Evaluation of CERES and CloudSat Surface Radiative Fluxes over the Southern Ocean

Roger T Marchand[1] and Laura M. Hinkelman[1]

[1]University of Washington

November 24, 2022

Abstract

Many studies involving surface radiative fluxes rely on surface fluxes retrieved by the Clouds and the Earth's Radiant Energy System (CERES) project, or derived from spaceborne cloud radar and lidar observations (CloudSat-CALIPSO). In particular, most climate models that participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) were found to have too little shortwave radiation being reflected back to space and excessive shortwave radiation reaching the surface over the Southern Ocean – an error with significant consequences for predicting both regional and global climate. There have been few evaluations of CERES or CloudSat retrievals over the Southern Ocean. In this article, CERES and CloudSat retrieved surface shortwave (SW) and longwave (LW) downwelling fluxes are evaluated using surface observations collected over the Southern Ocean during the Macquarie Island Cloud and Radiation Experiment (MICRE). Overall, biases (CERES – surface observations) in the CERES- surface fluxes are found to be slightly larger over Macquarie Island than most other regions, approximately +10 Wm for the SW and -10 Wm for the LW in the annual mean, but with significant seasonal and diurnal variations. If the Macquarie observations are representative of the larger SO, these results imply that CMIP5 model errors in SW surface fluxes are (if anything) somewhat larger than previous evaluation studies suggest. The bias in LW surface flux shows a marked increase at night, which explains most of the total LW bias. The nighttime bias is due to poor representation of cloud base associated with low clouds.

Evaluation of CERES and CloudSat Surface Radiative Fluxes over the Southern Ocean

Laura M. Hinkelman[1] and Roger Marchand[2]

[1]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, 98195.

[2]Department of Atmospheric Sciences, University of Washington, Seattle, WA, 98195.

Corresponding author: Dr. Roger Marchand (rojmarch@u.washington.edu)

Key Points:

- CERES and CloudSat-CALIPSO surface shortwave (SW) and longwave (LW) fluxes are compared with surface measurements over the Southern Ocean.
- Mean CERES surf fluxes are larger in SW (+10 Wm-2) & smaller in LW (-10 Wm-2) than observed with significant seasonal and diurnal variations.
- LW surface fluxes are larger at night (-16 Wm^{-2}), which explains most of the total bias, and is due to incorrect cloud base for low clouds.

Abstract –

Many studies involving surface radiative fluxes rely on surface fluxes retrieved by the Clouds and the Earth's Radiant Energy System (CERES) project, or derived from spaceborne cloud radar and lidar observations (CloudSat-CALIPSO). In particular, most climate models that participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) were found to have too little shortwave radiation being reflected back to space and excessive shortwave radiation reaching the surface over the Southern Ocean – an error with significant consequences for predicting both regional and global climate. There have been few evaluations of CERES or CloudSat retrievals over the Southern Ocean. In this article, CERES and CloudSat retrieved surface shortwave (SW) and longwave (LW) downwelling fluxes are evaluated using surface observations collected over the Southern Ocean during the Macquarie Island Cloud and Radiation Experiment (MICRE). Overall, biases (CERES – surface observations) in the CERES-surface fluxes are found to be slightly larger over Macquarie Island than most other regions, approximately +10 Wm^{-2} for the SW and -10 Wm^{-2} for the LW in the annual mean, but with significant seasonal and diurnal variations. If the Macquarie observations are representative of the larger SO, these results imply that CMIP5 model errors in SW surface fluxes are (if anything) somewhat larger than previous evaluation studies suggest. The bias in LW surface flux shows a marked increase at night, which explains most of the total LW bias. The nighttime bias is due to poor representation of cloud base associated with low clouds.

Plain Language Summary –

We compare satellite estimates for the amount of sunlight (solar) and thermal (infrared) energy reaching the surface, with surface observations collected at Macquarie Island. Macquarie Island is located in the Southern Ocean (SO) about halfway between New Zealand and Antarctica. The satellite-based estimates have seen little evaluation over the Southern Ocean. This is a concern because climate models, when compared with the satellite estimates, are not reflecting enough sunlight to space over the SO, which has important implications for simulating the current climate and climate changes. The comparison shows that the satellite estimates are reasonably good, but the differences between the satellite estimates and the surface measurements are somewhat larger at Macquarie than at most other locations, and suggests that (if anything) the satellite data are underestimating the model error associated with having too little reflected sunlight. In the infrared, the satellite errors are due to a systematic overestimation of the altitude of cloud base, and in general, the errors in both the solar and infrared have strong seasonal and diurnal variations.

1. Introduction

The Southern Ocean plays a large role in global oceanic heat and carbon uptake, in large measure because this is where much of the world's deep oceanic water returns to the surface, and on long timescales, much of the world's oceanic water passes through the Southern Ocean overturning circulation (Frölicher et al. 2015, Sallée et al. 2013). A number of studies over the past few years have identified a large excess in absorbed shortwave radiation (ASR) at the top of the atmosphere and in downwelling shortwave (SW) surface fluxes over the Southern Ocean in both climate models and reanalyses (e.g. Trenberth and Fasullo 2010, Ma et al. 2015, Kay et al. 2016, Zhang et al. 2016). In an analysis of the surface energy budget, Schneider and Reusch (2016) found that most climate models that participated in the Coupled Model Intercomparison Project Phase 5 (CMIP5) have excessive shortwave radiation reaching the surface over the Southern Ocean in early summer and midsummer as a result of having an insufficient shortwave (SW) cloud radiative effect (clouds do not reflect enough sunlight back toward space), which causes a warm bias in surface air temperatures during late summer; while in winter, most CMIP5 models have a negative longwave (LW) bias due to insufficient longwave cloud radiative forcing. On average, the water masses of the Southern Ocean in the CMIP5 models are too warm and light, also likely due in part to excess heat uptake (Sallée et al. 2013). These model radiative errors and associated excess heat uptake are of profound importance to global climate, including influencing the position of the Southern Hemisphere midlatitude jet and the Inter-Tropical Convergence Zone (ITCZ), as well as cross-hemispheric energy transports (Ceppi et al. 2012, 2013, Hwang and Frierson 2013, Kay et al 2016).

All of the above evaluations of model radiative fields rely on satellite top-of-atmosphere (TOA) or surface fluxes derived by the Clouds and the Earth's Radiant Energy System (CERES) project, specifically, the Energy Balanced and Filled TOA product (EBAF-TOA) (Loeb et al. 2018) and the EBAF-Surface flux product (Kato et al 2018). While CERES EBAF-Surface and related products have been evaluated against surface observations over some land regions and using data from (primarily tropical) buoys (e.g., Rutan et al. 2015, Kato et al 2018, Zhang et al. 2016), there has been little evaluation over the Southern Ocean. An exception is Rutan et al. (2018) who compared CERES retrievals with observed SW and LW downward surface fluxes measured from several Australian research vessels, including the Australian *Aurora Australis* ice-breaker. We summarize and discuss uncertainties estimated from these and other evaluation studies in detail later in this manuscript.

In response to the need for additional measurements of surface radiative fluxes, as well as precipitation, cloud and aerosol properties over the Southern Ocean, the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) program, the Australian Antarctic Division (AAD) and the Australian Bureau of Meteorology (BoM) collaborated in deploying a variety of ground-instrumentation to Macquarie Island between March 2016 and March 2018. Macquarie Island is located at 54.5° S, 158.9° E and has a small research station operated by AAD that is staffed year-round, in part by the BoM. The station supports a variety of research activities and includes a long history of surface weather and radiosonde observations (Hande et al. 2012, Wang et al. 2015).

The primary objective of the March 2016 to March 2018 deployment, hereafter the Macquarie Island Cloud and Radiation Experiment (MICRE), was to collect observations of surface radiation, precipitation, cloud and aerosol properties in order to evaluate satellite datasets and to improve knowledge of diurnal and seasonal variations in these properties, especially as pertains to the vertical structure of boundary layer clouds, precipitation, and the pervasive supercool liquid clouds which occupy this region.

In this article, CERES synoptic (SYN) 1 degree hourly SW and LW downwelling surface fluxes and monthly CERES EBAF-Surface fluxes are evaluated using surface observations collected during MICRE. The hourly CERES-SYN fluxes are derived using both Moderate Resolution Imaging Spectroradiometer (MODIS) and geostationary satellite imagery (Doelling et al 2013), and are subsequently used in the generation of the CERES EBAF-Surface fluxes, which includes corrections and adjustment to the SYN data to ensure consistency with CERES EBAF-TOA fluxes. This evaluation also briefly examines SW and LW surface fluxes derived operationally from spaceborne cloud radar (CloudSat) and lidar (Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations, CALIPSO) observations by the CloudSat project (Henderson et al. 2013). Section 2 summarizes the surface and satellite datasets used.

Results are given in section 3 and summarized in the context of previous surface-based evaluations in section 4, with conclusions and additional discussion given in section 5. Overall, biases (CERES – surface observations) in the CERES-SYN and EBAF downwelling surface fluxes are found to be slightly larger over the Macquarie Island than most other regions, approximately +10 Wm^{-2} for SW and -10 Wm^{-2} for the LW in the annual mean, but with significant seasonal and diurnal variations. Of particular note is that bias in LW surface flux shows a marked increase in bias (to about -16 Wm^{-2}) at night, which explains most of the total LW bias. The nighttime bias is found to be due to poor representation of cloud base associated with low clouds.

2. Description of Data

2.1 Surface dataset

In this manuscript we use observations of surface broadband SW and LW fluxes collected by the ARM broadband radiometers (ARM SKYRAD datastream mcqskyrad60sS1.b1, DOI: 10.5439/1025281), and in the later analysis, cloud base from the ARM ceilometer (ARM datastream mcqceilS1.b1, DOI: 10.5439/1181954) collected during MICRE. The shortwave radiometer calibration is traceable to the World Radiometric Reference and follows the Broadband Outdoor Radiometer CALibration (BORCAL) methods developed at the U.S. National Renewable Energy Laboratory, while calibration of the longwave radiometers is traceable to the interim World Infrared Standard Group standard (Andreas et al. 2018). The measurement uncertainty is expected to be about +/- 4% for the total downwelling shortwave flux and +/- 2% for the total downwelling longwave flux. The uncertainty in the field maybe slightly larger than the expected values, and we will return to this topic in the later discussion.

While data collection for most MICRE instrumentation began near the end of March or beginning of April 2016, there was unfortunately, a wire/grounding problem with the pyrgeometer (LW flux) measurements that was not corrected until August 15, 2016. Thus, the analysis presented is section 3 include SW flux measurements from April 3, 2016 to March 13, 2018 (a little over 23 months) and LW flux measurements from August 15, 2016 to March 13, 2018 (just under 19 months).

2.2 Satellite Datasets

In this study downwelling shortwave (SW) and longwave (LW) radiative fluxes from (1) the hourly CERES SYN 1 degree Edition 4A product (Doelling et al., 2013; Rutan et al., 2015), (2) the monthly CERES Energy Balanced and Filled (EBAF) Surface Product Edition 4 (Kato et al. 2018), and (3) the CloudSat Fluxes and Heating Rate with Lidar (FLXHR-LIDAR) version R05 (Henderson et al. 2013) are examined. The analysis includes comparison of hourly SYN data from the grid cell that contains Macquarie Island, and uses larger regional scale (mean fluxes) taken over a 10° x 10° area for the purpose of evaluating the EBAF and FLXHR-LIDAR products. We briefly summarize each dataset, below.

2.2.1 CERES SYN Product

Among other parameters, the CERES SYN 1 degree product provides hourly surface LW and SW fluxes computed based on cloud and aerosol properties derived from several satellites (MODIS Terra, MODIS Aqua, and Geostationary imagers), meteorological profiles from the NASA Global Modeling and Assimilation Office (GMAO), and surface properties from several sources (Rutan et al. 2015, Kato et al. 2019). Here we use the Edition 4.0 product (Doelling 2017, DOI: 10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A). At MODIS Terra and MODIS Aqua overpass times, cloud properties (more details below) are derived from MODIS data, while at other times of the day cloud properties are based on geostationary satellite observations (calibrated against MODIS) using similar retrieval algorithms (Doelling et al. 2013). Cloud properties from MODIS and geostationary imager data are collected (averaged) on the SYN 1 degree grid using four groups defined by the cloud *top* pressure. The four cloud-top-height categories are surface-to-700 hPa, 700-500 hPa, 500-300 hPa, and less-than-300 hPa. We note that the cloud *bases* are independent of the cloud *tops*, meaning that the cloud *base* in each category can be *below* the *top* or even the *base* of other cloud-top-height categories. For example, the cloud *base* of the 700-to-500 hPA category may be at a pressure-altitude that is larger than 700 hPA. The four cloud categories are randomly overlapped, as described in Kato et al. (2019, see Appendix A), creating 16 possible cloud vertical configurations. We provide some additional details and clarifications to the description given by Kato et al. 2019 in Appendix A of this document. Radiative fluxes are computed using a gamma-weighted two-stream model applied to (as many as) 4 of the 16 vertical configurations (as explained in Appendix A).

In general, the method of determining cloud geometric and optical properties in a given MODIS or geostationary pixel depends on the wavelength bands of the available observations, as described in Minnis et al. (2011). When the solar zenith angle is less than 82° (defined as "daytime"), several shortwave and infrared channels are used to retrieve cloud top effective temperature, phase, optical depth, and particle size using an iterative method. Beyond 82° (defined as "nighttime"), only infrared wavelengths are used. The nighttime algorithm includes an iterative process for clouds determined to be optically thin (based on thermal channel brightness temperature differences) that accounts for the cloud emissivity being less than 1 (i.e. not opaque in the infrared) and retrieves the cloud microphysics (optical depth and effective radius), while optically thick clouds are taken to be opaque and cloud microphysics are set to fixed values at night (depending on the inferred phase and cloud top temperature). Cloud base height is particularly important in determining downwelling LW surface flux. Because only passive remotely sensed data are used to construct the cloud profiles, the location of cloud base is not directly measured and not well constrained. The location of cloud base is determined from the retrieved cloud-top temperature and cloud thickness, where the cloud thickness is estimated in one of two ways. For liquid clouds, a relationship between optical depth and thickness derived from satellite and field data are applied (Minnis et al. 2011), while for ice clouds, a new latitude-dependent parameterization is used in SYN edition 4A. The new parameterization was developed using cloud property profiles constructed from the active remote sensors CloudSat and CALIPSO.

2.2.2 CERES Energy Balanced and Filled (EBAF) Surface Product

As described by Kato et al. (2018), the CERES project derives both top of atmosphere and surface radiative fluxes at several temporal and spatial scales, with the top-of-atmosphere (TOA) and surface irradiances determined separately. The TOA fluxes are derived directly from radiances measured by CERES instruments, and includes the Energy Balanced and Filled (EBAF) TOA product, which applies an algorithm that adjusts SW and LW TOA fluxes (within their uncertainties) in order to remove inconsistency between average global net TOA flux and heat storage in the earth–atmosphere system (Loeb et al. 2018). Surface fluxes, on the other hand, are computed using radiative transfer calculations following the discussion for the SYN product. The SYN TOA fluxes do not necessarily match those from the CERES EBAF-TOA product. In the CERES EBAF-Surface product, the atmospheric properties used to calculate the SYN fluxes are *bias corrected* and *adjusted* so that they produce TOA fluxes that match closely the EBAF-TOA fluxes. Here we use the CERES EBAF-Surface product Edition 4.0 (Loeb et 2017, DOI: 10.5067/TERRA+AQUA/CERES/EBAF-SURFACE_L3B004.0). The *bias correction* and *adjustment* procedures are complex, and a lengthy description is given by Kato et al. (2018). Very briefly, AIRS, CloudSat, and CALIPSO are used to estimate bias errors in some SYN inputs at the monthly scale. Specifically upper-tropospheric (200-500 hPa) temperature and specific humidity, low-level cloud fraction as viewed from space (over ocean), and total cloud fraction and cloud-base as viewed from the surface are bias corrected on spatial scales of 1 degree (but with some smoothing that includes the use of zonal averages in some cases). Following the bias correction (which nominally is correcting for errors in the SYN inputs), the monthly mean computed SYN-bias-corrected fluxes and EBAF-TOA fluxes are compared and differences are then minimized utilizing a Lagrange multiplier, which further adjusts temperature, water vapor, cloud, aerosol, and surface properties (within their uncertainties) in order to bring the computed TOA fluxes into close agreement with EBAF-TOA fluxes.

2.2.3 CloudSat Fluxes and Heating Rate with Lidar (FLXHR-LIDAR)

The CloudSat 2B-FLXHR-LIDAR product (Henderson et al. 2013) provides vertical profiles of SW and LW radiative fluxes and heating rates. The fluxes and heating rates are calculated using a two-stream plane-parallel doubling-adding radiative transfer model (L'Ecuyer et al. 2008, Henderson et al. 2013), based on vertical profiles constructed from radar and lidar backscatter from the CloudSat Cloud Profiling Radar (CPR) and the Cloud–Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard CALIPSO, respectively, along with auxiliary cloud information from MODIS, and environmental information from ECMWF. The radar and lidar data enable an explicit representation of vertical cloud properties, and in particular the representation of multi-layered cloud structures has been shown to improve the impact of cloud impacts on TOA and surface radiation (L'Ecuyer et al, 2019, Hang et al, 2019). This article uses the most recently released Revision 05 (R05) 2B-FLXHR-LIDAR data, available from the CloudSat data processing center (http://www.cloudsat.cira.colostate.edu/data-products/level-2b/2b-flxhr-lidar). The R05 data includes several improvements in land surface characteristics (i.e., surface emissivity and albedo) compared with R04 and is based on cloud properties from R05 CloudSat and V4 CALIPSO datasets. Major changes related to the R05 retrieval are described in Matus and L'Ecuyer (2017). Of particular note is that cloud properties for cirrus are now based on the CloudSat 2C-ICE product (Deng et al. 2013) and mixed phase clouds are more explicitly represented (Van Tricht et al. 2016), which has improved surface flux comparisons against ground sites in Greenland (McIlhattan et al. 2017). More generally, Matus and L'Ecuyer (2017) demonstrate that the improvements in R05 yield better agreement with respect to TOA global and regional fluxes when compared to the CERES CloudSat, CERES, and MODIS (CCCM) product (Kato et al. 2010).

3. Results

3.1 Comparison of coincident SYN hourly data with surface observations during MICRE

Figure 1 compares hourly CERES SYN SW and LW downwelling surface fluxes with hourly averages of measured values at Macquarie Island during MICRE. Specifically, Figure 1 shows the frequency of occurrence for a given pair of satellite-derived and surface-measured values. Here the frequency of occurrence values have been scaled logarithmically because the frequency is very large in some sections of the plots and low in others. Nominally, both observations would be equal and fall along the one-to-one line (shown in black), but with some scatter (departure from the line) due to the different spatial-scale in each dataset (the surface radiometers observe a much smaller area). We note nighttime values (defined here as times when the SW flux is less than 10 W/m^2) are not used in the SW comparison in Figure 1. The SYN SW values have a bias of 21.5 Wm^{-2} relative to the ground measurements during daylight. The bias would be roughly half this value if both day and night time samples were included, as is often done when reporting monthly or annual means. There are more than 8000 samples in this comparison and even considering serial correlation, the bias is significant at the 95% level of confidence. Nonetheless the points in the SW histogram appear to fall reasonably symmetrically about the one-to-one line.

The comparison of ground-measured and SYN LW fluxes in Figure 1 shows a bias of about -8.3 Wm^{-2}. We note that the SW and LW biases are in the opposite direction such that the bias in the total (SW + LW) radiative flux is small, with a magnitude of less than 2 Wm^{-2} in the daily (day +

night) average. We will discuss SW and net radiative flux in more detail later in the manuscript, and focus for the time being on the downwelling LW flux.

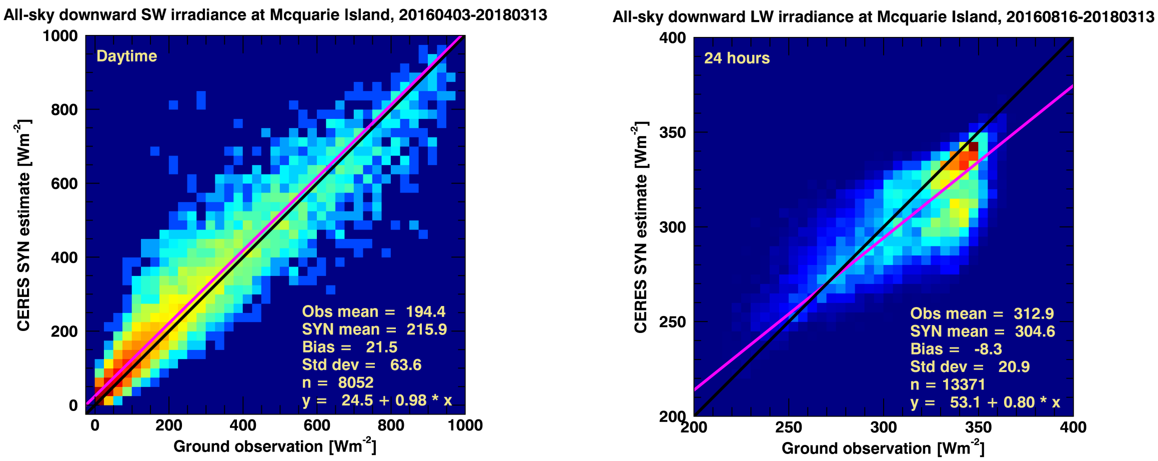

Figure 1. Comparisons between downwelling radiative fluxes from CERES SYN 1degree-hourly and ground-based measurements: a) Shortwave, b) Longwave. Nighttime values are not included in the SW comparisons.

Much of the LW bias is due to a cluster of points where SYN has LW fluxes near 300 Wm^{-2} and the ground observations have values near 340 Wm^{-2}, which stands out from the otherwise fairly linear distribution. Separating the LW data into daytime and nighttime populations (Figure 2) reveals that the offset cluster consists of fluxes occurring at night. While the magnitude of the SYN1deg daytime LW bias is only -1.4 Wm^{-2}, it increases to -16 Wm^{-2} at night.

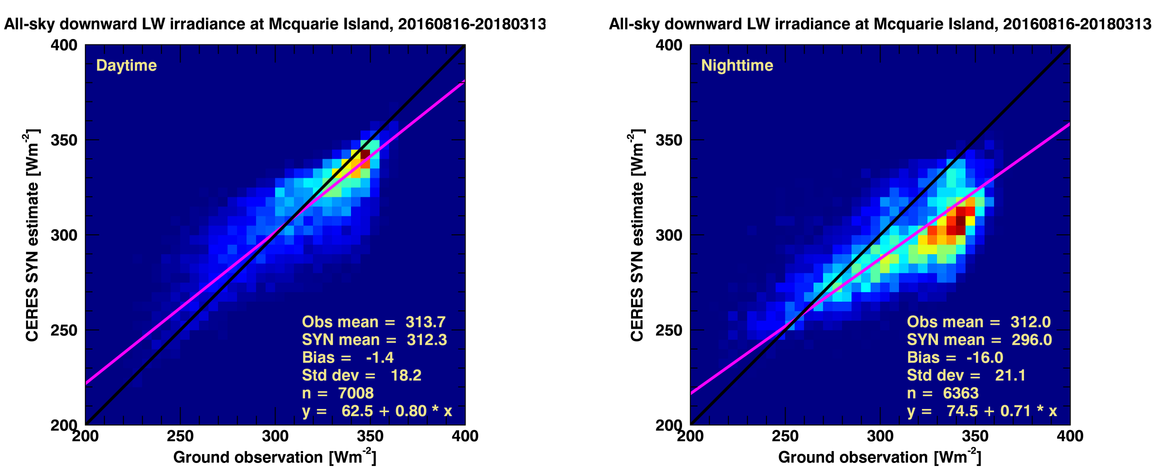

Figure 2. Comparisons between downwelling longwave fluxes from CERES SYN1deg-hourly and ground-based measurements. a) Daytime data. b) Nighttime data.

The vertical profiles of temperature, water vapor, and cloud base height are the primary determinants of downwelling LW flux at the Earth's surface. Given the relatively weak diurnal variations in temperature and water vapor in this region, and the algorithmic differences in the treatment of clouds between daytime and nighttime conditions in the satellite retrievals, one expects that the differences between the panels in Figure 2 are likely due to errors in cloud base. The SYN product provides cloud base pressure for each of the four cloud top height categories discussed in section 2.2, which we have converted to altitude above ground level using monthly mean profiles from radiosonde observations and accounted for cloud overlap (Appendix A). In Figure 3, we compare the distribution of the SYN lowest cloud base height to the distribution of the lowest cloud base determined by a Vaisala laser ceilometer (lidar) deployed during MICRE. The bars on the far right show the fraction of clear-sky (no clouds in the column). The ceilometer suggests somewhat more clear sky than the satellite. While this is not surprising given the ceilometer observes a smaller area than the satellite imager pixels used in SYN, it is also likely due in part to the inability of the ceilometer to detect clouds above 5 km and some optically thin ice clouds. Regardless of these issues, it is clear the CERES SYN data substantially under represents the occurrence of clouds with bases below 900 hPa at the Macquarie Island site.

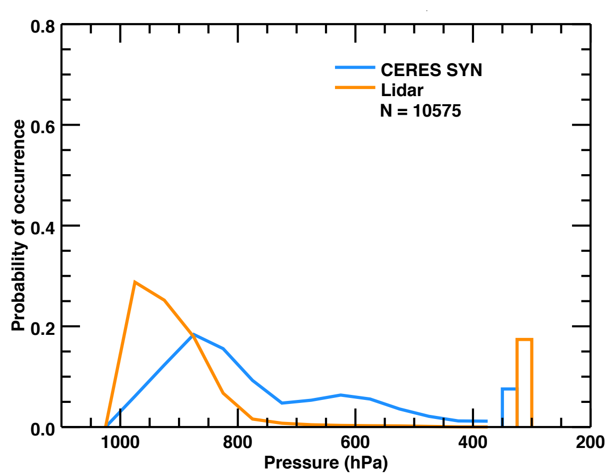

Figure 3. Cloud base distributions from CERES (blue) and ceilometer (orange). The bars on the far right show the fraction of clear-sky.

Figure 4 further divides the cloud base height distributions into cases where (top) the LW difference (SYN – ARM surface radiometer) is large and negative (less than -30 Wm^{-2}), (middle) the LW difference is small (within 10 Wm^{-2}), and (bottom) the LW difference is large and positive (greater than 30 Wm^{-2}). The top panel further demonstrates that large underestimates in the SYN LW flux occur when the ceilometer data is dominated by low clouds. This "large error" condition occurs about 13% of the time. There is very little clear during these 1-hour periods (in either dataset), which shows that this occurs when extensive low cloud cover is present. Examination of W-band radar shows that much of the time, the clouds in this category are multilayered, but also includes periods that are apparently dominated by only low-altitude stratocumulus (though it is possible the radar is failing to detect some high altitude cloud).

Not surprisingly, the middle panel shows that when the LW flux bias is small (less than 10 Wm^{-2}), the cloud base distributions are much more similar. Even considering the roughly 10% difference in the amount of clear sky (which again may be due to higher altitude clouds missed by the ceilometer), it is clear that the presence of clouds with bases below 900 hPA remains too small, and the occurrence of cloud-base above about 950 hPA is too large. In short, the same cloud base issue still occurs but is less severe.

In the bottom panel, biases greater than +30 Wm^{-2} are relatively uncommon, occurring 3% of the time, and are dominated by cases where the laser ceilometer does not detect any cloud over the one hour periods being analyzed. It is likely that most of the LW difference here is due to regional variability. Simply put, there are fewer clouds over the island (during the 1 hour over which ceilometer data is aggregated) as compared with the 1 degree region surrounding the site. Setting aside the difference in the amount of cloud, the distribution of cloud-base during these apparently-broken-low-cloud periods appears to be well captured by SYN. It is perhaps also noteworthy that the broken clouds (in this category) have an overall higher cloud base than the clouds in the other two categories.

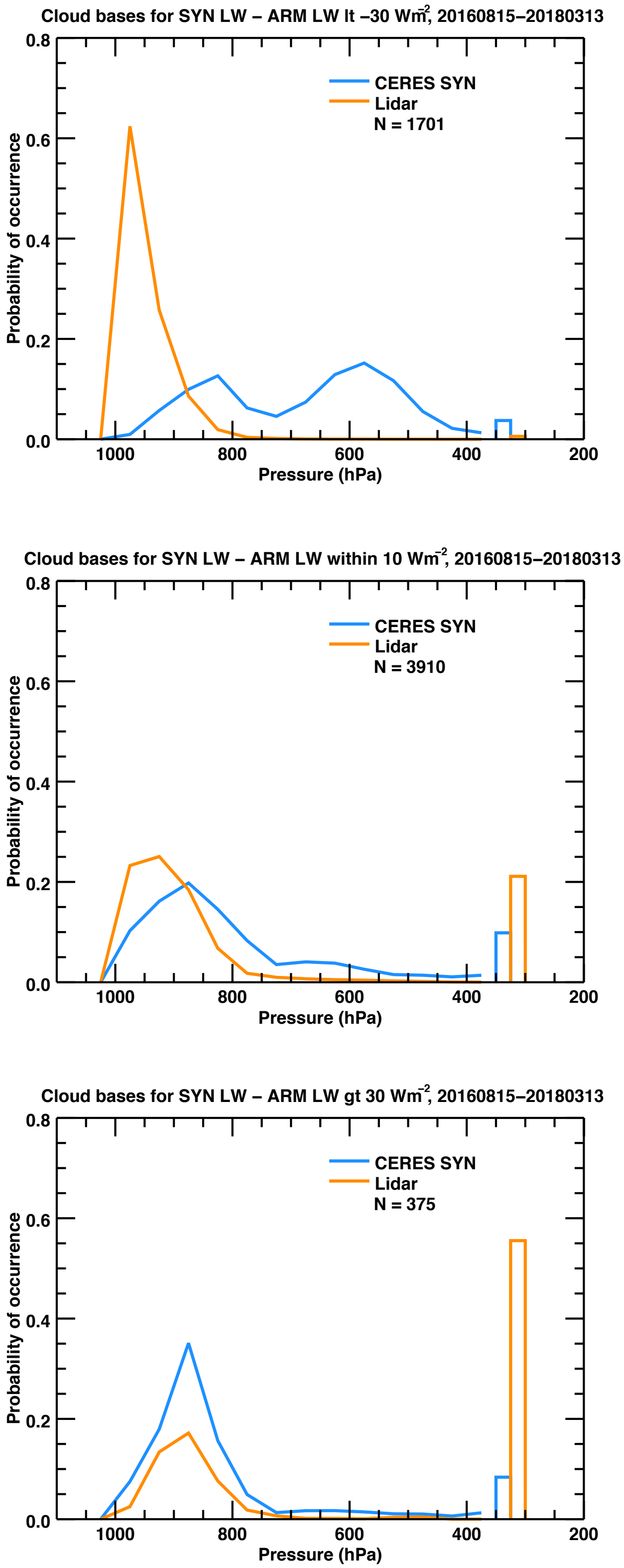

Figure 4. Same as figure 3 except limited to cases where (top) ΔLW = (SYN – ARM Ground) < -30 Wm^{-2}, (middle) -10 Wm^{-2} < ΔLW < 10 Wm^{-2}, (bottom) ΔLW > 30 Wm^{-2}.

As mentioned in the description of the SYN product, there are four different retrieval paths used in SYN. These are distinguished by whether MODIS or geostationary imager data are used, and by whether daytime (solar zenith angle less than 82^{o}) or nighttime algorithms are used. In Figure 5, we further examine the satellite and lidar cloud base distributions according to the four retrieval paths. Here, MODIS-based retrievals are given in the upper two panels and Geostationary-based retrievals (Himawari for the region and time-period being studied) in the lower two panels, while the two panels on the left are daytime retrievals (retrievals using visible and infrared channels) and on the right are nighttime retrievals (infrared channels only). Regardless of retrieval path, SYN under represents the presence of cloud bases below 900 hPA. However, the SYN and lidar cloud base height distributions do agree better during the daytime (left panels) than at night (right panels), regardless of whether MODIS or Geostationary data are being used. The overall similarity of the MODIS and Geostationary-based satellite retrievals indicates the flux differences are not being driven by problems with the calibration of the Geostationary data (at least for Himawari-8 in this region).

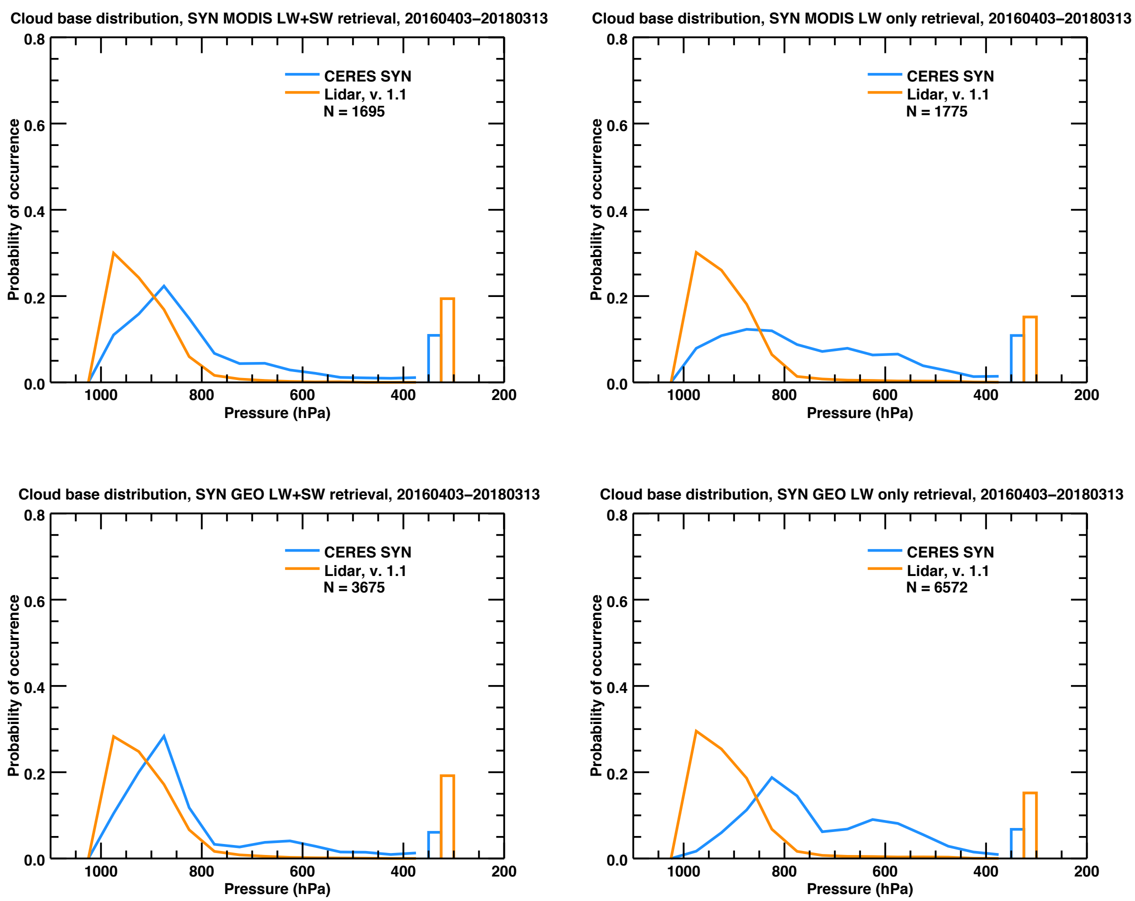

Figure 5. Same as Figure 3, except: (top left) MODIS daytime, (top right) MODIS nighttime, (bottom left) Geostationary (Himawari) daytime and (bottom right) Geostationary nighttime.

3.2 The Diurnal Cycle and the CloudSat FLXHR-LIDAR

Figure 6 plots the diurnal cycle of mean surface fluxes for all of the coincident SYN and observed surface fluxes during MICRE. Here the orange and light blue shading indicates the two-sigma sampling uncertainty (95% confidence) interval (given by twice the standard deviation divided by the square root of the number of samples) for the ceilometer and SYN data, respectively. Values from the CloudSat FLXHR-LIDAR product (R05) are shown as black dots. CloudSat and Calipso are sun-synchronous polar orbiting satellites which pass near Macquarie island at about 2 pm and 12:30 am local time. The CloudSat data shown here is the mean taken over the period August 15, 2006 until December 30, 2009, during which time CloudSat was operating nominally during both daytime and nighttime overpasses. While CloudSat did collect data between March 2016 and December 2017, data for this period (coincident with MICRE) has not yet been processed. In addition, owing to problems with the CloudSat satellite battery, CloudSat has only been able to collect data during the afternoon (2 pm), daylight overpass (at Macquarie Island) since April of 2011. Thus some level of statistical comparison becomes necessary. Comparing CloudSat and SYN data in this way requires there be little variation in the mean fluxes between the two time periods examined. In this regard we note the standard deviation of annual mean SW and LW flux in the CERES SYN product between 2001 and 2017 is only 1.7 and 2.1 Wm^{-2}, respectively.

The left panel in Figure 6, shows that the diurnal cycle of the SYN downwelling SW fluxes compares well with the observed SW fluxes. The largest difference occurs at about 11 am, where the difference is about 38 Wm^{-2}. The two-sigma (95% confidence) intervals barely overlap at 11 am, suggesting the difference between SYN and the surface SW fluxes are not likely due to sampling limitations. However, a calibration error of 4% in the surface observations would create greater overlap between the uncertainty shading, such that the possibility of a combination of calibration error and sampling differences cannot be rejected. Nonetheless, the fact that the difference has a diurnal cycle (in which differences are larger at 11 am than 1 pm, for instance) suggests that a large calibration error is not likely. We will discuss this result and SW fluxes in more detail later in section 5. At the time of the CloudSat afternoon overpass, the SYN, CloudSat, and measured values all agree within the sampling uncertainty. The sampling uncertainty in the CloudSat result is comparable, but slighter larger, than the size of the dot used to denote the CloudSat fluxes. In the LW (right panel), the SYN LW downwelling fluxes compare well with the surface measurements between about 9 am and 3 pm, but poorly overnight – consistent with the results shown in Figure 2 and associated discussion. The CloudSat downwelling fluxes, on the other hand, compare well with the surface measurements during both afternoon and night overpasses.

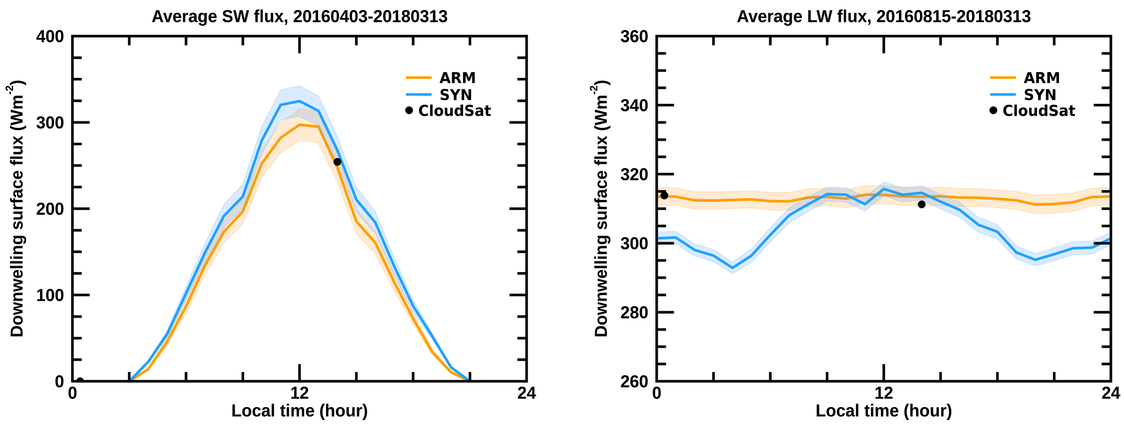

Figure 6. Diurnal cycle of the mean SYN retrieved and measured surface fluxes during MICRE. Left panel shows downwelling SW flux and right panel shows downwelling LW flux. Shading indicates sampling uncertainty in the mean. Black dots denote mean fluxes from the CloudSat FLXHR-LIDAR product (R05) based on data from 2006 through 2009 (note a different time period, see text). Sampling uncertainty of CloudSat data is comparable to the size of the dot.

Figure 7 shows that the same pattern found in figure 6 for all coincident data, is found in each individual season.   Seasonal and annual means and biases are given in Table 2. As there is less night during the Spring and Summer, it is not surprising to find the SYN LW biases (averaged over the day in Table 2), are less in Spring and Summer than during the Fall and Winter. Likewise, disagreement between SW SYN and surface fluxes are largest in Spring and Summer. Differences between CloudSat fluxes and surface values are within or near the sampling uncertainty (depicted by bars shown on Figure 7). In Summer, the CloudSat mean SW fluxes is high relative to the observations, but the bias remains within or close to the sampling uncertainty and we note there is additional uncertainty related to the differing time periods (these are not coincident data) which is not represented by the uncertainty bars in Figure 7. The sampling uncertainty for the CloudSat fluxes is much larger than that for CERES because CloudSat observes the region far less frequently than CERES.

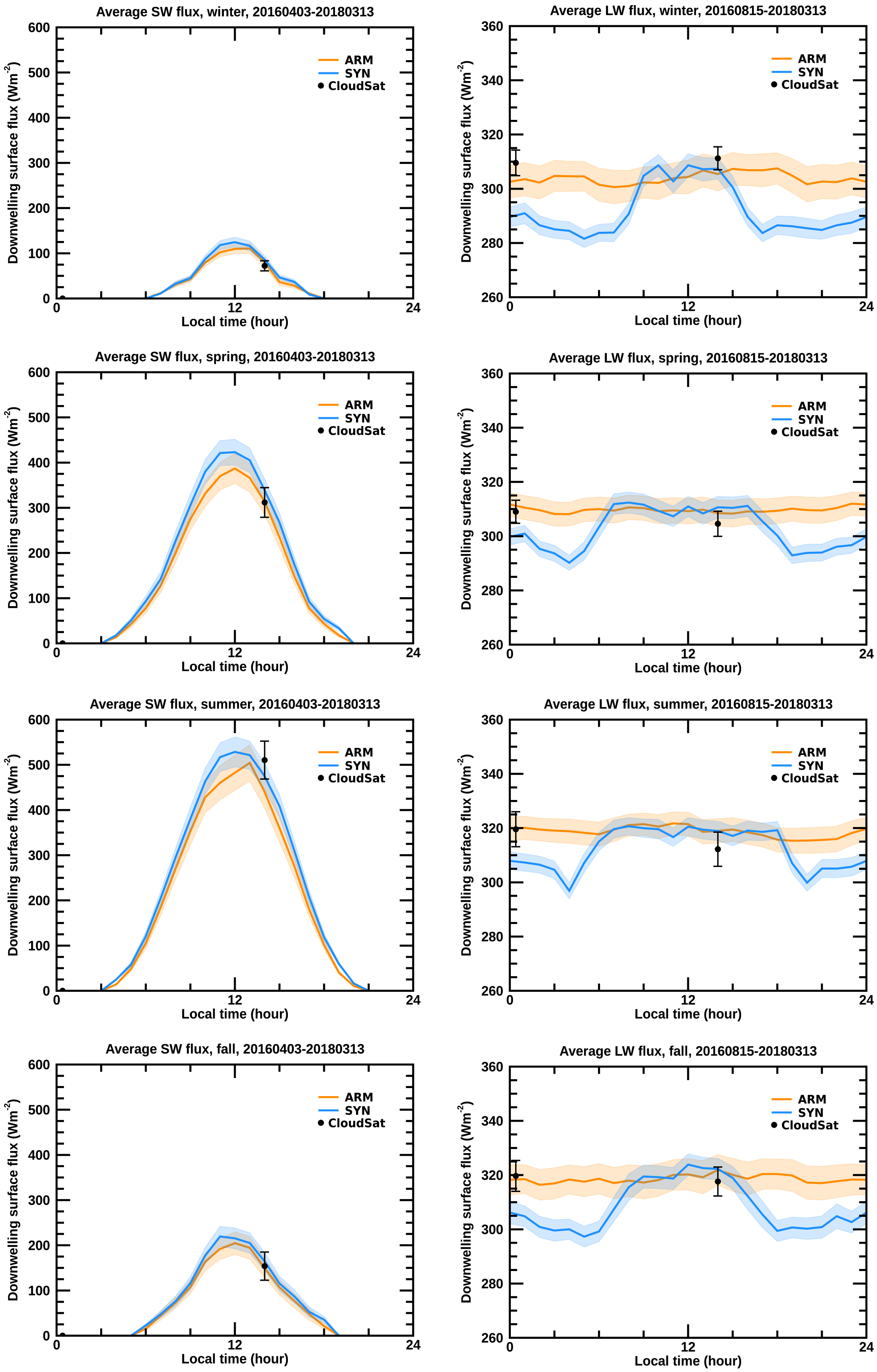
Average SW flux, winter, 20160403-20180313
Average LW flux, winter, 20160815-20180313
Average SW flux, spring, 20160403-20180313
Average LW flux, spring, 20160815-20180313
Average SW flux, summer, 20160403-20180313
Average LW flux, summer, 20160815-20180313
Average SW flux, fall, 20160403-20180313
Average LW flux, fall, 20160815-20180313
Downwelling surface flux (Wm^{-2})
Local time (hour)
ARM
SYN
CloudSat

Figure 7. Same as Figure 6, except showing result for Southern Hemisphere Winter (JJA), Spring (SON), Summer (DJF), and Fall (MAM). CloudSat data are from different time period (see text). Bars on CloudSat dots show sampling uncertainty (and do not include uncertainty due to interannual variability).

W/m^2	*Surface obs.*	*CERES SYN (1°) Coincident*	*CERES SYN (10°) 2001-2017*	*CERES EBAF (10°) 2001-2017*	*CloudSat (1°)* ***day**** *2006-2009*	*CloudSat (1°)* ***night**** *2006-2009*
Winter (JJA)						
SW mean	*24.5*	*27.0*	*27.1 [1.5]*	*27.4 [1.3]*	*72.2**	*-*
SW bias	*-*	*2.6*	*2.6*	*2.9*	*-8.2**	
LW mean	*302.3*	*291.1*	*294.4 [2.8]*	*291.5 [3.0]*	*311.2**	*309.5**
LW bias	*-*	*-11.2*	*-7.9*	*-10.8*	*5.8**	*6.9**
Spring (SON)						
SW mean	*122.8*	*138.4*	*138.8 [4.3]*	*141.7 [4.0]*	*311.8**	*-*
SW bias	*-*	*15.6*	*16.0*	*18.9*	*-1.6**	
LW mean	*309.7*	*302.5*	*302.3 [2.1]*	*297.7 [2.1]*	*304.6**	*309.0**
LW bias	*-*	*-7.1*	*-7.4*	*-12.0*	*-3.9**	*-2.7**
Summer (DJF)						
SW mean	*177.0*	*195.3*	*192.6 [4.8]*	*197.0 [4.3]*	*510.4**	*-*
SW bias	*-*	*18.4*	*15.6*	*20.0*	*70.5**	
LW mean	*318.8*	*312.5*	*315.5 [2.5]*	*310.2 [2.5]*	*312.2**	*319.7**
LW bias	*-*	*-6.3*	*-3.3*	*-8.6*	*-6.8**	*-0.1**
Fall (MAM)						
SW mean	*55.5*	*61.1*	*60.3 [2.6]*	*62.8 [2.6]*	*153.9**	*-*
SW bias	*-*	*5.6*	*4.8*	*7.3*	*4.3**	
LW mean	*318.9*	*308.2*	*307.6 [3.3]*	*303.8 [3.5]*	*317.6*	*319.7**
LW bias	*-*	*-10.7*	*-11.3*	*-15.1*	*-4.3**	*1.4**
Annual						
SW mean	*94.9*	*105.5*	*104.7 [1.7]*	*107.2 [1.1]*	*262.1**	*-*
SW bias	*-*	*10.5*	*9.8*	*12.3*	*16.3**	
LW mean	*312.4*	*303.6*	*304.9 [2.1]*	*300.8 [2.1]*	*311.4**	*314.5**
LW bias	*-*	*-8.8*	*-7.5*	*-11.6*	*-2.3**	*1.43**

*Table 2 – Seasonal and annual means in downwelling SW and LW fluxes. All values have units of Wm^{-2}. Biases given with respect to mean surface fluxes (e.g. SYN – Surface Obs). Surface and CERES values are 24-hour averages. *CloudSat values are NOT 24-hour averages, but the average value at the time of the CloudSat overpass (and biases are the difference with surface obs during the same hour). When present, parentheses "[ ]" show year-to-year standard deviation. Data for CERES SYN is given for coincident points in the 1 degree grid cell that contains Macquarie Island ground site, and for 10 x 10 degree (lat/lon) region centered on the island.*

3.3 Monthly SYN and EBAF-Surface Fluxes

CERES EBAF-Surface fluxes are only available on monthly (and longer) time scales. Thus rather than compare EBAF to coincident surface data (which would include only 18 to 24 points), we instead compare the CERES SYN and CERES EBAF for the 17 years of Edition 4 data available at the present time. EBAF Monthly values are close to SYN values in the 10 x 10 degree region surrounding Macquarie Island. Figure 8 shows the distribution of EBAF – SYN (Edition 4) downwelling surface fluxes for the period 2001 to 2017. As discussed in section 3.3, CERES EBAF-Surface fluxes contain both *bias corrections* and *adjustments,* which nominally include bias corrections for cloud base (see section 2.3). However, the net effect of the corrections appears to be in the wrong direction. EBAF SW fluxes are typically somewhat larger than SYN fluxes by a small amount (2.5 Wm^{-2} on average), when the surface measurements suggest the SYN fluxes are already too large. And similarly, EBAF LW fluxes are typically smaller then SYN fluxes (by 4.1 Wm^{-2} on average), when the surface measurements suggest the SYN fluxes are already too small. In short, CERES EBAF fluxes appear to have (if anything) slightly larger biases in the region surrounding Macquarie Island.

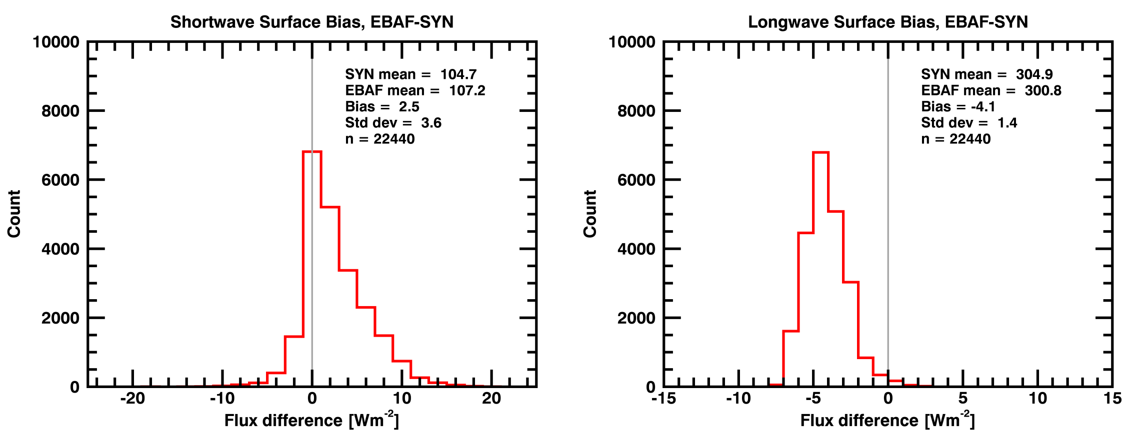

Figure 8. Distribution of monthly SW (left) and LW (right) downwelling surface flux differences between Edition 4 CERES EBAF and CERES SYN 1^o data (EBAF –SYN) for all grid points within 10^o of Macquarie Island between 2001 and 2017.

4. Results from Previous CERES Evaluation Studies

Several studies have evaluated CERES SYN and EBAF surface fluxes against surface observations. Rutan et al. (2015) evaluated the SYN Edition 3 and EBAF Edition 2.7 surface fluxes, comparing both to other satellite based estimates and measurements recorded at 37 globally distributed land-based sites and 48 buoys (over subtropical and tropical oceans) between 2000 to 2007. There were no buoy sites at mid or high latitudes in this data set. Relative to the surface measurements, SYN monthly mean downwelling SW fluxes were biased +1.8 Wm^{-2} over the land sites and +4.9 Wm^{-2} over the ocean sites with a standard deviation (in the monthly means) of about 12 Wm^{-2}, while EBAF Edition 2.7 surface fluxes were found to have a bias of only -0.5 Wm^{-2} over land and +5.0 over ocean with again a standard deviation of about 12 Wm^{-2} in the monthly means. For LW fluxes, Rutan et al. (2015) found biases for SYN of -4.2 Wm^{-2} over land and -3.6 Wm^{-2} over the ocean, with a standard deviation (in the monthly means) of about 10 Wm^{-2}, while for EBAF (Edition 2.7) the bias was just 1.2 Wm^{-2} over land and -3.5 Wm^{-2} over ocean, with a standard deviation in the monthly means of about 10 Wm^{-2}.

Kato et al. (2018) present a similar assessment for the EBAF Ed. 4.0 surface fluxes, using many of the same sites as Rutan et al. (2015) but using data from March 2000 through February 2016. The EBAF monthly mean SW downwelling fluxes were found to have biases of -0.8 Wm^{-2} and 4.8 Wm^{-2} over land and ocean, respectively (quite similar to values found by Rutan et al. 2015 using the previous edition of the EBAF data), while LW downwelling were improved with overall (taken across all station) mean biases of only -0.04 Wm^{-2} over land and 1.0 Wm^{-2} over ocean. Standard deviations in the monthly means remained about 10 Wm^{-2}.

Studies by Ma et al. (2015) and Zhang et al. (2016) also compared CERES EBAF downwelling SW fluxes with data from a wider range of sites/networks including (i) the Global Energy Balance Archive (GEBA) with sites located primarily in Europe and Japan but also some sites in Austrailia, Asia, South America, and Africa, (ii) the Greenland Climate Network (GC-NET) and (iii) Climate Data Center of Chinese Meteorological Administration (CDC/CMA). Most of these additional sites had mean biases less than 5 Wm^{-2}, with the large set of GEBA sites having an average bias of less than 2.5 Wm^{-2} in both summer and winter seasons.

In all of the above studies, some individual stations were found to have larger biases, and in some locations biases exceeded 20 Wm^{-2}. In most cases, these large differences were associated with suspect measurements (e.g., measurements which may suffer from dust contamination) or occur in mountainous or snow and ice covered regions, where larger differences might be due to spatial heterogeneity (of clouds and surface conditions) and challenges presented by snow and ice covered surfaces to the retrieval of both cloud and surface albedos (Hinkelman et al. 2015, Riihelä et al. 2017, Kato et al. 2018). Macquarie Island is far removed from Antarctic Sea Ice, and while it does snow on the island, it is largely free of snow (especially during Spring and Summer when SW flux errors are largest) and covers only a very small fraction of the area. As such, snow and ice is not a concern in this study. The possibility that the biases we find at Macquarie Island could be caused by spatial heterogeneity is more difficult to assess and we will return to this topic in the next section.

The poster presentation by Rutan et al. (2018) provides the only other direct evaluation of CERES surface fluxes over the Southern Ocean of which we are aware. This study utilized radiometric data collected from New Zealand and Australian research vessels over the period 2008-2016. This included data collected from the Australian *Aurora Australis* ice breaker during its resupply mission to Macquarie Island and Australian Antarctic stations, as well as data collected from the New Zealand Research Vessel (R/V) *Tangaroa,* which also include a few voyages which passed south 50° S. In total, the number of hourly samples gathered over the 8-year ship record is roughly equivalent to what was collected during MICRE over two years. In the latitude range between 50° to 60° S, Rutan et al. 2018 show differences between the SYN and surface LW fluxes between about 5-10 Wm^{-2} from both the *Aurora Australis* and R/V *Tangaroa,* consistent with what we find from Macquarie Island. Rutan et al. 2018 also generally find noteworthy day to night differences in the LW bias. For the *Aurora Australis* (which spent most of its time south 50° S) the day to night difference was about -8 Wm^{-2} in Summer and Fall, but near zero in Spring during which both day and night have a bias near -10 Wm^{-2}. On a minor note, the original poster presented by Rutan et al. 2018 contained an error in the seasonal-and-diurnal bias bar charts, and we thank Dr. Rutan for kindly providing us with corrected figures. The R/V *Tangaroa* data suggest similarly large day-to-night differences in Spring, Fall and Winter, but the seasonal-to-diurnal bias analysis was not subdivided by latitude, and a large fraction of the data being gathered from the R/V *Tangaroa* was gathered North of 45° S (and not over the Southern Ocean). In summary, the LW results of Rutan et al. (2018) are reasonably consistent with the present analysis.

In the SW, however, Rutan et al. (2018) found no significant bias in the annual mean between 50° to 60° S from the *Aurora Australis* (though there appears to be a bias south of 60° S in these data), but do show a SW bias from the R/V *Tangaroa* in this latitude range, which is consistent with the data collected at Macquarie. Taken over all SO latitudes, data collected from *Aurora Australis* suggest a seasonal SW bias of about: +5 Wm^{-2} in the Spring, -2 Wm^{-2} in the Summer, and +8 Wm^{-2} in the Fall with sampling uncertainty in each season of about 5 Wm^{-2}. During its voyages, the R/V *Tangaroa* passed by Macquarie Island whereas much of the *Aurora Australis* data was collected further to the east. Thus one possibility for the differences between the two ship datasets is that there is a longitudinal variation. However, we note that the ship cruises do not randomly sample Southern Ocean meteorological conditions. For good and obvious reasons, the resupply transits try to avoid Southern Ocean cyclones. We speculate that sampling uncertainty and conditional meteorological sampling more likely explain the differences. A regime-based analysis of the MICRE and ship datasets might prove worthwhile but such is beyond the scope of this first analysis, and as discussed in the next section it is possible the MICRE data could also be biased by island effects (local surface heterogeneity).

5. Discussion and Conclusions

We find the annual mean bias in the CERES SYN and EBAF SW downwelling flux during MICRE to be about +10 Wm^{-2} with a larger bias occurring in the Spring and Summer (15 to 20 Wm^{-2}), see Table 2. This is larger than the mean bias of about +5 Wm^{-2} found from using measurements from ocean buoys (primarily located in the subtropics and tropics) by Rutan et al. (2015) and Kato et al. (2108). This bias is also larger than the 95% sampling uncertainty of about 2 to 3 Wm^{-2} and the expected calibration uncertainty +/- 4% or a little over 4 Wm^{-2} in the annual mean for the surface measurements. Nonetheless, while neither sampling uncertainty nor calibration alone can account for the bias, in combination these two factors could account for much of the apparent bias. Another possibility is that the SW bias we find might be due to an "island effect", where clouds reflect more sunlight back toward space at the measurement site (because there is more cloud cover or clouds are more reflective over the island site) than over the surrounding ocean. If so, the results presented here suggest that this occurs preferentially in Spring and Summer and preferentially between roughly 9 am and noon. An analysis of cloud properties form MICRE (which will include analysis of ground-based cloud radar and depolarization lidar data, as well as satellite retrievals) is ongoing, and may provide some insight into the existence and cause of the SW bias.

As regards climate models, in many CMIP3 and CMIP5 models the downwelling SW surface flux is too large as compared against CERES-EBAF, with multimodel averages have differences that range between 10 and 25 Wm^{-2} over much of the Southern Ocean (Trenberth and Fasullo 2010, Ma et al. 2015, Kay et al. 2016, Zhang et al. 2016). If the Macquarie observations are correct and representative of the larger SO, the CERES SW fluxes are too small by roughly 10 Wm^{-2} in the SW and the model errors are larger than these previous studies have found. This suggests that additional measurements and analysis should be undertaken at Macquarie Island and other locations, to more firmly establish the SW bias we have found at Macquarie Island, and to determine the extent to which the Macquarie data are representative of other parts of the Southern Ocean, and nominally, to identify the underlying cause of the CERES bias.

We find the annual mean bias in the CERES SYN LW downwelling flux during MICRE is also of similar magnitude but opposite in sign, about -10 Wm^{-2} (see Table #2), with slightly larger values in the Fall and Winter than in the Spring and Summer. Unlike the situation in the SW, it is clear that the LW bias is not due to calibration or sampling. Rather an examination of the diurnal cycle shows the LW bias occurs almost entirely at night, which in turn is clearly related to the cloud-base being too high (and too cold) in the CERES SYN flux retrievals at night. In most respects, this result is not surprising. Comparison of LW fluxes from the previous version of CERES-MODIS retrievals (used in SYN) with retrievals based on a combination of CloudSat (radar), Calipso (lidar) and MODIS by Kato et al. (2011, see their figure 3) show a seasonally varying zonal bias in LW surface fluxes over the Southern Hemisphere, with values that range between about -3 to -7 Wm^{-2} at the latitude occupied by Macquarie Island. Kato et al. (2011) likewise identified the LW surface bias as being due primarily to problems with cloud base. As regards the current version of SYN (Edition 4), Kato et al. (2019, see their figure 1) suggest the near surface cloud occurrence profile (the volume of atmosphere containing cloud) remains too low near the surface in Edition 4 as compared with active sensors (radar and lidar profiles) from CloudSat and Calipso, and show there is a stark reduction in near surface cloud at night (Kato et al. 2019, their figure 2). Given the algorithmic nature of the error, which originates from errors in the cloud-base retrieval, it is likely that this LW bias affects much of the Southern Ocean, though the magnitude will likely vary with the amount of low cloud. Again additional measurements should be undertaken to establish the Macquarie results are correct and to examine the degree to which variations in sea surface temperature, cloud type and other factors matter.

We note that while the distribution of cloud bases in the SYN product is better during the daytime (when satellite visible channels are used in the cloud property retrievals), we find cloud bases below 900 hPA are still underrepresented (just not as severely as at night). Indeed Figure 2 (left panel) shows that during the day SYN LW fluxes tend to be too small (below the 1-to-1 line) when the observed fluxes are above 300 Wm^{-2} (because low based clouds are present) and too large (above the 1-to-1 line) when the observed fluxes are below about 300 Wm^{-2}. This suggests that the low daytime LW bias of less then 2 Wm^{-2} at Macquaire reported here is likely due in some part to a fortuitous cancellation of errors with other factors, and analysis of surface temperatures and boundary layer thermodynamic profiles (based on radiosonde data) should perhaps be undertaken to explore this issue, further.

The results presented in section 3, also demonstrate that the CERES-EBAF SW and LW fluxes track the CERES-SYN values closely in (at least) the 10 degree region surrounding Macquarie Island. While the *bias corrections* and *adjustments* applied to monthly EBAF data appear to have reduced biases in other regions (Kato et al. 2018), such does not seem to be the case at this location.

Overall, the LW flux comparison undertaken here reinforces the need for further improvements in CERES SYN (including CERES-MODIS retrievals) and EBAF treatments of low clouds and low cloud base at night, in particular. As mentioned briefly in section 3, our initial impression is that much of the LW error occurs when multilayer clouds are present, and an ongoing analysis of cloud properties from MICRE should provide additional details in this regard. Regardless, the relative success of LW fluxes during the day suggests that the nighttime problem is inherently rooted in the loss of information contained in the visible-radiances used in the daytime retrievals, and it may well be that what is needed is a greater reliance on climatological constraints or other *aprori* knowledge during the night. For example, for regions with small diurnal cycles in boundary layer thermodynamics, precipitation and clouds such as the Southern Ocean (Hande et al. 2012, Wang et al 2012), a simple approach might be to consider using statistical retrievals (tuned regressions) rather than the current "physical retrievals" to ensure the cloud geometric and microphysical properties at night match those during the day (for a given set of infrared channel measurements and perhaps meteorological variables).

While the most obvious (and arguably best) route to improving LW flux would be to focus on improving CERES SYN and CERES-MODIS retrievals that flow into CERES EBAF, an alternative might be to applied EBAF cloud base *bias corrections* separately to data collected during nighttime. That is, EBAF *corrections* could still be based on monthly data, but monthly-daytime and monthly-nighttime averages could be calculated and corrected separately, before being combined to calculate the 24-hour averaged monthly mean.

The CERES SYN and EBAF surface SW and LW biases nearly cancel (sum to near zero) in the annual mean. As far as we can conceive this is a coincidence, and we stress that it is true only in the annual mean. There is a significant imbalance on seasonal scales in the net radiation, with too much net radiative heating of the surface occurring in the Spring and Summer (because the magnitude of the positive SW bias is larger during these seasons and greater than the magnitude of the negative LW bias); and there is net radiative cooling of the surface in the Fall and Winter (because the magnitude of the negative LW bias is larger in these seasons and greater than the magnitude of the positive SW bias); and likewise in the diurnal cycle where there is too much SW heating during the day (which is strongest in the summer) and too little LW heating at night. Accordingly, evaluations of model output on seasonal or diurnal time scales with CERES SYN and EBAF datasets should consider these differences in seasonal and diurnal biases.

Acknowledgements: Thank you to Seiji Kato, David Rutan, and Fred Rose for your very helpful comments and explanations. This work was supported by NASA through grant NNX16AM05G. CERES SYN and EBAF data were obtained from the NASA Langley Research Center Atmospheric Science Data Center (https://eosweb.larc.nasa.gov/project/ceres) with DOI for the specific Edition 4 product used given in Section 2. Likewise, CloudSat FLXHR-LIDAR

data are available via the CloudSat Data Processing center (http://www.cloudsat.cira.colostate.edu/data-products/level-2b/2b-flxhr-lidar). MICRE observations made by the U.S DOE ARM program instrumentation (including surface radiation and ceilometer data sets) are available through the DOE ARM program data archive (https://adc.arm.gov/) (DOIs are given in section 2.1) and can also be obtained by request to Roger Marchand at the University of Washington (rojmarch@u.washington.edu).

Appendix A: Cloud Overlap Treatment in CERES SYN/EBAF-Surface Edition 4

The cloud overlap scheme described below is applied in CERES Edition 4, and was not used in earlier editions. As described in Kato et al. (2019) there are four cloud type categories, which are defined by the cloud-top pressure: low = surface-to-700 hPa, mid-low= 700-500 hPa, mid-high = 500-300 hPa, and high = less-than-300 hPa. These 4 cloud type categories are overlapped (in Edition 4) to give 15 different combinations of cloud overlap plus one clear scene. Table A1 shows the cloud overlap combinations. The cloud fraction associated with each of the 15 combinations is obtained assuming random overlap of the 4 cloud-types and simply given by the product of the "true" cloud fraction (or clear fraction) associated with each layer, such that

$$c1 = C1*C2*C3*C4$$

$$c2 = C1*C2*C3*(1-C4) \ldots$$

$$c15 = (1-C1)*(1-C2)*(1-C3)*C4$$

We stress that C1 to C4 are nominally the "true" cloud fraction for each pressure category NOT the cloud fraction observed from space (S1 to S4). In CERES processing C1 to C4 are derived from S1 to S4 assuming random overlap, see equations B1 to B4 in Kato et al. 2019.

Table A1 – Cloud Overlap Categories.

	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15
C1 high	x	x	x	x	x	x	x	x							
C2 mid-high	x	x	x	x					x	x	x	x			

C3 mid-low	x	x			x	x			x	x			x	x	
C4 low	x		x		x		x		x		x		x		x

As described in Kato et al (2019) at most four of these 15 overlap categories are used in the flux calculations, but the rules for selecting the 4 profiles were not clearly described. It is not simply the largest four values taken over the collection c1 to c15, but rather up to four values are chosen within the following subsets:

High_cloud_profile = the vertical profile associated with high clouds is represented by the category with the largest value between c1 to c8. If all are zero then no profile with high-cloud is used.

Mid-high_profile = the profile with a mid-high top is represented by the category with the largest value between c9 to c12 (if not all zero).

Mid_low_profile = the mid-low profile follows that with the largest value between c13 and c14 (if not both zero).

Low_cloud_profile = c15, if not zero.

In short, there are up to 4 cloud profiles used in the RT calculations, but with one profile associated with each of the original high, mid-high, mid-low and low categories. The cloud fraction assigned to each of these 4 categories remains that of the original category (S1 to S4). The overlap values c1 to c15 are only used to select a single profile for each of the four categories. Note the cloud-base associated with each profile is taken from the lowest layer.

Last, but not least, if the retrieved optical depth associated with any of the original four categories (high, mid-high, mid-low, low) is less than six, the overlap is ignored. Meaning the vertical profile of the cloud is assumed to have a cloud base equal to that of the original group.

So, for example, suppose we have a scenario in which S1 = 0, S2 = 0, S3=0.4 and S4=0.4 (meaning no high or mid-high clouds only mid-low and low clouds are present), with a cloud base for layer 3 (CB3) of 850 hPa and for layer 4 (CB4) of 780 hPA, and a cloud optical depth for layer 3 (OD3) of 10 and layer 4 (OD3) of 3. Yes, it is possible for CB3 to be lower (closer to the surface) than CB4 (each are retrieved independently).

In this scenario, (following Kato 2019 equations B1 to B4) one obtains C1 = C2 = 0 and C3 = 0.4 and C4 = S4/(1-S3) = ~ 0.67. Consequently, c1 to c8 = 0, and c9 to c12 = 0 and ONLY two cloud profiles of the possible four will be used in the radiative transfer (RT) calculations.

c13 = (1-C1)*(1-C2)*C3*C4 = 0.267
c14 = (1-C1)*(1-C2)*C3*(1-C4) = 0.133
c15 = (1-C1)*(1-C2)*(1-C3)*C4 = 0.4

Since c13 is larger than c14 and OD3 is larger than 6, profile c13 will be used for the mid-low category in the RT calculations with a total area covered by the c13 profile set to 0.4 (S3), with a cloud base set to CB4 or 780 hPA, and a optical depth of 10 (OD3). If OD3 had been 5, then the overlap would be ignored (equivalent in this case selecting profile c14), with a resulting cloud fraction of 0.4 (S3), cloud base at 850 hPA (CB3) and the same optical depth 10 (OD3). The second profile used in the RT calculation would be a single-layer low cloud with a cloud fraction of 0.4 (S4), with a base at 780 hPA (CB4) and an optical depth of 3 (OD4).

References

Andreas, Afshin, Dooraghi, Mike, Habte, Aron, Kutchenreiter, Mark, Reda, Ibrahim, and Sengupta, Manajit, 2018: Solar Infrared Radiation Station (SIRS), Sky Radiation (SKYRAD), Ground Radiation (GNDRAD), and Broadband Radiometer Station (BRS) Instrument Handbook, doi:10.2172/1432706.

Ceppi, P., Y.-T. Hwang, D. M. W. Frierson, and D. L. Hartmann, 2012: Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing, *Geophys. Res. Lett.*, 39, L19708, doi:10.1029/2012GL053115.

Ceppi, P., Y.-T. Hwang, X. Liu, D. M. W. Frierson, and D. L. Hartmann (2013), The relationship between the ITCZ and the Southern Hemispheric eddy-driven jet, J. Geophys. Res. Atmos., 118, 5136–5146, doi:10.1002/jgrd.50461.

Deng, M., G. G. Mace, Z. Wang, and R. P. Lawson, 2013: Evaluation of several A-Train ice cloud retrieval products with in situ measurements collected during the SPARTICUS campaign. J. Appl. Meteor. Climatol., 52, 1014–1030

Doelling, D. R., and Coauthors, 2013: Geostationary enhanced temporal interpolation for CERES flux products. J. Atmos. Oceanic Technol., 30, 1072–1090, doi:10.1175/JTECH-D-12-00136.1.

Doelling, D. (2017). CERES Level 3 SYN1deg-1Hour Terra-Aqua-MODIS HDF4 file - Edition 4A [Data set]. NASA Langley Atmospheric Science Data Center DAAC. https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A

Frölicher, T.L., J.L. Sarmiento, D.J. Paynter, J.P. Dunne, J.P. Krasting, and M. Winton, 2015: Dominance of the Southern Ocean in Anthropogenic Carbon and Heat Uptake in CMIP5 Models. J. Climate, 28, 862–886, doi: 10.1175/JCLI-D-14-00117.1.

Hande, L. B., S. T. Siems, M. J. Manton, and D. Belusic (2012), Observations of wind shear over the Southern Ocean, J. Geophys. Res., 117, D12206, doi:10.1029/2012JD017488.

Hang, Y, T. S. L'Ecuyer, D. S. Henderson, A. V. Matus, and Z. Wang, 2019: Reassessing the effect of cloud type on Earth's energy balance in the age of active spaceborne observations. Part II: Atmospheric heating. J. Climate, 32, 6219–6236,

Henderson, D. S., T. L'Ecuyer, G. Stephens, P. Partain, and M. Sekiguchi, 2013: A multisensor perspective on the radiative impacts of clouds and aerosols. J. Appl. Meteor. Climatol., 52, 853–871

Hinkelman, L. M., K. E. Lapo, N. C. Cristea, and J. D. Lundquist, 2015: Use of CERES SYN surface irradiance data as forcing for snowmelt simulation in complex terrain, J. Hydrometeor., 16, 2133-2152, doi:10.1175/JHM-D-14-0179.1.

Hwang, Y.-T., and D.M.M. Frierson, 2013: Link between the double-Intertropical Convergence Zone problem and cloud biases over the Southern Ocean. Proc. Natl. Acad. Sci., 110, 4935-4940. doi: 10.1073/pnas.1213302110

Kato, S., S. Sun-Mack, W. F. Miller, F. G. Rose, Y. Chen, P. Minnis, and B. A. Wielicki, 2010: Relationships among cloud occurrence frequency, overlap, and effective thickness derived from CALIPSO and CloudSat merged cloud vertical profiles. J. Geophys. Res., 115

Kato, S., et al., 2011: Improvements of top-of-atmosphere and surface irradiance computations with CALIPSO-,CloudSat-, and MODIS-derived cloud and aerosol properties, *J. Geophys. Res.*, 116, D19209, doi:10.1029/2011JD016050.

Kato, S., et al., 2018: Surface irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product, *J. Climate*, 31, 4501- 4527, doi:10.1175/JCLI-D-17-0523.1.

Kato, S., Rose, F. G., Ham, S. H., Rutan, D. A., Radkevich, A., Caldwell, T. E., Sun-Mack, S., Miller, W.F., and Chen, Y., 2019: Radiative heating rates computed with clouds derived from satellite-based passive and active sensors and their effects on generation of available potential energy. Journal of Geophysical Research: Atmospheres, 124, 1720–1740 doi:10.1029/2018JD028878.

Kay, J.E., C. Wall, V. Yettella, B. Medeiros, C. Hannay, P. Caldwell, and C. Bitz, 2016: Global Climate Impacts of Fixing the Southern Ocean Shortwave Radiation Bias in the Community Earth System Model (CESM). J. Climate, 29, 4617–4636, doi: 10.1175/JCLI-D-15-0358.1

L'Ecuyer, T. S., Y. Hang, A. V. Matus, and Z. Wang, 2019: Reassessing the effect of cloud type on Earth's energy balance in the age of active spaceborne observations. Part I: Top of atmosphere and surface. J. Climate, 32, 6219–6236

Loeb, N.G., D.R. Doelling, H. Wang, W. Su, C. Nguyen, J.G. Corbett, L. Liang, C. Mitrescu, F.G. Rose, and S. Kato, 2018: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. J. Climate, 31, 895–918, doi: 10.1175/JCLI-D-17-0208.1.

Loeb, N. (2017). CERES Level 3B EBAF-Surface Terra+Aqua netCDF file - Edition 4.0 [Data set]. NASA Langley Atmospheric Science Data Center DAAC. https://doi.org/10.5067/TERRA+AQUA/CERES/EBAF-SURFACE_L3B004.0

Ma, Q., K. Wang, and M. Wild, 2015: Impact of geolocations of validationdata on the evaluation of surfaceincident shortwave radiation from Earth System Models, *J. Geophys. Res. Atmos.,*120, 6825–6844, doi:10.1002/2014JD022572.

Mace, G. G. (2010), Cloud properties and radiative forcing over the maritime storm tracks of the Southern Ocean and North Atlantic derived from A ‐ Train, J. Geophys. Res., 115, D10201, doi:10.1029/2009JD012517.

Matus, A. V., and T. S. L'Ecuyer, 2017: The role of cloud phase in earth's radiation budget. J. Geophys. Res. Atmos., 122, 2559–2578

McIlhattan, E. A., T. S. L'Ecuyer, and N. B. Miller, 2017: Observational evidence linking Arctic supercooled liquid cloud biases in CESM to snowfall processes. J. Climate, 30, 4477–4495

Minnis, P., et al., 2011: CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part I: Algorithms, *IEEE Trans. Geosci. Remote Sens.*, 49, 4374-4400, doi:10.1109/TGRS.2011.2144601.

Riihelä, A., J. R. Key, J. F. Meirink, P. Kuipers Munneke, T. Palo, and K.-G. Karlsson, 2017: An intercomparison and validation of satellite-based surface radiative energy flux estimates over the Arctic, *J. Geophys. Res*, 122, 4829–4848, doi:10.1002/2016JD026443.

Rutan, D. A., S. Kato, D. R. Doelling, F. G. Rose, L. T. Nguyen, T. E. Caldwell, and N. G. Loeb, 2015: CERES synoptic product: Methodology and validation of surface radiant flux. *J. Atmos. Oceanic Technol.*, 32, 1121–1143, doi:10.1175/JTECH-D-14-00165.1.

Rutan, D., Rose, F., Smith, B., Jr., and S. Kato, 2018: Southern Ocean surface meteorology and radiative flux, past observations and comparison to CERES SYN1deg data product, paper presented at 2018 Joint ARM User Facility and ASR PI meeting, Department of Energy, Tysons, VA. Available at: https://asr.science.energy.gov/meetings/stm/posters/poster_pdf/2018/P001948.pdf

Sallée, J.-B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, T. J. Bracegirdle, Z. Wang, and T. Roy (2013), Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response, J. Geophys. Res. Oceans, 118, 1830–1844, doi:10.1002/jgrc.20135.

Schneider, D.P. and D.B. Reusch, 2016: Antarctic and Southern Ocean Surface Temperatures in CMIP5 Models in the Context of the Surface Energy Budget. J. Climate, 29, 1689–1716, doi: 10.1175/JCLI-D-15-0429.1

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

Wang, Z., S.T. Siems, D. Belusic, M.J. Manton, and Y. Huang, 2015: A Climatology of the Precipitation over the Southern Ocean as Observed at Macquarie Island. J. Appl. Meteor. Climatol., 54, 2321–2337, doi: 10.1175/JAMC-D-14-0211.1.

Zhang, Xiaotong; Liang, Shunlin; Wang, Guoxin; Yao, Yunjun; Jiang, Bo; Cheng, Jie. 2016: Evaluation of the Reanalysis Surface Incident Shortwave Radiation Products from NCEP, ECMWF, GSFC, and JMA Using Satellite and Surface Observations, Remote Sens., 8(3), 225, doi:10.3390/rs8030225.

Zhang, T., P. W. Stackhouse Jr., S. J. Cox, J. C. Mikovitz, and C. N. Long, 2019: Clear-sky shortwave downward flux at the Earth's surface: Ground-based data vs. satellite-based data, *J. Quant. Spectrosc. Rad. Trans.*, 224, 247-260.

Van Tricht, K., and Coauthors, 2016: Clouds enhance Greenland ice sheet meltwater runoff. Nat. Commun., 7, 10266