# Comparison of TOA and BOA LW radiation fluxes inferred from ground-based sensors, A-Train satellite observations and ERA reanalyses at the High Arctic Station Eureka over the 2002 to 2020 period

Yann Blanchard[1], Jacques Pelon[2], Christopher J Cox[3], Julien Delanoë[4], Edwin W Eloranta[5], and Taniel Uttal[6]

[1]ESCER Centre
[2]LATMOS/IPSL
[3]CIRES/NOAA-ESRL
[4]Laboratoire ATmosphere, Milieux, et Observations Spatiales (LATMOS), IPSL/UVSQ/CNRS
[5]University of Wisconsin
[6]NOAA ETL

November 22, 2022

**Abstract**

This paper focuses on the accuracy of longwave radiation flux retrievals at the top and bottom of the atmosphere at Eureka station, Canada, in the high Arctic. We report comparisons between seven products derived from (1) calculations based on a combination of ground-based and space-based lidar and radar observations, (2) standard radiometric observations from the CERES satellite, (3) direct observations at the surface from a broadband radiation station and (4) the ERA-Interim and ERA5 reanalyses. Statistical, independent analyses are first performed to look at recurring bias and trends in fluxes at Top and Bottom of the Atmosphere. The analysis is further refined comparing fluxes derived from coincident observations decomposed by scene types. Results show that radiative transfer calculations using ground-based lidar-radar profiles derived at Eureka agree well with TOA LW fluxes observed by CERES and with BOA LW fluxes reference. CloudSat-CALIPSO also show good agreement with calculations from ground-based sensor observations, with a relatively small bias. This bias is shown to be largely due to low and thick cloud occurrences that the satellites are insensitive to owing to attenuation from clouds above and surface clutter. These conditions of opaque low clouds, cause an even more pronounced bias for CERES BOA flux calculation in winter, due to the deficit of low clouds identified by MODIS. ERA-I and ERA5 fluxes behave differently, the large positive bias observed with ERA-I is much reduced in ERA5. ERA5 is closer to reference observations due to a better behaviour of low and mid-level clouds.

# Comparison of TOA and BOA LW radiation fluxes inferred from ground-based sensors, A-Train satellite observations and ERA reanalyses at the High Arctic Station Eureka over the 2002 to 2020 period

**Yann Blanchard[1], Jacques Pelon[2], Christopher J. Cox[3,4], Julien Delanoë[2], Edwin W. Eloranta[5] and Taniel Uttal[4]**

[1]Centre pour l'Étude et la Simulation du Climat à l'Échelle Régionale (ESCER), Université du Québec à Montréal, Montréal, Québec, Canada.

[2]Laboratoire Atmosphères, Milieux, Observations Spatiales, SU-UVSQ-CNRS, Paris, France.

[3]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, National Oceanic and Atmospheric Administration (NOAA), Boulder, Colorado, USA.

[4]Physical Sciences Division, NOAA, Boulder, Colorado, USA.

[5]Space Science and Engineering Center, University of Wisconsin–Madison, Madison, Wisconsin, USA.

Corresponding author: Yann Blanchard (yann.blanchard@usherbrooke.ca)

**Key Points:**

- Top-Of-Atmosphere flux observations, model calculations and reanalyses agree within the measurement uncertainty at a High Arctic Station
- Large Bottom-Of-Atmosphere biases are observed among datasets and are mainly due to incorrect vertical representation of low opaque clouds
- Ground-based active sensors are essential to complement space-based cloud observations in order to better understand Arctic climate change

## Abstract

This paper focuses on the accuracy of longwave radiation flux retrievals at the top and bottom of the atmosphere at Eureka station, Canada, in the high Arctic. We report comparisons between seven products derived from (1) calculations based on a combination of ground-based and space-based lidar and radar observations, (2) standard radiometric observations from the CERES satellite, (3) direct observations at the surface from a broadband radiation station and (4) the ERA-Interim and ERA5 reanalyses. Statistical, independent analyses are first performed to look at recurring bias and trends in fluxes at Top and Bottom of the Atmosphere. The analysis is further refined comparing fluxes derived from coincident observations decomposed by scene types. Results show that radiative transfer calculations using ground-based lidar-radar profiles derived at Eureka agree well with TOA LW fluxes observed by CERES and with BOA LW fluxes reference. CloudSat-CALIPSO also show good agreement with calculations from ground-based sensor observations, with a relatively small bias. This bias is shown to be largely due to low and thick cloud occurrences that the satellites are insensitive to owing to attenuation from clouds above and surface clutter. These conditions of opaque low clouds, cause an even more pronounced bias for CERES BOA flux calculation in winter, due to the deficit of low clouds identified by MODIS. ERA-I and ERA5 fluxes behave differently, the large positive bias observed with ERA-I is much reduced in ERA5. ERA5 is closer to reference observations due to a better behaviour of low and mid-level clouds.


## Plain Language Summary

Satellite and reanalysis datasets are widely used for climate and process studies in the Arctic in order to complement sparse ground-based measurements. This study compares ground-based observations of Arctic clouds and longwave fluxes at a Canadian High Arctic station with satellite and reanalysis products. Both statistical and coincident analyses show a good top of the atmosphere agreement, but reveal biases in surface fluxes that are due to the underestimation of the occurrence of low and thick clouds, frequent in the Arctic. The results allow for an evaluation of flux product uncertainties and for an assessments of their limitations. The outcomes of this study can be applied over the entire Arctic region and can inform the instrumentation choices at various polar ground-based sites.

## 1 Introduction

Interest in the Arctic climate has increased as the effects of global warming have begun to manifest in the region over the several decades (IPCC, 2013). These manifestations include increases in surface temperature that are larger than those observed at lower latitudes (McBean, 2005; Comiso et Hall, 2014); significant decreases in sea-ice extent and thickness (Palm et al., 2010, Serreze and Barry, 2011; Lang et al., 2017); and changes in Arctic cloud cycle and interactions (Sedlar et al., 2011; Liu et al., 2012; Abe et al., 2016). Large scale meteorological dynamical forcings on a more fragile sea-ice interface impact surface energy budgets and modify ice properties. Transport of aerosols (Rahn, 1981, Ancellet et al., 2014, Igel et al., 2017) and larger water vapor intrusions from lower latitudes (Doyle et al., 2011, Boisvert et al., 2015; Liu et al., 2018) can affect cloud cycle and precipitation, as well as cloud radiative effects (Cox et al., 2015). Cloud radiation, especially from low-level clouds is a key component of the energy budget at the surface (Serreze and Barry, 2014; Sedlar et al., 2011; Sedlar et al, 2012; Shupe et al., 2013, English et al., 2015), and such clouds are directly impacted by the aforementioned atmospheric variability. A better understanding of feedbacks controlling Arctic change and the need for improved models (English et al., 2015, Kay et al., 2016, Li and Xu, 2020) emphasize the need to better constrain models with observations. To achieve this, particular attention must be given to the autumn-winter-spring period, during which transport of warmer and moister air masses from mid-latitudes may enhance sea-ice decline (Graham et al., 2017), including through modulation of longwave and shortwave cloud effects (Cox et al. 2016).

There are only a small number of surface land stations in the pan-Arctic region dedicated to atmospheric research (Uttal et al., 2016), and only a subset of these regularly make measurements using active instrumentation such as radar and lidar, which are necessary to retrieve cloud properties with vertical resolution. These retrieved profiles are valuable for understanding the vertical distribution and properties of cloud layers necessary to accurately model radiative transfer through the atmosphere (Shupe et al., 2013; Shupe et al., 2015a; Shupe et al., 2015b). The stations are located over land and many are coastal. Thus, data may be subject to spatial heterogeneity characteristic of such environments (e.g., orographic effects, specific atmospheric or ocean circulation flows, variable surface reflectivity) and so may not be representative at the regional scale (Eastman and Warren 2010, Shupe et al. 2011a). New stations (drifting buoys) are being implemented over the Arctic ocean (Provost et al., 2015; Mariage et al., 2017), that should bring new information on aerosol and cloud profiles as well as the Surface Radiation Budget (SRB), together providing a regional support to characterize SRB in combination with space observations. However, observations from Clouds and the Earth's Radiant Energy System–Energy Balanced and Filled (CERES-EBAF) (Loeb et al., 2009) have long been the only available radiative flux information over the Arctic Ocean. The advent of polar-orbiting satellite active sensors, with the success of CALIPSO/CloudSat missions (Stephens et al., 2018), allows for a more precise estimation of the regional Arctic cloud cover, quantification of cloud type vertical distribution, and inference of radiative fluxes at the regional scale (Kay and L'Ecuyer, 2013; Kay et al., 2016). The upcoming EarthCARE mission is designed to pursue and reinforce this progress through a continued instrumental synergy (Illingworth et al., 2015). However, while satellites provide the spatial coverage lacking from the surface stations, they do not directly observe the surface radiation budget and so must be validated.

Extensive characterization of Arctic SRB therefore necessitates a combination of the ground-based and satellite retrievals and a more accurate evaluation of all biases through comprehensive intercomparisons between observations. Previous work emphasized that the use of passive instruments (e.g. MODIS) alone in insufficient because of underestimation of cloud fraction in winter and autumn (Liu, 2010, Blanchard et al., 2014, hereinafter B14). As cloud products from satellite are commonly used to contribute to atmospheric reanalysis and to compute cloud radiative forcing, errors in cloud detection or biases in cloud products, as shown in Liu and Key (2016), may lead to errors in flux calculations. Consequently, B14 concluded that spaceborne lidar–radar synergy is essential for a complete representation of the cloud vertical profile, but that both surface and space observations are needed to reduce biases in all observations. Near-surface clouds are frequent in central Arctic (Uttal et al., 2002; Mariage et al, 2017). Below about 1 km in altitude, space-based radar observations are inhibited by ground clutter (Palerme et al., 2019). Conversely, lidar sensitivity may be limited below clouds by attenuation, enhanced in presence of supercooled layers at cloud top. B14 found that the characterization of low clouds as well as boundary layer events (composed of aerosols and/or precipitating ice crystals) are two of the principal challenges for spaceborne observations and the determination of radiation fluxes at the surface.

Longwave (LW) radiation is an important component of the energy budget in the Arctic and is indeed the only radiative flux during polar night. LW is additionally particularly sensitive to the profile of atmospheric and cloud properties and therefore products such as CERES-EBAF are very sensitive to errors in the profiles from which the calculations of the fluxes are made. In this paper, we focus on the retrieval of LW radiation fluxes both at the top of atmosphere (TOA) and at the bottom of atmosphere (BOA), as observed from the ground and from space and simulated from radiative transfer models over Eureka, Nunavut, Canada (80 °N, 86 °W). Eureka is a high-Arctic surface observatory with 5 years of overlapping measurements from the necessary instrumentation and is representative of a particularly dry region of the Arctic (Cox et al. 2012) where clouds are distributed over a wider range of heights than other locations (Shupe et al. 2011a). The instrumentation and the time series of records at Eureka site has allowed a significant number of studies related to climate (Lesins et al., 2010; Cox et al., 2012), comparisons of cloud cover (de Boer, 2009; Shupe, 2011a, 2011b, B14) and downwelling fluxes (Cox et al., 2012). According to this latter study, the yearly average downwelling LW cloud radiative effect (difference of cloudy and clear air downwelling fluxes) at the bottom of the atmosphere (BOA) at Eureka is about 27 $W/m^2$. The aim of this work is to analyze radiative flux comparisons following an approach similar to the one developed in B14. Namely, we perform two main analyses on a statistical basis using independent and coincident observations involving vertical profiles and cloud retrievals from a synergistic use of lidar and radar data. We interpret results generally to draw conclusions applicable to the performance of the products under particular atmospheric regimes.

We first present upwelling and downwelling LW fluxes derived from satellite and surface observations, including calculations using cloud profile measurements at Eureka. All observations are compared to the ERA-Interim and ERA5 reanalyses of the European Center for Medium Weather Forecast (ECMWF). Comparisons of seasonal variations of fluxes from statistical analyses and cloud vertical distribution based on independent datasets are detailed in section 3. On the basis of the coincident data, comparison of flux distributions and their differences are then discussed in section 4. Finally, we discuss the results of the comparisons and identify biases and limitations.

## 2 Description of observation site and datasets

The focus surface observation site are the Zero Altitude PEARL Auxiliary Laboratory (0PAL) and the Surface and Atmospheric Flux, Irradiance and Radiation Extension (SAFIRE), both part of the Polar Environment Atmospheric Research Laboratory (PEARL; Fogal et al., 2013) in Eureka, Nunavut, Canada, which is one of the high-latitude stations of the Network for the Detection of Atmospheric Composition Change (NDACC, http://www.ndsc.ncep.noaa.gov/sites/stat_reps/eureka/). It is also part of the IASOA network (Uttal et al. 2016), and, located at SAFIRE, a World Radiation Monitoring Center Baseline Surface Radiation Network (WCRP-BSRN, http://bsrn.awi.de/) station during the study period, 2007 and 2011 (Dreimel et al. 2018). Ground-truthing of satellite studies is one of the principle objectives of BSRN. Note that while the radiosonde launch facility, as well as the lidar and radar instruments are located close to sea level at 0PAL, SAFIRE is located approximately 5 km northeast at 85 m (asl). Eureka is situated in the northernmost part of the Canadian Arctic Archipelago, a region having complex topography and variable surface type. However, despite this heterogeneity, the station offers a critical mass of observations and because of its latitude also a high frequency of satellite overpasses, enabling a larger number of coincident samples to be analyzed (B14). Figure 1 shows a map of Ellesmere and Axel Heiberg islands with the location of the Eureka station marked on the western coast of Ellesmere as well as a photograph of the BSRN station highlighting flat, open area chosen for siting SAFIRE.

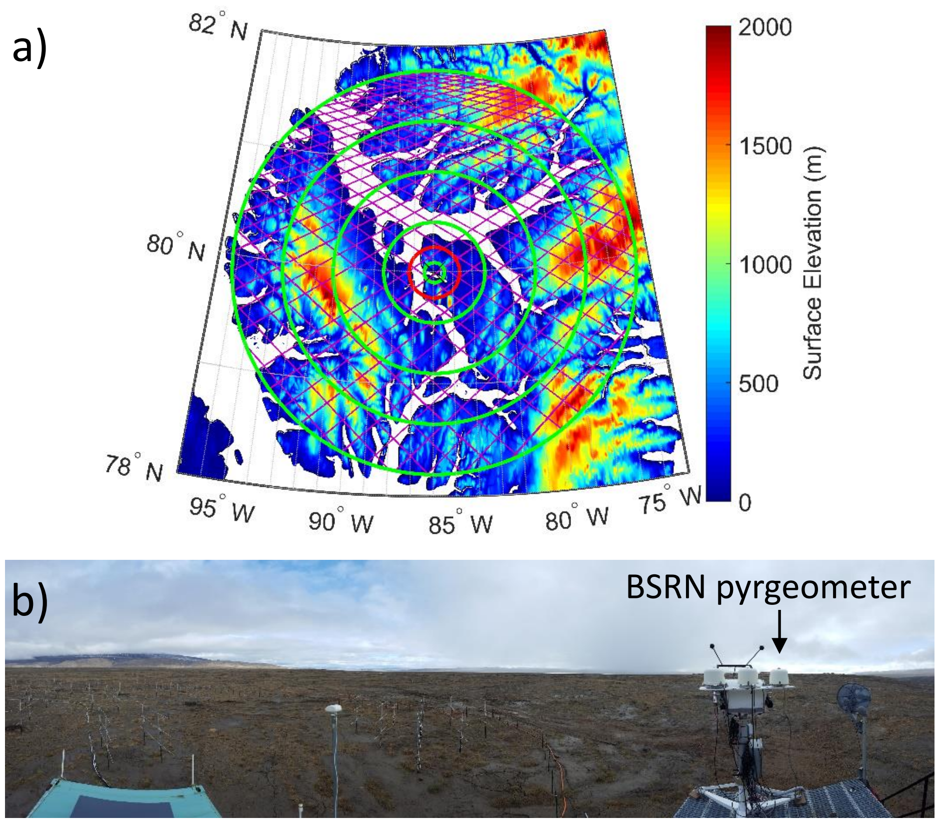



**Figure 1**. (a) A-Train tracks (in magenta) close to the Eureka station (black cross) during January 2007 superposed over the digital elevation model Global 30 arc s elevation dataset (GTOPO30) used for CALIPSO data analysis. The green concentric circles (radius of 10, 25, 50, 100, 150 and 200 km) denote the area of this study. The 25 (red) circle delimits domains where surface orography and heterogeneity are minimized. (b) Panoramic view of the Eureka radiation site (SAFIRE) where the pyrgeometer is located.

Kovacs and McCormick (2005) suggest that for cloud-comparison purposes a length scale of a few tens of kilometers and a time scale of a few minutes is sufficient for identifying coincident observations. Based on this recommendation, we will define 25 km from Eureka as the maximum distance for the current study, which is also similar to the grid size of ERA-I, ERA5 and CERES. The region where spaceborne observations are analyzed is also shown in Figure 1.

CERES and MODIS have collected measurements since 1999 and 2002 onboard TERRA and AQUA, respectively. CloudSat and CALIPSO (hereafter referred to as C-C) were launched in 2006 and are part of the constellation of satellites formed with AQUA (A-Train). CloudSat has made measurements only during daytime since 2011 due to a battery anomaly and the production of CloudSat 2B-FLXHR-LIDAR products (hereafter C-C-FLX) was discontinued at that time. Two releases of this product (R04 and R05) are however available from 2006 and almost over the same period. Although considering the whole period from 2002 to 2020 for the overall statistical analysis, the period of detailed analysis on coincident observations is limited to the overlap period spanning from June 2006 to May 2010, due to the availability of radiation products.

Table 1 summarizes the main characteristics of the datasets and methods used in this study.

**Table 1**. Satellite, ground-based and reanalysis data sets and methods used in this study

| Name | CERES | C-C-FLX | ERA-I / ERA5 | EUR-LR | BSRN |
|---|---|---|---|---|---|
| Long name | CERES_SSF_Aqua-XTRK_Edition4A | CloudSat 2B-FLXHR-LIDAR | ECMWF ERA-Interim / ERA5 | EUREKA-LIDAR-RADAR | Baseline Surface Radiation Network |
| Version | Edition 4A | Release 04 and 05 | | | |
| Temporal resolution | 2 – 4 s | 0.16 s | 3h | 3 min | 1 min |
| Vertical resolution | N/A | 240 m | 137 levels | 30 m | N/A |
| Footprint | 20 km x 20 km | 1.4 km x 1.7 km | 0.125°x 0.125° | N/A | N/A |
| Cloud properties | MODIS Collection 5 cloud products | CloudSat and CALIPSO | Reanalysis | From radar-lidar synergy | N/A |
| TOA fluxes | Observed | BugsRad RTM | Reanalysis | Streamer RTM | |
| BOA fluxes | Longwave Model B | BugsRad RTM | Reanalysis | Streamer RTM | Observed |
| References | Wielicki et al., 1996 Loeb et al., 2018 | Henderson et al., 2013 | Dee et al., 2011 Hersbach and Dee, 2016 | Donovan and van Lammeren (2001); Shupe (2007) | McArthur, 2005; Driemel et al., 2018 |

## 2.1 Fluxes from ground-based observations

Profiles of cloud properties are regularly measured above Eureka from combined radar and lidar measurements (B14). Here, we use these data as reference for heights of cloud layers over the site. TOA and BOA fluxes based on the lidar and radar measurements at Eureka (hereafter “EUR-L-R”) were calculated using the Streamer radiative transfer code (Key and Schweiger, 1988). The input parameters include atmospheric profiles (interpolated from twice daily radiosonde measurements), aerosol optical depth from the Eureka sunphotometer (part of AERONET, https://aeronet.gsfc.nasa.gov/), and cloud layer information (type of layer, altitude, layer optical depth, and mean effective radius) from ground-based lidar and radar (as detailed in Blanchard et al., 2017). Cloud type was derived from a multisensor classifier (Shupe, 2007) and particle sizes were retrieved from the ratio of radar and lidar backscatter cross-section (Eloranta et al., 2007) using default processing on the web site http://hsrl.ssec.wisc.edu. The purpose of this product is to be a comparable analog to the CloudSat and CALIPSO products, but from the perspective of the surface.

The LW flux data from the BSRN station are 1-minute averages based on 1 Hz samples collected by a shaded Eppley pyrgeometer mounted on a sun tracker (Driemel et al., 2018). Grachev et al. (2018) reported on the intercomparability of the BSRN LW for an overlap period with another radiometer approximate 700 m east of the BSRN station at Eureka and found a negligible bias (~1 W/m$^2$) and with a standard deviation of 10.5 W/m$^2$ in the differences of hourly means.

The BSRN pyrgeometer was maintained approximately at daily intervals. In cold climates this maintenance includes manual removal of ice from the sensor windows, which commonly occurs. The pyrgeometer was ventilated, which helps maintain temperature stability and mitigate the formation of ice. Unfortunately, despite these procedures, icing of the window frequently occurred on the pyrgeometer throughout the study period. Because the specific post-processing procedures used on the data archived at BSRN are undocumented, we began with the raw data set and conducted our own quality control, including implementing the procedures recommended by Long and Shi (2008) as well as visual screening for signs of icing. The signal from the iced window is similar to the signal from clouds, making it difficult to identify. For an upward-facing LW measurement, the bias caused by the ice is generally positive, and is large when the sky is clear and small when the sky is cloudy. Manual removal of ice by the technicians causes a change in the signal that is very fast compared to natural variability and this non-physical signal is easily identifiable, as is the decrease in radiance following the growth curve of the developing ice that precedes the cleaning backward in time. By identifying and removing these features, the visual screening likely removed most of the ice that occurred when the sky was radiatively clear and the bias was large, but the subsequent absence of the radiatively clear time periods in the record produces a climatological bias in the monthly means. Monthly means are only used in this study for qualitative purposes so it was more important to have a representative estimate than a direct measurement for these periods. We therefore filled the gaps from the data removed because of icing with a calculation of the clear-sky downwelling LW following Long and Turner (2008), which is based on Brutsaert's equation and requires only the radiometric measurements and collocated meteorology. Time periods that use these estimates are not incorporated into the validation analysis of this study. The subset of observations coincident with the satellite overpasses that are used for comparison received further scrutiny individually, including analysis of logbook records, radar and lidar data, meteorology and the other radiometric data in order to identify and remove additional suspect data that remained. This procedure had the added benefit of being well-suited to identify times where comparisons were likely to be influenced by cloud cover that was within the ~160º effective FOV of the pyrgeometer at times when the skies directly over Eureka were clear. Appendix A presents results from this data screening.

### 2.2 Fluxes from satellite

Radiation measurements from the CERES instrument on AQUA and TERRA provide a direct observation of the upwelling TOA radiances (Wielicki et al., 1996), that are converted into fluxes using angular distribution models that provide a stable time series (Loeb et al., 2012). Specific comparisons of the Clouds and the Earth's Radiant Energy System–Energy Balanced and Filled (CERES-EBAF) TOA fluxes (Loeb et al., 2009) have been performed at high latitudes over the Arctic (Kay et al., 2013, Huang et al., 2017b). CERES provides access to a long and homogeneous radiation database, and has been used in numerous analyses. Based on these considerations, the AQUA CERES-SSF TOA fluxes (V4.0) were taken as a reference for comparisons in the present study. The best retrieval of the LW CERES surface (BOA) fluxes is achieved using cloud properties derived from MODIS and processed to agree with observed LW TOA fluxes (Gupta et al., 1992). They have been compared to other observations and validated against surface radiation measurements (Gupta et al., 2010, Kratz et al., 2020) for mid and low latitudes. In the Single Satellite Footprint (SSF) product Level2, adding MODIS cloud retrievals and Goddard Meteorological Assimilation Office (GMAO) atmospheric profiles, the BOA fluxes at the surface are also available in a 20 km x 20 km grid. CERES BOA fluxes are used as part of the Arctic Observation and Reanalysis Integrated System (ArORIS) gathering several datasets for climate studies in this region (Christensen et al., 2016). Initial comparisons over Greenland showed small dispersion (Christensen et al., 2016) confirmed by further studies over the whole Arctic (Huang et al., 2017b).

The CLOUDSAT-2B-FLXHR-LIDAR (hereafter named as C-C-FLX) products provide a direct estimation of TOA and BOA fluxes consistent with the liquid and ice water content and the cloud vertical profiles obtained from CloudSat, CALIOP and MODIS measurements, using atmospheric profiles from ECMWF (Henderson et al., 2013, L'Ecuyer et al., 2008). The TOA and BOA flux amount are defined as the first and the last non-zero value of FU and FD parameters in the last available version product (R05) presently used from the CloudSat data center (http://www.cloudsat.cira.colostate.edu/data-products/level-2b/). In this study we use the currently available R05 products (http://www.cloudsat.cira.colostate.edu/data-products/level-2b/2b-flxhr-lidar), publicly available since March 2020, and discuss differences with previous version R04. As the CLOUDSAT-2B-FLXHR-LIDAR R05 product is expected be less prone to atmospheric biases due to cloud phase, well identified by lidar (Hu et al., 2009), we considered it to better represent clouds in flux calculations.

The C-C-FLX (R05) time series starts in June 2006 and ends in August 2010 (April 2011 for R04) while CERES on AQUA begins in July 2002. Both products are expected to differ due to factors related to cloud vertical distribution (Matus and L'Ecuyer, 2017). Whereas CERES is based on MODIS data inversion, C-C-FLX input is a direct retrieval of vertical profiles from CloudSat and CALIPSO active sensors. The cloud profiles are usually better constrained with active instruments, even if some biases remain (Chan and Comiso, 2010). Moreover, it has been shown that C-C misses some low clouds (B14). A second issue may be due to the spatial distribution as C-C-FLX fluxes are given at the radar footprints (1.4 km) along a track whereas the CERES grid is 20 km x 20 km, somewhat smoothing spatial variability. For both datasets, we will discuss in sections 4.1.1 and 4.2.1 the representativeness of taking the nearest pixels to the station versus an average of all the values located at less than 25 km from the station.

### 2.3 Fluxes from re-analysis

The ECMWF ERA-Interim (hereafter ERA-I) project is based on meteorological reanalysis that were assimilated from various datasets (Dee et al., 2011). ECMWF Integrated Forecast System uses a four-dimensional variational data assimilation (4DVar). In this study, we considered monthly average from a 0.125° x 0.125° grid interpolated from the original 80 km x 80 km special resolution, which represents approximately 14 km x 2 km at the latitude of Eureka. For coincident comparison purposes, ECMWF's ERA-I 3-hour reanalysis products were used, which corresponds to a delay of 60 and 45 minutes with A-Train overpasses near 11:00 and 15:45 UTC respectively. The ECMWF most advanced reanalysis product, ERA5, was recently released and provides several improvements compared to ERA-I, as detailed by Hersbach and Dee (2016), and uses a more advanced 4DVar assimilation scheme, and higher vertical (137 vs. 60 levels) and horizontal resolutions (31 km vs. 79 km).

## 3 Statistics from independent datasets

In this section, as well as in the next section dealing with coincident measurements, we will first analyze TOA LW fluxes followed by downwelling fluxes at the surface.

### 3.1 TOA fluxes

LW TOA monthly fluxes from CERES observations, CALIPSO-CloudSat flux calculations and reanalysis from ERA5 are shown in Figure 2 for a period which extends from 2002 to 2020.

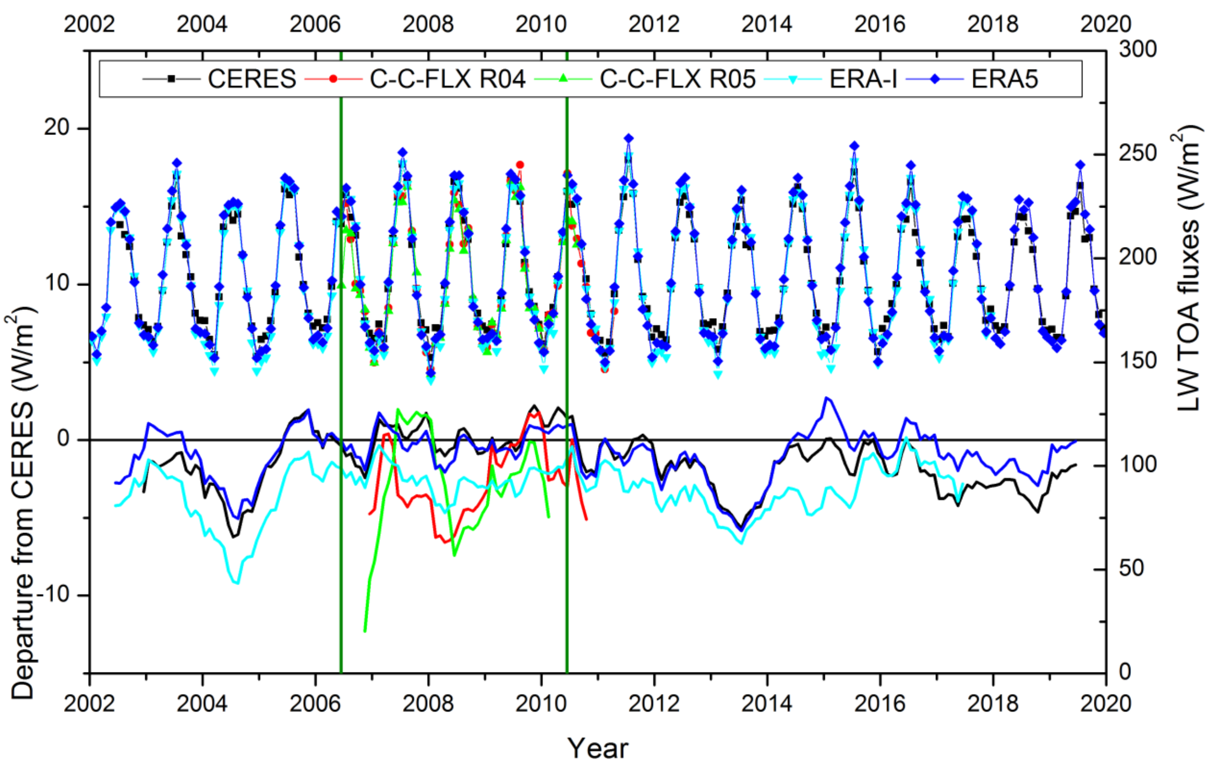



**Figure 2**. Monthly variation of TOA LW fluxes (right side scale) measured and derived from satellite (CERES in black and CloudSat-2B-FLXHR-LIDAR R04 (in red), R05 (in green) and re-

analysis (ERA-I in cyan and ERA5 in blue) at Eureka from 2002 to 2020. 12-month moving average departures (left hand scale) from 2006/06-2010/05 CERES average are shown in the bottommost graph. The common observation period of this study is bounded by vertical green lines.

Monthly variations are similar amongst the datasets, minimum values being observed in winter and maximum values in summer, with the range of the annual cycle being about 70-80 W/m$^2$, depending on dataset. ERA5 and CERES are available over the full period from 2002 to 2020, whereas C-C-FLX R05 are limited to a period of 4 years (5 years for R04), as mentioned in the previous section. Plots in the lower graph of Figure 2 show the departure of 12-month moving average for each dataset from CERES 2006-2010 multi-year mean, which is the common observation period. It is seen that C-C-FLX R05 data are biased low on average by about 5 W/m$^2$ as compared to CERES data. R04 shows similar values except the low values at the end of 2006. Two periods (respectively 2004 and 2013) in the whole CERES and ERA sequence were significantly different from the 2006-2010 average with departure larger than the overall standard deviation of 2.0 W/m$^2$. More particularly, those periods show values comparable to or larger than 3 times the standard deviation. They occurred before and after the reference period in 2004 and 2013 for both CERES and ERA5 datasets. Note that the 12-month moving average from ERA5 are generally in good agreement with those from CERES, except a small difference of about -2 W/m$^2$ since 2014. R05 data show significant differences and appear to be biased low with respect to CERES and ERA5.

Average values, seasonal variations, trends and standard deviations are reported in Table 2 for both 18 and 4-year periods. Departures from CERES, considered here as the reference for TOA LW fluxes, are shown in the second part of Table2. ERA-I statistics are also included to be discussed along with newly available ERA5.

**Table 2**. Annual and seasonal variations of LW TOA fluxes for CERES, C-C-FLX (R04 and R05), ERA-I and ERA5 over the whole dataset period and coincident period (July 2006 to May 2010) based on monthly means. Linear trends are bolded when considered significant (more than 2 sigma). Colour shading is representative of the difference with the reference (red when they are above CERES (darker) by more than 5W/m$^2$, blue below, green is within 2 W/m$^2$)

| TOA | Time period | **Annual mean** | | **DJF** | | **MAM** | | **JJA** | | **SON** | | **ONDJFM** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean (σ) | Trend (W/m² / year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) |
| STATISTICAL ANALYSIS OVER THE WHOLE OBSERVATION PERIOD | | | | | | | | | | | | | |
| CERES | 07/ 2002 - 12/ 2019 | 191.2 (2.0) | -0.1 (0.1) | 164.3 (2.8) | 0.1 (0.1) | 187.6 (3.0) | 0.1 (0.2) | 226.3 (6.6) | -0.5 (0.3) | 186.9 (2.0) | -0.1 (0.1) | 169.3 (1.8) | 0.0 (0.1) |
| C-C-FLX R04 | 07/ 2006 - 04/ 2011 | 189.4 (2.4) | 0.8 (1.2) | 158.5 (4.8) | 0.8 (1.7) | 184.0 (4.0) | **3.0 (0.6)** | 226.4 (8.5) | 1.5 (3.0) | 184.7 (2.8) | -0.4 (1.0) | 165.6 (3.2) | 0.3 (1.1) |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C-C-FLX R05 | 06/ 2006 - 08/ 2010 | 190.7 (4.0) | -0.1 (2.2) | 158.3 (5.5) | 3.2 (2.0) | 183.4 (4.5) | **3.2 (1.0)** | 220.0 (11.8) | 2.4 (4.1) | 188.4 (10.1) | -1.2 (5.4) | 165.9 (2.5) | 1.7 (0.6) |
| ERA-I | 01/ 2002 - 12/ 2017 | 189.4 (2.0) | 0.1 (0.1) | 155.9 (3.0) | -0.1 (0.2) | 184.6 (2.9) | 0.3 (0.1) | 229.9 (4.7) | 0.1 (0.3) | 187.0 (2.5) | 0.0 (0.1) | 163.5 (2.4) | 0.1 (0.1) |
| ERA5 | 01/ 2002 - 12/ 2019 | 191.7 (1.8) | 0.0 (0.1) | 159.2 (2.7) | 0.0 (0.1) | 189.2 (3.0) | 0.1 (0.1) | 232.5 (5.8) | 0.0 (0.3) | 185.8 (2.6) | -0.1 (0.1) | 165.1 (2.1) | 0.0 (0.1) |
| STATISTICAL ANALYSIS OVER THE COMMON OBSERVATION PERIOD | | | | | | | | | | | | | |
| | Time period | Annual mean | | DJF | | MAM | | JJA | | SON | | ONDJFM | |
| | | Mean (σ) | Minus CERES | Mean (σ) | Minus CERES | Mean (σ) | Minus CERES | Mean (σ) | Minus CERES | Mean (σ) | Minus CERES | Mean (σ) | Minus CERES |
| CERES | 06/ 2006 - 05/ 2010 | 192.7 (2.1) | | 164.8 (2.5) | | 187.7 (4.5) | | 230.7 (6.4) | | 187.4 (1.1) | | 169.5 (2.3) | |
| C-C-FLX R04 | 06/ 2006 - 05/ 2010 | 188.8 (4.3) | -3.8 (2.5) | 159.3 (5.1) | -5.6 (3.1) | 184.0 (4.0) | -3.7 (3.3) | 227.4 (9.4) | -4.0 (5.9) | 185.2 (2.9) | -2.2 (3.4) | 166.1 (3.4) | -3.4 (1.2) |
| C-C-FLX R05 | 06/ 2006 - 05/ 2010 | 188.6 (6.2) | -5.4 (3.2) | 158.3 (5.5) | -6.1 (2.7) | 183.4 (4.5) | -4.3 (3.2) | 221.4 (13.1) | -9.3 (7.5) | 188.4 (10.1) | -1.3 (6.2) | 165.9 (2.5) | -3.6 (0.5) |
| ERA-I | 06/ 2006 - 05/ 2010 | 190.1 (0.6) | -2.6 (1.5) | 156.4 (3.1) | -8.4 (1.1) | 184.1 (2.6) | -3.6 (2.1) | 231.9 (4.5) | 1.2 (2.4) | 187.9 (2.5) | 0.5 (1.6) | 163.3 (1.8) | -6.1 (1.4) |
| ERA5 | 06/ 2006 - 05/ 2010 | 192.3 (1.2) | -0.4 (0.9) | 159.3 (3.3) | -5.5 (0.9) | 188.7 (3.1) | 1.0 (1.5) | 234.9 (6.1) | 4.2 (1.0) | 186.2 (2.5) | -1.2 (1.5) | 165.0 (2.2) | -4.4 (1.1) |

Table 2 shows that it is also the case during the selected intensive common observation period, where both C-C-FLX releases are about 5 W/m² smaller than CERES in winter and less than 4 W/m² the rest of the seasons. The difference of 5.3 W/m$^2$ in JJA between C-C-FLX R04 and R05 may be explained by more data available in R05 in 2006 summer, but also by the longwave land emissivity that varies by surface type in R05 (Henderson and L'Ecuyer, 2020). While C-C-FLX is always smaller than CERES. The statistics over the whole observation period show that ERA-I has slightly smaller TOA LW fluxes than CERES but that ERA5 are similar to CERES and closer on the annual means. Both ERA datasets are close to CERES, except in winter where they are closer to C-C-FLX showing an 8.4 W/m² and 5.5 W/m² deficit relative to CERES. The use of ERA5 over ERA-I (with developments in model physics, core dynamics, assimilation system, higher spatial and temporal resolution) leads to an increase in LW fluxes by several W/m² and significantly reduces the bias with CERES, except in summer where the bias is increased. Autumn shows the opposite behaviour with decreased fluxes in ERA5, and degraded agreement. Note that the warm ground temperature bias in ERA reanalyses (Wang et al., 2019) is not relevant in our study because weather observations at Eureka are assimilated.

CERES, ERA-I and ERA5 do not show significant trends in LW TOA fluxes in the last eighteen years, neither on average nor on a seasonal analysis. The trends are significant only for C-C-FLX dataset in spring when trend is larger than 2 sigma. C-C-FLX annual data show a much larger variability between 2006 and 2010, as compared to the three other datasets that have a larger footprint.

To summarize, TOA averages over the whole periods appear to be in good agreement for all datasets (within about 5 W/m²). Only small differences (CERES can be larger by about 6 to 8 W/m² than C-C-FLX, ERA5 and ERA-I in DJF) are observed on seasonal TOA fluxes, and these differences are consistent with previous work (Loeb et al., 2018).

### 3.2 BOA fluxes

Figure 3 presents the comparison of the LW BOA downwelling fluxes from CERES, ERA-I and ERA5 over the full 2002-2020 period, and along with the C-C-FLX, the reference ground-based broad-band radiation dataset over a more restricted period of time. Table 3 gives the yearly and seasonal average values as well as the estimated trends.

Note that the screening method described in section 2.1 and Appendix A helps to remove suspicious measurements (184572 cases representing about 8% of the whole dataset). But this filtering resulted in an under-representation of clear sky and then higher seasonal LW (up to 2.8 $W/m^2$ in winter, see Table 3). Therefore, those values are not expected to be used for climatological analysis, but as a point of comparison with other datasets.

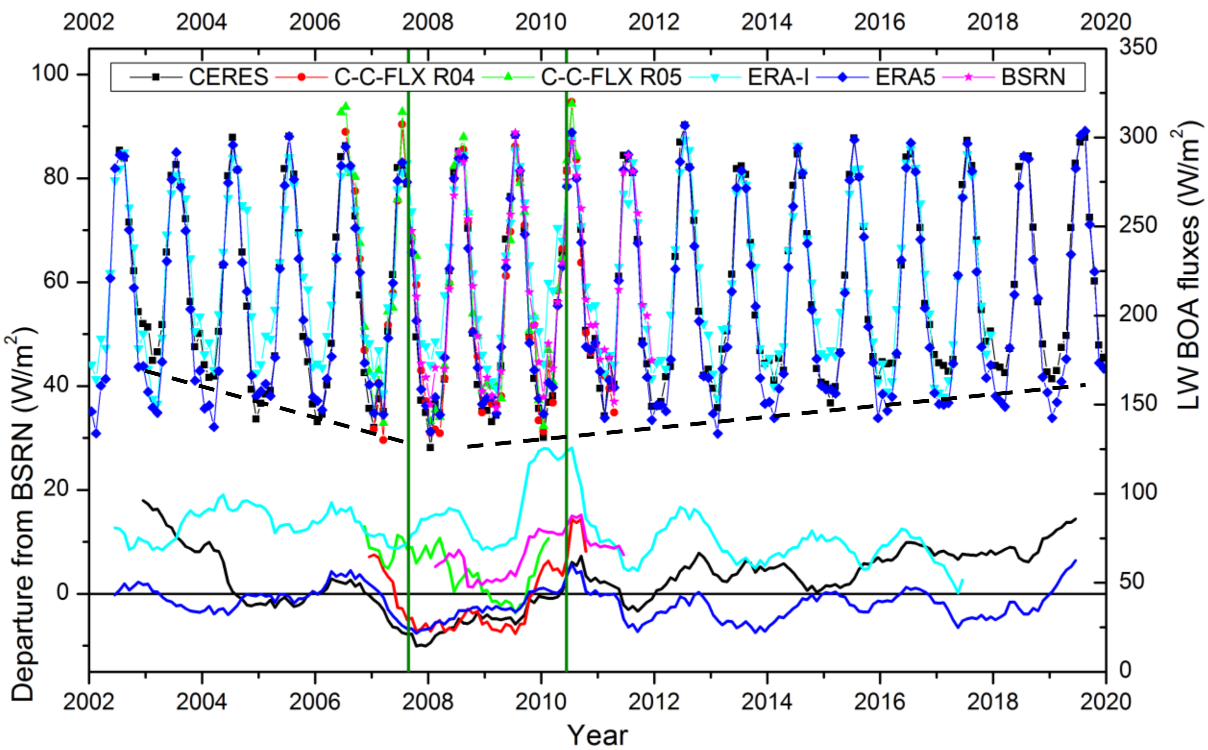



**Figure 3**. Monthly variation of BOA LW fluxes (right side scale) measured and derived from the ground (filtered BSRN), from satellite (CERES and CloudSat-2B-FLXHR-LIDAR R04 and R05) and re-analysis (ERA-I and ERA5) at Eureka from 2002 to 2020. Black dashed lines represent estimated trends on CERES annual minimum. 12-month moving average departures (left hand scale) from 2007/09-2010/05 BSRN average are shown in the bottommost graph. The common observation period of this study is bounded by vertical green lines.

Figure 3 shows a good agreement between all LW BOA fluxes for the summer, where annual cycle maxima among datasets vary within 5 W/m$^2$, except for C-C-FLX R05 between 2006 and 2010, where it exceeds 20 W/m$^2$. Large differences between the datasets are observable in winter with divergence at minimum as large as 30 W/m$^2$ in 2003. Variation of winter fluxes for CERES shows a V-shape trend (see Fig. 3). The average decrease is about 30 W/m$^2$ from 2002 to the middle of the period analyzed e.g. between 2008 and 2009, and an increase after (representing about 20 W/m$^2$). This increase in CERES fluxes may be due to the increase in low cloud winter temperature or changes in low cloud fraction over this part of the Arctic. This trend is however not seen on ERA5 (and ERA-I) datasets as was evidenced in meteorological trends observed in Arctic (Jun et al., 2016, Graham et al., 2019a). This V-shape trend observed in winter over the 18-year period tends to reduce the overall trend as reported in Table 3. Some residual low frequency (about 10 years) modulation is apparent, in particular during the winter months, with a small peak-to-peak amplitude (about 3 W/m$^2$). BSRN data tend to agree with the uniformity of LW BOA fluxes measured in summer at Eureka (standard deviation of 1.8 W/m$^2$ between 2007 and 2011, see Table 3), and the low values (~165 W/m$^2$) measured in the winters from 2008 to 2010.

From Table 3 (see also Fig. 3), it is seen that CERES statistics over the 18-year period is about 5 W/m$^2$ smaller than BSRN averages, but over the common period, this difference increases to up to 10 W/m$^2$. C-C-FLX is also smaller over this period but R05 shows reduced bias in all seasons compared to R04. Major changes were made in C-C-FLX R05 to improve the representation of cloud properties (cloud detection, supercooled liquid and ice clouds microphysical properties) along with updated data ingested. The annual difference is still lower than BSRN by 2.4 W/m$^2$ over the common observation period and is mainly attributed to differences observed in winter (DJF) when differences are about 10 W/m$^2$, consistent with the aforementioned sampling limitations in the BSRN during the icing season. C-C-FLX R04 shows higher bias during the polar night (ONDJFM) with peak differences being observed between C-C-FLX R04 and BSRN in winter that reach -17.3 W/m$^2$.

The reanalysis averages reported in Table 3 show interesting features with a correction of fluxes in ERA5, that reduces annual fluxes by 14 W/m² on average with respect to ERA-I. This reduction in LW downwelling fluxes occurs almost all year, except in June and July. As for TOA comparison, several reasons could explain the better performance of ERA5, e.g. a more detailed data assimilation system with higher vertical resolution and better surface and radiation models. Although ERA5 is closer to BSRN than ERA-I, it still underestimates LW BOA by ~15 W/m² in ONDJFM, relatively to BSRN measurements. As seen from Fig. 3, ERA5 and BSRN are in good agreement over the period of minimal winter fluxes, but although in good agreement with CERES in summer, ERA5 does not show winter flux increases seen by CERES.

**Table 3**. Annual and seasonal variations of LW BOA fluxes for CERES, C-C-FLX (R04 and R05), ERA-I, ERA5 and BSRN over the whole dataset period and coincident period (September 2007 to May 2010) based on monthly means. Linear trends are bolded when considered significant (more than 2 sigma). Colors indicate differences with BSRN fluxes taken as a reference (orange when they are above BSRN by more than 2 W/m$^2$, blue below -2 W/m$^2$, green in between). Darker colors are used above +/- 10 W/m$^2$).

| BOA | Time period | **Annual mean** | | **DJF** | | **MAM** | | **JJA** | | **SON** | | **ONDJFM** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean (σ) | Trend (W/m² / year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) | Mean (σ) | Trend (W/m² /year) |
| STATISTICAL ANALYSIS OVER THE WHOLE OBSERVATION PERIOD | | | | | | | | | | | | | |
| | 09/ 2007 - 12/ 2011 | 216.1 (3.7) | 1.0 (1.9) | 168.9 (11.0) | **7.7 (2.6)** | 189.5 (10.0) | -1.5 (5.4) | 282.4 (1.8) | 1.0 (0.7) | 219.0 (4.9) | 2.3 (1.2) | 178.9 (7.6) | **5.4 (1.8)** |
| unfiltered | 09/ 2007 - 12/ 2011 | 214.8 (3.7) | 0.6 (2.0) | 166.1 (8.4) | **5.9 (2.0)** | 188.8 (10.0) | -2.0 (5.3) | 282.4 (1.8) | 1.0 (0.7) | 217.2 (5.0) | **2.5 (1.1)** | 176.2 (6.5) | **4.5 (1.6)** |
| C-C-FLX R04 | 07/ 2006 - 04/ 2011 | 205.9 (6.5) | 3.0 (2.8) | 152.7 (12.8) | 6.0 (3.2) | 182.0 (12.9) | 6.9 (5.1) | 286.0 (9.2) | -0.3 (3.4) | 211.0 (10.1) | **-5.5 (2.0)** | 165.1 (7.1) | 2.2 (2.3) |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C-C-FLX R05 | 07/ 2006 - 08/ 2010 | 211.4 (7.0) | **-6.6 (2.5)** | 155.5 (7.9) | 0.4 (4.3) | 185.8 (11.7) | 6.6 (4.4) | 289.9 (13.0) | -2.5 (4.5) | 227.1 (12.7) | **-8.6 (3.4)** | 171.2 (4.1) | -1.1 (2.1) |
| CERES | 07/ 2002 - 12/ 2019 | 211.2 (6.1) | 0.5 (0.3) | 160.2 (14.5) | 0.7 (0.7) | 192.1 (7.1) | 0.1 (0.4) | 286.0 (4.4) | 0.3 (0.2) | 208.1 (8.1) | 0.3 (0.4) | 170.1 (12.1) | 0.6 (0.6) |
| ERA-I | 01/ 2002 - 12/ 2017 | 220.0 (6.0) | -0.4 (0.3) | 174.9 (9.3) | -0.9 (0.5) | 201.4 (10.8) | -0.2 (0.6) | 279.4 (5.2) | 0.1 (0.3) | 224.4 (7.8) | **-1.1 (0.3)** | 186.8 (7.6) | -0.7 (0.4) |
| ERA5 | 01/ 2002 - 12/ 2019 | 206.5 (3.6) | 0.0 (0.2) | 153.5 (5.4) | -0.2 (0.3) | 185.0 (4.7) | 0.1 (0.2) | 284.1 (4.9) | 0.1 (0.2) | 203.6 (6.6) | -0.1 (0.3) | 164.0 (4.2) | -0.1 (0.2) |
| STATISTICAL ANALYSIS OVER THE COMMON OBSERVATION PERIOD | | | | | | | | | | | | | |
| | | Annual mean | | DJF | | MAM | | JJA | | SON | | ONDJFM | |
| | Time period | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN |
| filtered | 09/ 2007 - 05/ 2010 | 205.8 (16.2) | | 163.9 (6.1) | | 192.4 (10.0) | | 280.9 (1.1) | | 217.5 (6.4) | | 175.8 (5.2) | |
| BSRN unfiltered | 09/ 2007 - 05/ 2010 | 204.7 (16.1) | -1.1 (0.2) | 162.5 (5.4) | -1.4 (0.7) | 192.0 (9.4) | -0.4 (0.7) | 280.9 (1.1) | 0.0 (0.0) | 215.3 (6.1) | -2.2 (0.6) | 173.6 (4.7) | -2.1 (0.6) |
| C-C-FLX R04 | 09/ 2007 - 05/ 2010 | 195.9 (18.7) | -8.3 (2.2) | 146.6 (5.3) | -17.3 (6.8) | 183.0 (15.7) | -3.1 (3.1) | 279.5 (9.7) | -1.4 (8.7) | 209.4 (2.1) | -8.2 (7.0) | 160.7 (3.4) | -13.5 (3.9) |
| C-C-FLX R05 | 09/ 2007 - 05/ 2010 | 203.1 (15.2) | -2.4 (4.4) | 153.7 (8.5) | -9.7 (9.3) | 187.1 (14.0) | 0.4 (2.4) | 281.1 (15.3) | 0.2 (14.2) | 223.6 (13.1) | -0.1 (8.7) | 169.5 (3.0) | -4.7 (5.6) |
| CERES | 09/ 2007 - 05/ 2010 | 193.9 (20.0) | -10.8 (4.0) | 142.8 (2.1) | -21.2 (6.3) | 187.8 (10.5) | 0.8 (3.8) | 284.9 (3.3) | 4.0 (2.3) | 198.2 (3.6) | -19.3 (2.9) | 154.1 (2.6) | -21.3 (4.2) |
| ERA-I | 09/ 2007 - 05/ 2010 | 218.6 (15.1) | 13.8 (5.9) | 181.6 (12.0) | 17.6 (6.1) | 212.6 (21.0) | 25.4 (12.5) | 279.3 (4.4) | -1.6 (3.3) | 225.9 (4.4) | 8.4 (2.3) | 191.1 (11.4) | 16.0 (6.0) |
| ERA5 | 09/ 2007 - 05/ 2010 | 196.4 (18.4) | -8.1 (2.7) | 150.8 (4.0) | -13.2 (5.9) | 188.3 (8.2) | 2.2 (6.9) | 283.8 (1.3) | 2.9 (0.2) | 198.9 (1.6) | -18.6 (4.8) | 159.9 (3.1) | -15.1 (4.2) |

As a first conclusion on these seasonal average analysis of BOA fluxes (Table 3), CERES and C-C-FLX averages derived from space observations and ERA5 reanalysis are in rather good agreement although about 10 $W/m^2$ smaller than BSRN. They can even be larger than -20 $W/m^2$ during the polar night (ONDJFM). Conversely, ERA-I appears to be biased high with respect to all observations except in JJA, where all results are in agreement within 4 $W/m^2$. The observed bias of ERA-I is coherent with previous analyses, where an over-estimation of low-cloud cover causes higher LW BOA in winter, whereas ERA-I LW are subject to a dry bias in summer (Zygmuntowska et al., 2012; Zib et al., 2012; Chernokulsky and Mokhov, 2012; Lenaerts et al., 2017; Huang et al., 2017a). Although relatively few studies with ERA5 evaluation in the Arctic are available, it seems that several biases of ERA-I are better addressed in ERA5, in terms of representation of temperature and humidity profiles and wind speed near surface (Graham et al., 2019a; Graham et al., 2019b; Betts et al., 2019). However it is not clear if low cloud fraction is better represented.

In order to further analyze the origin of these differences, we come back to cloud vertical information as it was identified in B14 as a source of difference in sensitivities.

### 3.3 Cloud vertical structure and type at Eureka

Over the June 2006 to May 2010 period, the number of vertical profiles was 329,204 for EUR-LR (1 profile every 3 minutes), 57,976 for MODIS on AQUA (taking all the pixels whose center is less than 25 km from Eureka), 16,927 for DARDAR (481 overpasses; this dataset, DARDAR-MASK-v1.1.4 is based on CloudSat and CALIPSO synergy, as described in Delanoë and Hogan, 2010 and Ceccaldi et al., 2013), 17,024 for CloudSat-CLDCLASS-LIDAR R05 (labelled as C-C in Fig.4) and 5,844 reanalyses for ERA-I and ERA5.

DARDAR, C-C and EUR-LR have similar vertical distributions of cloud layers above 3 km. As detailed in B14 and Liu et al. (2017), very low clouds are difficult to address from space, and this is confirmed here from DARDAR, C-C and MODIS for which cloud fractions are much lower than EUR-LR ground-based observations below 2 km, as evidenced in Fig. 4. Compared to EUR-LR, DARDAR and C-C are close in all seasons, except for high clouds ($z > 8$ km) in autumn and winter (that may be due to the use of a better vertical resolution in DARDAR, which is sampled at CALIOP vertical resolution). DARDAR and C-C give close results although DARDAR gives a higher amount of ice clouds and less mixed-phase clouds (see Appendix B). We find that the cloud fraction in reanalysis is generally biased low below 8 km, which is consistent with the findings of Liu and Key (2016) for a larger region of the Arctic.

In general, there is good agreement in the vertical profiles of cloud fraction excepting MODIS, ERA5 and ERA-I, which are systematically smaller than the other datasets (and this is particularly the case for ERA5). In spring, although slightly smaller than ground-based observations, cloud fractions from all sources agree above 5 km, but significant discrepancies are observed below this altitude and even more below 2 km, in agreement with previous findings from B14.

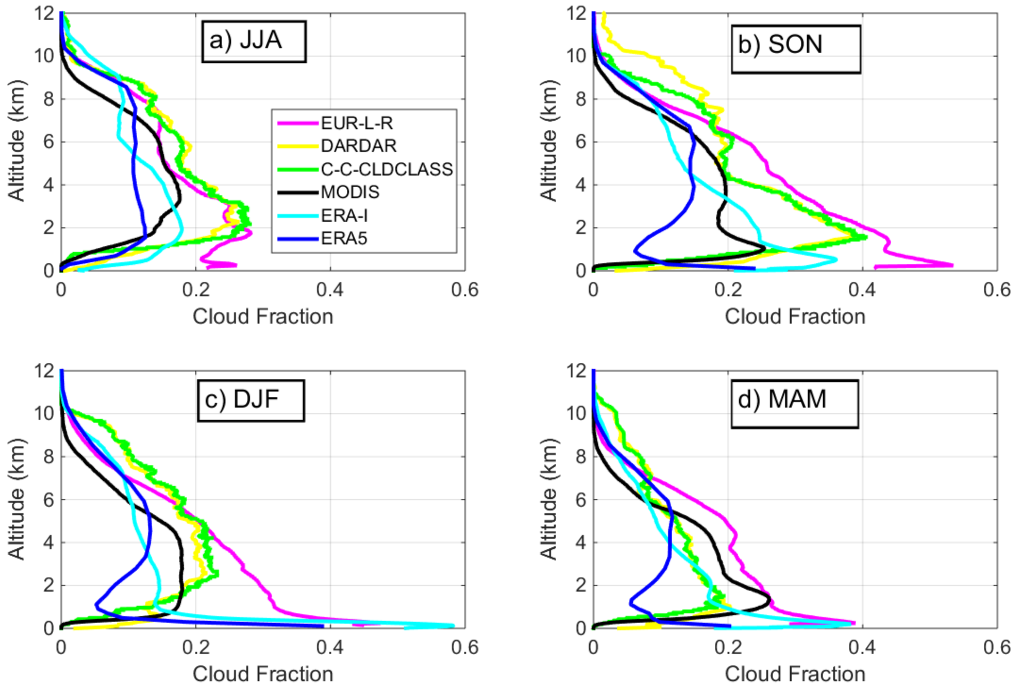



**Figure 4**. Cumulated seasonal vertical scene-type distribution between June 2006 and May 2010 all the independent datasets at less than 25 km from the station.

Low clouds detected by EUR-LR are more frequent in all seasons but summer (JJA), especially in winter (DJF) when the difference with the other datasets is larger (Fig. 4). An occurrence peak is observed by EUR-LR near the surface (within the first 500 to 1000 m) during winter (and spring), only captured by ERA5 and ERA-I. MODIS strongly underestimates low clouds in summer, and, to a lesser extent, in other seasons. B14 also showed that MODIS underestimates cloud fraction during winter (from October to February). This is a known issue (Liu et al., 2010). MODIS cloud fraction biases vary with season, as the complete darkness during the polar night prevents the use of visible channels for cloud retrievals and implies the use of a nighttime cloud detection algorithm (Liu et al., 2004). MODIS distribution peaks at 1km in MAM, when the temperature inversion is the strongest, about 10°C on average. ERA-I misses mid and high clouds in all seasons with a large increase in the near surface cloud fraction from September to May. This was also discussed by Zygmuntowska et al., 2012 and Zib et al., 2012). ERA5 better captures mid- and high-level clouds, which are mainly ice clouds, but largely misses low clouds at SON, DJF and MAM, especially above 500 to 1000m, where near-surface temperature inversion usually occurs.

In DJF, satellite observations and analyses dramatically lack low level clouds (between 0 and 4 km), where most of Arctic clouds occur (Shupe et al., 2011a). This is compensated in ERA-I and ERA5 by an excess of near surface clouds. In all seasons DARDAR, C-C and MODIS lack low clouds below 1 km. Spaceborne radar detection suffers from surface contamination echo, and lidar detection efficiency is decreased by attenuation in liquid water clouds. In most seasons, EUR-LR is missing some high clouds, due to the attenuation of lidar signal in opaque clouds and due to decreasing radar sensitivity with range. The better agreement (above 7 km) is obtained in JJA and the larger dispersion in this altitude range is observed in SON.

Large differences in BOA fluxes, in Table 3, were observed during winter, which is the period when significant mismatch appears in cloud vertical distribution. This deficit of low clouds lowers BOA fluxes calculated from satellite dataset, compared to BSRN. The smaller contribution of MODIS low clouds at this season can explain the lower LW BOA value of CERES compared to C-C-FLX. This is the opposite for MAM. The excess of near surface clouds for ERA5 (and to a lower extent for ERA-I) in DJF and MAM, combined with a warmer temperature profile compared to radiosondes (not shown), can also explain the over-estimation of downwelling fluxes for ERA5 (and ERA-I).

To go further in the comparison of fluxes, we looked to coincident datasets following the B14 approach.

## 4 Coincident measurements from independent datasets

In this section we will focus on TOA and BOA flux analyses for the subsets of coincident observations in space and time. We consider here the datasets that directly provide radiative fluxes (CERES and C-C-FLX) and datasets for which we have calculated fluxes (EUR-LR) using Streamer RTM. Note that ERA-I and ERA5 data points are not strictly coincident with A-Train overpasses, but as they are within approximately 1 hour, we included them in the analysis. For TOA we compared all datasets CERES, C-C-FLX, EUR-LR, ERA-I and ERA5 keeping CERES as a reference, whereas for BOA, the same datasets were considered with BSRN acting as the reference.

### 4.1 TOA fluxes

#### 4.1.1 Mean seasonal TOA fluxes

Here, we will analyze the evolution of seasonal fluxes at TOA for the different datasets with respect to CERES and discuss correlations and histograms of spread. We will further study differences by type of scene. This analysis includes 249 coincident samples, seasonally distributed as DJF=56; MAM=67; JJA=47; SON=79. Mean seasonal fluxes from coincident measurements highlights any systematic bias between datasets. Table 3 summarizes the average values determined for all the seasons and annual mean.

**Table 4**. Seasonal LW TOA average fluxes for coincident CERES and other retrievals for the period spanning from 09/2006 to 04/2010. C-C-FLX R05 is put aside at the end of the table due to smaller data points compared. Standard deviations are in brackets. Colors are reported as in Table 2.

| **TOA** | **Total** | | **DJF** | | **MAM** | | **JJA** | | **SON** | |
|---|---|---|---|---|---|---|---|---|---|---|
| # of cases | 249 | | 56 | | 67 | | 47 | | 79 | |
| | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km |
| CERES < 25km | 192.1 (30.4) | | 164.0 (13.1) | | 189.7 (21.5) | | 239.9 (16.1) | | 185.6 (16.9) | |
| CERES nearest | 192.2 (30.6) | 0.1 | 163.7 (13.5) | -0.2 | 189.9 (21.6) | 0.2 | 239.9 (17.0) | 0.0 | 185.8 (16.9) | 0.3 |
| EUR-LR | 191.1 | -1.0 | 163.6 | -0.4 | 187.2 | -2.5 | 237.1 | -2.8 | 186.6 | 1.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (31.4) | | (16.7) | | (21.0) | | (19.2) | | (22.1) | |
| C-C-FLX R04 < 25km | 188.3 (31.4) | -3.8 | 158.7 (12.6) | -5.3 | 187.4 (22.5) | -2.2 | 237.8 (14.4) | -2.2 | 180.5 (17.4) | -5.1 |
| C-C-FLX R04 nearest | 187.9 (32.3) | -4.1 | 156.9 (13.1) | -7.1 | 186.8 (22.9) | -2.8 | 238.4 (16.0) | -1.5 | 180.9 (18.1) | -4.6 |
| ERA-I | 187.4 (32.0) | -4.7 | 156.9 (15.4) | -7.1 | 183.8 (24.3) | -5.9 | 237.2 (12.1) | -2.7 | 182.4 (17.2) | -3.2 |
| ERA5 | 191.3 (34.7) | -0.7 | 160.1 (13.7) | -3.8 | 191.5 (24.5) | 1.8 | 246.2 (14.5) | 6.3 | 180.6 (20.6) | -4.9 |
| | | | | | | | | | | |
| # of cases | 221 | | 46 | | 65 | | 47 | | 63 | |
| | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km | Mean (σ) | Minus CERES 25 km |
| C-C-FLX R05 < 25km | 189.2 (31.4) | -5.2 | 157.3 (14.0) | -6.3 | 186.1 (22.2) | -3.6 | 234.6 (14.6) | -5.3 | 181.8 (17.5) | -6.0 |
| C-C-FLX R05 nearest | 188.5 (32.9) | -5.9 | 154.7 (15.0) | -8.8 | 185.3 (23.1) | -4.4 | 235.8 (16.5) | -4.1 | 181.1 (17.8) | -6.7 |

Table 4 allows for clarification of conclusions drawn from Table 2. There is a good agreement between all datasets within 5 W/m$^2$ for annual LW TOA, with values in general larger for CERES, the bias being higher during wintertime and, to a lesser extent, in autumn. Streamer calculations, based on EUR-LR retrievals, agree well with CERES, except in spring and summer, where they are smaller (but by less than 3 W/m$^2$). C-C-FLX R05 shows larger deficit than R04 in all seasons with a total bias of -5.9 W/m$^2$. Several changes in R05 could explain the difference with R04, namely longwave land emissivity and cloud properties (Henderson and L'Ecuyer, 2020).

We propose to take a closer look at those differences in terms of seasonal differences, depending on key parameters.

### 4.1.2 Seasonal flux differences and spatial heterogeneity

To assess the role of spatial heterogeneity, we compared the nearest and 25km circle-mean value of each coincident measurement from CERES and C-C-FLX. Figure 5 displays LW TOA departures from CERES (< 25 km) as reported in Table 4. It shows a systematic underestimation (mean annual bias of -3.1 W/m$^2$ with a standard deviation of 2.1 W/m$^2$) of LW for all datasets compared to CERES. Part of this difference could be explained by the spatial sampling over an heterogeneous and steep terrain (such as shown for Eureka in Fig. 1), where surface temperatures are hard to precisely account for at the different pixel sizes among dataset (CERES: 20 x 20 km, C-C-FLX: 1km, ERA: 14 x 2km). In winter this effect is expected to be smaller because the region is ubiquitously snow and ice covered, limiting heterogeneity in surface emissivity.

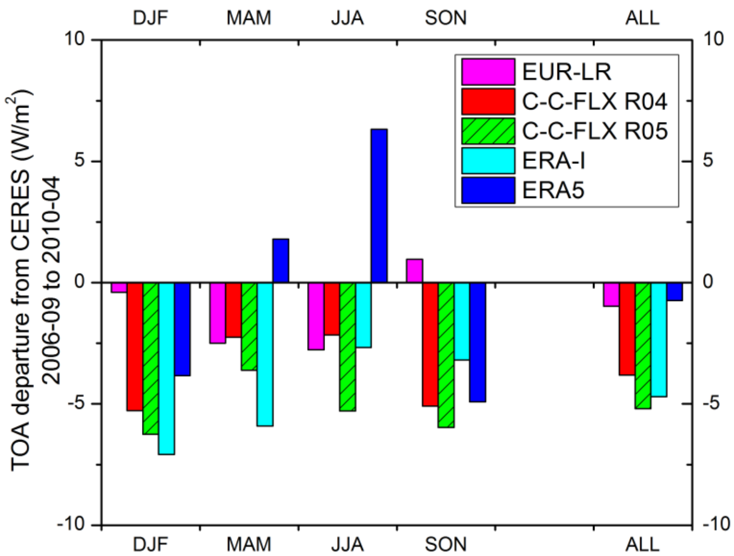



**Figure 5**. LW TOA departure from CERES as reported in Table 4. C-C-FLX R05 comparison with CERES is based on only 221 points (249 for R04 and other datasets).

Best agreement is seen from comparisons between CERES and the ground-based EUR-LR flux calculations with an absolute difference being maximum in MAM and JJA (about 2.5 W/m$^2$). Heterogeneity has a small impact for CERES (less than 0.3 W/m$^2$ because CERES measurements are to a certain extent smoothed in the 20 x 20 km pixel) and slightly higher for C-C-FLX (up to 1.8 W/m$^2$ for R04 and 2.5 W/m$^2$ for R05 in winter).

During this time of year, thick liquid and mixed-phase clouds obstruct higher clouds, from a ground-based perspective (Fig. 4). It further underscores the importance of the cloud vertical distribution in increasing accuracy of the radiative transfer calculations. As in Table 2, we see in Table 4 and Fig. 5 that the C-C-FLX and CERES differences are negative for all seasons. A good agreement is however obtained (about -3.8 W/m$^2$) between C-C-FLX and CERES. It is comparable to the one obtained in the statistical study although results for DJF and SON are degraded (see Fig. 6). Compared with Table 2, ERA5 bias with CERES is still small in average, but is degraded in summer and fall. This could be due to the temporal sampling of 3-hour re-analysis in seasons when atmospheric properties can rapidly change. ERA-I is behaving slightly differently with larger departures observed in DJF and MAM, consistently with Table 2. Overall the sampling effect does not appear as a first-order reason that can explain the differences in TOA between datasets.

Figure 6 shows scatter plots of the coincident retrieved LW TOA fluxes for EUR-LR, C-C-FLX R04, ERA-I and ERA5 and histograms of their differences with respect to CERES dataset. Note that C-C-FLX R05 plots are not shown here as the conclusions are similar to R04 and the number of points is smaller (see Table 5 were results are reported). In the scatter plots we have identified both seasonal and overall correlation coefficients. In the histograms and Table 5, we have identified the biases and half widths at 60% of the maximum and one fourth of the full width at 1/e2 of the maximum (e.g. one standard deviation - s - of a gaussian distribution), and the number of points outside 3 s to give an indication of the outliers.

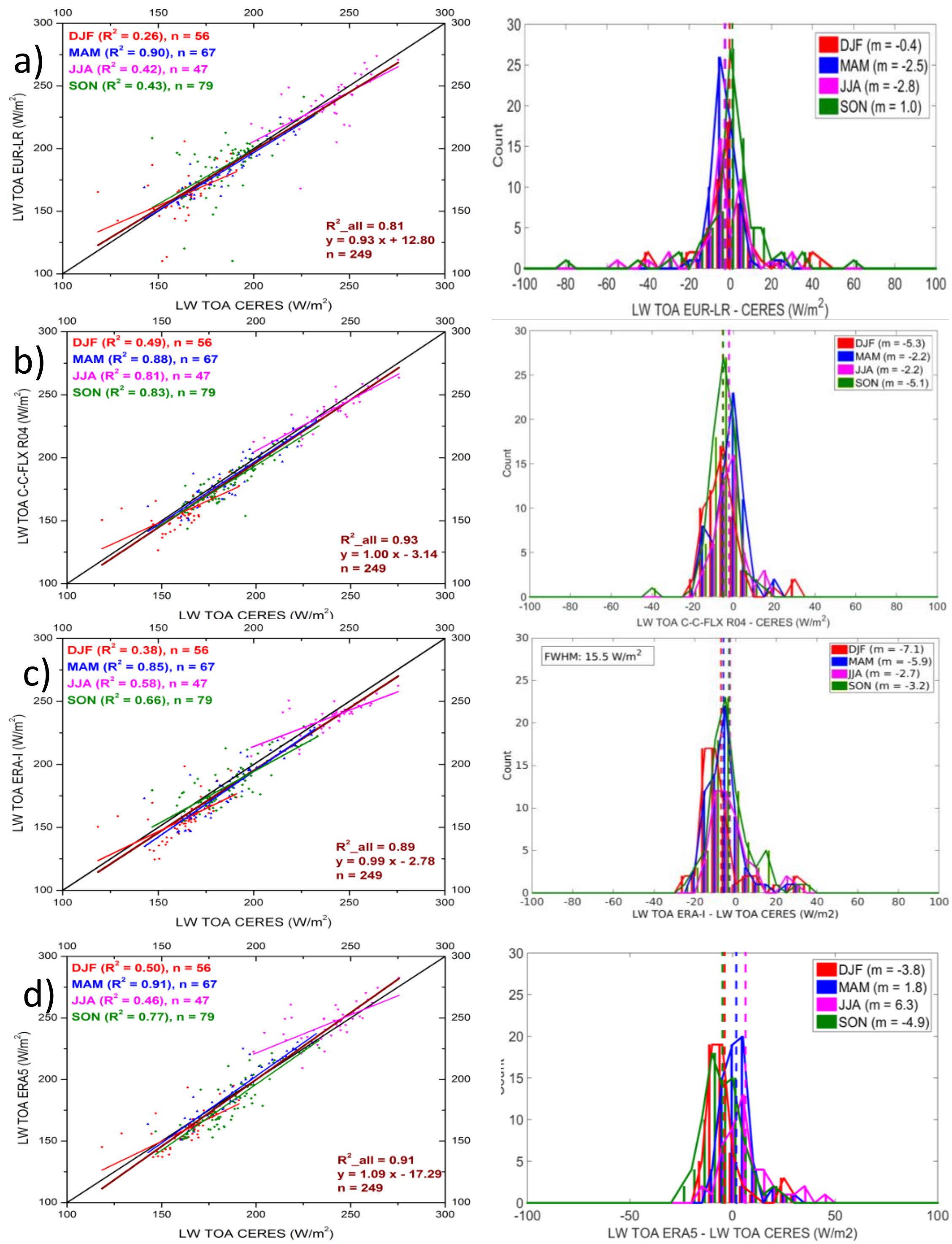



**Figure 6**. Seasonal TOA upwelling LW fluxes for CERES and EUR-LR (a), C-C-FLX R04 (b), ERA-I (c) and ERA5 (d) for 249 measurements at the same time, between September 2006 and April 2011.

**Table 5**. Gaussian fit statistics of TOA differences between EUR-LR, C-C-FLX, ERA-I, ERA5 and CERES. Note that a smaller number of points were available for R05 (see Table 4).

| | EUR-LR | C-C-FLX R04 - | C-C-FLX R05 - | ERA-I - | ERA5 - |
|---|---|---|---|---|---|

| | - CERES | CERES | CERES | CERES | CERES |
|---|---|---|---|---|---|
| Mean | -1.3 | -4.0 | -5.6 | -7.4 | -2.7 |
| Half width at 60% | 7.0 | 6.8 | 6.3 | 7.3 | 8.7 |
| FWHM | 16.3 | 15.8 | 14.8 | 17.0 | 20.3 |
| σ (¼ of full width at 1/e2) | 6.9 | 6.7 | 6.3 | 7.2 | 8.6 |
| Number of outliers (> 3 σ) | 26 | 5 | 5 | 14 | 8 |

Looking at EUR-LR plots and histograms, a few large seasonal outliers are evidenced in Fig. 6a and Fig. 6b, except in MAM which correspond to a smaller amount of high clouds (see Fig. 4). These outliers are homogeneously distributed below and above the mean. The worst correlations occurred in DJF, SON and JJA with differences up to 80 W/m$^2$ as one can see from the outliers of the histogram. In the case of opaque clouds, ground-based instruments are not able to correctly resolve the vertical profiles of cloud fraction, particle size and extinction at upper levels due to transmission losses. As a result, mean cloud temperature is set too high and this causes an overestimation of LW TOA. Another critical scenario is the presence of high clouds, sometimes above opaque clouds. Due to the decreasing radar sensitivity with range and the fact that the lidar signal can be totally attenuated in opaque clouds, it is likely to miss those high clouds and then underestimates LW TOA, with a bias that depends on cloud layer optical depth.

The good agreement noticed before between C-C-FLX R04 (and R05) and CERES corresponds to good correlation slopes but a higher dispersion with a weaker number of outliers than for EUR-LR. Differences are larger in winter (DJF) and in autumn (SON), where the slope is smaller than 1 and the outliers are more numerous, especially with large positives. In the histograms DJF and SON are indeed characterized by a larger bias with respect to CERES (about -5 W/m$^2$). For these two seasons a rather broad dispersion is observed with a few outliers at +30 W/m$^2$. In winter and spring, the difference shows a secondary peak at -10 W/m$^2$. This is not statistically significant, but it can be due to the under-estimation of low ice clouds, if one looks at cloud vertical fraction with respect to EUR-LR, but the overall number of points remains small for that to be significant.

ERA-I appears slightly biased low by about 7.4 W/m$^2$ on average, and the number of outliers is large. However their distribution is different from the EUR-LR one with a large number of positive values creating a secondary peak at +30W/m$^2$, not shown here. Finally, we find that ERA5 LW TOA is on average in relatively better agreement if we consider all points with a global correlation of R2=0.91, close to what is obtained for ERA-I. Seasonal scatter plot and histogram (Fig. 6d) highlight opposite pattern in winter and summer, with a much smaller correlation. It can be in part explained by the fact that ERA5 overestimates cloud cover (especially low cloud) in the Arctic in wintertime, in a way similar to (but less than) the one already shown for ERA-I (see Fig. 4), and consistent with previous work, done with ERA-I (Chernokulsky and Mokhov, 2012; Zygmuntowska et al., 2012). The modest correlation in summer (R2=0.46) may be linked to the fact that ECMWF reanalyses underestimate by half the liquid water content of summer clouds (Zygmuntowska et al., 2012; Huang et al., 2017a). This will be further discussed in the next subsection.

4.1.2 TOA differences decomposed by cloud optical depth, type of scenes and height of lower layer

To validate the hypothesis of section 4.1.2, that most differences are due to clouds, we now plot the LW differences depending on the total visible optical depth (Fig. 7a), the phase of cloud layers (Fig. 7b) and the top altitude of the lowest layer (Fig. 7c). We only keep ERA5 in these plots for sake of clarity, as showing comparable behavior, and will briefly discuss main differences with ERA-I. To compare with the same number of cases (249), only C-C-FLX-R04 is shown here but we discuss below the comparison between R04 and R05 with respect to CERES.

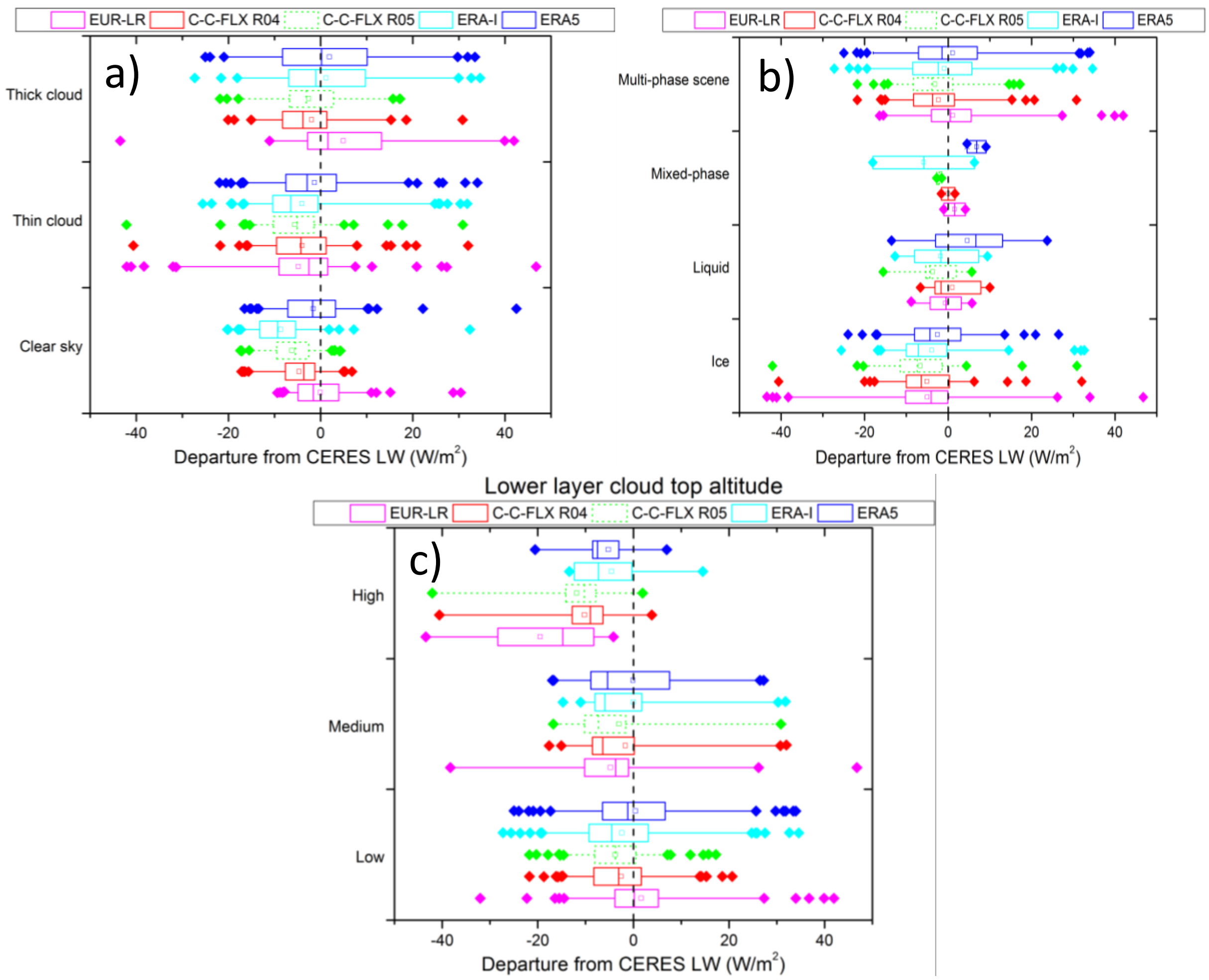



**Figure 7**. Difference of TOA LW fluxes between DATASET (EUR-LR, C-C-FLX R04, R05, ERA-I and ERA5) minus CERES depending on (a) total visible optical depth, (b) phase of cloud layers and (c) top height of the lower layer. Boxes correspond to 25%, median and 75% values, thin bars show 5 and 95% and squares are used to show the mean. Outliers are also reported as coloured diamonds.

Note that the type of scene classification (either clear, thin or thick clouds), phase of clouds (liquid water, ice, mixed-phase or multiple phase scene) and top height of the lower layer are based on EUR-LR observations and therefore depends on instrument sensitivity and can be biased in the case of opaque clouds and very thin clouds. Thick/Thin clouds threshold is set to total visible optical depth of 2. Multiple phase scene indicates that layers with different phase are present in the column.

Figure 7a confirms the relatively good agreement for TOA LW for clear-sky scenes. With a decreased departure from -8.7 to -2.6 W/m$^2$, ERA5 reanalyses of clear sky are improved compared to ERA-I, where surface emissivity, surface temperature or atmospheric absorption have been identified as possible source of discrepancies, as shown in Huang et al. (2017a). The comparison between both C-C-FLX datasets shows degraded statistics for R05 relative to R04, especially for clear sky (median bias of -3.6 W/m$^2$ for R04 and -5.5 W/m$^2$ for R05). This could be due to changes in R05 implementation of longwave land emissivity, which is relatively complex to parametrize in heterogeneous and steep terrain like Eureka. There is a warm bias for EUR-LR due to the presence of thick clouds (COD>2), when lidar signal is extinguished and the cloud layer top altitude is not precisely found. Therefore, the EUR-LR cloud layer are wrongly positioned (too low, too warm). TOA departures based on cloud type are fairly similar amongst datasets. EUR-LR fails to get correct LW TOA when high clouds are present, due to a decrease in radar sensitivity for small particles (as discussed in Grenier et al., 2009 and Blanchard et al., 2017).

Ice layers are very frequent and cause a large spread in TOA differences. There are very few liquid-phase clouds only (7) and mixed-phase only (2) cases. Figure 7c shows that all types of clouds are mainly biased low with respect to CERES TOA measurements, with an emphasis on high clouds. This is rather surprising for C-C-FLX R04 and R05, because of high sensitivity of lidar to cirrus clouds as evidenced in the high C-C-FLX cloud occurrence reported in Fig. 4. A possible reason could be inaccurate estimations of ice water content and microphysics in flux calculations, but this evaluation is beyond the scope of this study. It has also to be noted that some additional discrepancies could occur due to the temperature inversion layer which could be badly captured with GMAO or ERA5 coarse vertical resolutions.

Overall two main issues are confirmed here: the bias in high clouds for EUR-LR, and the bias in clear air identified for ERA-I is now corrected in ERA5.

## 4.2 BOA fluxes

### 4.2.1 Mean seasonal BOA fluxes

Between September 2007 and May 2010, both active instruments and BSRN sensor were operational at Eureka. Repeating the same methodology as in section 4.1, we first discussed annual and seasonal statistics.

**Table 6**. As for table 4, seasonal variation of LW BOA for BSRN, EUR-LR, CERES, C-C-FLX (R04 and R05), ERA-I and ERA5 for the period from 09/2007 to 04/2010. Standard deviations are in brackets. Colors are used as in Table 3.

| BOA | Total | | DJF | | MAM | | JJA | | SON | |
|---|---|---|---|---|---|---|---|---|---|---|
| # of points | 149 | | 29 | | 48 | | 19 | | 53 | |
| | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN |
| BSRN Filtered | 205.4 | | 170.3 | | 188.1 | | 272.7 | | 216.0 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (50.0) | | (37.4) | | (40.4) | | (39.2) | | (40.3) | |
| EUR-LR | 206.2 (50.8) | 0.9 | 171.4 (36.9) | 1.0 | 189.8 (42.0) | 1.7 | 273.4 (37.2) | 0.7 | 216.2 (43.0) | 0.1 |
| CERES < 25km | 194.8 (52.7) | -10.5 | 151.0 (32.7) | -19.3 | 188.7 (44.5) | 0.7 | 281.9 (35.0) | 9.1 | 193.1 (35.3) | -23.0 |
| CERES nearest | 195.7 (53.3) | -9.7 | 151.1 (34.9) | -19.3 | 190.4 (43.8) | 2.4 | 280.4 (39.1) | 7.6 | 194.5 (37.7) | -21.5 |
| C-C-FLX R04 < 25km | 201.4 (51.4) | -4.0 | 160.2 (38.4) | -10.1 | 189.4 (41.5) | 1.3 | 272.6 (35.2) | -0.1 | 209.1 (41.7) | -6.9 |
| C-C-FLX R04 nearest | 200.3 (52.3) | -5.1 | 159.0 (40.2) | -11.4 | 186.9 (42.8) | -1.1 | 269.8 (37.6) | -2.9 | 210.0 (42.2) | -6.0 |
| ERA-I | 222.6 (47.2) | 17.3 | 191.9 (44.9) | 21.5 | 215.4 (42.7) | 27.3 | 279.1 (38.2) | 6.4 | 225.8 (36.8) | 9.8 |
| ERA5 | 200.4 (47.3) | -4.9 | 159.8 (33.5) | -10.5 | 190.6 (33.0) | 2.6 | 279.4 (27.8) | 6.6 | 203.3 (34.9) | -12.8 |
| | | | | | | | | | | |
| # of points | 130 | | 24 | | 48 | | 19 | | 39 | |
| | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN | Mean (σ) | Minus BSRN |
| C-C-FLX R05 < 25km | 209.6 (50.7) | 1.3 | 169.3 (36.0) | 0.5 | 193.7 (42.4) | 5.7 | 270.8 (38.8) | -2.0 | 224.0 (39.7) | -2.0 |
| C-C-FLX R05 nearest | 207.7 (52.8) | -0.6 | 164.8 (37.4) | -4.0 | 192.1 (43.0) | 4.0 | 269.2 (37.6) | -3.6 | 223.3 (45.2) | -2.7 |

In Table 6, we can see that there is a wider span of annual LW BOA averages amongst datasets, from 194.8 W/m$^2$ (CERES) to 222.6 W/m$^2$ (ERA-I). We restate that the lack of availability of ground-based measurements during the 2006 and 2008 summers, and icing screening from BSRN can induce sampling seasonal effects. Therefore, those values are not expected to be used for climatological analysis. We see that the spatial sampling effect of CERES and C-C-FLX fluxes (labelled as 25 km and nearest) is relatively small compared to the difference with BSRN and can be explained by a mixture of cloud edges or transition with clear sky. Differences remain high and comparable for all fluxes excepting CERES. The dispersion on average annual values are of the order of 5 W/m$^2$, excepting CERES and ERA-I data, but those on seasonal values can be about twice larger in winter, when the number of cases is reduced to 29. One can see that differences are larger than for summer (about 6 W/m$^2$) when the number of points is even more reduced (19). In all cases standard deviations remain high, and residual uncertainties on average values (standard deviation divided by the square root of the number of points) are 4 W/m$^2$ (annual average) to 8 W/m$^2$ (summer). These values have to be kept in mind in the discussion of observed differences.

BSRN and EUR-LR agree well (within 1.7 W/m$^2$ over all seasons and better than 1W/m$^2$ on average), confirming the high level of confidence of the combination of active measurements with the Streamer simulations. BSRN field-of-view angular integration and EUR-LR time integration are also contributing to this agreement.

While CERES and C-C-FLX R04were in good agreement in Table 3, the coincident comparison showed that differences can be up to 16 W/m$^2$ in autumn and about 6.5 W/m$^2$ in annual mean (CERES being biased low with respect to C-C-FLX R04 by about 5W/m$^2$). C-C-FLX R05 shows reduced bias in all seasons except in spring and summer. The better representation of ice and mixed-phase clouds in R05 could explain this improvement and this hypothesis will be discussed in the next section. Satellite observations are lower than BSRN in all seasons and more particularly in DJF (-10 to -20 W/m$^2$), except for C-C-FLX R05, while ERA-I is systematically much higher than BSRN for all seasons (between 6.4 to 27 W/m$^2$) and more than 15 W/m$^2$ on average, which is consistent with the overestimation of cloud fraction at low altitude (Zib et al., 2012). The several modifications implemented for ERA5 have a significant impact as it decreases BOA LW fluxes by 22 W/m$^2$ with respect to ERA-I, and even more in winter. We found that ERA5 is in general in much better agreement with other datasets.

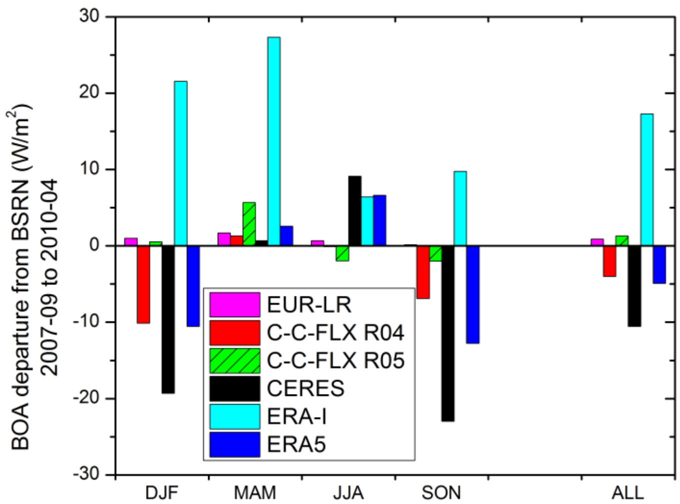
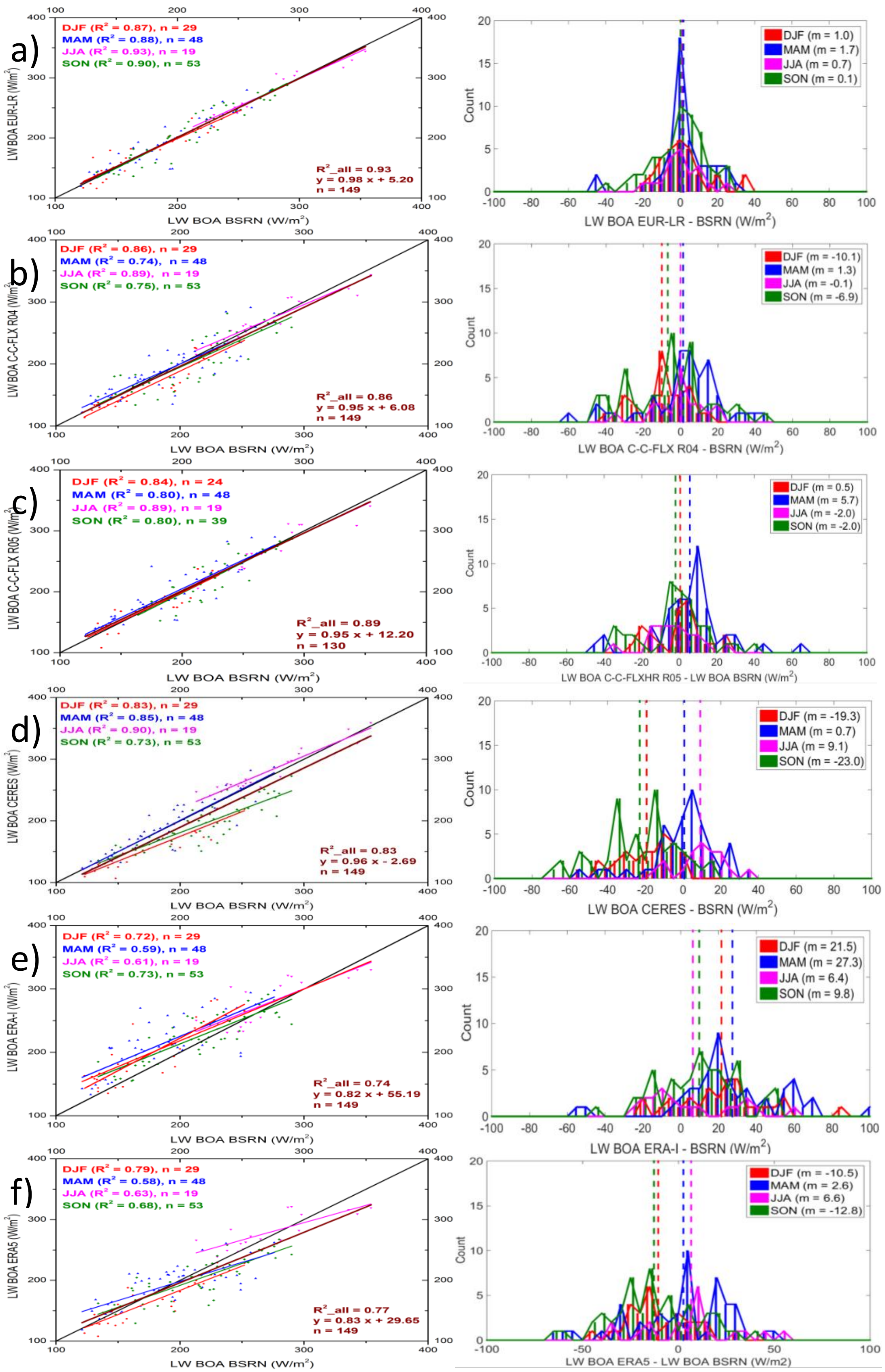



**Figure 8**. LW BOA departure from BSRN for coincident measurements as reported in Table 6. C-C-FLX R05 bars are dashed because the comparison with BSRN is based on only 130 points.

Figure 8 is reporting the differences observed in Table 6. It evidences that the largest differences are observed for CERES and ERA-I, with annual biases of -11 and +17 W/m², respectively. Largest differences in autumn and winter (and spring for ERA-I), when the cloud vertical distribution was divergent. The difference observed between spring and summer and the two other seasons between C-C-FLX and CERES is statistically meaningful. A closer look at those differences will help to understand the biases.

4.2.2 Seasonal flux differences

Looking at coincident fluxes in a way similar to TOA analysis helps to identify systematic seasonal, methodological or instrumental biases compared to BSRN reference. We must be aware however that due to the footprint of satellite observations, it may be possible that although the separation distance is kept small, ground-based active instruments may not be looking at the same cloud. The reduced number of cases makes the multi-parameter analysis more difficult in terms of quantification. Fig. 9 is reporting (as in Fig. 6 for TOA fluxes) one-to-one flux comparisons and histograms of flux differences however evidences significant differences. Table 7 summarizes main parameters reported in histograms.


a)
DJF (R² = 0.87), n = 29
MAM (R² = 0.88), n = 48
JJA (R² = 0.93), n = 19
SON (R² = 0.90), n = 53
R²_all = 0.93
y = 0.98 x + 5.20
n = 149
LW BOA EUR-LR (W/m²)
LW BOA BSRN (W/m²)
DJF (m = 1.0)
MAM (m = 1.7)
JJA (m = 0.7)
SON (m = 0.1)
Count
LW BOA EUR-LR - BSRN (W/m²)
b)
DJF (R² = 0.86), n = 29
MAM (R² = 0.74), n = 48
JJA (R² = 0.89), n = 19
SON (R² = 0.75), n = 53
R²_all = 0.86
y = 0.95 x + 6.08
n = 149
LW BOA C-C-FLX R04 (W/m²)
LW BOA BSRN (W/m²)
DJF (m = -10.1)
MAM (m = 1.3)
JJA (m = -0.1)
SON (m = -6.9)
Count
LW BOA C-C-FLX R04 - BSRN (W/m²)
c)
DJF (R² = 0.84), n = 24
MAM (R² = 0.80), n = 48
JJA (R² = 0.89), n = 19
SON (R² = 0.80), n = 39
R²_all = 0.89
y = 0.95 x + 12.20
n = 130
LW BOA C-C-FLX R05 (W/m²)
LW BOA BSRN (W/m²)
DJF (m = 0.5)
MAM (m = 5.7)
JJA (m = -2.0)
SON (m = -2.0)
Count
LW BOA C-C-FLXHR R05 - LW BOA BSRN (W/m²)
d)
DJF (R² = 0.83), n = 29
MAM (R² = 0.85), n = 48
JJA (R² = 0.90), n = 19
SON (R² = 0.73), n = 53
R²_all = 0.83
y = 0.96 x - 2.69
n = 149
LW BOA CERES (W/m²)
LW BOA BSRN (W/m²)
DJF (m = -19.3)
MAM (m = 0.7)
JJA (m = 9.1)
SON (m = -23.0)
Count
LW BOA CERES - BSRN (W/m²)
e)
DJF (R² = 0.72), n = 29
MAM (R² = 0.59), n = 48
JJA (R² = 0.61), n = 19
SON (R² = 0.73), n = 53
R²_all = 0.74
y = 0.82 x + 55.19
n = 149
LW BOA ERA-I (W/m²)
LW BOA BSRN (W/m²)
DJF (m = 21.5)
MAM (m = 27.3)
JJA (m = 6.4)
SON (m = 9.8)
Count
LW BOA ERA-I - BSRN (W/m²)
f)
DJF (R² = 0.79), n = 29
MAM (R² = 0.58), n = 48
JJA (R² = 0.63), n = 19
SON (R² = 0.68), n = 53
R²_all = 0.77
y = 0.83 x + 29.65
n = 149
LW BOA ERA5 (W/m²)
LW BOA BSRN (W/m²)
DJF (m = -10.5)
MAM (m = 2.6)
JJA (m = 6.6)
SON (m = -12.8)
Count
LW BOA ERA5 - LW BOA BSRN (W/m2)

772

**Figure 9**. Seasonal BOA downwelling LW fluxes for BRSN and EUR-LR (a) C-C-FLX R04 (b), C-C-FLX R05 (c), CERES (d), ERAI (e) and ERA5 (f) for the 149 measurements (except for R05) at the same time, between September 2007 and April 2010.

**Table 7**. Gaussian fit statistics of BOA differences between EUR-LR, C-C-FLX R04 and R05, CERES, ERA-I, ERA5 and BSRN. Note that a smaller number of points were available for R05 (see Table 6).

| | EUR-LR - BSRN | C-C-FLX R04 - BSRN | C-C-FLX R05 - BSRN | CERES - BSRN | ERA-I - BSRN | ERA5 - BSRN |
|---|---|---|---|---|---|---|
| Mean | 0.8 | -0.5 | 2.8 | -6.8 | 2.6 | -4.7 |
| Half width at 60% | 10.0 | 16.3 | 12.5 | 21.3 | 20.9 | 25.2 |
| FWHM | 23.5 | 38.0 | 29.2 | 49.6 | 48.6 | 58.7 |
| σ (¼ of full width at 1/e2) | 9.9 | 16.1 | 12.4 | 21.1 | 10.6 | 24.9 |
| Number of outliers (> 3 σ) | 5 | 1 | 6 | 3 | 5 | 0 |

It is apparent from the histograms of BOA flux differences given in Fig. 9 that all comparisons show very large dispersions except between EUR-LR and BSRN. No significant bias and very few outliers (close to +/- 40 W/m$^2$, as evidenced from the narrower distribution) are observed in this last case. The correlation between EUR-LR and BSRN is indeed high (R2=0.93, and s = 9.9 W/m$^2$), but comparison of individual coincident times can be off by up to 50 W/m$^2$. Those outliers are likely explained by the fact that the effective spatial resolution of active instruments after time averaging remains small compared to BSRN pyranometer/pyrgeometer, located 2.3 km away from the station, which measures hemispheric (160 degrees) LW fluxes.

The overall distributions are widely spread in almost all other cases from -60 to +60 W/m$^2$. Results between C-C-FLX (R04 and R05) and BSRN show a dispersion of seasonal differences rather contained, limiting the overall bias, slightly better for R05. Figure 9c shows a smaller dispersion of C-C-FLX R05 bias vs BSRN, as compared to CERES. This confirms the advantage of lidar-radar synergy from space. But there are still high seasonal variations, especially in spring, when C-C are missing clouds below 5 km (Fig. 4). Differences between R04 and R05 are minor but an overall better agreement in found for R05 and especially in winter (the R04 bias of -10 W/m$^2$ is reduced to + 0.5 W/m$^2$ for R05). Some improvements in R05 algorithm regarding ice and mixed-phase clouds could explain those differences.

CERES and ERA5 show the wider spreads with a dispersion (2 sigmas) of about 50 W/m$^2$. The mean biases vary from about -20 W/m$^2$ in winter and autumn to about + 10 W/m$^2$ in summer for CERES. In those winter and autumn seasons, MODIS is missing almost half of low and mid clouds (see Fig. 4). ERA5 is also biased low in winter and autumn compared to BSRN and larger in summer. This result is in agreement to what was reported in Zib et al. (2012) for Ny-Ålesund and Barrow, but using ERA-Interim. Note that those stations are, however, coastal and their cloud fraction variation is different from the one at Eureka (Shupe et al, 2011a; Shupe 2011b). ERA5 corrects the BOA LW positive bias for ERA-I in winter explained by the overestimation of very low cloud in reanalyses (as reported in several papers, see Zygmuntowska et al., 2012; Zib et al., 2012; Huang et al., 2017a) as seen in Fig.4. C-C-FLX and ERA5 average values show biases in winter remain high (larger than -10W/m$^2$) with broad distributions.

### 4.2.3 BOA differences decomposed by type of scenes

In this subsection, we discuss the overall relative difference of BOA fluxes to BSRN measurements as a function of the scene type as supported by ground-based lidar/radar observations.

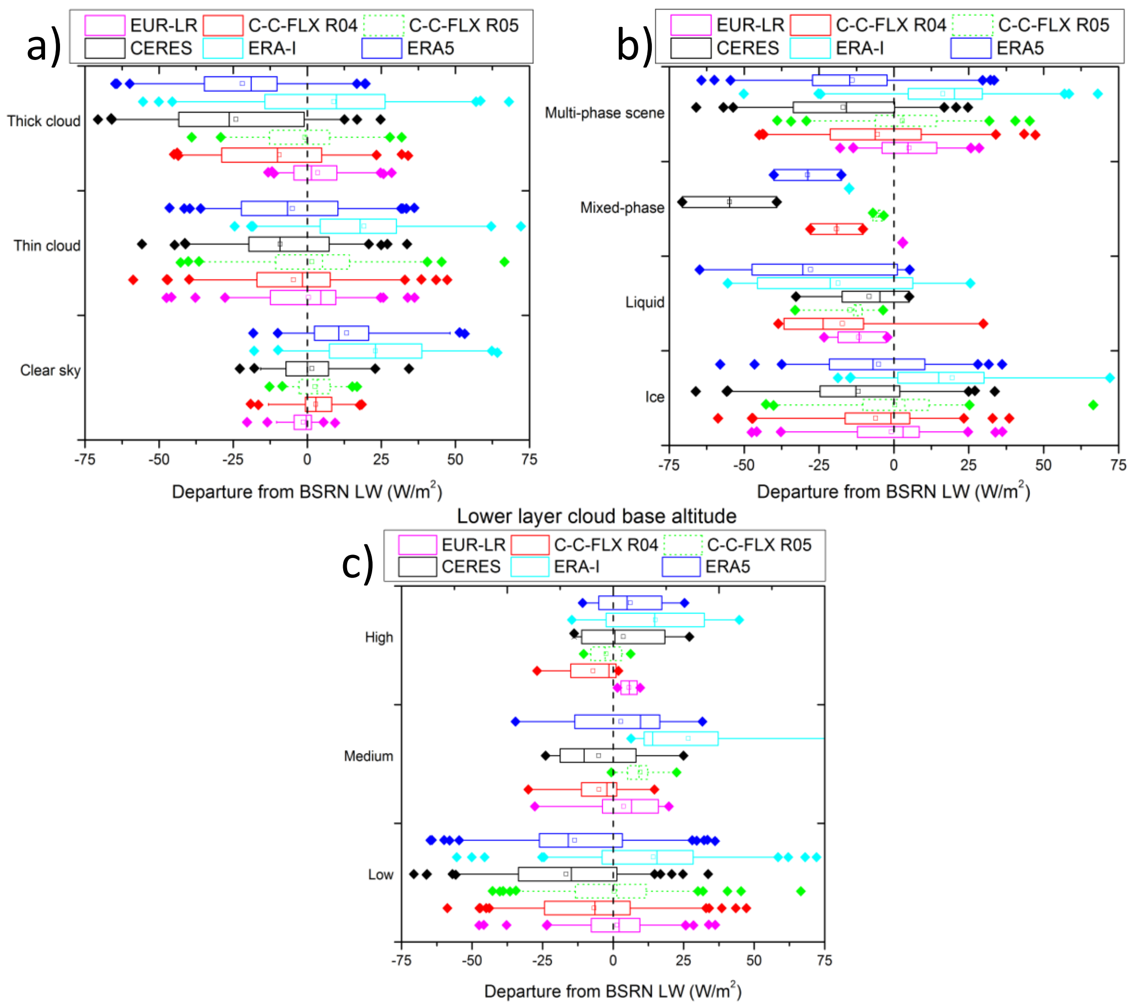



**Figure 10**. Difference of BOA LW fluxes depending on the total visible optical depth (a), phase of cloud layers (b) and bottom height of the lower layer (c). Boxes correspond to 25%, median and 75% values, thin bars show 5 and 95% and squares are used to show the mean. Outliers are also reported as coloured diamonds.

Clear sky events (Fig 10a) are well captured by CERES, C-C-FLX and EUR-LR while the mean bias is about +13 W/m$^2$ for ERA5. For opaque clouds (COD > 2), CERES and C-C-FLX are biased low because they miss correct cloud base heights as identified in Fig. 4. Such clouds are expected to be mid- and low-level clouds. Indeed, high clouds that were missed by EUR-LR above opaque clouds (in the discussion about TOA) don't have a significant impact when looking at downwelling fluxes. However, it is confirmed that all other (CERES, ERA5 and to a smaller extend CC-FLX R05) LW BOA fluxes are biased low for low clouds by more than 20 W/m$^2$ with respect to BSRN BOA measurements, which appears to be the main driver of biases. The direct comparison of C-C-FLX releases (see Appendix C) confirms that R05 significantly reduces the strong biases identified for low and thick clouds. Fig. 10b shows that there is a negative bias for ice layers for all dataset considered here except EUR-LR, as evidenced in Table 6. Mixed-phase (supercooled) clouds appear to be challenging for CERES as previously emphasized (Matus and L'Ecuyer, 2017). This also appears to be the case for ERA5 and ERA-I, but there are rather few mixed-phase cases here to draw any definitive conclusion (Fig. 10c). As mentioned for TOA, ECMWF reanalyses are still struggling to get the water content of liquid clouds correct. This remains true for ERA5 as was the case for ERA-I (Zygmuntowska et al., 2012).

### 4.3 Summary

Further to B14, we have applied in this paper the approach that was laid out for cloud occurrence to the analysis of LW radiation budget at top and bottom of the atmosphere at Eureka station. The statistical analyses are enforced by an approach of separating statistical independent analysis and coincident confrontation of observations constrained by scene types. This approach controls for some sampling and observational biases that affect the analysis, and the horizontal heterogeneity was found to be a small factor. The results indicate that there is rather good agreement in TOA fluxes (within a few W/m$^2$), but considerably less agreement in the arguably more important BOA fluxes. Main findings are summarized in Table 8.

**Table 8**. Findings for each dataset

| | **Cloud vertical distribution** | **TOA** | **BOA** |
|---|---|---|---|
| BSRN | N/A | N/A | Used as a reference |
| EUR-LR | Detects large number of hydrometeors close to the ground in winter. Is not able to detect high features above opaque clouds (in MAM and JJA). | Relatively good agreement (about 1 W/m$^2$). Issue when high clouds are not detected | Very good agreement with BSRN (bias less than 1 W/m$^2$) |
| C-C-FLX R04 | Misses a significant amount of low ice clouds in winter. Good agreement with ground-based above 2km. | Always smaller than CERES (about -4 W/m$^2$). Could be due to footprint compared to CERES (20 km x 20 km), that smooths TOA fluxes | Overall bias close to -5 W/m$^2$. Better than CERES. Bigger bias in winter (-10 W/m$^2$) |
| C-C-FLX R05 | Same as C-C-FLX R04 | Smaller than CERES with larger bias compared to R04, probably due to different surface emissivity | Good agreement with BSRN (bias less than 1.5 W/m$^2$). Clouds are better addressed than R04 but too many mixed-phase are detected (in spring). |
| CERES | Misses clouds in all seasons, but this is more dramatic in winter over snow surface. | Used as a reference | Overall bias of about -10W/m$^2$). Differences are high in winter (about -20 W/m$^2$) and autumn due to a wrong detection of low clouds. |
| ERA-I | Underestimates cloud fraction by a factor of 2. This is | Bias for clear sky due to a coarse temperature | Biggest overall bias (about + 20 W/m$^2$) larger in winter, as |

| | | | |
|---|---|---|---|
| | somewhat compensated by the detection of clouds very close to the surface. | profile that misses temperature inversion. | caused by a bad re-analysis of cloud vertical distribution. Issues with water, ice clouds and clear sky. |
| ERA5 | Similar to ERA-I, but a lower amount of low clouds. | In good agreement with CERES, except in summer due to inaccurate liquid water content | Overall bias close to -5 W/m$^2$. Larger bias in winter (-10 W/m$^2$). Clear sky bias is still present (13 W/m$^2$) but reduced compared to ERA-I. Low liquid and ice clouds are the main sources of errors |

For the TOA, we used CERES as a reference. From the statistical independent analysis, we found results comparable to what has been previously obtained, with good agreement (better than 5 W/m$^2$) between datasets and low biases.

Observations of broadband longwave radiation using a surface passed pyrgeometer as part of BSRN were used as a reference for BOA analysis. A careful examination of each coincident case was undertaken to improve the quality and confidence in the measurements incorporated into the analysis. Comparison with fluxes determined using Streamer code using inputs from ground-based lidar-radar vertical profiles of cloud properties and meteorological data gave a very high agreement (with a standard deviation of less than 5 W/m$^2$), comparable to the agreement obtained for TOA fluxes. This is a remarkable result in the comparison of the BOA fluxes, for which deviation among dataset is much larger, as with satellite and reanalysis data. Low opaque clouds in wintertime are found to be the most challenging to detect for passive radiometry due to small temperature difference with the underlying snow surface. Those clouds are not well identified by CERES, as MODIS underestimates cloud fraction especially in winter and autumn (Fig. 4). This remains an issue for active sensors as well, although to a smaller extent, but ground clutter, smaller droplets for optically thick water clouds for the radar and transmission decrease for the lidar are significant issues limiting overall performance. Recent reanalyses ERA5 are improved, as differences from references are reduced compared to ERA-Interim. Some bias, however, are persistent for clear sky and cloud vertical profile, that shows the needs for improving model resolutions.

Statistical and coincident analysis revealed comparable agreement in TOA with biases smaller than 5 W/m$^2$ for all observations and analyses with respect to CERES observations. No obvious trend was found on the statistical dataset. Narrow distributions are observed for satellite observations, but a larger dispersion is seen on analyses, with a larger number of outliers for ERA5. The difference observed appears to be mainly due to high clouds. Their occurrence is slightly smaller for CERES at higher altitudes. This may be due to the fact that the altitude attribution is underestimated by MODIS.

The results for BOA fluxes show more differences. Ground-based lidar-radar inputs to radiative transfer code (streamer) give at the same time unbiased fluxes with the lower dispersion with respect to BSRN reference. All other (CC-FLX, CERES and ERA5) show biases ranging from 1 (C-C-FLX R05) to -10 W/m$^2$ (CERES and ERA5) analyzed as due to poor representation of low (mixed-phase) cloud properties (liquid water content). ERA5 corrects the very large positive longwave bias of +17 W/m$^2$ observed with ERA-I, but cloud distribution remains biased with respect to ground-based observations. Further improvements thus remain to be done in both the retrieved fluxes from observations and analyses to better address liquid water content of complex Arctic low-level clouds observed in cold seasons.

## 5 Conclusions

Existing TOA flux observing and modeling strategies are in good agreement and seem sufficient. BOA fluxes on the other hand are more problematic and while there is agreement between the ground-based broad-band observations and ground-based radar-lidar retrievals, these are only for infrequent, single observatory sites and model and satellite methodologies to characterize BOA fluxes are still insufficient for monitoring or characterizing the Arctic system.

It is essential for future operations that active sensors at ground-based sites be operated in polar regions to complement space observations in order to correctly identify cloud vertical profiles. Without integration of the ground-based, ongoing reference datasets into observing strategies it seems unlikely that space-based observations or model projections will be able to independently measure or calculate the BOA fluxes that are a critical component for characterizing and monitoring the extreme environmental changes occurring in the Arctic environment.

## Acknowledgments

Acknowledgments are due to the Canadian Network for the Detection of Atmospheric Change (CANDAC), Study of Environmental Arctic Change (SEARCH) and Environment Canada for their operational support as well as radiosondes and observations data at Eureka. We are grateful to Kenneth Moran from NOAA and Edwin Eloranta from University of Wisconsin–Madison for providing MMCR and AHSRL data, respectively. The authors thank the CALIPSO team at NASA Langley Research Center (https://www-calipso.larc.nasa.gov/), the ICARE data Center in Lille/France (https://www.icare.univ-lille.fr/) and the CloudSat team at Colorado State University for the availability of the level-2 data (http://www.cloudsat.cira.colostate.edu/data-products/). This research was supported by the Centre National d'Etudes Spatiales (CNES). Acknowledgments also go to ECMWF for the ERA5 and ERA-I reanalyses and to the NASA CERES team for the production of the CERES data. NASA CERES–SSF-TOA and SSF-Surface products are available at http://ceres.larc.nasa.gov/. The BSRN dataset can be accessed from World Radiation Monitoring Center at http://www.bsrn. awi.de/.

## References

Abe, M., Nozawa, T., Ogura, T., & Takata, K. (2016). Effect of retreating sea ice on Arctic cloud cover in simulated recent global warming. *Atmospheric Chemistry and Physics*, *16*(22), 14343-14356. doi:10.5194/acp-16-14343-2016

Ancellet, G., Pelon, J., Blanchard, Y., Quennehen, B., Bazureau, A., Law, K. S., & Schwarzenboeck, A. (2014). Transport of aerosol to the Arctic: analysis of CALIOP and French aircraft data during the spring 2008 POLARCAT campaign. *Atmospheric Chemistry and Physics*, *14*(16), 8235–8254. https://doi.org/10.5194/acp-14-8235-2014

Betts, A. K., Chan, D. Z., & Desjardins, R. L. (2019). Near-Surface Biases in ERA5 Over the Canadian Prairies. *Frontiers in Environmental Science, 7*. doi:10.3389/fenvs.2019.00129

Blanchard, Y., Pelon, J., Eloranta, E. W., Moran, K. P., Delanoë, J., & Sèze, G. (2014). A Synergistic Analysis of Cloud Cover and Vertical Distribution from A-Train and Ground-Based Sensors over the High Arctic Station Eureka from 2006 to 2010. *Journal of Applied Meteorology and Climatology, 53*(11), 2553-2570. doi:10.1175/jamc-d-14-0021.1

Blanchard, Y., Royer, A., O'Neill, N. T., Turner, D. D., & Eloranta, E. W. (2017). Thin ice clouds in the Arctic: Cloud optical depth and particle size retrieved from ground-based thermal infrared radiometry. *Atmospheric Measurement Techniques*, *10*(6), 2129-2147. doi:10.5194/amt-10-2129-2017

de Boer, G., Eloranta, E. W., & Shupe, M. D. (2009). Arctic Mixed-Phase Stratiform Cloud Properties from Multiple Years of Surface-Based Measurements at Two High-Latitude Locations. *Journal of the Atmospheric Sciences, 66*(9), 2874-2887. doi:10.1175/2009jas3029.1

Boisvert, L. N., & Stroeve, J. C. (2015). The Arctic is becoming warmer and wetter as revealed by the Atmospheric Infrared Sounder. *Geophysical Research Letters, 42*(11), 4439-4446. doi:10.1002/2015gl063775

Ceccaldi, M., Delanoë, J., Hogan, R. J., Pounder, N. L., Protat, A., & Pelon, J. (2013). From CloudSat-CALIPSO to EarthCare: Evolution of the DARDAR cloud classification and its comparison to airborne radar-lidar observations. *Journal of Geophysical Research: Atmospheres, 118*(14), 7962-7981. doi:10.1002/jgrd.50579

Chan, M. A., & Comiso, J. C. (2011). Cloud features detected by MODIS but not by CloudSat and CALIOP. *Geophysical Research Letters*, *38*(24). doi:10.1029/2011gl050063

Chernokulsky, A., & Mokhov, I. I. (2012). Climatology of Total Cloudiness in the Arctic: An Intercomparison of Observations and Reanalyses. *Advances in Meteorology*, *2012*, 1–15. https://doi.org/10.1155/2012/542093

Comiso, J. C., & Hall, D. K. (2014). Climate trends in the Arctic as observed from space. *Wiley Interdisciplinary Reviews: Climate Change*, *5*(3), 389–409. https://doi.org/10.1002/wcc.277

Cox, C. J., Walden, V. P., & Rowe, P. M. (2012). A comparison of the atmospheric conditions at Eureka, Canada, and Barrow, Alaska (2006-2008). *Journal of Geophysical Research: Atmospheres*, *117*(D12), n/a. https://doi.org/10.1029/2011jd017164

Cox, C. J., Walden, V. P., Rowe, P. M., & Shupe, M. D. (2015). Humidity trends imply increased sensitivity to clouds in a warming Arctic. *Nature Communications, 6*(1). doi:10.1038/ncomms10117

Cox, C. J., Uttal, T., Long, C. N., Shupe, M. D., Stone, R. S., & Starkweather, S. (2016). The Role of Springtime Arctic Clouds in Determining Autumn Sea Ice Extent. *Journal of Climate, 29*(18), 6581-6596. doi:10.1175/jcli-d-16-0136.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553-597. doi:10.1002/qj.828

Delanoë, J., & Hogan, R. J. (2010). Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds. *Journal of Geophysical Research, 115*. doi:10.1029/2009jd012346

Doyle, J. G., Lesins, G., Thackray, C. P., Perro, C., Nott, G. J., Duck, T. J., . . . Drummond, J. R. (2011). Water vapor intrusions into the High Arctic during winter. *Geophysical Research Letters, 38*(12). doi:10.1029/2011gl047493

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., … and König-Langlo, G.: (2018). Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), *Earth Syst. Sci. Data*, 10, 1491–1501, https://doi.org/10.5194/essd-10-1491-2018

Eastman, R., & Warren, S. G. (2010). Interannual Variations of Arctic Cloud Types in Relation to Sea Ice. *Journal of Climate, 23*(15), 4216-4232. doi:10.1175/2010jcli3492.1

Eloranta, E. W., Uttal, T., & Shupe, M. (2007). Cloud particle size measurements in Arctic clouds using lidar and radar data. *2007 IEEE International Geoscience and Remote Sensing Symposium*. doi:10.1109/igarss.2007.4423292

English, J. M., Gettelman, A., & Henderson, G. R. (2015). Arctic Radiative Fluxes: Present-Day Biases and Future Projections in CMIP5 Models. *Journal of Climate, 28*(15), 6019-6038. doi:10.1175/jcli-d-14-00801.1

Fogal, P. F., Leblanc, L. M., & Drummond, J. R. (2013). The Polar Environment Atmospheric Research Laboratory (PEARL): Sounding the Atmosphere at 80º North. *Arctic, 66*(3). doi:10.14430/arctic4321

Grachev, A. A., Persson, P. O., Uttal, T., Akish, E. A., Cox, C. J., Morris, S. M., . . . Repina, I. A. (2017). Seasonal and latitudinal variations of surface fluxes at two Arctic terrestrial sites. *Climate Dynamics, 51*(5-6), 1793-1818. doi:10.1007/s00382-017-3983-4

Graham, R. M., Cohen, L., Petty, A. A., Boisvert, L. N., Rinke, A., Hudson, S. R., . . . Granskog, M. A. (2017). Increasing frequency and duration of Arctic winter warming events. *Geophysical Research Letters, 44*(13), 6974-6983. doi:10.1002/2017gl073395

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., . . . Hudson, S. R. (2019a). Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer. *Journal of Climate*, *32*(14), 4121-4143. doi:10.1175/jcli-d-18-0643.1

Graham, R. M., Hudson, S. R., & Maturilli, M. (2019b). Improved Performance of ERA5 in Arctic Gateway Relative to Four Global Atmospheric Reanalyses. *Geophysical Research Letters, 46*(11), 6138-6147. doi:10.1029/2019gl082781

Grenier, P., Blanchet, J.-P., & Muñoz-Alpizar, R. (2009). Study of polar thin ice clouds and aerosols seen by CloudSat and CALIPSO during midwinter 2007. *Journal of Geophysical Research, 114*(D9). doi:10.1029/2008jd010927

Gupta, S. K., Darnell, W. L., & Wilber, A. C. (1992). A Parameterization for Longwave Surface Radiation from Satellite Data: Recent Improvements. *Journal of Applied Meteorology*, *31*(12), 1361-1367. doi:10.1175/1520-0450(1992)0312.0.co;2

Gupta, S. K., Kratz, D. P., Stackhouse, P. W., Wilber, A. C., Zhang, T., & Sothcott, V. E. (2010). Improvement of Surface Longwave Flux Algorithms Used in CERES Processing. *Journal of Applied Meteorology and Climatology, 49*(7), 1579-1589. doi:10.1175/2010jamc2463.1

Henderson, D. S., L'Ecuyer, T., Stephens, G., Partain, P., & Sekiguchi, M. (2013). A Multisensor Perspective on the Radiative Impacts of Clouds and Aerosols. *Journal of* Applied Meteorology and Climatology, 52(4), 853-871. doi:10.1175/jamc-d-12-025.1

Henderson, D. S., & L'Ecuyer, T. (2020). CloudSat Level 2B Fluxes and Heating Rates with Lidar [2B-FLXHR-LIDAR] Process Description and Interface Control Document [Available online at http://www.cloudsat.cira.colostate.edu/sites/default/files/products/files/2B-FLXHR-LIDAR_PDICD.P1_R05.rev0_.pdf]

Hersbach, H., & Dee, D. (2016). ERA5 reanalysis is in production, *ECMWF Newsletter*, 147, ECMWF, Reading, UK

Hu, Y., Winker, D., Vaughan, M., Lin, B., Omar, A., Trepte, C., Flittner, D., Yang, P., Nasiri, S. L., Baum, B., Holz, R., Sun, W., Liu, Z., Wang, Z., Young, S., Stamnes, K., Huang, J., & Kuehn, R. (2009). CALIPSO/CALIOP Cloud Phase Discrimination Algorithm. *Journal of Atmospheric and Oceanic Technology*, *26*(11), 2293–2309. https://doi.org/10.1175/2009jtecha1280.1

Huang, Y., Dong, X., Xi, B., Dolinar, E. K., & Stanfield, R. E. (2017a). The footprints of 16 year trends of Arctic springtime cloud and radiation properties on September sea ice retreat. *Journal of Geophysical Research: Atmospheres, 122*(4), 2179-2193. doi:10.1002/2016jd026020

Huang, Y., Dong, X., Xi, B., Dolinar, E. K., Stanfield, R. E., & Qiu, S. (2017b). Quantifying the Uncertainties of Reanalyzed Arctic Cloud and Radiation Properties Using Satellite Surface Observations. *Journal of Climate, 30*(19), 8007-8029. doi:10.1175/jcli-d-16-0722.1

Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., . . . Zadelhoff, G. V. (2015). The EarthCARE Satellite: The Next Step Forward in Global Measurements of Clouds, Aerosols, Precipitation, and Radiation. *Bulletin of the American* Meteorological Society, 96(8), 1311-1332. doi:10.1175/bams-d-12-00227.1

IPCC (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jun, S., Ho, C., Jeong, J., Choi, Y., & Kim, B. (2016). Recent changes in winter Arctic clouds and their relationships with sea ice and atmospheric conditions. *Tellus A: Dynamic Meteorology and Oceanography, 68*(1), 29130. doi:10.3402/tellusa.v68.29130

Kay, J. E., & L'ecuyer, T. (2013). Observational constraints on Arctic Ocean clouds and radiative fluxes during the early 21st century. *Journal of Geophysical Research: Atmospheres, 118*(13), 7219-7236. doi:10.1002/jgrd.50489

Kay, J. E., L'Ecuyer, T., Chepfer, H., Loeb, N., Morrison, A., & Cesana, G. (2016). Recent Advances in Arctic Cloud and Climate Research. *Current Climate Change Reports, 2*(4), 159-169. doi:10.1007/s40641-016-0051-9

Key, J. R., & Schweiger, A. J. (1998). Tools for atmospheric radiative transfer: Streamer and FluxNet. *Computers & Geosciences, 24*(5), 443-451. doi:10.1016/s0098-3004(97)00130-1

Kovacs, T., & McCormick, P. (2005). Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) quidpro quo validation plan. [Available online at http://calipsovalidation.hamptonu.edu/QPQ_plan062206.htm.]

Kratz, D. P., Gupta, S. K., Wilber, A. C., & Sothcott, V. E. (2020). Validation of the CERES Edition-4A Surface-Only Flux Algorithms. *Journal of Applied Meteorology and Climatology, 59*(2), 281-295. doi:10.1175/jamc-d-19-0068.1

Lang, A., Yang, S., & Kaas, E. (2017). Sea ice thickness and recent Arctic warming. *Geophysical Research Letters, 44*(1), 409-418. doi:10.1002/2016gl071274

L'Ecuyer, T. S., Wood, N. B., Haladay, T., Stephens, G. L., & Stackhouse, P. W. (2008). Impact of clouds on atmospheric heating based on the R04 CloudSat fluxes and heating rates data set. *Journal of Geophysical Research, 113*. doi:10.1029/2008jd009951

Lenaerts, J. T., Van Tricht, K., Lhermitte, S., & L'ecuyer, T. S. (2017). Polar clouds and radiation in satellite observations, reanalyses, and climate models. *Geophysical Research Letters, 44*(7), 3355-3364. doi:10.1002/2016gl072242

Lesins, G., Duck, T. J., & Drummond, J. R. (2010). Climate trends at Eureka in the Canadian high arctic. *Atmosphere-Ocean, 48*(2), 59-80. doi:10.3137/ao1103.2010

Li, Z., & Xu, K. (2020). Arctic Clouds Simulated by a Multiscale Modeling Framework and Comparisons With Observations and Conventional GCMs. *Journal of Geophysical Research: Atmospheres, 125*(1). doi:10.1029/2019jd030522

Liu, Y., Key, J. R., Frey, R. A., Ackerman, S. A., & Menzel, W. (2004). Nighttime polar cloud detection with MODIS. *Remote Sensing of Environment, 92*(2), 181-194. doi:10.1016/j.rse.2004.06.004

Liu, Y., Ackerman, S. A., Maddux, B. C., Key, J. R., & Frey, R. A. (2010). Errors in Cloud Detection over the Arctic Using a Satellite Imager and Implications for Observing Feedback Mechanisms. *Journal of Climate, 23*(7), 1894-1907. doi:10.1175/2009jcli3386.1

Liu, Y., Key, J. R., Liu, Z., Wang, X., & Vavrus, S. J. (2012). A cloudier Arctic expected with diminishing sea ice. *Geophysical Research Letters, 39*(5). doi:10.1029/2012gl051251

Liu, Y., & Key, J. R. (2016). Assessment of Arctic Cloud Cover Anomalies in Atmospheric Reanalysis Products Using Satellite Data. *Journal of Climate, 29*(17), 6065-6083. doi:10.1175/jcli-d-15-0861.1

Liu, Y., Shupe, M. D., Wang, Z., & Mace, G. (2017). Cloud vertical distribution from combined surface and space radar–lidar observations at two Arctic atmospheric observatories. *Atmospheric Chemistry and Physics, 17*(9), 5973-5989. doi:10.5194/acp-17-5973-2017

Liu, Y., Key, J. R., Vavrus, S., & Woods, C. (2018). Time Evolution of the Cloud Response to Moisture Intrusions into the Arctic during Winter. *Journal of Climate, 31*(22), 9389-9405. doi:10.1175/jcli-d-17-0896.1

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., . . . Wong, T. (2009). Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget. *Journal of Climate, 22*(3), 748-766. doi:10.1175/2008jcli2637.1

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., . . . Kato, S. (2018). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate, 31*(2), 895-918. doi:10.1175/jcli-d-17-0208.1

Mariage, V., Pelon, J., Blouzon, F., Victori, S., Geyskens, N., Amarouche, N., . . . Provost, C. (2017). IAOOS microlidar-on-buoy development and first atmospheric observations obtained during 2014 and 2015 arctic drifts. *Optics Express, 25*(4). doi:10.1364/oe.25.000a73

McArthur L. J. B. (2005). Baseline Surface Radiation Netword (BSRN). Operations Manual Version 2.1, WCRP-121, WMO/TD-No. 1274

McBean, G., and Coauthors (2005). *Arctic climate: Past and present. Arctic Climate Impact Assessment Scientific Rep*., Cambridge University Press, 21–60. [Available online at http://www.acia.uaf.edu/PDFs/ACIA_Science_Chapters_Final/ACIA_Ch02_Final.pdf.]

Matus, A. V., & L'ecuyer, T. S. (2017). The role of cloud phase in Earth's radiation budget. *Journal of Geophysical Research: Atmospheres, 122*(5), 2559-2578. doi:10.1002/2016jd025951

Palerme, C., Claud, C., Wood, N. B., L'ecuyer, T., & Genthon, C. (2019). How Does Ground Clutter Affect CloudSat Snowfall Retrievals Over Ice Sheets? *IEEE Geoscience and Remote Sensing Letters, 16*(3), 342-346. doi:10.1109/lgrs.2018.2875007

Palm, S. P., Strey, S. T., Spinhirne, J., & Markus, T. (2010). Influence of Arctic sea ice extent on polar cloud fraction and vertical structure and implications for regional climate. *Journal of Geophysical Research, 115*(D21). doi:10.1029/2010jd013900

Provost C., Pelon, J., Sennéchael, N., Calzas, M., Blouzon, F., et al. (2015). IAOOS (Ice - Atmosphere - Arctic Ocean Observing System, 2011-2019). *Mercator Ocean Quarterly Newsletter*, Mercator Ocean, Special Issue with ICE-ARC, pp.13-15

Rahn, K. A. (1981). Relative importances of North America and Eurasia as sources of arctic aerosol. *Atmospheric Environment (1967), 15*(8), 1447-1455. doi:10.1016/0004-6981(81)90351-6

Sedlar, J., Tjernström, M., Mauritsen, T., Shupe, M. D., Brooks, I. M., Persson, P. O., . . . Nicolaus, M. (2010). A transitioning Arctic surface energy budget: The impacts of solar zenith angle, surface albedo and cloud radiative forcing. *Climate Dynamics*, *37*(7-8), 1643-1660. doi:10.1007/s00382-010-0937-5

Sedlar, J., Shupe, M. D., & Tjernström, M. (2012). On the Relationship between Thermodynamic Structure and Cloud Top, and Its Climate Significance in the Arctic. *Journal of Climate, 25*(7), 2374-2393. doi:10.1175/jcli-d-11-00186.1

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change, 77*(1-2), 85-96. doi:10.1016/j.gloplacha.2011.03.004

Serreze, M. C., & Barry, R. G. (2014). *The Arctic climate system*. New York, NY: Cambridge University Press.

Shupe, M. D. (2007). A ground-based multisensor cloud phase classifier. *Geophysical Research Letters, 34*(22). doi:10.1029/2007gl031008

Shupe, M. D., Walden, V. P., Eloranta, E., Uttal, T., Campbell, J. R., Starkweather, S. M., & Shiobara, M. (2011a). Clouds at Arctic Atmospheric Observatories. Part I: Occurrence and Macrophysical Properties. *Journal of Applied Meteorology and Climatology, 50*(3), 626-644. doi:10.1175/2010jamc2467.1

Shupe, M. D. (2011b). Clouds at Arctic Atmospheric Observatories. Part II: Thermodynamic Phase Characteristics. *Journal of Applied Meteorology and Climatology*, *50*(3), 645-661. doi:10.1175/2010jamc2468.1

Shupe, M. D., Persson, P. O., Brooks, I. M., Tjernström, M., Sedlar, J., Mauritsen, T., . . . Leck, C. (2013). Cloud and boundary layer interactions over the Arctic sea ice in late summer. *Atmospheric Chemistry and Physics*, *13*(18), 9379-9399. doi:10.5194/acp-13-9379-2013

Shupe, M. D., Tjernstrom, M., & Persson, P. O. G., (2015a). Challenge of Arctic clouds and their implications for surface radiation [in "State of the Climate in 2014"], *Bull. Amer. Meteor. Soc.*, 96(7), S130-S131

Shupe, M. D., Turner, D. D., Zwink, A., Thieman, M. M., Mlawer, E. J., & Shippert, T. (2015b). Deriving Arctic Cloud Microphysics at Barrow, Alaska: Algorithms, Results, and Radiative Closure. *Journal of Applied Meteorology and Climatology*, *54*(7), 1675-1689. doi:10.1175/jamc-d-15-0054.1

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., . . . Lebsock, M. (2018). CloudSat and CALIPSO within the A-Train: Ten Years of Actively Observing the Earth System. *Bulletin of the American Meteorological Society, 99*(3), 569-581. doi:10.1175/bams-d-16-0324.1

Uttal, T., Curry, J. A., Mcphee, M. G., Perovich, D. K., Moritz, R. E., Maslanik, J. A., . . . Grenfeld, T. C. (2002). Surface Heat Budget of the Arctic Ocean. *Bulletin of the American Meteorological Society*, *83*(2), 255-275. doi:10.1175/1520-0477(2002)0832.3.co;2

Uttal, T., Starkweather, S., Drummond, J. R., Vihma, T., Makshtas, A. P., Darby, L. S., . . . Intrieri, J. M. (2016). International Arctic Systems for Observing the Atmosphere: An International Polar Year Legacy Consortium. *Bulletin of the American Meteorological Society, 97*(6), 1033-1056. doi:10.1175/bams-d-14-00145.1

Wang, C., Graham, R. M., Wang, K., Gerland, S., & Granskog, M. A. (2019). Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and evolution. *The Cryosphere, 13*(6), 1661-1679. doi:10.5194/tc-13-1661-2019

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bulletin of the American Meteorological Society, 77*(5), 853-868. doi:10.1175/1520-0477(1996)0772.0.co;2

Zib, B. J., Dong, X., Xi, B., & Kennedy, A. (2012). Evaluation and Intercomparison of Cloud Fraction and Radiative Fluxes in Recent Reanalyses over the Arctic Using BSRN Surface Observations. *Journal of Climate, 25*(7), 2291-2305. doi:10.1175/jcli-d-11-00147.1

Zygmuntowska, M., Mauritsen, T., Quaas, J., & Kaleschke, L. (2012). Arctic Clouds and Surface Radiation – a critical comparison of satellite retrievals and the ERA-Interim reanalysis. *Atmospheric Chemistry and Physics, 12*(14), 6667-6677. doi:10.5194/acp-12-6667-2012

## Appendix A: Screening frost events

LWD measured by the BSRN station at Eureka was frequently affected by the presence of ice on the sensor dome, likely in the form of frost or rime. Though the signal caused by ice is difficult to distinguish from the signal caused by clouds, occurrences of icing in the data set were identifiable by the characteristic "growth curve” (a rapid increase in signal that plateaus as the coverage of ice over and optical depth of the ice increases) over the course of 12-48 hours followed by abrupt decreases in flux when the domes were cleaned by the tech (e.g., Figure A1a). The data set was visually screened for these occurrences and the suspect data removed. Figure A1b shows the composite percentage of data for each month of the year when data was removed, indicating that ice occurs ~10-25% of the time from October through March, but that it rarely occurs in other months.

To further ensure data quality of the BSRN data used for the direct comparisons with the satellite products, additional screening was performed on coincident cases used for analysis. This screening revisited to possibility of icing and also investigated whether cloud cover was sufficiently uniform so as to be representative of the fields of view from both the surface and satellite measurements. Outliers in the comparison (examples in Figure A1c) received this further scrutiny; the signal was reviewed again for evidence of icing and the tech’s logbook notes were reviewed. Additionally, observations from a nearby vertically-pointing cloud radar (refer to Shupe et al. 2011a for instrument details) were reviewed for evidence of temporally-heterogeneous cloud cover (e.g., frontal systems passing near in time to overpass) that would suggest a scene mismatch as an explanation for the discrepancy between the surface and satellite observations.

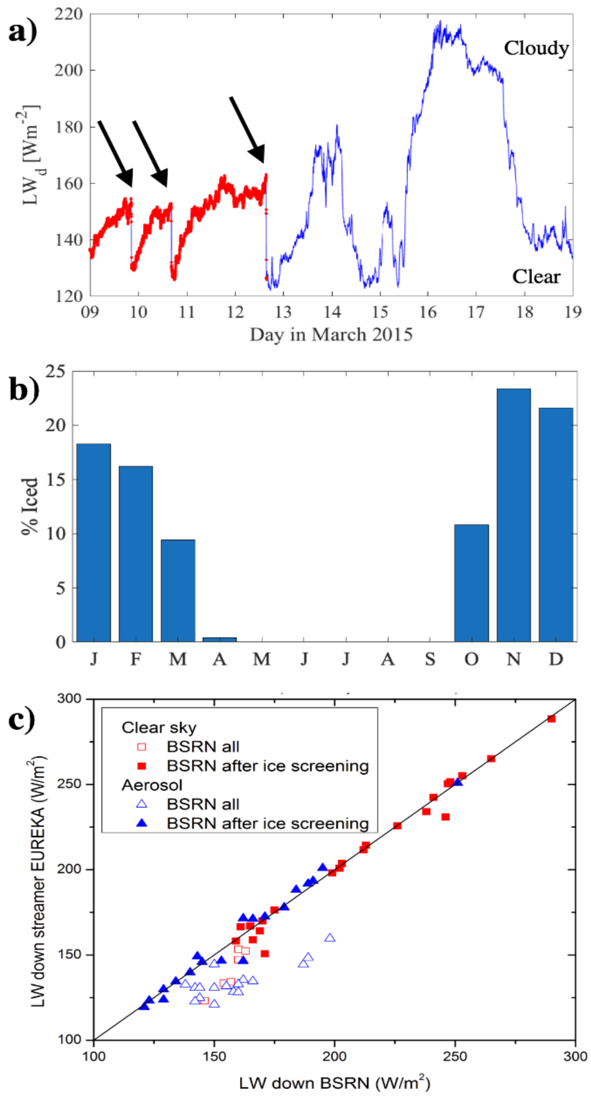



**Figure A1**. (a) Example of the signal caused by icing of the upward-facing pyrgeometer (LWD measurement) occurring during clear sky on three consecutive days in March 2015 at Eureka. Red dots denote the identifications of ice. The arrows point to the times when the technician cleaned the ice from the dome. The magnitude of the signal caused by the icing is similar to the variability cause by alternating clear-sky and cloudy periods between day 13 and 19. (b) Percent of time in each month (aggregate 2007-2011) when ice was identified and removed. (c) Ice screening results of coincident measurements from BSRN and simulations EUR-LR. The colors indicate the scene classification based on ground-based observations, where aerosol meaning that an aerosol layer with an optical depth larger than 0.2 was observed. Suspect behavior (open symbols) were revisited. The screening method helps to remove suspicious measurements (184572 cases representing about 8% of the whole dataset). The screened observations are not used in our trend analysis, but allows to get a reliable reference dataset for comparisonsLW TOA departure from CERES as reported in Table 4. C-C-FLX R05 comparison with CERES is based on only 221 points (249 for R04 and other datasets).

## Appendix B: Annual cloud vertical distribution

Cloud vertical distribution and cloud type are indeed key parameters in flux calculations. We here compare input vertical profiles from satellite and ground/based measurements and reanalysis between June 2006 and May 2010 (green lines in Figure 2). In this paper, as a conclusion from B14, the EUR-LR is considered to be the reference for the low-level clouds, whereas space radar-lidar data are considered as such for upper level data (> 6 km).

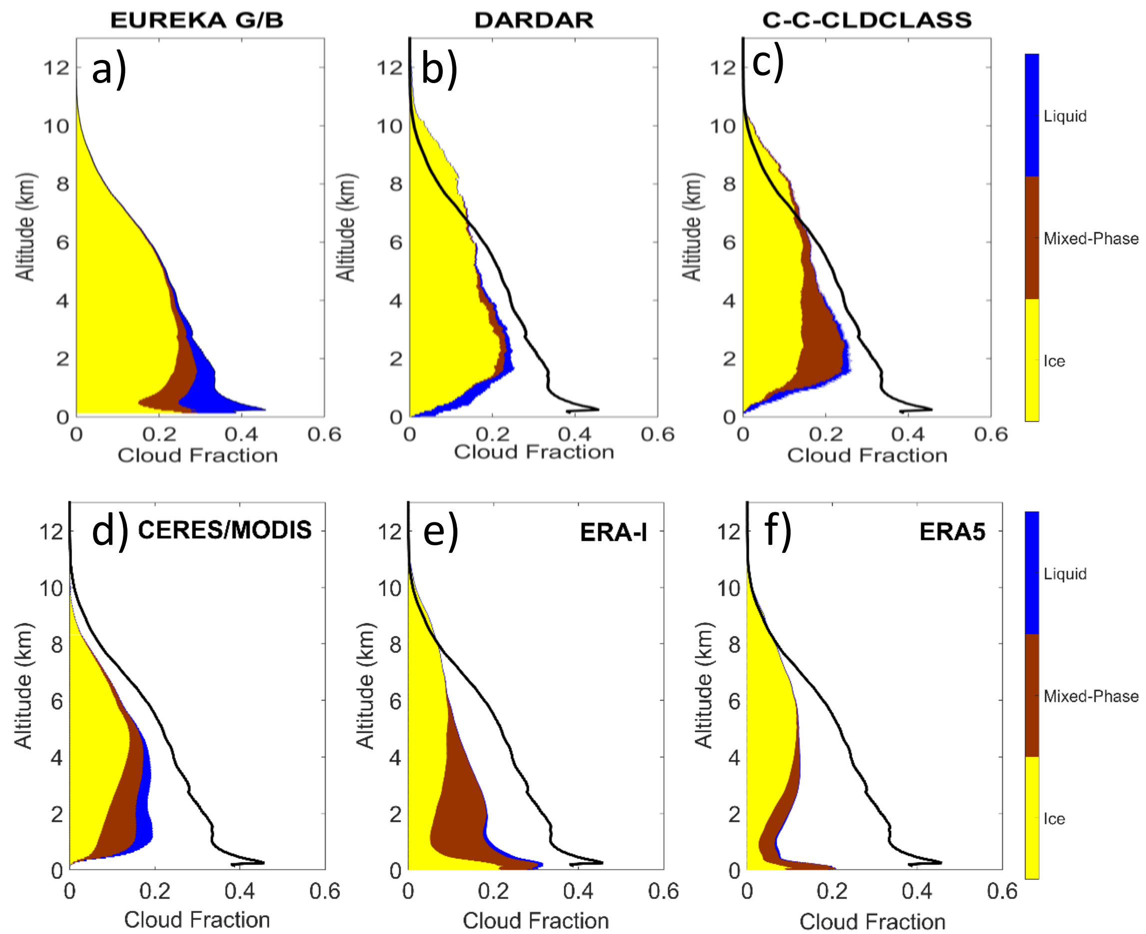



**Figure B1**. Cumulated vertical cloud-type distribution between June 2006 and May 2010 for all the independent datasets (a) EUR-LR; b) DARDAR; c) CloudSat 2B-CLDCLASS-LIDAR R05 (C-C); d) MODIS; e) ERA-I; f) ERA5 at less than 25 km from the station as compared to ground-based observations at EUREKA (solid line) given as the total cloud occurrence in Fig. 4a. We have reported EUR-LR total cloud fraction from FigB1a in all other figures B1b, c, d, e, f

DARDAR and C-C cloud vertical distributions are very close to EUR-LR above 2 km. MODIS are close to DARDAR, although MODIS is slightly more biased below 4 km. As detailed in B14, very low clouds are difficult to address from space, and this is confirmed here from DARDAR, C-C and MODIS for which cloud fractions are much lower than EUR-LR observations below 2 km. DARDAR and C-C are close although DARDAR gives a higher amount of ice clouds and less mixed-phase clouds. The finer vertical resolution for DARDAR (60 m) compared to C-C (240 m) might explain this difference as C-C would not be able to distinguish different water phases within one radar gate. ERA-I misses a large fraction of cloud below 8 km (Liu and Key, 2016). ERA5 appears to have a bias larger than ERA-I, and the fraction of mixed-phased clouds is observed to be much smaller. All behaviours are however rather similar, with more or less important bias in the vertical cloud fraction but significant biases below 2 km.

## Appendix C: Comparison of LW BOA CloudSat flux products R04 and R05

A comparison is made on cloud types and cloud properties for 130 cases of observations with C-C-FLX R04 and R05 products.

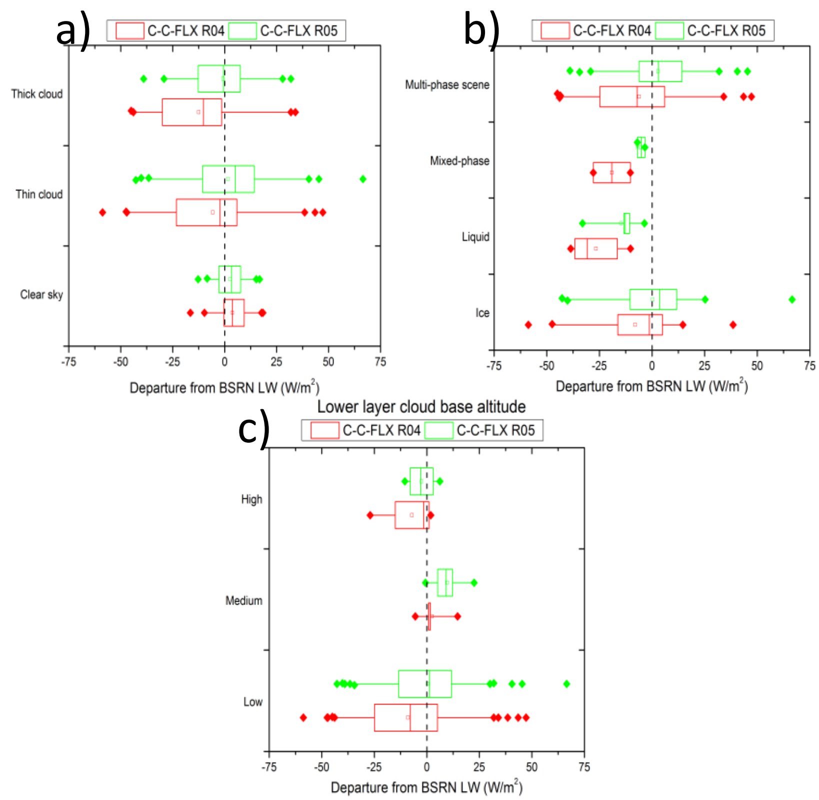



**Figure C1**. Difference of BOA LW fluxes of C-C-FLX R04 (red) and C-C-FLX R05 (green) depending on the total visible optical depth (a), phase of cloud layers (b) and bottom height of the lower layer (c). Boxes correspond to 25%, median and 75% values, thin bars show 5 and 95% and squares are used to show the mean. Outliers are also reported as coloured diamonds.

Figure C1 shows box plots of CloudSat products departure from BSRN for 130 coincident cases to evaluate the impact of the changes in R05. Those changes were made to improve the representation of cloud properties (cloud detection, supercooled liquid and ice clouds microphysical properties) and consistency with CALIPSO cloud products along with updated data ingested. We found that R05 significantly reduces the strong biases identified for low and thick clouds. It confirms the importance of resolving cloud phase vertical distribution in surface flux calculations.