493(1992)120<0303:acrsmf>2.0.co;2), 1992.
- Roemmich, D. and Team, A. S.: Argo: the challenge of continuing 10 years of progress, *Oceanography*, 22, 46–55, 2009.
- Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., Garzoli, S. L., Schmid, C., and Ignaszewski, M.: The Argo Program: Observing the global ocean with profiling floats, *Oceanography*, 22, 34–43, <https://doi.org/10.5670/oceanog.2009.36>, 2009.
- Rosoldi, M., Madonna, F., Pappalardo, G., Hey, J. V., and Zheng, Y.: The lesson learnt during interact-I and INTERACT-II actris measurement campaigns, in: EPJ Web of Conferences, vol. 176, p. 11002, EDP Sciences, 2018.
- Rossov, W. B. and Schiffer, R. A.: ISCCP cloud data products, *Bulletin of the American Meteorological Society*, [https://doi.org/10.1175/1520-0477\(1991\)072<0002:ICDP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2), 1991.
- Rossov, W. B. and Schiffer, R. A.: Advances in understanding clouds from ISCCP, *Bulletin of the American Meteorological Society*, 80, 2261–2288, [https://doi.org/10.1175/1520-0477\(1999\)080<2261:AIUCFI>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2), 1999.
- Rossum, G.: Python reference manual, 1995.
- Salomonson, V. V., Barnes, W., Xiong, J., Kempler, S., and Masuoka, E.: An overview of the Earth Observing System MODIS instrument and associated data systems performance, in: Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International, vol. 2, pp. 1174–1176, IEEE, <https://doi.org/10.1109/IGARSS.2002.1025812>, 2002.
- Sandford, M., Allan, P., Caldwell, M., Delderfield, J., Oliver, M., Sawyer, E., Harries, J., Ashmall, J., Brindley, H., Kellock, S., et al.: The geostationary Earth radiation budget (GERB) instrument on EUMETSAT's MSG satellite, *Acta Astronautica*, 53, 909–915, [https://doi.org/10.1016/S0094-5765\(02\)00209-6](https://doi.org/10.1016/S0094-5765(02)00209-6), 2003.
- Sato, K., Inoue, J., Alexander, S. P., McFarquhar, G., and Yamazaki, A.: Improved Reanalysis and Prediction of Atmospheric Fields Over the Southern Ocean Using Campaign-Based Radiosonde Observations, *Geophysical Research Letters*, <https://doi.org/10.1029/2018GL079037>, 2018.
- Satoh, M., Stevens, B., Judd, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global cloud-resolving models, *Current Climate Change Reports*, 5, 172–184, <https://doi.org/10.1007/s40641-019-00131-0>, 2019.
- Schiffer, R. and Rossow, W. B.: The International Satellite Cloud Climatology Project (ISCCP)- The first project of the World Climate Research Programme, *American Meteorological Society, Bulletin*, 64, 779–784, <https://doi.org/10.1175/1520-0477-64.7.779>, 1983.
- Schlosser, E., Haumann, F. A., and Raphael, M. N.: Atmospheric influences on the anomalous 2016 Antarctic sea ice decay, *The Cryosphere*, 12, 1103–1119, <https://doi.org/10.5194/tc-12-1103-2018>, 2018.
- Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., and Saha, S.: Practice and philosophy of climate model tuning across six US modeling centers, *Geoscientific Model Development*, 10, 3207, <https://doi.org/10.5194/gmd-2017-30-ac1>, 2017.
- Schuddeboom, A., McDonald, A., Parsons, S., Morgenstern, O., and Harvey, M.: Examining the NZESM Cloud representation with Self Organizing Maps, in: EGU General Assembly Conference Abstracts, vol. 19, p. 11287, 2017.
- Schuddeboom, A., McDonald, A. J., Morgenstern, O., Harvey, M., and Parsons, S.: Regional Regime-Based Evaluation of Present-Day General Circulation Model Cloud Simulations Using Self-Organizing Maps, *Journal of Geophysical Research: Atmospheres*, 123, 4259–4272, <https://doi.org/10.1002/2017JD028196>, 2018.
- Schuddeboom, A., Varma, V., McDonald, A. J., Morgenstern, O., Harvey, M., Parsons, S., Field, P., and Furtado, K.: Cluster-Based Evaluation of Model Compensating Errors: A Case Study of Cloud Radiative Effect in the Southern Ocean, *Geophysical Research Letters*, 46, 3446–3453, <https://doi.org/10.1029/2018GL081686>, 2019.
- Schulzweida, U.: CDO User Guide Version 1.9.5, 2018.
- Silber, I., Verlinde, J., Eloranta, E. W., Flynn, C. J., and Flynn, D. M.: Polar liquid cloud base detection algorithms for high spectral resolution or micropulse lidar data, *Journal of Geophysical Research: Atmospheres*, 123, 4310–4322, <https://doi.org/10.1029/2017JD027840>, 2018.
- Simmonds, I.: Modes of atmospheric variability over the Southern Ocean, *Journal of Geophysical Research: Oceans*, 108, <https://doi.org/10.1029/2000jc000542>, 2003.
- Simmonds, I. and Keay, K.: Mean Southern Hemisphere extratropical cyclone behavior in the 40-year NCEP–NCAR reanalysis, *Journal of Climate*, 13, 873–885, [https://doi.org/10.1175/1520-0442\(2000\)013<0873:mshceb>2.0.co;2](https://doi.org/10.1175/1520-0442(2000)013<0873:mshceb>2.0.co;2), 2000.
- Simmonds, I., Keay, K., and Lim, E.-P.: Synoptic activity in the seas around Antarctica, *Monthly Weather Review*, 131, 272–288, [https://doi.org/10.1175/1520-0493\(2003\)131<0272:saita>2.0.co;2](https://doi.org/10.1175/1520-0493(2003)131<0272:saita>2.0.co;2), 2003.
- Simpson, J., Kummerow, C., Tao, W.-K., and Adler, R. F.: On the tropical rainfall measuring mission (TRMM), *Meteorology and Atmospheric Physics*, 60, 19–36, <https://doi.org/10.1007/BF01029783>, 1996.
- Sinclair, M. R.: An objective cyclone climatology for the Southern Hemisphere, *Monthly Weather Review*, 122, 2239–2256, [https://doi.org/10.1175/1520-0493\(1994\)122<2239:aocft>2.0.co;2](https://doi.org/10.1175/1520-0493(1994)122<2239:aocft>2.0.co;2), 1994.
- Sinclair, M. R.: A climatology of cyclogenesis for the Southern Hemisphere, *Monthly Weather Review*, 123, 1601–1619, [https://doi.org/10.1175/1520-0493\(1995\)123<1601:acocft>2.0.co;2](https://doi.org/10.1175/1520-0493(1995)123<1601:acocft>2.0.co;2), 1995.

- Smith, G., Priestley, K., Loeb, N., Wielicki, B., Charlock, T., Minnis, P., Doelling, D., and Rutan, D.: Clouds and Earth Radiant Energy System (CERES), a review: Past, present and future, *Advances in Space Research*, 48, 254–263, <https://doi.org/10.1016/j.asr.2011.03.009>, 2011.
- Smith, R.: A scheme for predicting layer clouds and their water content in a general circulation model, *Quarterly Journal of the Royal Meteorological Society*, 116, 435–460, <https://doi.org/10.1002/qj.49711649210>, 1990.
- Smith, S. A. and Jonas, P. R.: Observations of the turbulent fluxes in fields of cumulus clouds, *Quarterly Journal of the Royal Meteorological Society*, 121, 1185–1208, <https://doi.org/10.1002/qj.49712152602>, 1995.
- Smith, W., Hickey, J., Howell, H. B., Jacobowitz, H., Hilleary, D., and Drummond, A.: Nimbus-6 earth radiation budget experiment, *Applied optics*, 16, 306–318, <https://doi.org/10.1364/ao.16.000306>, 1977.
- Spinhirne, J. D.: Micro pulse lidar, *IEEE Transactions on Geoscience and Remote Sensing*, 31, 48–55, <https://doi.org/10.1109/36.210443>, 1993.
- Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., et al.: The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation, *Bulletin of the American Meteorological Society*, 83, 1771–1790, <https://doi.org/10.1175/BAMS-83-12-1771>, 2002.
- Stokes, G. M. and Schwartz, S. E.: The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the cloud and radiation test bed, *Bulletin of the American Meteorological Society*, 75, 1201–1222, [https://doi.org/10.1175/1520-0477\(1994\)075<1201:TARMPP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2), 1994.
- Stuecker, M. F., Bitz, C. M., and Armour, K. C.: Conditions leading to the unprecedented low Antarctic sea ice extent during the 2016 austral spring season, *Geophysical Research Letters*, 44, 9008–9019, <https://doi.org/10.1002/2017gl074691>, 2017.
- Sundqvist, H.: A parameterization scheme for non-convective condensation including prediction of cloud water content, *Quarterly Journal of the Royal Meteorological Society*, 104, 677–690, <https://doi.org/10.1002/qj.49710444110>, 1978.
- Sundqvist, H., Berge, E., and Kristjánsson, J. E.: Condensation and Cloud Parameterization Studies with a Mesoscale Numerical Weather Prediction Model, *Monthly Weather Review*, 117, 1641–1657, [https://doi.org/10.1175/1520-0493\(1989\)117<1641:CACPSW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2), 1989.
- Swales, D. J., Pincus, R., and Bodas-Salcedo, A.: The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2, *Geoscientific Model Development*, 11, 77–81, <https://doi.org/10.5194/gmd-2017-148>, 2018.
- Tange, O. et al.: Gnu parallel—the command-line power tool, *The USENIX Magazine*, 36, 42–47, 2011.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>, 2012.
- Taylor, P. C., Kato, S., Xu, K.-M., and Cai, M.: Covariance between Arctic sea ice and clouds within atmospheric state regimes at the satellite footprint level, *Journal of Geophysical Research: Atmospheres*, 120, 12 656–12 678, <https://doi.org/10.1002/2015JD023520>, 2015.
- Telford, P., Braesicke, P., Morgenstern, O., and Pyle, J.: Description and assessment of a nudged version of the new dynamics Unified Model, *Atmospheric Chemistry and Physics*, 8, 1701–1712, <https://doi.org/10.5194/acp-8-1701-2008>, 2008.
- Tiedtke, M.: Representation of Clouds in Large-Scale Models, *Monthly Weather Review*, 121, 3040–3061, [https://doi.org/10.1175/1520-0493\(1993\)121<3040:ROCILS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2), 1993.
- Torvalds, L.: Linux: a portable operating system, Master's thesis, University of Helsinki, 1997.
- Trenberth, K. E. and Fasullo, J. T.: Simulation of present-day and twenty-first-century energy budgets of the southern oceans, *Journal of Climate*, 23, 440–454, <https://doi.org/10.1175/2009JCLI13152.1>, 2010.
- Tsaknakis, G., Papayannis, A., Kokkalis, P., Amiridis, V., Kambezidis, H., Mamouri, R., Georgoussis, G., and Avdikos, G.: Inter-comparison of lidar and ceilometer retrievals for aerosol and Planetary Boundary Layer profiling over Athens, Greece, *Atmospheric Measurement Techniques*, 4, 1261–1273, <https://doi.org/10.5194/amt-4-1261-2011>, 2011.
- Turner, J., Phillips, T., Marshall, G. J., Hosking, J. S., Pope, J. O., Bracegirdle, T. J., and Deb, P.: Unprecedented springtime retreat of Antarctic sea ice in 2016, *Geophysical Research Letters*, 44, 6868–6875, <https://doi.org/10.1002/2017gl073656>, 2017.
- Van Der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy array: a structure for efficient numerical computation, *Computing in Science & Engineering*, 13, 22–30, <https://doi.org/10.1109/MCSE.2011.37>, 2011.
- Van Tricht, K., Gorodetskaya, I., Lhermitte, S., Turner, D., Schween, J., and Van Lipzig, N.: An improved algorithm for polar cloud-base detection by ceilometer over the ice sheets, *Atmospheric Measurement Techniques*, 7, 1153–1167, <https://doi.org/10.5194/amt-7-1153-2014>, 2014.
- Vergara-Temprado, J., Miltenberger, A. K., Furtado, K., Grosvenor, D. P., Shipway, B. J., Hill, A. A., Wilkinson, J. M., Field, P. R., Murray, B. J., and Carslaw, K. S.: Strong control of Southern Ocean cloud reflectivity by ice-nucleating particles, *Proceedings of the National Academy of Sciences*, 115, 2687–2692, <https://doi.org/10.1073/pnas.1721627115>, 2018.
- Vignesh, P. P., Jiang, J. H., Kishore, P., Su, H., Smay, T., Brighton, N., and Velicogna, I.: Assessment of CMIP6 Cloud Fraction and Comparison with Satellite Observations, *Earth and Space Science*, 7, e2019EA000 975, <https://doi.org/10.1029/2019EA000975>, 2020.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors: SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python, arXiv e-prints, arXiv:1907.10121, 2019.

- Wall, C. J., Hartmann, D. L., and Ma, P.-L.: Instantaneous linkages between clouds and large-scale meteorology over the Southern Ocean in observations and a climate model, *Journal of Climate*, 30, 9455–9474, <https://doi.org/10.1175/JCLI-D-17-0156.1>, 2017.
- Wallace, J. M. and Hobbs, P. V.: *Atmospheric science: an introductory survey*, vol. 92, Elsevier, 2006.
- Walters, D., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., et al.: The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations, *Geoscientific Model Development*, 10, 1487–1520, <https://doi.org/10.5194/gmd-12-1909-2019>, 2017.
- Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, *Geoscientific Model Development*, 12, 1909–1963, <https://doi.org/10.5194/gmd-12-1909-2019>, 2019.
- Wang, Z. and Sassen, K.: Cloud type and macrophysical property retrieval using multiple remote sensors, *Journal of Applied Meteorology*, 40, 1665–1682, [https://doi.org/10.1175/1520-0450\(2001\)040<1665:CTAMPR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2), 2001.
- Watson-Parris, D., Schutgens, N., Cook, N., Kipling, Z., Kershaw, P., Gryspeerd, E., Lawrence, B., and Stier, P.: Community Intercomparison Suite (CIS) v1. 4.0: a tool for intercomparing models and observations, *Geoscientific Model Development*, 9, 3093–3110, <https://doi.org/10.5194/gmd-9-3093-2016>, 2016.
- Webb, M., Senior, C., Bony, S., and Morcrette, J.-J.: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, *Climate Dynamics*, 17, 905–922, <https://doi.org/10.1007/s003820100157>, 2001.
- Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., et al.: The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6, *Geoscientific Model Development*, 10, 359–384, <https://doi.org/10.5194/gmd-10-359-2017>, 2017.
- Welton, E. J., Voss, K. J., Gordon, H. R., Maring, H., Smirnov, A., Holben, B., Schmid, B., Livingston, J. M., Russell, P. B., Durkee, P. A., et al.: Ground-based lidar measurements of aerosols during ACE-2: Instrument description, results, and comparisons with other ground-based and airborne measurements, *Tellus B*, 52, 636–651, <https://doi.org/10.1034/j.1600-0889.2000.00025.x>, 2000.
- Welton, E. J., Voss, K. J., Quinn, P. K., Flatau, P. J., Markowicz, K., Campbell, J. R., Spinhirne, J. D., Gordon, H. R., and Johnson, J. E.: Measurements of aerosol vertical profiles and optical properties during INDOEX 1999 using micropulse lidars, *Journal of Geophysical Research: Atmospheres*, 107, INX2–18, <https://doi.org/10.1029/2000JD000038>, 2002.
- Welton, E. J., Campbell, J. R., Berkoff, T. A., Valencia, S., Spinhirne, J. D., Holben, B., Tsay, S.-C., and Schmid, B.: The NASA Micro-Pulse Lidar Network (MPLNET): an overview and recent results, *Opt. Pur. Apl*, 39, 67–74, 2006.
- Werner, M.: Shuttle radar topography mission (SRTM) mission overview, *Frequenz*, 55, 75–79, <https://doi.org/10.1515/FREQ.2001.55.3-4.75>, 2001.
- Wiegner, M. and Gasteiger, J.: Correction of water vapor absorption for aerosol remote sensing with ceilometers, *Atmospheric Measurement Techniques*, pp. 3971–3984, <https://doi.org/10.5194/amt-8-3971-2015>, 2015.
- Wiegner, M. and Geiß, A.: Aerosol profiling with the Jenoptik ceilometer CHM15kx, *Atmos. Meas. Tech*, 5, 1953–1964, <https://doi.org/10.5194/amt-5-1953-2012>, 2012.
- Wiegner, M., Madonna, F., Biniotoglou, I., Forkel, R., Gasteiger, J., Geiß, A., Pappalardo, G., Schäfer, K., and Thomas, W.: What is the benefit of ceilometers for aerosol remote sensing? An answer from EARLINET, *Atmospheric Measurement Techniques*, 7, 1979–1997, <https://doi.org/10.5194/amt-7-1979-2014>, 2014.
- Wiegner, M., Mattis, I., Pattantyús-Ábrahám, M., Bravo-Aranda, J. A., Poltera, Y., Haefele, A., Hervo, M., Görsdorf, U., Leinweber, R., Gasteiger, J., Haefelin, M., Wagner, F., Cermak, J., Komínková, K., Brettle, M., Münkel, C., and Pönitz, K.: Aerosol backscatter profiles from ceilometers: validation of water vapor correction in the framework of CeilLinEx2015, *Atmospheric Measurement Techniques*, 12, 471–490, <https://doi.org/10.5194/amt-12-471-2019>, 2019.
- Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee III, R. B., Louis Smith, G., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment, *Bulletin of the American Meteorological Society*, 77, 853–868, [https://doi.org/10.1175/1520-0477\(1996\)077<0853:CATERE>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2), 1996.
- Williams, D. N., Ananthakrishnan, R., Bernholdt, D., Bharathi, S., Brown, D., Chen, M., Chervenak, A., Cinquini, L., Drach, R., Foster, I., et al.: The Earth System Grid: Enabling access to multimodel climate simulation data, *Bulletin of the American Meteorological Society*, 90, 195–206, <https://doi.org/10.1175/2008BAMS2459.1>, 2009.
- Williams, J., Morgenstern, O., Varma, V., Behrens, E., Hayek, W., Oliver, H., Dean, S., Mullan, B., and Frame, D.: Development of the New Zealand Earth System Model: NZESM, *Weather and Climate*, 36, 25–44, <https://doi.org/10.2307/26779386>, 2016.
- Williams, K. and Webb, M.: A quantitative performance assessment of cloud regimes in climate models, *Climate dynamics*, 33, 141–157, <https://doi.org/10.1007/s00382-008-0443-1>, 2009.
- Williams, K., Ringer, M., Senior, C., Webb, M., McAvaney, B., Andronova, N., Bony, S., Dufresne, J.-L., Emori, S., Gudgel, R., et al.: Evaluation of a component of the cloud response to climate change in an intercomparison of climate models, *Climate Dynamics*, 26, 145–165, <https://doi.org/10.1007/s00382-005-0067-7>, 2006.
- Williams, K., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S., Senior, C., and Williamson, D.: The

- Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models, *Journal of Climate*, 26, 3258–3274, <https://doi.org/10.1175/JCLI-D-12-00429.1>, 2013.
- Williams, K. D. and Bodas-Salcedo, A.: A multi-diagnostic approach to cloud evaluation, *J. Geophys. Res.*, submitted, 10, <https://doi.org/10.5194/gmd-10-2547-2017>, 2017.
- Wilson, D., Bushell, A., and Morcrette, C.: Unified Model Documentation Paper 030: The PC2 Cloud Scheme, UK Met Office, 2015.
- Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., and Morcrette, C. J.: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description, *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134, 2093–2107, <https://doi.org/10.1002/qj.333>, 2008a.
- Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., Morcrette, C. J., and Bodas-Salcedo, A.: PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations, *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134, 2109–2125, <https://doi.org/10.1002/qj.332>, 2008b.
- Winker, D., Pelon, J., Coakley Jr, J., Ackerman, S., Charlson, R., Colarco, P., Flamant, P., Fu, Q., Hoff, R., Kittaka, C., et al.: The CALIPSO mission: A global 3D view of aerosols and clouds, *Bulletin of the American Meteorological Society*, 91, 1211–1230, <https://doi.org/10.1175/2010BAMS3009.1>, 2010.
- Winker, D. M., Pelon, J. R., and McCormick, M. P.: The CALIPSO mission: Spaceborne lidar foStephens200r observation of aerosols and clouds, in: *Third International Asia-Pacific Environmental Remote Sensing Remote Sensing of the Atmosphere, Ocean, Environment, and Space*, pp. 1–11, International Society for Optics and Photonics, <https://doi.org/10.1117/12.466539>, 2003.
- Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, *Journal of Atmospheric and Oceanic Technology*, 26, 2310–2323, <https://doi.org/10.1175/2009JTECHA1281.1>, 2009.
- Wiscombe, W. J.: Mie scattering calculations: Advances in technique and fast, vector-speed computer codes, Tech. rep., National Center for Atmospheric Research Boulder, Colorado, 1979.
- Wiscombe, W. J.: Improved Mie scattering algorithms, *Applied optics*, 19, 1505–1509, <https://doi.org/10.1364/AO.19.001505>, 1980.
- WMO: WMO Statement on the State of the Global Climate in 2019, 2020.
- Zadra, A., Williams, K., Frassoni, A., Rixen, M., Adames, Á. F., Berner, J., Bouysse, F., Casati, B., Christensen, H., Ek, M. B., et al.: Systematic Errors in Weather and Climate Models: Nature, Origins, and Ways Forward, *Bulletin of the American Meteorological Society*, 99, ES67–ES70, <https://doi.org/10.1175/BAMS-D-17-0287.1>, 2018.
- Zdunkowski, W., Trautmann, T., and Bott, A.: *Radiation in the atmosphere: a course in theoretical meteorology*, Cambridge University Press, 2007.
- Zelinka, M. D., Randall, D. A., Webb, M. J., and Klein, S. A.: Clearing clouds of uncertainty, *Nature Climate Change*, 7, 674–678, <https://doi.org/10.1038/nclimate3402>, 2017.
- Zhang, J., Xia, X., and Chen, H.: A comparison of cloud layers from ground and satellite active remote sensing at the Southern Great Plains ARM site, *Advances in Atmospheric Sciences*, 34, 347–359, <https://doi.org/10.1007/s00376-016-6030-1>, 2017.
- Zhang, Y., Xie, S., Klein, S. A., Marchand, R., Kollias, P., Clothiaux, E. E., Lin, W., Johnson, K., Swales, D., Bodas-Salcedo, A., et al.: The ARM Cloud Radar Simulator for Global Climate Models: Bridging Field Data and Climate Models, *Bulletin of the American Meteorological Society*, 99, 21–26, <https://doi.org/10.1175/BAMS-D-16-0258.1>, 2018.
- Zheng, Y. and Li, Z.: Episodes of Warm-Air Advection Causing Cloud-Surface Decoupling During the MARCUS, *Journal of Geophysical Research: Atmospheres*, <https://doi.org/10.1029/2019JD030835>, 2019.