A Comparison of Top-of-Atmosphere Radiative Fluxes from CERES and ARISE

Patrick Charles Taylor[1], Kyle Frederick Itterly[2], Joe Corbett[2], Anthony Bucholtz[3], Sergio Sejas[4], Wenying Su[4], David R. Doelling[4], and Seiji Kato[4]

[1]National Aeronautics and Space Administration (NASA)
[2]Science Systems and Applications, Inc.
[3]Naval Postgraduate School
[4]NASA Langley Research Center

November 24, 2022

Abstract

Uncertainty in Arctic top-of-atmosphere (TOA) radiative flux observations stems from the low sun angles and the heterogeneous scenes. Advancing our understanding of the Arctic climate system requires improved TOA radiative fluxes. We compare Cloud and Earth's Radiant Energy System (CERES) TOA radiative fluxes with Arctic Radiation-IceBridge Sea and Ice Experiment (ARISE) airborne measurements using two approaches: grid box averages and instantaneously-matched footprints. Both approaches indicate excellent agreement in the longwave and good agreement in the shortwave, within 2 uncertainty considering all error sources (CERES and airborne radiometer calibration, inversion, and sampling). While the SW differences are within 2 uncertainty, both approaches show a ~-10 W m^{-2} average CERES-aircraft flux difference. Investigating the source of this negative difference, we find a substantial sensitivity of the flux differences to the sea ice concentration dataset. Switching from imager-based to passive microwave-based sea ice data in the CERES inversion process reduces the differences in the grid box average fluxes and in the sea ice partly cloudy scene anisotropy in the matched footprints. In the long-term, more accurate sea ice concentration data are needed to reduce CERES TOA SW flux uncertainties. Switching from imager to passive microwave sea ice data, in the short-term, could improve CERES TOA SW fluxes in polar regions, additional testing is required. Our analysis indicates that calibration and sampling uncertainty limit the ability to place strong constraints ($<\pm 7\%$) on CERES TOA fluxes with aircraft measurements.

Hosted file

arisepaper_supplementary material.docx available at https://authorea.com/users/531603/articles/603718-a-comparison-of-top-of-atmosphere-radiative-fluxes-from-ceres-and-arise

A Comparison of Top-of-Atmosphere Radiative Fluxes from CERES and ARISE

Patrick C. Taylor[1,*], Kyle F. Itterly[2], Joe Corbett[2], Anthony Bucholtz[4], Sergio Sejas[2], Wenying Su[1], Dave Doelling[1], and Seiji Kato[1]

[1]NASA Langley Research Center, Hampton, VA, USA

[2]Science Systems and Applications, Inc, Hampton, VA, USA

[3]Naval Postgraduate School, Monterey, CA, USA

*Corresponding author email (Patrick.c.taylor@nasa.gov)

Abstract

Uncertainty in Arctic top-of-atmosphere (TOA) radiative flux observations stems from the low sun angles and the heterogeneous scenes. Advancing our understanding of the Arctic climate system requires improved TOA radiative fluxes. We compare Cloud and Earth's Radiant Energy System (CERES) TOA radiative fluxes with Arctic Radiation-IceBridge Sea and Ice Experiment (ARISE) airborne measurements using two approaches: grid box averages and instantaneously-matched footprints. Both approaches indicate excellent agreement in the longwave and good agreement in the shortwave, within 2σ uncertainty considering all error sources (CERES and airborne radiometer calibration, inversion, and sampling). While the SW differences are within 2σ uncertainty, both approaches show a ~-10 W m^{-2} average CERES-aircraft flux difference. Investigating the source of this negative difference, we find a substantial sensitivity of the flux differences to the sea ice concentration dataset. Switching from imager-based to passive microwave-based sea ice data in the CERES inversion process reduces the differences in the grid box average fluxes and in the sea ice partly cloudy scene anisotropy in the matched footprints. In the long-term, more accurate sea ice concentration data are needed to reduce CERES TOA SW flux uncertainties. Switching from imager to passive microwave sea ice data, in the short-term, could improve CERES TOA SW fluxes in polar regions, additional testing is required. Our analysis indicates that calibration and sampling uncertainty limit the ability to place strong constraints (<±7%) on CERES TOA fluxes with aircraft measurements.

1. Introduction

The Arctic is one of the most rapidly changing regions on the planet. The ongoing changes span the full complement of Arctic climate sub-systems: the atmosphere, ocean, cryosphere, land, and ecosystems (e.g., Taylor et al. 2017). The energy exchanges between these subsystems may also be changing, affecting Arctic climate system evolution. Thus, measuring the energy flows is critical for advancing our understanding of the Arctic climate system by enabling the diagnosis of the factors driving system change. Using available observations and meteorological reanalysis output, recent research indicates that the energy flows within the Arctic climate system have changed (e.g., Riihelä et al. 2013; Duncan et al. 2020) and will continue to change (e.g., Boeke et al. 2021). However, obtaining accurate energy flux data is a challenge.

Top-of-atmosphere (TOA) energy budget data from the Clouds and Earth's Radiant Energy System (CERES; Wielicki et al. 1996; Loeb et al. 2018) has been instrumental in quantifying Arctic energy changes (e.g., Riihelä et al. 2013; Kay and L'Ecuyer 2013; Duncan et al. 2020). Six CERES instruments onboard the Terra, Aqua, Soumi-NPP, and NOAA-20 polar orbiting satellites have provided the most spatially and temporally complete record of global shortwave (SW) and longwave (LW) TOA radiative fluxes beginning in 2000. The CERES record has enabled many advances in Arctic climate science including quantifying the sea ice albedo feedback (e.g., Pistone et al. 2014) and evaluating contemporary climate models (e.g., Boeke and Taylor 2016; Wei et al. 2021). However, CERES radiative fluxes are most uncertain in polar regions (Kato et al. 2013; Su et al. 2015a,b). Some studies suggest the possibility of biases in reflected CERES SW Arctic fluxes and surface albedo relative to *in situ* data (Riihela et al. 2017; Huang et al. 2022). It is important to understand these sources of uncertainty to improve the radiation budget record.

An understanding of the methodology used to generate TOA fluxes from CERES observations is critical to formulating investigations of the uncertainty. CERES instruments do not measure radiative flux; rather, CERES instruments measure broadband radiances that are inverted to determine radiative fluxes. CERES calibrated radiances are first corrected for the effects of the instrument optical train (called "spectral unfiltering"; Loeb et al. 2001). Next, TOA fluxes are obtained by applying a scene-dependent radiance-to-flux inversion algorithm to the unfiltered radiances, called angular distribution models (ADMs). ADMs are constructed using empirical and theoretical approaches (Su et al. 2015a) and defined for many scene types. The scene type is determined using atmospheric state information from data assimilation and surface and cloud properties from other satellite instruments (Su et al. 2015a,b). The heterogeneous mixture of clouds, sea ice, and ocean within the Arctic region makes radiance-to-flux inversion and scene type identification especially challenging increasing TOA flux uncertainty (Su et al. 2015b).

There has been a lack of data available to evaluate CERES TOA fluxes in the Arctic. Prior validation efforts over the polar regions have relied on indirect methods (Su et. al. 2015b) that only determine uncertainties in the radiance-to-flux inversion process, not the absolute uncertainty. Additionally, past comparisons of CERES against *in situ* data have primarily used surface-based measurements (e.g., Rutan et al. 2015; Riihelä et al. 2017; Huang et al. 2022) making it challenging to draw conclusions about TOA fluxes. The lack of other broadband instruments in a polar orbit prevents the inter-comparison with other sensors over the Arctic (and Antarctic) that would provide a better evaluation of absolute uncertainty under polar conditions.

We leverage a unique opportunity to evaluate CERES data against *in situ* aircraft observations in the Arctic to address this gap. A similar approach was attempted at mid-latitudes during the Atmospheric Radiation Measurement Enhanced Shortwave Experiment (ARESE) campaign comparing GOES-8 derived broadband albedo with measurements using a radiometer onboard the NASA ER-2 (Pope et al. 2002; Valero et al. 2003). They found agreement between GOES-8 and the radiometer to within measurement uncertainty. The 2001 Chesapeake Lighthouse and Aircraft Measurements for Satellites (CLAMS) experiment off the U.S. East Coast was also designed to test the inputs and data products from instruments onboard NASA's Terra spacecraft (Smith et al. 2005). Evaluation of CERES irradiances, however, has been confined to comparisons between low altitude aircraft and surface irradiance measurements with the CERES computed irradiance dataset (Charlock 2004). During the ARISE Campaign (Smith et al. 2017; Section 2), a unique, statistical sampling strategy was employed to evaluate CERES time-averaged TOA upwelling LW and SW fluxes. This study reports on the results of these CERES-ARISE flux comparisons.

2. Background: ARISE Campaign

The Arctic Radiation-IceBridge Sea and Ice Experiment (ARISE) was a joint mission between the radiation, cloud microphysics, and cryosphere communities. The general aim was to measure the SW and LW radiation while characterizing the atmospheric state, clouds, and sea ice conditions. Complete details of the ARISE campaign can be found in Smith et. al. (2017).

Briefly, the NASA Wallops C-130 was outfitted with a suite of radiation measuring instruments, cloud probes, and the Land, Vegetation, and Ice Sensor (LVIS) used during NASA Operation IceBridge. The radiation suite consisted of a set of upward and downward looking pyranometers and pyrgeometers, an upward and downward looking solar spectral flux radiometer (SSFR), and a sun-tracking photometer (4STAR). In all, 15 scientific flights were performed, extensively sampling the Beaufort Sea region of the Arctic Ocean. Generally, the flight paths were pre-planned and designed to maximize the sampling over a variety of conditions and intersect with polar orbiting satellite tracks. A specific aim was to perform *in situ* validation of the CERES TOA irradiances using the broadband radiometers on the aircraft.

3. Data and Methodology

a. CERES

Most CERES instruments consist of a broadband scanning radiometer measuring the radiance in a SW channel (0.3-5μm), a total channel (0.3-100μm), and a window channel (8-12μm). For the instrument aboard NOAA-20, the window channel is replaced by a LW channel (5-50μm). The LW radiance for the Flight Model 1 (FM1) through FM5 instruments is derived from the difference between the total and SW channels. The nominal footprint size at nadir is ~20 km^2, expanding to 70x50km at oblique scan angles. Irradiances are derived from measured radiances using scene type dependent ADMs (Wielicki et al. 1996). Scene types are classified by a combination of the surface type (e.g., land type, snow-covered, ocean, or sea ice) and cloud properties. ADM scene identification additionally depends on cloud phase and optical depth in the SW and on surface skin temperature and surface-cloud temperature difference in the LW. Cloud information is from co-incident imagers—the MODIS instrument on Terra and Aqua and VIIRS on Soumi-NPP and NOAA-20 (Trepte et al. 2019; Minnis et al. 2020). Skin temperature data is from the imager retrieval for clear-sky and the GEOS 5.4.1 meteorological analysis for cloudy scenes. Land type is from the International Geosphere-Biosphere Programme (IGBP) surface classification (Loveland and Belward 1997; Belward et al. 1999) and snow and sea ice coverage from the National Snow and Ice Data Center's Near-Real Time Snow and Ice Extent (NISE) dataset and an imager-derived snow and sea ice concentration (SIC) product (Su et al. 2015a).

CERES TOA flux measurement uncertainty originates from two main sources: calibration and radiance-to-flux inversion. Calibration uncertainty is 1.0% in the SW and 0.5% in the LW

(Wielicki et al. 1996; Loeb et al. 2008). CERES calibration uncertainty and stability is verified using a suite of methods including an onboard blackbody source and cold space views in the LW and a tungsten lamp and solar and lunar calibration looks in the SW (Wielicki et al. 1996; Loeb et al. 2016). The multi-pronged approach for CERES calibration verification includes intercomparisons with other space instruments. The result is a set of stable CERES instruments all referenced to FM1 with calibration drift less than ±0.5 Wm^{-2} $decade^{-1}$, a factor of 3-4 better than anticipated (Loeb et al. 2007; Loeb et al. 2012)

The CERES radiance-to-flux inversion algorithm is the second primary source of uncertainty. Su et al. (2015a) provides a description of the Edition 4 ADMs and Su et al. (2015b) evaluates the inversion procedure. Briefly, ADMs are constructed from CERES measurements between 2000 and 2005 when one of the instruments was operating in rotating-azimuth-plane mode. This mode allows the instruments to observe a given scene type from a wide range of viewing geometries. The data are then composited into solar zenith, viewing zenith, and relative azimuth angle bins and further by scene type to empirically determine the anisotropic factors for inversion. The LW radiance-to-flux inversion procedure mainly accounts for limb-darkening effects and has a negligible dependence on solar zenith angle (SZA) and relative azimuth angle (Loeb et al. 2003; Loeb et al. 2005). In the SW, however, the anisotropy has a large dependence on viewing geometry and scene type.

For the ADM uncertainties in this study, the most suitable are the Single Scanner Footprint – MISR flux consistency root mean square error values converted into a flux uncertainty for the SW and the CERES-MODIS flux consistency for the LW. This is done by multiplying the flux consistency values by 0.6 (Su et al. 2015b). All-sky ADM uncertainty values are used for ocean and sea ice with two modifications. First, the value for the ocean scenes in Su et al. (2015b; 1.14%)

is derived using global measurements. To better estimate polar region uncertainty, calculations are repeated using only ocean measurements poleward of 60°N/S resulting in an uncertainty estimate of 3.8% for Polar Oceans. Secondly, Su et al. (2015b) reports the uncertainty for combined snow and ice scenes (5%). To estimate uncertainty for sea ice-only scenes, we apply the same methodology to sea ice-only scenes resulting in an uncertainty of 4.8%. For the LW, the values provided in Su et al (2015b) are used: 1.14% for snow/ice scenes and 1.5% for ocean scenes.

Scene identification errors are a critical aspect of uncertainty in the radiance-to-flux inversion process. Scene identification errors impact ADM uncertainty in two ways: (1) during development by determining which footprints are aggregated to build the ADMs and (2) during ADM application by determining which anisotropic factor is used to compute the flux. With respect to ARISE, scene identification errors stem from errors in cloud properties and the SIC data. Cloud retrieval errors influence ADMs in complex ways due to the different VZA dependencies of the anisotropic factor by cloud type. Thus, scene misidentification due to the cloud retrieval errors can result in an over or an underestimation of the TOA fluxes depending on VZA (Su et al. 2015b). In addition, Su et al. (2015b) found that the ADM impacts of cloud property retrieval related scene identification errors are strongest in sea ice regions.

CERES data streams are distinguished by the level of processing. CERES Level 2 refers to the single scanner footprint (SSF) data and includes ADM inverted footprint radiative fluxes. CERES level 3 refers to the Synoptic (SYN) product that applies time and space averaging procedures to put the FM instrument operating in cross-track scan mode SSF data onto a 1°x1° grid (Doelling et al. 2013). Edition 4a (Ed4a) SSF and SYN data are used. The SSF1deg level 3 are single satellite gridded instantaneous products. Over polar regions, the SYN1deg product combines the Terra and Aqua SSF instantaneous gridded observations and temporally interpolates and averages the fluxes into hourly GMT intervals. The SSF1deg Edition 4a (Ed4a) and SYN1deg Ed4a data are used in the gridbox comparison while the instantaneous footprint comparison utilizes the SSF Ed4a level 2 data from the cross-track and programmable azimuthal plane scan (PAPS) mode.

b. Airborne Broadband Radiometers (BBR)

Broadband Radiometers (BBRs) mounted on the top and bottom of the aircraft during ARISE measured the down- and upwelling global solar (SW) irradiance (0.2–3.6 μm); and the down- and upwelling infrared (LW) irradiance (4.5–42 μm) (Smith et. al. 2017). These BBRs were Kipp and Zonen CM-22 pyranometers (Kipp and Zonen 2004) and CG-4 pyrgeometers (Kipp and Zonen 2001), modified for aircraft use (Bucholtz et al. 2010). The modifications included new sealed back housings to prevent condensation and freezing inside the domes, with the connector on the bottom of the housing for easier aircraft mounting. The new housings retained the front-end optics and electronics of the original instruments but allowed an amplifier to be mounted directly below the sensor. With the voltage signal amplified at the sensor, the instruments were operated in current loop mode to minimize electronic noise.

The BBR radiometers were calibrated pre- and post-mission. The SW radiometer calibration was performed using the standard alternating sunshade method (ASTM 2005), where the given sensor is compared to the true direct solar irradiance measured by an Eppley automatic Hickey–Frieden absolute cavity radiometer. The sensitivities for the SW radiometers from pre- and post-mission calibrations agreed to within 1%. The LW radiometers were calibrated using a blackbody immersed in a variable temperature alcohol bath. The calibration coefficients for the LW radiometers from pre- and post-mission calibrations agreed to within 2%. Thus, the stability of the SW and LW radiometers during ARISE was excellent. The total BBR calibration uncertainty is estimated as 4% in the SW and 5% in the LW.

Even before modification, the Kipp and Zonen radiometers have features that make them attractive for aircraft use and for use in the Arctic. For example, the CM-22s use 4 mm quartz domes with high thermal conductivity and good thermal coupling to the body to minimize the zero offset problem associated with these types of detectors (Kipp and Zonen 2003a; Ji and Tsay 2000). They also have a low tilt response (<0.2%) for tilt angles up to 90^{o}, meaning there is almost no change in signal when tilted.

The CG-4s use a silicon dome that has a solar blind filter and a meniscus shape with a full 180^{o} field-of-view with a good cosine response. Due to the construction methods used, any solar radiation absorbed by the window is effectively conducted away, allowing accurate measurements in full sunlight, and eliminating the need for a shading disk. In addition, excellent dome to body thermal coupling eliminates the need for a dome thermistor and the calculation of the dome to body temperature offset that is required by other pyrgeometers (Kipp & Zonen 2003b; Philipona et al. 1995). The CG-4s also have a low tilt angle response.

For Arctic work, a key feature of Kipp and Zonen radiometers is their low temperature response. Specifically, both the CM-22 (Kipp and Zonen 2004) and the CG-4 (Kipp and Zonen 2001) have an applied thermistor compensation circuit, optimized for each sensor that suppresses the dependence of the sensor sensitivity to the temperature of the instrument giving them a low temperature response (CM-22 < 0.5%, CG-4 < 1%) over the temperature range of -20^{o}C to +50^{o}C, as confirmed by Su et al. (2008). They also have an internal temperature sensor for the body of the instrument to correct for temperature effects in post-processing, if required.

During the portions of the ARISE flights used in this paper, the temperatures of the BBRs at altitude never fell below -25^{o}C and were typically -20^{o}C or higher. They, therefore, stayed within the range of the temperature compensation circuitry and no temperature adjustments to the sensitivities determined from the pre- and post-mission calibrations were required. This is further shown by Chen et al. (2019) where a comparison was made between the SW measurements from BBR and the Solar Spectral Flux Radiometer (SSFR) that also flew on the aircraft for ARISE. The SSFR is a moderate resolution flux (irradiance) spectrometer (Pilewskie et al. 2003) whose sensitivity is not temperature dependent. Chen et al. (2019) compared the upwelling and downwelling SW measurements from BBR and SSFR for an above cloud case on 11 September 2014 and a below-cloud case on 13 September 2014. Excellent agreement (within a few W m^{-2}) was found between BBR and SSFR SW fluxes for these cases.

c. Comparison Methodology

CERES fluxes are determined at the satellite level and then set to a 20 km reference level (Loeb et al. 2002). To provide an accurate comparison between the BBR measurements at the aircraft level and the CERES measurements, the scattering and absorption of the radiation above the aircraft must be considered. To do this, we use the Langley Fu-Liou radiative transfer code (Fu and Liou 1993; Fu et al. 1998; Kratz and Rose 1999; Kato et al. 1999; 2005; Rose et al. 2013). This code requires inputs of cloud fraction, top and bottom heights, phase, particle size, and optical depth as well as atmospheric profiles of temperature, water vapor, and ozone, and surface skin temperature and surface type information to determine spectral albedo and emissivity. The atmospheric profile information in this study comes from the Goddard Earth Observing System (GEOS) Reanalysis version 5.4.1 (Suarez et al. 2005), currently used in CERES data production. Cloud properties are produced using MODIS and VIIRS data by the CERES Cloud Working Group (Trepte et al. 2019; Minnis et al. 2020). Surface type is determined by the ASI AMSR2 SIC (Spreen et al. 2008; Melsheimer and Spreen 2019) dataset—a high spatial and temporal resolution passive microwave derived dataset.

The process for adjusting BBR measurements to their TOA equivalent is as follows:

1) Match the BBR measurement in time and space with the GEOS atmospheric profile.
2) Determine MODIS cloud properties and SIC by spatially collocating the BBR measurement within the MODIS swath from the overpass closest in time. Average cloud properties observed by BBR are determined using a solid angle weighting of all the pixels.
3) The radiative transfer code is run and the upward LW and SW flux profiles are stored. $BBR_{TOA}(\lambda)$, where λ represents the LW or SW flux, is calculated as: $BBR_{TOA}(\lambda) = BBR(\lambda) * (F(\lambda, TOA)/F(\lambda, P_{BBR}))$ (1), where P_{BBR} is the aircraft pressure level.
4) BBR_{TOA} values are set to the 20 km CERES reference level using the inverse square law.
5) BBR_{TOA} and SSF SW level 2 footprint fluxes are adjusted to the SYN1deg gridbox hourly-integrated SZA to control for SZA differences as: $BBR_{TOA,adj}(SW) = BBR_{TOA}(SW) * (\cos(SZA_{SYN1deg})/\cos(SZA_{BBR}))$ (2), where $BBR_{TOA,adj}(SW)$ is the SZA-weighted flux. Same equation is applied to the SSF1deg SW fluxes.

4. Results

a. Grid box comparison

A series of grid box experiments (Fig. 1) captured the spatial and temporal variation of the TOA upwelling fluxes at scales comparable to the CERES gridded products (1°x1°). These grid boxes consist of flying five legs of ~200km in length spaced ~20 km apart. These legs bound an area of 200 x 80 km but due to the hemispheric nature of the irradiance measurements a slightly larger area was sampled. In the analysis, these large boxes are split making two ~100x80 km boxes roughly the size of a 1°x1° grid box at the equator. Three of these grid boxes were attempted, one over marginal sea ice (on 9/7/2014), one over a region of high SIC (on 9/11/2014), and one over ice-free ocean (on 9/15/2014). The planned grid boxes on the 7^{th} and 11^{th} were executed successfully. On the 15th, high clouds occurred in the planned flight area and prevented the aircraft from sampling above cloud. Instead, a different, smaller grid box ~100 x 100 km was flown to the southeast. Figure 1 shows the grid box positions.

The surface and cloud properties of the five grid boxes varied substantially. The grid boxes are pictured in Fig. 2 as the yellow boxes overlaid on a true color MODIS image. The true color images demonstrate the cloud field differences between flight days as well as the evolution of the scene over the ~2-hour sampling window. The average properties and standard deviations within each grid box are shown in Table 1 and scene property box and whisker plots for each CERES instrument and the MODIS pixels found within 20 km of the aircraft flight track in Fig. 3.

The grid boxes on September 7th (GB071 and GB072; Fig. 2a) are both located within the marginal ice zone (MIZ) and contain moderately thick, overcast low clouds (Table 1; Fig. 3). GB072 had a higher NISE SIC (17%) than GB071 (10%). The SIC distribution (Fig. 3) indicates similar ranges within the two grid boxes; the median of the SIC distribution in GB071 is between ~5% whereas for GB072 it is between 15-20%. Both grid boxes were overcast for the duration of the sampling time, however, GB071 had a lower average MODIS-retrieved cloud optical depth than GB072 (7 and 10, respectively). The distribution of cloud optical depth (Fig. 3) between GB071 and GB072 differed substantially where GB071 contained a larger number of lower cloud optical depths (< 3) and GB072 a larger number of higher cloud optical depth values (>10), accounting for the large SW flux differences between the grid boxes. The clouds were low with cloud top pressures 868 hPa in GB071 and 904 hPa in GB072. In addition, the cloud conditions changed slightly over the sampling period in each grid box as thicker clouds moved northwest from GB072 to GB071 (Fig. 2). The effect of the sampling differences and scene changes on the comparison are described and quantified in Section 5a.

The grid boxes on September 11th (GB111 and GB112; Fig. 2b) are located over a region of high SIC (86% and 77%, respectively). The SIC depends on the dataset used. The ASI-AMSR2 (not shown) dataset has concentrations near 100% for both grid-boxes, whereas the NISE dataset has a lower value for each grid box. CERES imager-based clear-sky weighted SIC shows higher values than NISE. Each grid box was partly cloudy with GB111 having a smaller cloud fraction (43% vs. 84%) especially as sampling progressed. The larger cloud fraction in GB112 results from the larger number of overcast footprints (Fig. 3). The clouds in both grid boxes were low with cloud top pressures of 829 and 872 hPa. The mean MODIS-retrieved cloud optical depths are 4 and 9 in GB111 and GB112, respectively.

The grid box on September 15th (GB151; Fig. 2c) is located over a mostly ice-free region with a small amount of sea ice in the western corner. GB151 contained high, multi-layer (average cloud top pressure, 594 hPa), and thick clouds with a mean optical depth of 38 (Table 1; Fig. 3). The distribution of SIC and cloud fraction within GB151 is homogeneous (Fig. 3q,r), whereas the cloud top pressure and cloud optical depth showed greater heterogeneity (Fig. 3s,t). In addition, the scene on this day did not change appreciably as the sampling progressed.

Figure 4 summarizes the grid box mean SW and LW fluxes (Table 1 and Fig. 4a,c) from CERES SSF1deg, CERES SYN1deg, and BBR_{TOA} and the CERES-BBR differences in Fig. 4c,d. In the LW, the differences between CERES SSF1deg, CERES SYN1deg, and BBR_{TOA} are small, less than ±10 Wm^{-2} for all comparisons and within the 2σ uncertainty (combined calibration + inversion) for all five grid boxes. The average SSF1deg-BBR difference in the LW flux is +2.5 W m^{-2}. The SYN1deg-BBR differences are similar to SSF1deg-BBR with some slight deviations due to spatial mismatching arising from the gridded and temporally averaged nature of SYN1deg. Figure 5 shows the box and whisker plots for SSF1deg and BBR_{TOA} fluxes for each grid box. Comparing individual CERES instruments, it is evident that the flux distributions can differ substantially indicating a sensitivity of the fluxes to grid box sampling. Discussed further below, the differences in the FM1 and FM2 box and whisker plots result from the increased number of footprints within the grid box for FM2 from the use of the PAPS mode to "stare" at the grid boxes and the increased footprint size due to larger VZAs; FM2 samples a larger area than FM1. Hourly fluxes and scene properties are summarized in Fig. S1 and Tables S1-S3.

Despite the strong agreement in the LW, the SW results suggest a ~-10 Wm^{-2} difference in four of the five grid box experiments; the average differences between SSF1deg and BBR across all experiments is -13.0 W m^{-2}. Of the five experiments, agreement was within the combined 2σ total calibration+inversion uncertainty for three (GB071, GB112, and GB151). These grid boxes represent a wide range of scenes: overcast MIZ, partly cloudy sea ice, overcast sea ice, and overcast ocean. GB072 and GB111, overcast MIZ scene and partly cloudy sea ices scenes, show mean SSF1deg-BBR differences larger than the 2σ total calibration+inversion uncertainty. Figure 5 shows the distribution of SW flux observations for CERES and BBR indicating that the BBR fluxes are shifted to slightly higher SW values. Figure 4 also shows the mean flux values from the hourly averaged fluxes, 1°x1° gridded fluxes from SYN1deg showing similar results to the SSF1deg fluxes.

There are two key takeaways from the grid box comparison. First, we find excellent agreement between CERES and BBR in the LW. Second, we find agreement for most grid boxes within 2σ uncertainty between CERES and BBR in the SW, however the results show a large negative difference (~-10 W m^{-2}) for four of the five experiments and is consistent with previous results (Rihella et al. 2017; Huang et al. 2022). While this bias is within the assessed uncertainty for four-of-five grid boxes, the nature of this negative difference warrants further investigation.

b. Instantaneous footprint comparison

A second approach to comparing the CERES SSF instantaneous footprint and BBR fluxes. The SSF footprints are collocated in space by a nearest neighbor finding the minimum Haversine distance and within ±15 minutes of the one-minute averaged BBR measurements. One-minute average BBR measurements assuming a flight speed of 140 m s^{-1} yield a track length of 8.4 km and a "footprint size" of ~24 km. The matched measurements are limited to periods when the aircraft is above the highest cloud layer; this reduces error caused by incorrect cloud properties. The benefit of the instantaneous approach, relative to the grid box approach, is that sampling differences are minimized. The downside of the instantaneous approach is that the reduced sampling increases noise.

This instantaneous matching approach yields 39 footprints over a variety of scene types. Figure 6 shows the summary of the instantaneous comparison between BBR_{TOA} and SSF SW (Fig. 6a,c) and LW (Fig. 6b,d) fluxes. High correlation is found between the SSF and BBR LW and SW fluxes. The mean SSF-BBR_{TOA} difference in the SW is -7.7 Wm^{-2} (-3.6%) and -0.6 Wm^{-2} (0.3%) in the LW. The results in Fig. 6a indicate a dependence on the SW flux magnitude such that brighter scenes show larger negative differences. Figure 6c,d and Table 2 suggest a spatial dependence of the SSF-BBR differences corresponding to surface type, although positive and negative differences are found over low and high SIC regions.

Grouping SSF footprints into common scene types, Table 2 demonstrates a scene dependence of the SSF-BBR differences. The largest SW differences are found over sea ice partly cloudy scenes, an SSF-BBR difference of -17.1 Wm^{-2} (Table 2). This is not surprising due to the dependence on the imager-based SIC, which tends to overestimate SIC and has increased uncertainty for cloudier conditions. However, these differences are not statistically robust due to the small sample size (N=9). In-atmosphere 3D effects are unaccounted for in this comparison and likely to be largest over partly cloudy scenes for SW fluxes (Ham et al. 2014). Both ocean cloudy and sea ice overcast, two prevalent Arctic scene types, show better agreement with differences of -0.8 Wm^{-2} and -9.1 Wm^{-2}, respectively. Results for these scene types are also limited by sample size. As with the grid box experiments, the agreement in the LW is better, with -1.9 Wm^{-2} over ocean cloudy and -0.9 Wm^{-2} for sea ice overcast.

A concern with the instantaneous approach is that the CERES field-of-view varies with VZA, increasing from ~20 km^2 at nadir to 70x50 km at 70°. Using MODIS radiances, we investigate this effect by selecting the pixels surrounding the matched aircraft location with an increasing area from a 4x4 km region to a 40x40 km region. We apply a narrowband-to-broadband scheme (Doelling et al. 2013; 2016) to calculate the broadband radiance and then apply the CERES inversion (called CERES-like fluxes). This provides CERES-like fluxes at different spatial scales with a footprint size independent of VZA. The results show a small effect of the variable CERES footprint size on the comparison. Figure 6a,b illustrates this for a representative set of results for a 20x20 km region (CERES-like) around the CERES-BBR collocation point. Moreover, no statistically significant correlation is found between the CERES-BBR differences and VZA, further evidence that footprint size differences are not strongly influencing the results.

In summary, the takeaways from the instantaneous comparison match those from the grid box comparison. First, better agreement is found between the CERES and BBR fluxes in the LW than the SW. Second, scene-type dependent CERES-BBR flux differences are found in both the LW and SW, although larger in the SW. Surprisingly, the mean differences between CERES and ARISE are similar for the grid box and instantaneous comparisons. Given the larger and negative average CERES-BBR difference in the SW that is consistent with previous results, we describe additional analyses to understand the potential contributions. For the remainder of this study, the comparisons will focus on SSF-BBR differences.

5. Discussion of CERES-BBR Flux differences

a. Sampling

In Section 4a, we compare grid box averaged fluxes from CERES overpasses that occur within ±15 minutes of the BBR measurements made during a ~two-hour aircraft sampling period. Thus, CERES and BBR grid box average fluxes are obtained using different sampling approaches. Average BBR fluxes are computed by averaging the two hours of measurements made while traversing the grid box. Alternatively, CERES grid box average fluxes are computed using nearly instantaneous snapshots from ~4 satellite overpasses during the 2-hour period. Figure 7 illustrates the continuous aircraft sampling vs. the near-instantaneous satellite sampling (vertically stacked symbols). Considering a constant scene, the two sampling approaches would provide the same grid box average flux. In cases of substantial scene changes, the sampling differences result in substantial differences in the grid box mean fluxes and influence the comparison.

We estimate the influence of sampling on the grid box mean flux comparison by leveraging the complete spatial coverage of MODIS. In estimating sampling effects, we compute and compare grid box average fluxes in two ways: (1) sampling CERES-like fluxes at each overpass with aircraft sampling (matching only in space) and (2) sampling CERES-like grid box fluxes with continuous aircraft sampling (matching in space and time). Sampling strategy (1) represents the satellite sampling approach and (2) the aircraft sampling approach. The flux differences between the approach (1) and (2) are only caused by the differences in sampling and are used to provide an estimate of the sampling uncertainty. While CERES-like fluxes are less accurate, they capture the spatial variability of the scene and are appropriate for estimating sampling uncertainty. Figure 8 illustrates the SW and LW CERES-like flux distributions obtained from the different sampling approaches on 9/11/2014.

Estimation of the sampling effect on the CERES-BBR differences relies on comparing the grid box average flux differences between the two approaches. For each grid box there are 3 or 4 Terra and Aqua overpasses (Fig. 7) producing 19 satellite-aircraft sampling flux differences (Table 3). The results indicate that the sampling effects for individual grid boxes range between ±10 W m^{-2} in the SW and the LW. The sampling differences result in both positive and negative values suggesting that sampling is a random effect when considering multiple grid boxes indicated by the small mean sampling difference of 1.64 W m^{-2} in the SW and -0.20 W m^{-2} in the LW. The sampling uncertainty is estimated as the standard deviation across all 19 of the satellite and aircraft sampling flux differences and indicate an uncertainty of ±4.6 W m^{-2} (~1.8%) in the SW and ±3.7 W m^{-2} (~1.7%). Our results suggest that sampling differences provide a substantial influence on the CERES-BBR grid box comparison; after accounting for sampling uncertainty, all CERES-BBR grid box mean SW flux differences agree within 2σ uncertainty (Fig. 4). The analysis also provides evidence that the sampling effect is random, as the mean effect (Table 3) is not statistically different from zero at the 95% confidence level. Thus, sampling differences do not contribute to the negative CERES-BBR difference.

b. Scene ID

The identification of scene properties is necessary for creating the CERES data record, as it drives ADM development and selection. Thus, inaccurate scene identification can cause substantial errors in fluxes (Su et al. 2015a,b). *In situ* cloud properties, a key determinant of the scene, were not measured during the grid box TOA flux validation experiments since the requirement was to be above clouds. In the absence of *in situ* cloud properties, our analysis focuses on the influence of different SIC data sets.

We consider the effects of two different sea ice data sets on CERES fluxes, namely MODIS imager-based SIC (Su et al. 2015a) and NISE (Brodzik and Stewart 2016). The Near-real-time Ice and Snow Extent (NISE) provides daily SIC data and are intended to be the best estimate of current sea ice conditions. The MODIS imager-based SIC uses a collection of visible radiance channels to retrieve SIC in clear pixels (Su et al. 2015a). CERES uses NISE to determine if a footprint contains sea ice and the imager-derived SIC for ADM selection. To assess the influence of sea ice scene identification, we compute the instantaneous and grid box average CERES flux changes that result from using NISE for ADM selection. A caveat with this approach is that ADMs (Su et al. 2015a) were developed using the imager-derived SIC and hence our results may not fully represent the flux changes from using NISE during ADM development.

Figure 6e-f illustrates the changes in the SSF TOA fluxes for the 39 instantaneously matched footprints when NISE replaces the imager SIC. The results show an increase in the SSF SW flux for 9 of the 10 footprints where there was an impact; SW flux changes range from -2.1 to 44.0 Wm^{-2}. LW flux changes are much smaller ranging from -2.0 to 2.9 Wm^{-2}. Stratifying the SW flux changes by scene type indicates a substantial change (~26 W m^{2}) in the SSF-BBR flux differences for sea ice partly cloudy scenes when using NISE (Table 2). The negative bias relative to BBR turns positive for sea ice partly cloudy scenes (Table 2), which is consistent with the tendency of NISE data to underestimate SIC (Kern et al. 2019). Additionally, using NISE SIC data reclassifies three ocean cloudy scenes as sea ice partly cloudy (Table 2). The change in sea ice dataset does not impact the sea ice overcast scenes because SIC is not used in their selection. Thus, the use of the imager-based SIC can account for some of the persistent negative CERES-BBR difference for both sea ice and ocean scenes.

Secondly, we consider the impact of using the NISE SIC data for ADM selection on the grid box average fluxes (Table 4). Average flux values could only be computed for instruments in cross-track mode (e.g., FM1, FM3, and FM5). When NISE replaces the imager-based SIC, most changes in the grid box mean LW and SW fluxes are small, except GB112 where the difference is -12 W m^{-2}. This result brings the GB112 mean SSF1deg SW flux into better agreement with BBR. Further evidence that the SIC data can contribute to the negative difference.

The impact of the NISE SIC data on the GB112 SW flux led us to further investigate the surprisingly small impact on GB111, since both are primarily composed of sea ice partly cloudy scenes. Table 4 shows the grid box mean fluxes for FM1 and FM2 instruments showing a different result for each instrument. The FM1 flux changes are consistent with FM3 and FM5 (not shown). Using NISE SIC data results in a large increase in the FM1 grid box mean SW flux (~31 W m^{-2}) and a decrease in the FM2 SW flux (~6 W m^{-2}) for GB111. These instrument dependent changes largely offset due to the greater number of FM2 footprints within GB111. This may be a unique case, as FM1 and FM2 showed the same sign for the effect of using NISE SIC for GB112. Additional data is required to separate the influence of the SIC data set from the sensitivity to sampling differences. Overall, the analysis illustrates a substantial sensitivity of the CERES fluxes to the SIC data set and provides evidence that the use of the imager-based SIC data is contributing to the negative CERES-BBR SW difference.

c. Angular Distribution Models

In this section, we investigate the influence of ADMs on the CERES-BBR differences. The empirical nature of ADMs means that they provide an average anisotropy for a given scene; the limited samples used to build ADMs means there is a variation of the anisotropy within a defined scene. These ADM characteristics can cause instantaneous fluxes to be substantially higher or lower over a small area due to the high spatial autocorrelation of radiances, such as in this analysis. Over larger areas and longer time periods these variations tend to cancel, providing unbiased fluxes (Su et al. 2015a,b). As a result of these challenges, we take several approaches to investigate the influence of ADMs on the CERES-BBR SW flux differences.

The first approach used to investigate the influence on the SSF-BBR instantaneous flux differences is to compare the anisotropic factors for the instantaneously matched footprints. The 39 instantaneously matched footprints provide pairs of CERES observed radiances and BBR fluxes. The ratio of these terms is used to determine the "perfect" anisotropic factor ($R_{perfect}$), in other words the anisotropic factor that would give an exact CERES and BBR flux match.

$$R_{perfect} = \frac{\pi I}{BBR_{TOA}} \quad (3)$$

where I is the CERES measured unfiltered radiance, BBR_{TOA} is the aircraft measured flux, and $R_{perfect}$ is the perfect anisotropic factor. Figure 9 shows a comparison of the CERES anisotropic factors (R_{CERES}) with the $R_{perfect}$ for the instantaneously matched footprints. The results show a ~0.8 correlation between R_{CERES} and $R_{perfect}$ in the SW, a mean difference of <0.01, and a root mean square error of 0.09. The LW anisotropic factor comparison is shown for completeness indicating a weaker correlation than in the SW; however, the root mean square error and the mean difference are smaller in the LW. The wider range in the LW $R_{perfect}$ values may result from a partial mismatch between SSF and BBR sampling. Figure 9c,d shows the SW and LW CERES-BBR flux difference distribution across all instantaneous matches. The standard deviation of the SW CERES-BBR flux differences is 23.9 W m^{-2}. A caveat with this approach is that calibration differences and differences in footprint area between CERES and BBR influence the results; thus, this assessment of the ADM influence should be treated as qualitative.

Despite this caveat, there is utility in comparing the R_{CERES}-$R_{perfect}$ differences across scene types since the calibration and footprint area differences should be largely independent of scene type. The Fig. 9 inset legend summarizes the R_{CERES} and $R_{perfect}$ anisotropic factors for each scene type, showing that the R_{CERES} is systematically ~0.07 larger than $R_{perfect}$ for the sea ice partly cloudy scenes indicating that the anisotropy differences in these scenes contribute to a lower CERES SW flux and the negative SSF-BBR flux differences. All other scene types show better agreement.

This anisotropy difference is strongly influenced by the sea ice data set. Using NISE for ADM selection changes the R_{CERES} for these sea ice partly scenes from 0.812 to 0.748 (not shown), yielding better agreement with $R_{perfect}$. Only minor changes (<0.01) in the anisotropy occur in the other scene types. This result indicates that the anisotropy represented by the ADMs is robust and that the instantaneous flux uncertainty is mainly due to misidentified scenes.

A second approach to assess the potential influence of ADMs on the SSF-BBR SW flux differences is to evaluate the dependence of SSF SW fluxes on viewing geometry. This approach is motivated by the design of ADMs, which is meant to provide a flux inversion independent of viewing geometry. Under normal operations, the CERES instruments operate in cross track mode (perpendicular to the spacecraft motion direction) scanning from +70° to -70° VZA and only views a specific grid box from a narrow, repeatable set of viewing geometries. During ARISE, FM2 operated in PAPS mode, scanning through a point in each of the grid boxes as it passed within range providing more footprints from a broader set of viewing geometries. This is accomplished by changing the azimuth angle of the instrument scan after each zenith scan and stopped at nadir (θ=0°) during this operation. The FM1 and FM2 scan patterns for GB071 are shown in Fig. 10

illustrating that FM2 samples a larger range of relative azimuth angles from 135° to 180°, whereas FM1 samples only at ~114°. Evident from Fig. 10 is that FM2 obtains more samples of each grid box at higher VZAs than FM1. We analyze the CERES SW flux differences obtained from the varied FM1 and FM2 viewing geometries focusing on VZA to further investigate the influence of ADMs.

The comparison between FM1 and FM2 is influenced by the surface and cloud properties that each instrument observed. To place the influence of observed scene differences into context, we first compare the cloud and surface properties observed by FM1 and FM2 (Table 5; Fig. 3). For GB071 and GB072 FM2 observed slightly larger SIC and cloud optical depth. These scene differences contribute to FM2-FM1 SW flux differences for these grid boxes of 7.1 Wm^{-2} for GB071 and -4.5 W m^{-2} for GB072. For GB111 and GB112, FM1 and FM2 observed similar NISE SIC, cloud fraction, and cloud optical depth values. The average FM2-FM1 SW flux differences are 3.8 and -1.6 W m^{-2} for GB111 and GB112, respectively. Over the four grid boxes, FM2-FM1 differences are both positive and negative suggesting that the viewing geometry and footprint size differences are likely not having a systematic influence on the CERES-BBR differences. The results illustrate a sensitivity of the grid box mean fluxes to scene property differences that results from the combination of sampling differences and scene heterogeneity.

To further investigate the influence of spatial sampling differences on the FM2-FM1 differences, we resample FM1 like FM2. As previously mentioned, the FM1 and FM2 footprint sizes within the grid box are different due to the different VZAs. The footprint size differences could influence the FM2-FM1 flux differences due to scene type differences (cloud properties and sea ice) viewed by each instrument. To evaluate the footprint size influence on the FM2-FM1 comparison, we resample FM1 by including additional footprints with centroids found within the

FM2 sampling area (FM2-like). The FM2-like mean SW fluxes are summarized in Table 5. For GB071 and GB111, FM2-like sampling increases the SW flux relative to FM1, whereas for GB112 there is a flux decrease and no change for GB072 relative to FM1. These differences between FM1 and FM2-like fluxes are generally consistent with the changes in scene properties (Table 5). This mix of increased and decreased fluxes illustrates the influence of scene sampling and heterogeneity on the flux differences since the changes in sampling result in larger flux changes than the FM1-FM2 differences.

The final analysis of the role of ADMs in the CERES-BBR flux differences is to consider the angular dependence of the CERES fluxes within scene types (Fig. 11). To combine all the CERES FM1 and FM2 footprints into a single analysis and evaluate the VZA dependence of the inverted fluxes, we attempt to control for the influence of within grid box and within scene type variability by computing a CERES SW flux anomaly (defined as SSF flux minus the SSF1deg grid box mean flux). Assuming that within grid box scene properties are randomly distributed; thus, given enough footprints at a VZA, the differences in the scene properties should average to zero. This procedure enables the accumulation of all CERES footprints into a common phase space and results in a larger number of footprints.

The results suggest that the CERES SW fluxes show a dependence on VZA for specific scene types. Figure 11 illustrates the relationship between the CERES SW flux anomalies and VZA stratified into four different scene types. The sea ice overcast (green) and sea ice clear (black) scene types show no statistically significant dependence of the SW flux anomaly on VZA and no bias in anisotropy. The sea ice partly cloudy (gold) and the ocean cloudy (blue) scene types, however, show a statistically significant (95% confidence level) dependence of the SW flux anomalies on VZA. For sea ice partly cloudy scenes the SW flux anomalies decrease with increasing VZA suggesting a positive anisotropy bias consistent with the instantaneously matched results. The ocean cloudy scene type, however, indicates the opposite dependence with larger SW fluxes at larger VZAs. In assessing our randomly varying scene property assumption, we find that the ocean cloudy scene contains footprints with low NISE SIC values (<15%) that show a statistically significant (95% confidence level) increase with VZA; thus, the increase in SW fluxes with VZA for ocean cloudy scenes is explained by the increase in SIC with VZA missed by the imager-based SIC data due to the cloud cover. Considering sea ice partly cloudy scenes, we find no statistically significant (95% confidence level) VZA dependence of cloud properties or SIC. This result provides additional evidence for a bias in the anisotropy for sea ice partly cloudy scenes that contributes to the persistent negative CERES-BBR SW flux difference. These results are complicated by the dependence of footprint size on VZA; however, the footprint size effect is expected to similarly influence the other scene types and yet does not generate a negative correlation between the SW flux anomalies and VZA.

To further investigate the sea ice partly cloudy scene types, Figure 11b shows a stratification of the SW anomaly fluxes by cloud fraction. The results also indicate the tendency of the average SW anomaly flux to decrease with VZA for nearly all cloud fraction bins. The black dots illustrate the results of a difference of means test between adjacent boxes; a black dot is placed into a bin when its mean is statistically significantly different at 95% confidence with the bin mean to its right. Figure 11b suggests a small decrease in the SW flux with increasing VZA for CFs between 20 and 80% and VZA > 20°. This result is consistent with Fig. 11a and indicates that the dependence of the SW flux on VZA for sea ice partly cloudy scenes is independent of CF.

We have presented several analyses in this section to investigate the influence of the CERES inversion algorithm. The results do not suggest a bias in the CERES inversion process for the ocean cloudy, sea ice overcast, and sea ice clear scenes. The results indicate that the CERES inversion for sea ice partly cloudy scenes contributes to the negative CERES-BBR SW flux difference due to scene misidentifications from the imager-based sea ice data set. We find that the CERES-BBR differences are reduced when replacing the imager sea ice data set with NISE.

d. Radiative Transfer Model Flux Adjustment

To convert the aircraft measured irradiance into a TOA equivalent irradiance, we must use a radiative transfer code. Due to errors in the code inputs, such as cloud properties, atmospheric state properties, and surface albedo, this method could introduce biases into our comparisons. Unfortunately, we do not have simultaneous in-situ measurements that would allow us to validate the model inputs. The error is likely small because any potential bias is minimized by using the ratio approach to convert the aircraft level to TOA.

To evaluate the magnitude of any potential bias, we use the Langley Fu-Liou radiative transfer code to create a look-up table to compute the difference between the "true" TOA flux and the ratio-derived TOA flux as a function of the aircraft level flux and the difference between the aircraft-level flux and the simulated flux. These LUTs are then applied to the BBR measurements and the differences between the BBR and Fu-Liou fluxes are analyzed. This results in a small increase in the BBR TOA flux for each grid box. The largest increase is 2.6 Wm^{-2} for GB071 in the SW, the remaining grid boxes all have increases of $<1Wm^{-2}$. Additionally, the two different shades of blue dots in Fig. 7 indicate the flux results before and after the ratio is applied. The changes in the BBR SW flux observations after applying the ratio are small $<1Wm^{-2}$. The LW flux differences before and after the ratio is applied are larger and applying the ratio brings the fluxes into alignment with the independently determined CERES LW fluxes. These considerations suggest that the uncertainty in the radiative transfer code inputs to the BBR adjustment approach is not contributing significantly to the CERES-BBR differences.

6. Summary

The comparison between CERES and BBR measurements has given two main results. First, there is excellent agreement within 2σ uncertainty between the CERES and BBR in the LW. This result indicates that the CERES calibration and radiance-to-flux inversion are working well. The second result in the SW is less clear. Both the grid box and instantaneous comparisons show that CERES agrees with BBR to within the 2σ total uncertainty, after including sampling uncertainty. However, we find a tendency for the CERES TOA SW fluxes to be lower than BBR. There are several factors that could contribute to the average negative SW flux difference: 1) satellite vs. aircraft sampling differences, 2) anisotropy differences, 3) aircraft flux adjustment, and 4) total calibration.

The analysis provides evidence that these factors all increase the uncertainty but not all contribute to a negative SW flux difference. First, differences in satellite vs. aircraft sampling pattern led to an unbiased 1.8% uncertainty on the grid box SW average fluxes. Next, the flux adjustment from aircraft flight level to 20 km to account for above aircraft scattering and absorption has a small effect on the SW fluxes and does not indicate a bias.

Our results illustrate a substantial sensitivity of the CERES SW fluxes to the SIC data set and provide evidence that scene misidentification from the imager sea ice data set is contributing to the negative CERES-BBR SW difference. The imager sea ice data set tends to overestimate sea ice concentration leading to a positive anisotropy bias within sea ice partly cloudy scenes. When replacing the imager sea ice data with the NISE passive microwave sea ice data, the CERES-BBR SW differences are reduced. However, the NISE passive microwave sea ice data in this analysis only modified the ADM selection and not the ADMs themselves. The NISE passive microwave sea ice data should therefore be tested as a possible replacement for the imager-based SIC used in CERES by modifying the ADM development as well. Further analysis is required to better understand this potential bias in sea ice partly cloudy ADMs. Data collected using PAPS mode for FM2 during the summer leg of the MOSAiC campaign (May-September 2020) will allow further investigation of the contribution of ADMs to CERES SW fluxes over sea ice. This work is ongoing.

Calibration differences between CERES and BBR are likely contributing to the consistent negative difference. The full answer is unknown; however, we can use a case to describe the potential contribution. Considering the average fluxes for the GB072 (Table 1), the SW flux calibration accuracy uncertainties are ±2.3 W m^{-2} and ±9.7 W m^{-2} for CERES and BBR, respectively. Considering these numbers and the small number of grid boxes, the uncertainty in the CERES-BBR differences attributed to calibration is ±10.0 W m^{-2}. Considering a worst-case scenario where the BBR measurement during the period were on the high side of this calibration range (e.g., +9.7 W m^{-2}) and CERES measurements where on the low side (-2.3 W m^{-2}), a substantial portion (~12.0 W m^{-2}) of the difference (~17.8 W m^{-2}) could be explained by calibration. Further assessing the contribution of accuracy differences in the CERES-BBR differences requires additional measurements. The best approach to investigate biases in CERES SW fluxes would be through a comparison with an SI-traceable, absolutely calibrated, in-space instrument such as the planned CLARREO Pathfinder mission, although the orbit of this mission is confined between ~±50 ^{o} N and S precluding the observation of polar scenes.

Evaluating CERES fluxes in the Arctic is challenging because of the large variability and heterogeneity in the surface characteristics (e.g., surface albedo). In other words, the way the CERES scanner samples a grid box area can give very different fluxes depending upon the surface types captured. Thus, more data are needed to unravel the nature of the potential bias. Our results indicate that when considering quantifiable sources of uncertainty that the CERES and BBR fluxes agree within the 2σ uncertainty level. However, we identified several potential factors that could cause the CERES SW fluxes to be biased low: (1) total calibration and (2) errors in sea partly cloudy scene anisotropy due to scene misidentification by the sea ice data set. In the long-term, more accurate sea ice concentration data are needed to reduce CERES TOA SW flux uncertainties. Switching from imager to passive microwave sea ice data, in the short-term, could improve CERES TOA SW fluxes in polar regions. Our uncertainty analysis indicates that total calibration and sampling limit the ability to place strong constraints (better than ±7%) on CERES TOA fluxes with aircraft measurements. Our results, confirm that CERES TOA radiative flux data are suitable for polar climate science analysis.

7. References

ASTM (2005), Standard test method for calibration of a pyranometer using a pyrheliometer. ASTM G167-05, ASTM International, 21 pp, doi:10.1520/G0167-05.

Belward, A. S., Estes, J. E., and Kline, K. D. (1999), The igbp-dis global 1-km land-cover data set discover: A project overview. Photogrammetric Engineering and Remote Sensing, 65(9):1013–1020.

Boeke, R. C. and P. C. Taylor (2016), Evaluation of the Arctic surface radiation budget in CMIP5 models. *J. Geophys. Res.,* **121**, 8525-8548, doi: 10.1002/2016JD025099.

Boeke, R. C., P. C. Taylor, S. Sejas (2021), On the nature of the positive Arctic lapse rate feedback. *Geophys. Res. Lett.*, **48**, e2020GL091109. https://doi.org/10.1029/2020GL091109

Brodzik, M. J. and J. S. Stewart. 2016. *Near-Real-Time SSM/I-SSMIS EASE-Grid Daily Global Ice Concentration and Snow Extent, Version 4*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/3KB2JPLFPK3R. [Date Accessed, 10/1/2014].

Bucholtz, A., Hlavka, D. L., McGill, M. J., Schmidt, K. S., Pilewskie, P., Davis, S. M., Reid, E. A., and Walker, A. L., Directly Measured Heating Rates of a Tropical Subvisible Cirrus Cloud, *Journal of Geophysical Research-Atmospheres*, 115, 1-11 (2010)

Chen, H., S. Schmidt, M. D. King, G. Wind, A. Bucholtz, E. A. Reid, M. S-Rozenhaimer, W. L. Smith, P. C. Taylor, S. Kato, P. Pilewskie: *Shortwave Radiative Effect of Arctic Low-Level Clouds: Evaluation of Imagery-Derived Irradiance with Aircraft Observations*, Atmos. Meas. Tech., 14, 2673-2697, https://doi.org/10.5194/amt-14-2673-2021, (2021).

Doelling, D. R., N. G. Loeb, D. F. Keyes, M. L. Nordeen, D. Morstad, C. Nguyen, B. A. Wielicki, D. F. Young, and M. Sun (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. R., M. Sun, L. T. Nguyen, M. L. Nordeen, C. O. Haney, D. F. Keyes, P. E. Mlynczak (2016), Advances in Geostationary-Derived Longwave Fluxes for the CERES Synoptic (SYN1deg) Product, *J. Atmos. Ocean. Techn.*, **33**(3), 503-521. doi: 10.1175/JTECH-D-15-0147.1

Duncan, B. N., Ott, L. E., and co-authors (2020), Space-based observations for understanding changes in the arctic-boreal zone. *Reviews of Geophysics*, **58**, e2019RG000652. https://doi.org/10.1029/2019RG000652

Fu, Q, Lesins, G., Higgins, J., Charlock, T., Chylek, P., Michalsky, J. (1998), Broadband water vapor absorption of solar radiation tested using ARM data. *Geophysical Research Letters* **25**(8): 1169-1172.

Fu, Q., Liou, K. N. (1993), Parameterization of the radiative properties of cirrus clouds. *Journal of Atmospheric Sciences*, **50**(13): 2008-2025.

Ham, S.-H., S. Kato, H. W. Barker, F. G. Rose, and S. Sun-Mack (2014), Effects of 3-D clouds on atmospheric transmission of solar radiation: Cloud type dependencies inferred from A-train satellite data. *J. Geophys. Res. Atmos.*, **119**, 943–963, https://doi.org/10.1002/2013JD020683.

Huang, Y. P. C. Taylor, D. Rutan, F. Rose, M. Shupe (2022), Toward a more realistic representation of surface albedo in NASA CERES-derived surface radiative fluxes: A comparison with MOSAIC field campaign, *Elem. Sci. Anth.*, **10:1**, https://doi.org/10.1515/elementa.2022.00013.

Ji, Q., S. C. Tsay, 2000: On the Dome Effect of Eppley Pyrgeometers and Pyranometers, *Geophys. Res. Lett.*, 27, 7, 971-974.

Kato, S., Ackerman, T.P., Mather, J.H., Clothiaux, E.E. (1999), The k-distribution method and correlated-k approximation for a shortwave radiative transfer model. *J. Quant. Spec. Rad. Trans.*, **62**(1), 109-121.

Kato, S., Rose, F.G., Charlock, T.P. (2005) Computation of domain-averaged irradiance using satellite-derived cloud properties. *J. Atmos. Ocean. Tech.,* **22**(2), 146-164.

Kato, S., and co-authors (2013), Surface Irradiances Consistent with CERES-Derived Top-of-Atmosphere Shortwave and Longwave Irradiances. *J. Climate*, **26**(9), 2719-2740, https://doi.org/10.1175/JCLI-D-12-00436.1.

Kay, J. E., and T. L'Ecuyer (2013), Observational constraints on Arctic Ocean clouds and radiative fluxes during the early 21st century, *J. Geophys. Res. Atmos.*, **118**, 7219–7236, doi:10.1002/jgrd.50489.

Kern, S., T. Lavergne, D. Notz, L.T. Pedersen, R. T. Tonboe, R. Saldo, and A. M. Sørensen (2019), Satellite passive microwave sea-ice concentration data set intercomparison: closed ice and ship-based observations, *The Cryosphere*, **13**, 3261–3307, doi: 10.5194/tc- 13-3261-2019.

Kipp & Zonen (2001), CG4 pyrgeometer: Instruction manual. 64 pp. [Available online at www.kippzonen.com/Download/33/CG-4-Manual.]

Kipp & Zonen, 2003a, Instruction manual CM22 pyranometer, pp. 61.

Kipp & Zonen, 2003b, Instruction manual CG4 pyrgeometer, pp. 61.

Kipp and Zonen (2004), CM22 precision pyranometer: Instruction manual. 65 pp. [Available online at www.kippzonen.com/Download/55/CM-22-Pyranometer-Manual.]

Kratz, D.P., Rose, F.G. (1999), Accounting for molecular absorption within the spectral range of the CERES window channel. *J. Quant. Spec. Rad. Trans.*, **61**(1), 83-95.

Loeb, N. G. and co-authors (2001), Determination of Unfiltered Radiances from the Clouds and the Earth's Radiant Energy System Instrument. *J. Appl Meteor. Clim.*, **40**(4), 822-835, https://doi.org/10.1175/1520-0450(2001)040<0822:DOURFT>2.0.CO;2.

Loeb, N. G., Kato, S., & Wielicki, B. A. (2002). Defining Top-of-the-Atmosphere Flux Reference Level for Earth Radiation Budget Studies, Journal of Climate, 15(22), 3301-3309. https://doi.org/10.1175/1520-0442(2002)015<3301:DTOTAF>2.0.CO;2

Loeb, N. G., Manalo-Smith, N., Kato, S., Miller,W. F., Gupta, S. K., Minnis, P., and Wielicki, B. A. (2003), Angular Distribution Models for Top-of-Atmosphere Radiative Flux Estimation from the Clouds and the Earth's Radiant Energy System Instrument on the Tropical Rainfall Measuring Mission Satellite, Part I: Methodology, *J. Appl. Meteol.*, **42**, 240–265, 2003.

Loeb, N. G., Kato, S., Loukachine, K., and Manalo-Smith, N. (2005), Angular Distribution Models for Top-of-Atmosphere Radiative Flux Estimation from the Clouds and the Earth's Radiant Energy System Instrument on the Terra Satellite, Part I: Methodology, *J. Atmos. Ocean. Tech.*, 22, 338–351.

Loeb, N. G. and co-authors (2007), Multi-instrument comparison of Top-of-Atmosphere Reflected Solar Radiation. *J Climate*, **20**, 575-591, doi: https://doi.org/10.1175/JCLI4018.1

Loeb, N. G. and co-authors (2009), Toward Optical Closure of the Earth's Top-of-Atmosphere Radiation Budget. *J. Climate*, **22**, 748-766, doi: 10.1175/2008JCLI2637.1.

Loeb, N. G. and co-authors (2012), Advances in Understanding Top-of-Atmosphere Radiation Variability from Satellite Observations. *Survey Geophys.*, **33,** 359–385, DOI 10.1007/s10712-012-9175-1.

Loeb, N. G. and co-authors (2018), Clouds and Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *J. Climate*, **31**(2), 895-918, https://doi.org/10.1175/JCLI-D-17-0208.1.

Loveland, T. R. and Belward, A. S. (1997). The international geosphere biosphere programme data and information system global land cover data set (DISCover). Acta Astronautica, 41(4):681–689.

Melsheimer, Christian; Spreen, Gunnar (2019): AMSR2 ASI sea ice concentration data, Arctic, version 5.4 (NetCDF) (July 2012 – December 2018). *PANGAEA*, https://doi.org/10.1594/PANGAEA.898399

Meier, W. N., F. Fetterer, A. K. Windnagel, and J. S. Stewart (2021), *NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4*. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: https://doi.org/10.7265/efmz-2t65. [Date accessed, 11/1/2018].

Minnis, P., S. Sun-Mack, Y. Chen, F. Chang, C. R. Yost, W. L. Smith, P. W. Heck, R. F. Arduini, S. T. Bedka, Y. Yi, G. Hong, Z. Jin, D. Painemal, R. Palikonda, B. R. Scarino, D. A. Spangenberg, R. A. Smith, Q. Z. Trepte, P. Yang, Y. Xie, 2020: CERES MODIS Cloud Product Retrievals for Edition 4–Part I: Algorithm Changes. IEEE Transactions on Geoscience and Remote Sensing, 1-37. doi: 10.1109/TGRS.2020.3008866.

Philipona, R., C. Frohlich, and C. Betz, 1995: Characterization of pyrgeometers and the accuracy of atmospheric longwave radiation measurements, *Appl. Opt.*, **34**, 1598-1605.

NASA/LARC/SD/ASDC. (2014). CERES Single Scanner Footprint (SSF) TOA/Surface Fluxes, Clouds and Aerosols Terra-FM1 Edition4A [Data set]. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from https://doi.org/10.5067/TERRA/CERES/SSF_Terra-FM1_L2.004A

NASA/LARC/SD/ASDC. (2014). CERES Single Scanner Footprint (SSF) TOA/Surface Fluxes, Clouds and Aerosols Terra-FM2 Edition4A [Data set]. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from https://doi.org/10.5067/TERRA/CERES/SSF-FM2_L2.004A

NASA/LARC/SD/ASDC. (2014). CERES Single Scanner Footprint (SSF) TOA/Surface Fluxes, Clouds and Aerosols Aqua-FM3 Edition4A [Data set]. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from https://doi.org/10.5067/AQUA/CERES/SSF-FM3_L2.004A

NASA/LARC/SD/ASDC. (2017). CERES and GEO-Enhanced TOA, Within-Atmosphere and Surface Fluxes, Clouds and Aerosols 1-Hourly Terra-Aqua Edition4A [Data set]. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A

NASA/LARC/SD/ASDC. (n.d.). ARISE C-130 Aircraft Merge Data Files [Data set]. NASA Langley Atmospheric Science Data Center DAAC. Retrieved from https://doi.org/10.5067/AIRCRAFT/ARISE_Merge_Data_1

Pilewskie, P., J. Pommier, R. Bergstrom, W. Gore, S. Howard, M. Rabbette, B. Schmid, P. V. Hobbs, and S. C. Tsay, Solarspectral radiative forcing during the Southern African Regional Science Initiative, *J. Geophys. Res.*,108(D13), 8486,doi:10.1029/2002JD002411, 2003

Pistone, K., I. Eisenman, V. Ramanathan (2014), Albedo decrease due to Arctic sea ice loss. *Proceedings of the National Academy of Sciences*, 111(9) 3322 3326; DOI:10.1073/pnas.1318201111.

Pope, S. K., F. P. J. Valero, W. D. Collins, P. Minnis (2002): Comparison of ScaRaB, GOES 8, aircraft, and surface observationsof the absorption of solar radiation by clouds. *J. Geophys Res.*, **107**(D11), 4114,10.1029/2001JD001139

Riihelä, A., T. Manninen, and V. Laine (2013), Observed changes in the albedo of the Arctic sea-ice zone for the period 1982-2009. *Nature Clim Change*, **3**, 895-898, doi: 10.1038/NCLIMATE1963.

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. Atmos.*, **122**, 4829–4848, doi:10.1002/2016JD026443.

Rose, FG, Rutan, DA, Charlock, T, Smith, GL, Kato, S. 2013. An algorithm for the constraining of radiative transfer calculations to CERES-observed broadband top-of-atmosphere irradiance. *J. Atmos. Ocean. Tech.*, **30**(6), 1091-1106.

Rutan, DA, Kato, S, Doelling, DR, Rose, FG, Nguyen, LT, Caldwell, TE, Loeb, NG. 2015. CERES synoptic product: Methodology and validation of surface radiant flux. *Journal of Atmospheric and Oceanic Technology* **32**(6): 1121-1143.

Smith, W. L. and co-authors (2005), EOS *Terra* Aerosol and Radiative Flux Validation: An Overview of the Chesapeake Lighthouse and Aircraft Measurements for Satellites (CLAMS) Experiment, *J. Atmos. Sci.*, **62**(4), 903-918, https://doi.org/10.1175/JAS3398.1.

Spreen, G., L. Kaleschke, and G. Heygster (2008), Sea ice remote sensing using AMSR-E 89 GHz channels. *J. Geophys. Res.*, **113**, C02S03, doi:10.1029/2005JC003384.

Su, W., E. Dutton, T. P. Charlock, W. Wiscombe, (2008), Performance of Commercial Radiometers in Very Low Temperature and Pressure Environments Typical of Polar Regions and of the Stratosphere: A Laboratory Study, *J. Atmos. Oceanic Technol.*, **25**, 558-569

Su, W., Corbett, J., Eitzen, Z., and Liang, L. (2015), Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: methodology, Atmos. Meas. Tech., 8, 611–632, https://doi.org/10.5194/amt-8-611-2015.

Su, W., Corbett, J., Eitzen, Z., and Liang, L. (2015), Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: validation, Atmos. Meas. Tech., 8, 3297–3313, https://doi.org/10.5194/amt-8-3297-2015.

Suarez M.J., and co-authors (2005), Documentation and validation of the Goddard Earth Observing System (GEOS) Data Assimilation System—Version 4. M. J. Suarez, Ed., Technical Report Series on Global Modeling and Data Assimilation, Vol. 26, NASA Tech. Rep. NASA/TM-2005-104606, 187 pp.

Taylor, P. C., W. Maslowski J. Perlwitz, and D. J. Wuebbles (2017), Arctic Changes and their Effects on Alaska and the Rest of the United States. In: *Climate Science Special Report: Fourth National Climate Assessment, Volume I* [Wuebbles, D. J., S. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Steward, and T. K. Maycock (eds.)]. U.S. Global Climate Change Research Program, Washington, DC, USA, pp. 303-332, doi: 10.7930/J00863GK.

Trepte, Q. Z., P. Minnis, S. Sun-Mack, C. R. Yost, Y. Chen, Z. Jin, G. Hong, F. Chang, W. L. Smith, K. M. Bedka, T. L. Chee, 2019: Global Cloud Detection for CERES Edition 4 Using Terra and Aqua MODIS Data. IEEE Transactions on Geoscience and Remote Sensing, 1-40. doi: 10.1109/TGRS.2019.2926620

Valero, F. P, S. K. Pope, and co-authors (2003): Absorption of solar radiation by clear and cloudy atmosphere during the Atmospheric Radiation Measurement Enhanced Shortwave Experiments (ARESE) I and II: Observations and Models. *J. Geophys. Res.*, **108**(D1), 4016, doi: 10.1029/2001jd001384.

Wei, J., Wang, Z., Gu, M. *et al*. An evaluation of the Arctic clouds and surface radiative fluxes in CMIP6 models. *Acta Oceanol. Sin*. **40,** 85–102 (2021). https://doi.org/10.1007/s13131-021-1705-6

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper (1996), Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, *Bull. Amer. Meteor. Soc.*, **77**(5), 853-868, https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2.

Open Research

Data Availability Statement: All data are freely available and can be accessed at asdc.larc.nasa.gov. Aircraft flux data, DOI: 10.5067/SUBORBITAL/ARISE/DATA001. CERES radiative flux and MODIS cloud property data: DOI: 10.5067/AQUA/CERES/SSF-FM3_L2.004A; 10.5067/NPP/CERES/SSF_NPP-FM5_L2.002A; 10.5067/TERRA/CERES/SSF_Terra-FM1_L2.004A; 10.5067/TERRA/CERES/SSF-FM2_L2.004A; 10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A.

Tables

Table 1. Summary of average grid box scene characteristics and radiative fluxes. Within grid box variability is represented by the standard deviations shown in parentheses. CERES Ed4a SSF results include footprints from FM1, 2, 3, and 5.

Grid Box	NISE SIC (%) (Stdev)	CERES Ed4a Imager SIC (%) (Stdev)	Cloud fraction (%) (Stdev)	Cloud optical depth (Stdev)	Cloud top pressure (hPa) (Stdev)	CERES SSF Ed4a OLR (Stdev) (Wm^{-2})	CERES SSF Ed4a SW (Stdev) (Wm^{-2})	Number of SSF Footprints (#)	BBR OLR (Stdev) (Wm^{-2})	BBR SW (Stdev) (Wm^{-2})	BBR Count (# of 1-min averages)	CERES Ed4a SYN OLR (Wm^{-2})	CERES Ed4a SYN SW (Wm^{-2})
701	9.8 (9.9)	85.4 (35.3)	100.0 (0.2)	7.0 (2.7)	867.9 (85.7)	222.7 (6.7)	195.5 (19.1)	351	228.5 (2.6)	194.2 (13.7)	71	226.3	188.4
702	16.9 (9.1)	74.2 (38.9)	99.4 (0.7)	9.6 (2.6)	903.5 (36.2)	225.4 (5.5)	223.4 (17.5)	497	223.2 (3.2)	241.2 (24.5)	68	226.8	220.8
111	86.3 (3.1)	100.0 (0.1)	43.4 (28.9)	4.4 (1.6)	829.2 (195.3)	220.7 (3.6)	240.7 (13.3)	560	219.2 (2.3)	258.8 (14.2)	70	223.9	236.1
112	77.3 (6.5)	100.0 (0.0)	84.1 (27.6)	8.9 (2.6)	871.9 (21.2)	223.9 (2.7)	265.4 (19.6)	337	217.9 (1.6)	278.2 (14.3)	61	220.3	267.5
151	0.3 (0.8)	97.2 (16.4)	99.7 (1.0)	40.1 (31.9)	626.4 (120.4)	209.0 (9.2)	256.5 (24.9)	108	199.9 (7.1)	265.5 (26.1)	58	208.7	251.2

Table 2. Summary statistics of instantaneously matched CERES SSF_Ed4a-BBR flux comparisons. The number in parentheses in the SW NISE as imager mean difference column represents that footprint counts, which change when using NISE to classify scenes.

ADM GROUP	**N (count)**	**SW SSF-BBR Mean Difference (W m^{-2})**	**SW SSF STDEV (W m^{-2})**	**SW SSF-NISE as imager-BBR Mean Difference (W m^{-2})**	**SW SSF-NISE as imager STDEV (W m^{-2})**	**SSF LW Mean Difference (W m^{-2})**	**LW SSF-STDEV (W m^{-2})**
Ocean Cloudy	15	-0.8	17.6	-2.4 (12)	18.2	-1.9	11.2
Sea Ice Clear	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Sea Ice Partly Cloudy	9	-17.1	13.6	+9.2 (12)	17.7	1.9	7.4
Sea Ice Overcast	15	-9.1	29.3	-9.1 (15)	29.3	-0.9	11.0

Table 3. Summary of the grid box averaged fluxes for the aircraft-like sampling and differences between the satellite minus the aircraft-like sampling fluxes for individual overpasses from the sampling study. The column titles for individual overpasses indicates the satellite (Terra/Aqua), and the hour of the overpass in UTC (20/21/22).

Grid Box	SW aircraft -like	Terra_ 20	Aqua_ 21	Terra_ 22	Aqua_ 22	Avg. SW Diff	LW aircraft-like	Terra_ 20	Aqua_ 21	Terra_ 22	Aqua_ 22	Avg LW Diff
071	220.11	4.59	-0.01	10.73	4.01	4.83	225.51	4.78	9.21	-1.16	-7.6	1.31
072	255.13	3.29	1.88	-2.8	-8.12	-1.44	230.15	0.11	0.07	-2.29	-3.0	-1.28
111	251.76	5.44	2.49	0.75	-0.28	2.1	221.27	0.57	0.62	0.23	-0.4	0.26
112	293.32	-3.12	-2.72	-0.08	-0.39	-1.58	219.16	2.37	2.43	-0.32	-0.23	1.06
151	282.30	N/A	6.14	9.92	-0.58	5.16	209.88	N/A	-1.03	-7.54	-0.59	-3.05
					Mean Difference:	**1.64**					**Mean Difference:**	**-0.20**

Table 4. Summary of the grid box mean flux differences for CERES SSF Ed4a and when using NISE sea ice data for ADM selection.

Grid Box	NISE SIC (%)	Imager SIC (%)	LW SSF Ed4a minus NISE SIC (Wm^{-2})	SW SSF Ed4a minus NISE SIC (Wm^{-2})	SW FM1 SSF Ed4a minus NISE (Wm^{-2})	SW FM2 SSF Ed4a minus NISE (Wm^{-2})
071	9.8	85.4	0.1	0.0	0.0	0.0
072	16.9	74.2	-0.1	-0.3	-0.4	-0.2
111	86.3	43.4	0.0	0.1	-30.6	+6.2
112	77.3	84.1	0.0	-12.3	-19.7	-10.9
151	0.3	99.7	0.1	0.0	0.0	n/a

Table 5. Summary of the grid box mean fluxes and scene characteristics for FM1, FM2, and FM1 sampled like FM2 (FM2-like).

Grid Box	NISE SIC (%)			Cloud fraction (%)			CERES Ed4a SW (Wm^{-2})			Count (#)		
	FM1	FM2	FM2-like	FM1	FM2	FM2-like	FM1	FM2	FM2-like	FM1	FM2	FM2-like
701	8.1	14.5	17.5	100	100	100	188.1	195.2	209.8	46	216	138
702	13.5	18.1	22.2	99	99	99	226.5	222.0	226.1	46	371	202
111	86.4	86.1	78.2	42.9	44.2	68.6	238.3	242.1	239.5	49	448	185
112	76.6	80.3	76.8	86.6	90.4	83.8	270.7	269.1	261.2	43	239	104

Figures

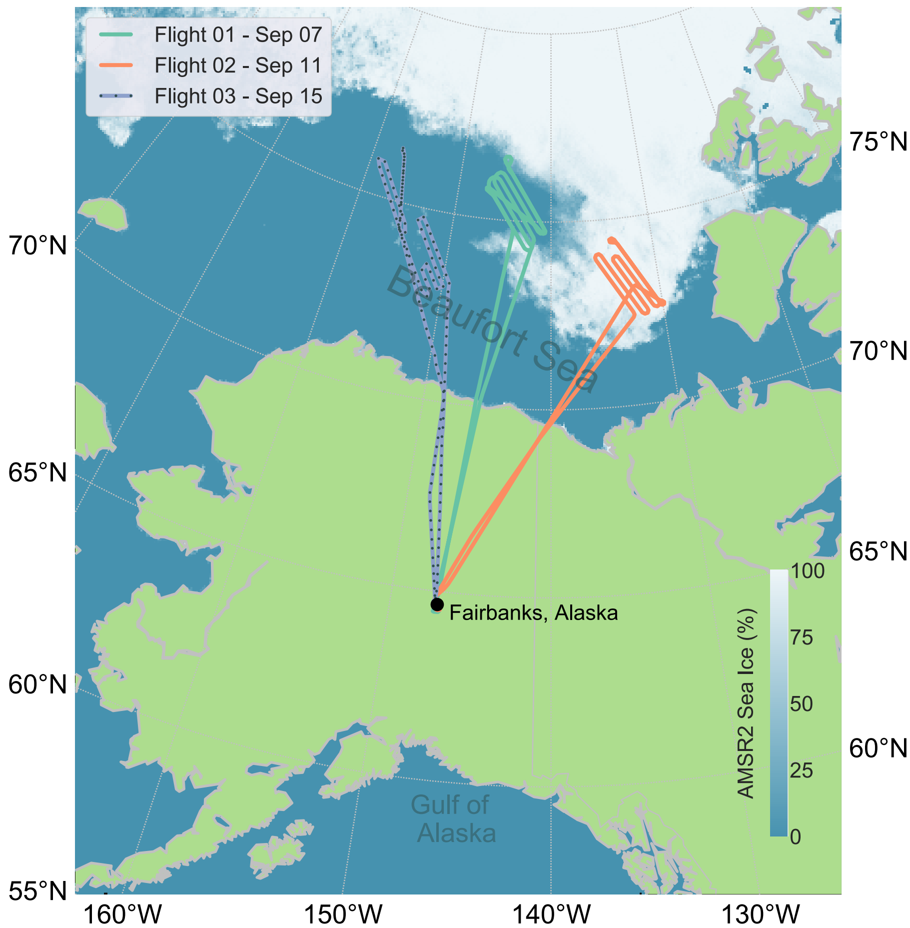

Fig. 1. C-130 aircraft flight tracks for the CERES top-of-atmosphere grid-box experiments. The sea ice concentration from the AMSR2-ASI on September 15, 2014 is shown for reference.

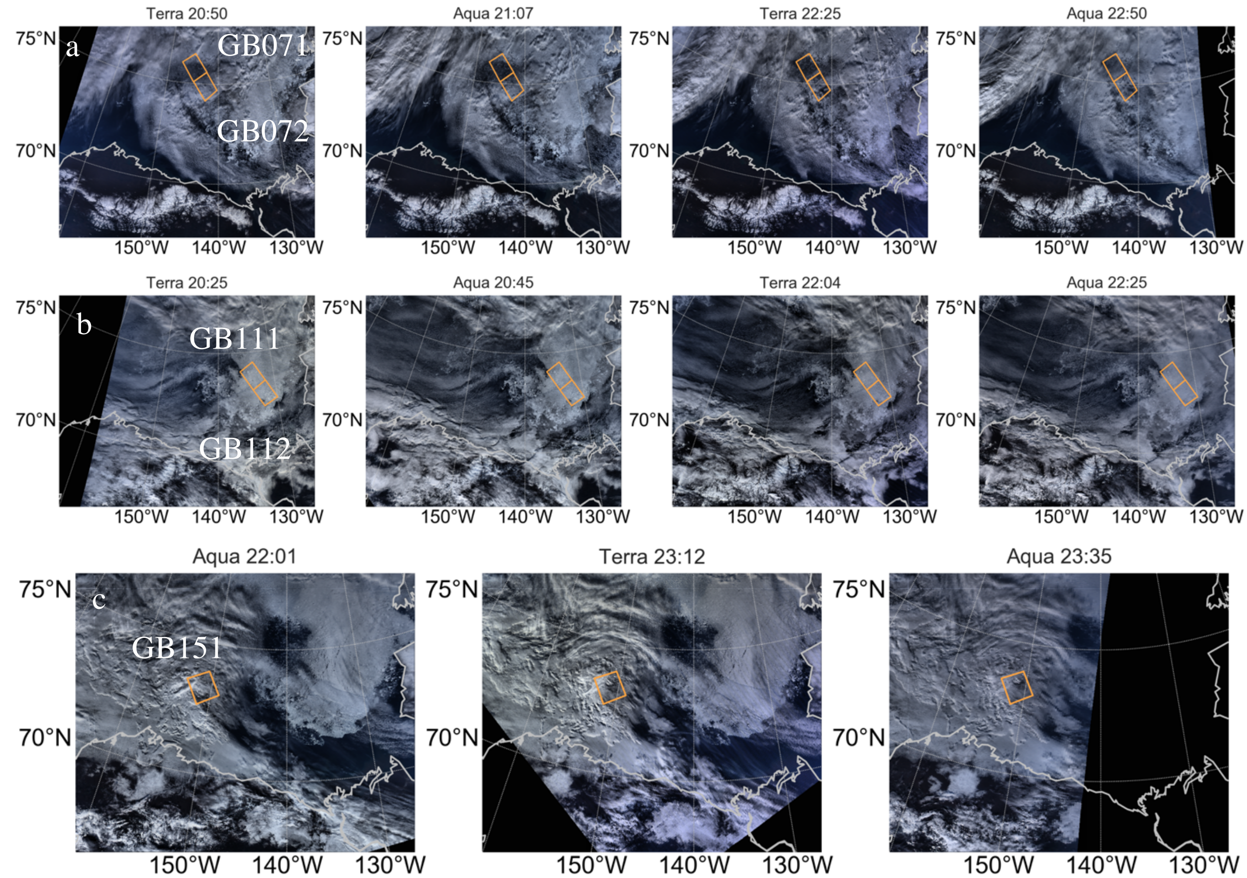

Fig. 2. True-color MODIS images from Terra and Aqua for the overpasses coincident with the aircraft sampling (a) 9/7/2014, (b) 9/11/2014, and (c) 9/15/2014. Overpass time stamp given in UTC. Yellow boxes show the borders of the CERES grid box sampling region on each day.

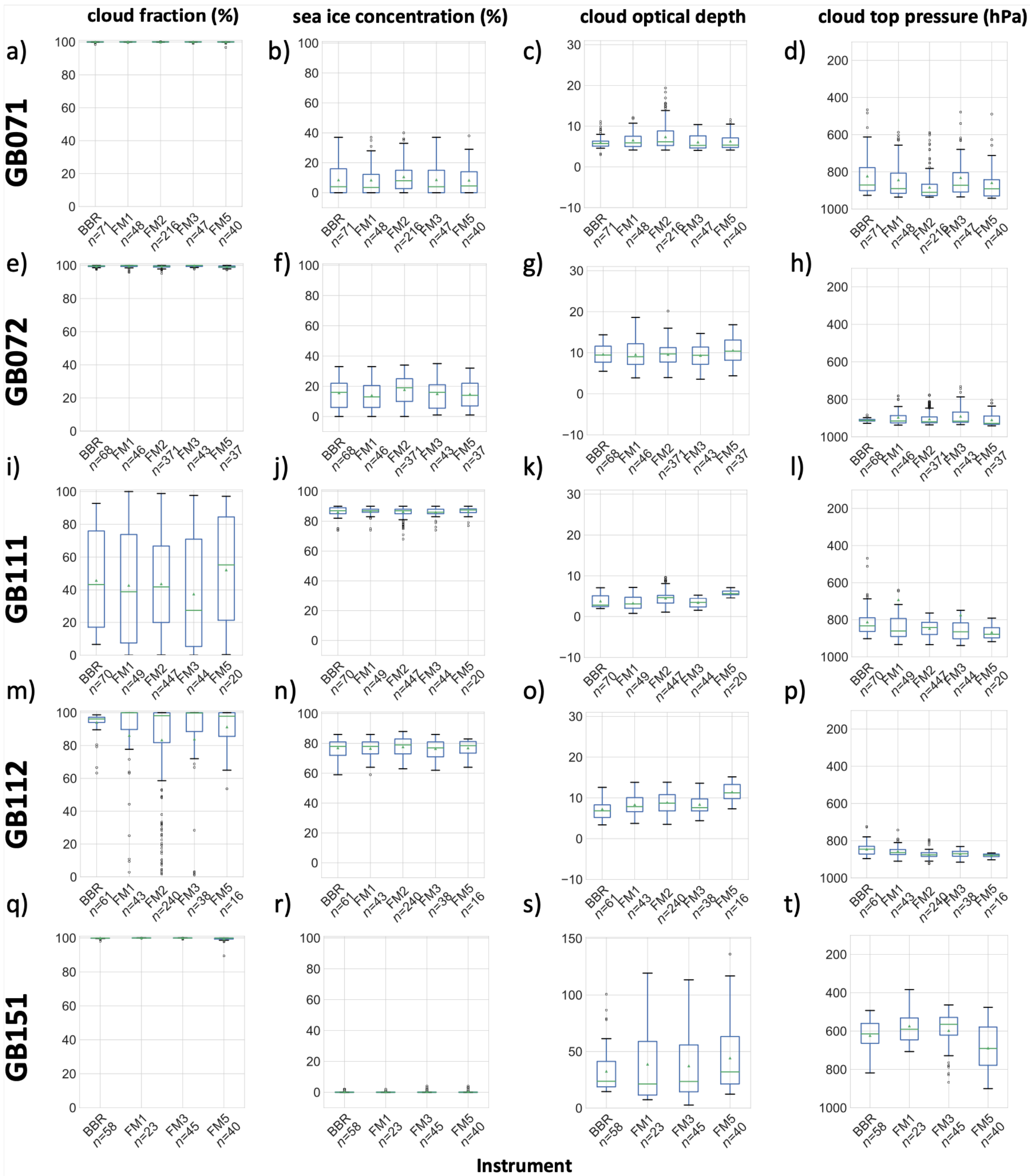

Fig. 3. Box and whisker plots of cloud fraction, sea ice concentration, cloud optical depth, and cloud top pressure height are shown for (a-d) GB071, (e-h) GB072, (i-l) GB111, (m-p) GB112, (q-t) GB151. Distributions of surface and cloud properties are shown separately using a 20-km^2 window about the aircraft from each CERES instrument: FM1, FM2, FM3, and FM5. The top and bottom of each box indicates the upper and lower quartiles, the horizonal line indicates the median, the triangle symbol indicates the mean, and outliers greater than 1.5 times the interquartile range are indicated by circles.

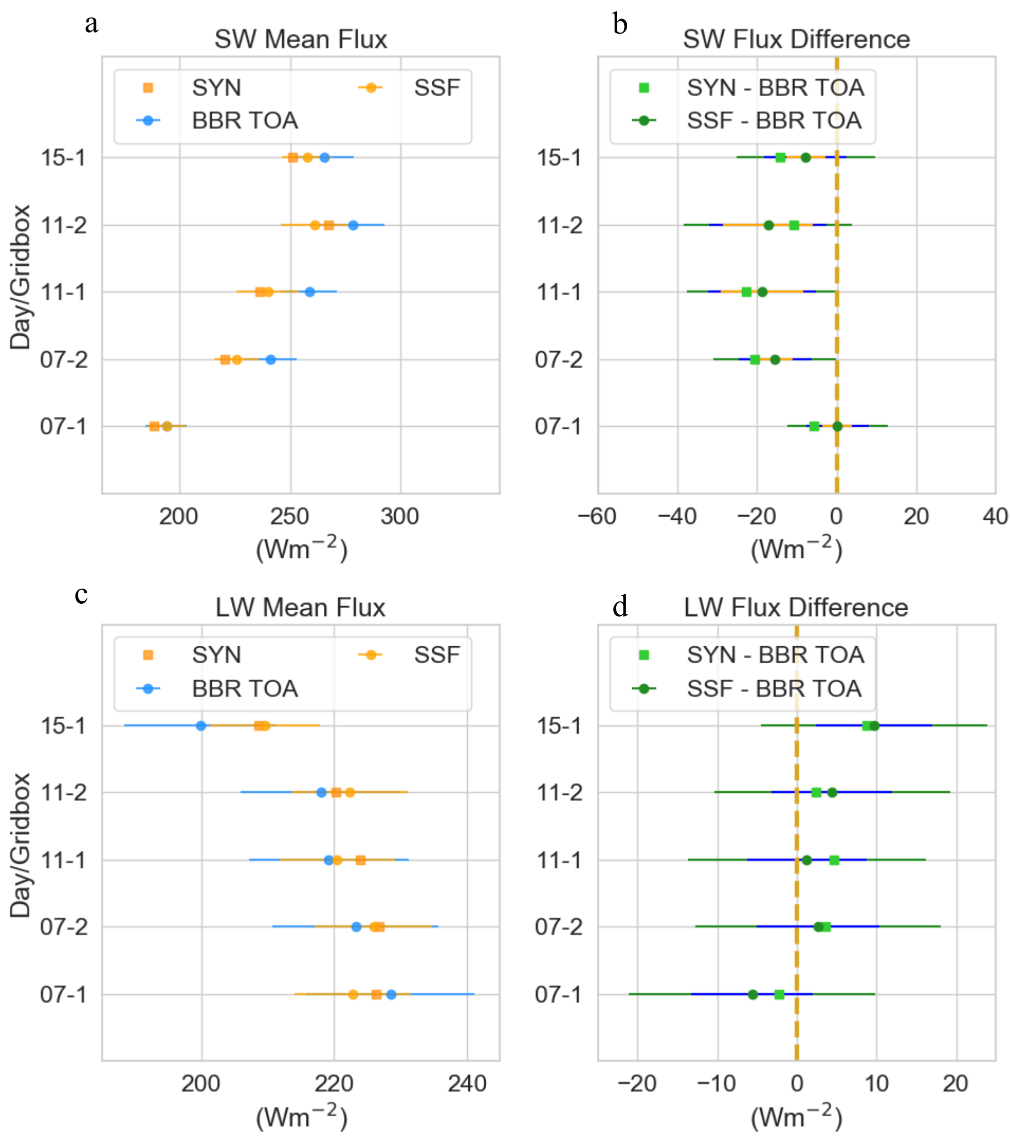

Fig. 4. Whisker plots showing the 2-hour grid box mean (a) SW and (c) LW fluxes for CERES E4a SYN, CERES Ed4a SSF (average of all FM1, 2, 3, and 5 footprints), and BBR_{TOA} and the (b) SW and (d) LW flux differences between BBR_{TOA} and CERES. Error bars represent the 2σ combined uncertainty from calibration (green), inversion (orange), and sampling (blue).

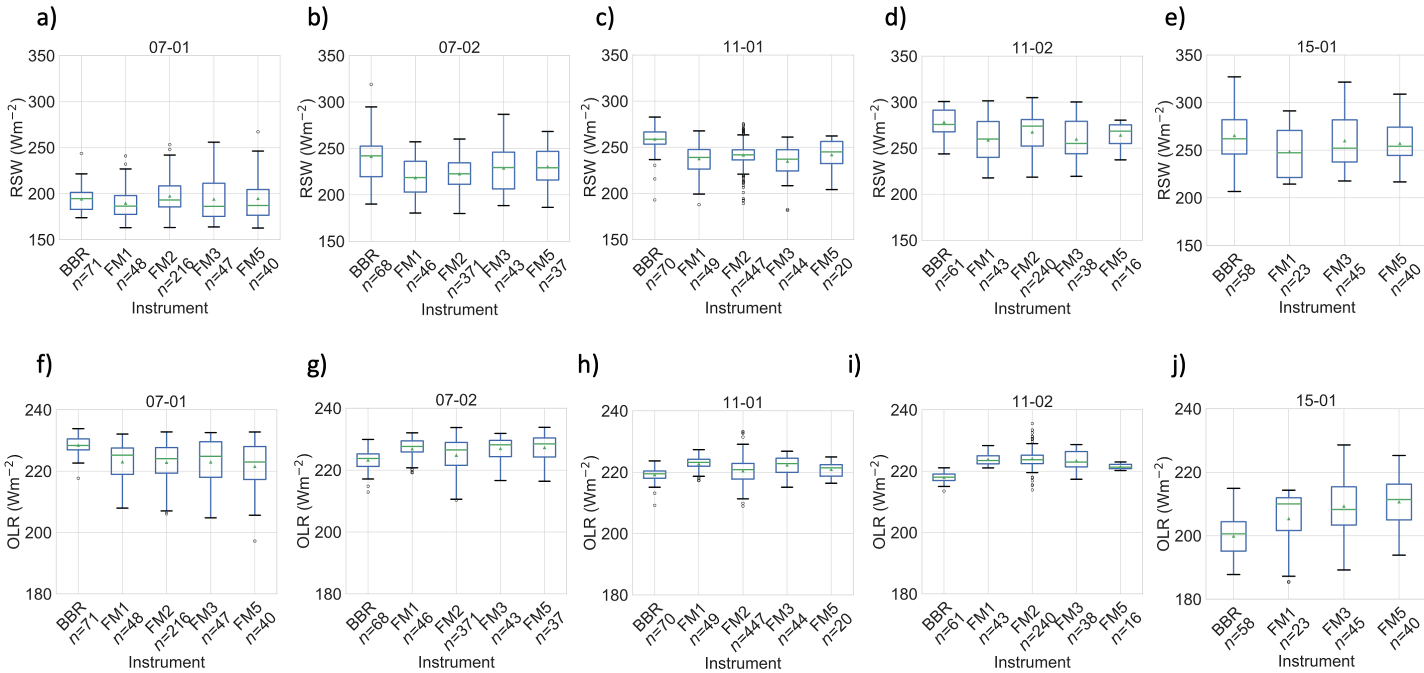

Fig. 5. Box and whisker plots of CERES instrument and BBR (a-e) SW and (f-j) LW fluxes for each grid box (a-b,f-g) 9/7/2014, (c-d,h-i) 9/11/2014, and (e,j) 9/15/2014. CERES observations are also shown for individual instruments: FM1, FM2, FM3, and FM5. The top and bottom of each box indicates the upper and lower quartiles, the horizonal line indicates the median, the triangle symbol indicates the mean, and outliers greater than 1.5 times the interquartile range are indicated by circles.

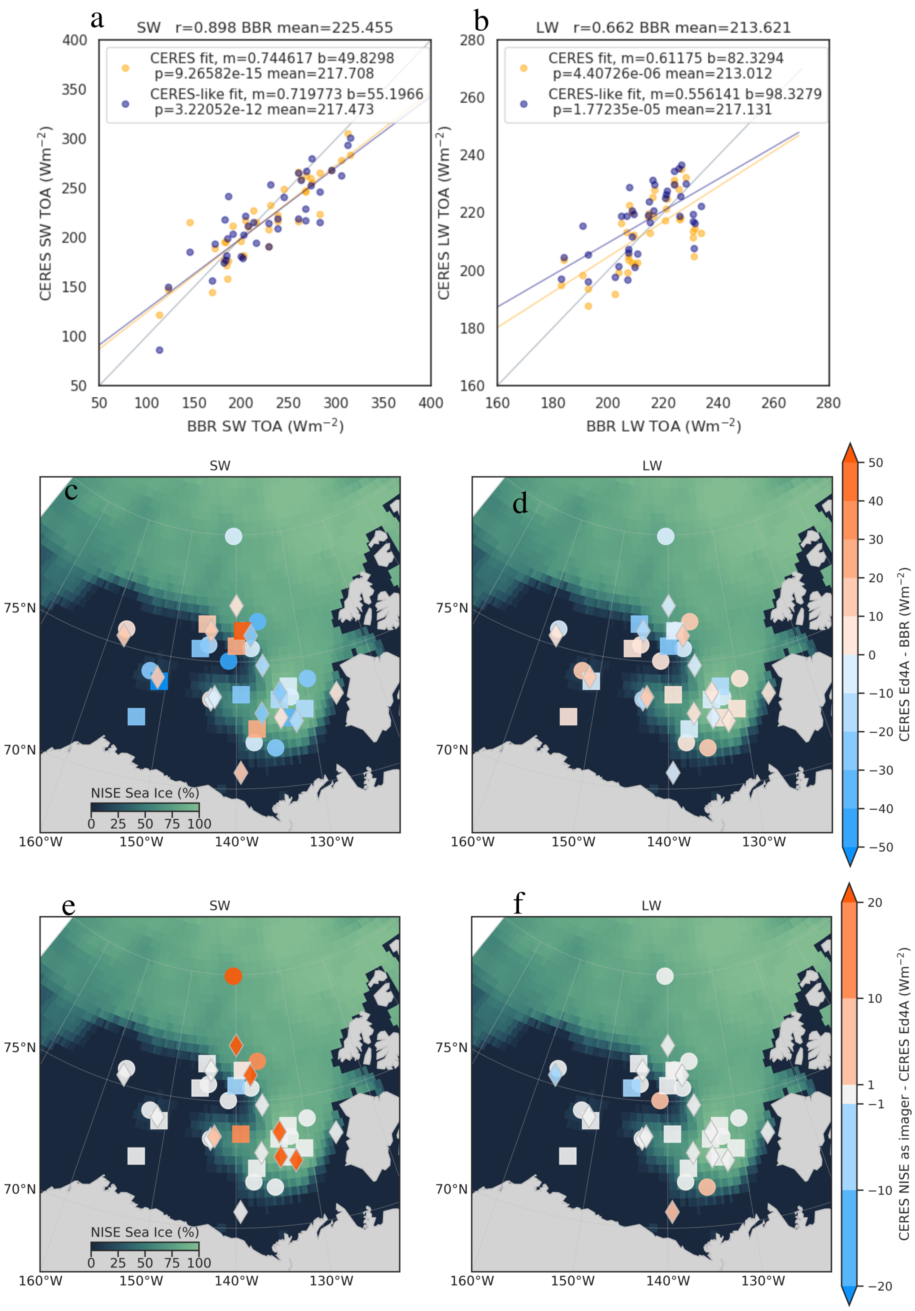

Figure 6. Comparison between the CERES and BBR instantaneously matched radiative flux measurements. Top panels show scatterplots of the (a) SW and (b) LW SSF and BBR_{TOA} fluxes. The yellow symbols show the CERES measurements, and the blue markers show the CERES-like fluxes (20 km x 20 km). Middle panels show the spatial distribution of the (c) SW and (d) LW SSF-BBR_{TOA} differences. Bottom panels show the changes in the (e) SW and (f) LW when the NISE SIC is used for ADMs selection as CERES NISE minus CERES Ed4a. The symbol shapes denote the satellite: Terra (circle), Aqua (square), and NPP (diamond).

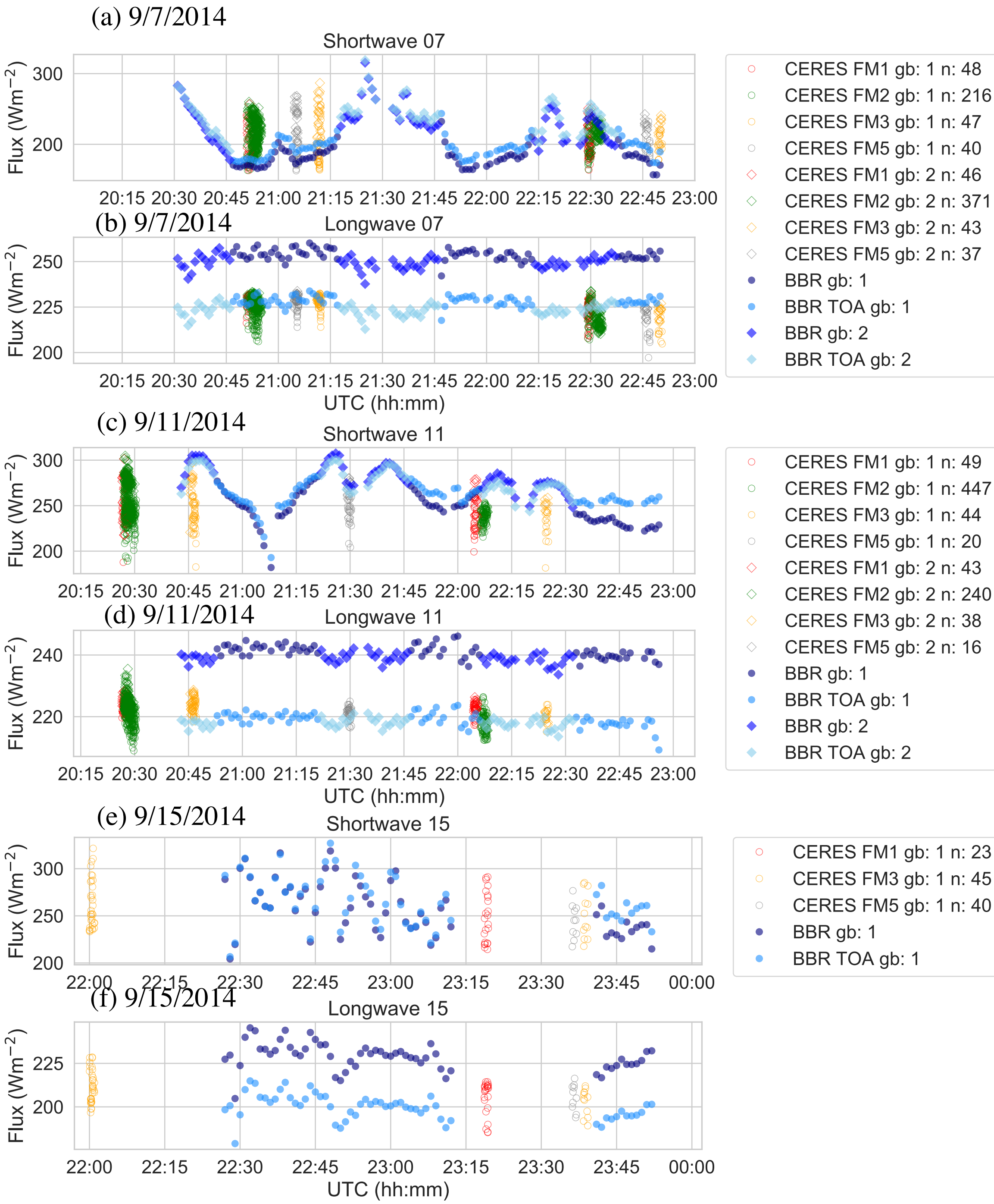

Fig. 7. Time series of LW and SW observed fluxes (blue dots) and the CERES overpass observations for individual satellite overpasses for FM1 (red), FM2 (green), FM3 (yellow), and FM5 (gray) for (a,b) 9/7/2014, (c,d) 9/11/2014, and (e,f) 9/15/2014. Different shades of blue represent the fluxes before and after adjustment from flight-level (darker shades) to TOA (lighter shades) and the symbols represent different grid boxes: circle for GB1 and diamond for GB2.

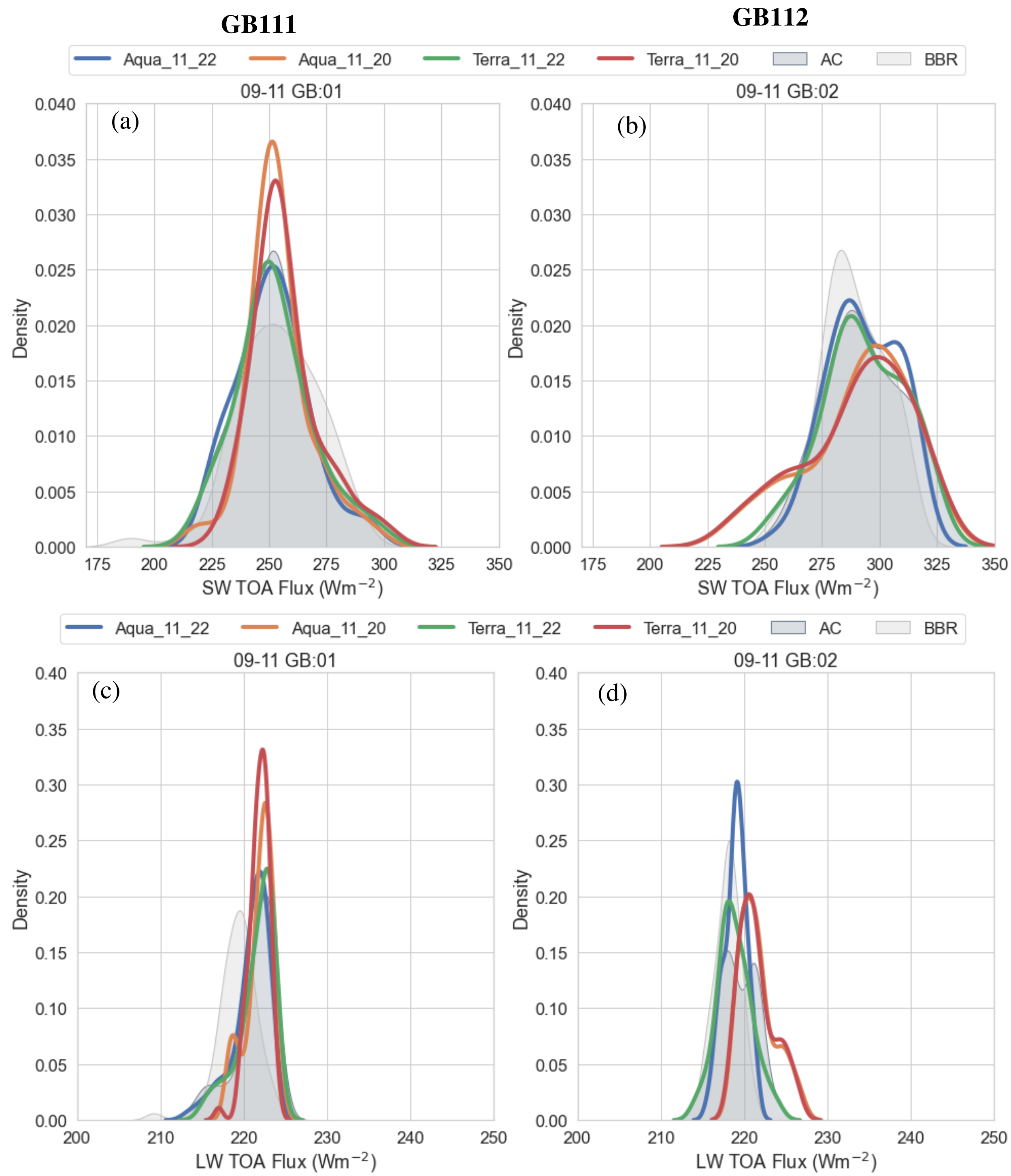

Fig. 8. Probability distribution of (a,b) SW and (c,d) LW fluxes for the satellite instantaneous sampling approach at each overpass time (blue, orange, green, and red) and the aircraft sampling (AC, dark gray), where the left column represents GB111 and the right column GB112. For context, probability distributions of the BBR measurements are shown in light gray.

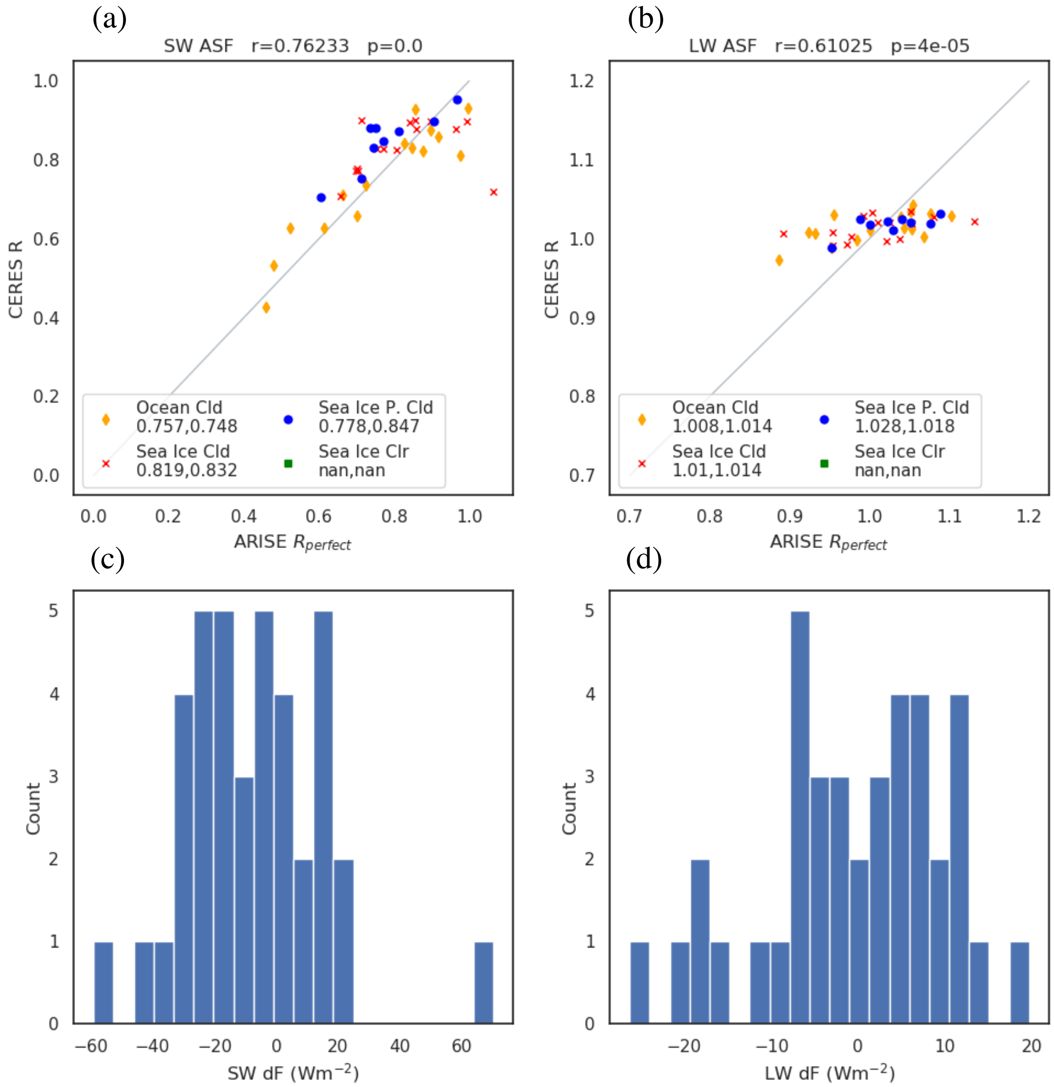

Fig. 9. Scatterplot of the R_{CERES} vs. $R_{perfect}$ for (a) SW and (b) LW. The inset legends in panels (a) and (b) shows the average value for $R_{perfect}$ on the left and R_{CERES} on the right for each scene-type. Histograms of the flux difference that resulted from the anisotropic factor differences (R_{CERES}-$R_{perfect}$) are shown in (c) SW and (d) LW.

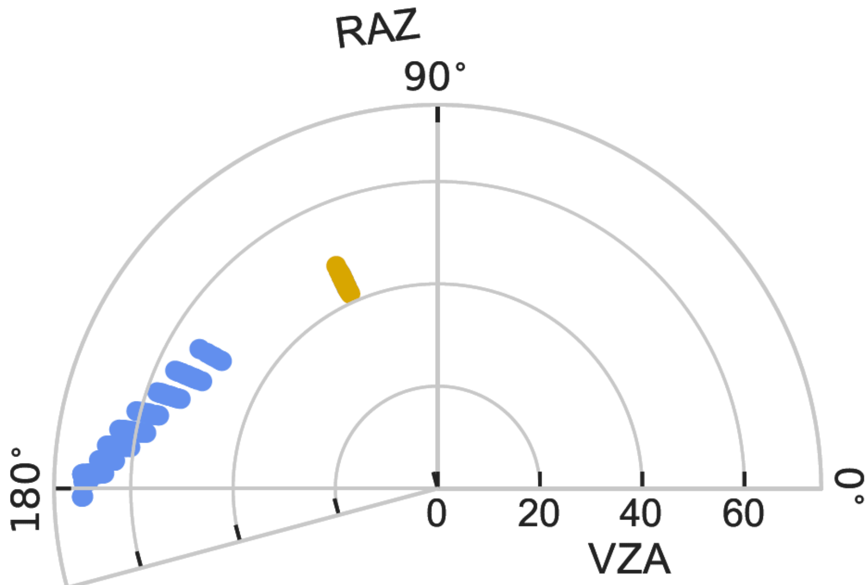

Fig. 10. Polar plot of the FM1 (yellow circles) and FM2 (blue circles) sampled viewing geometries in VZA and relative azimuth angle (RAZ) for 9/7/2014.

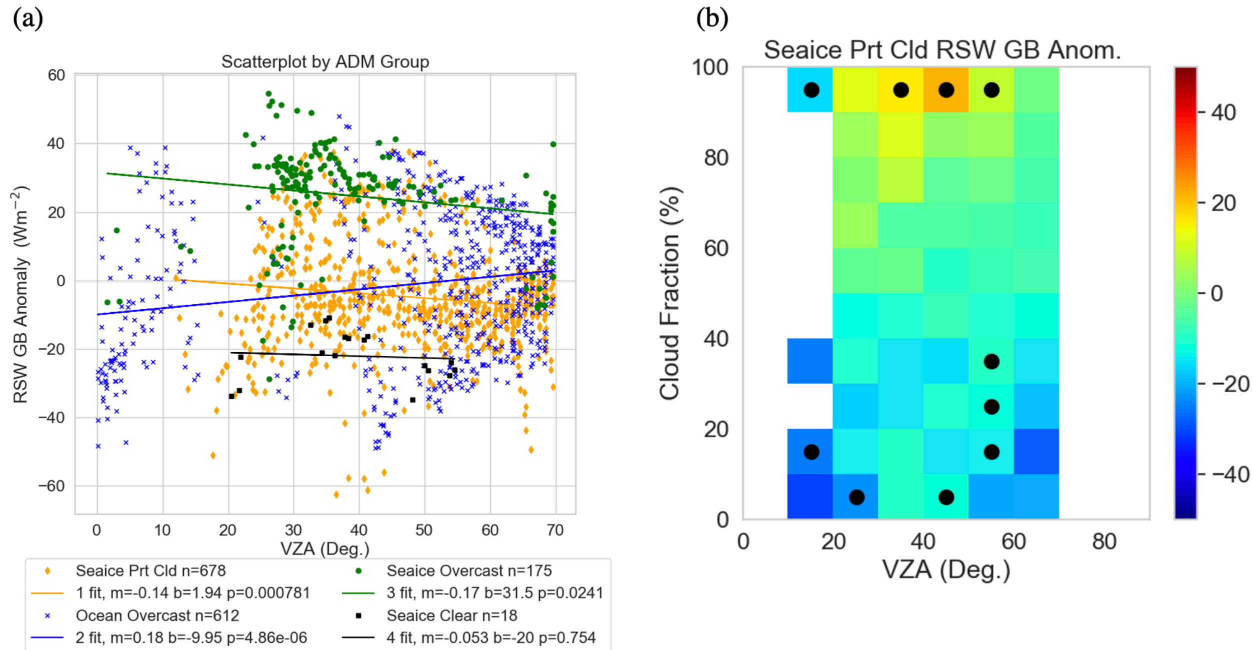

Fig. 11. Scatterplot illustrating the dependence of (a) CERES grid box SW flux anomalies (y-axis; defined as the CERES SW minus the CERES grid box mean SW flux) on VZA (x-axis) and (b) a joint distribution of the CERES grid box SW flux anomalies stratified by cloud fraction and VZA. Black dots represent the statistically significant mean differences at 95% confidence with the box to the adjacent to the right.