Supplementary Material for "Four decades of global surface albedo estimates in the third edition of the CLARA climate data record"
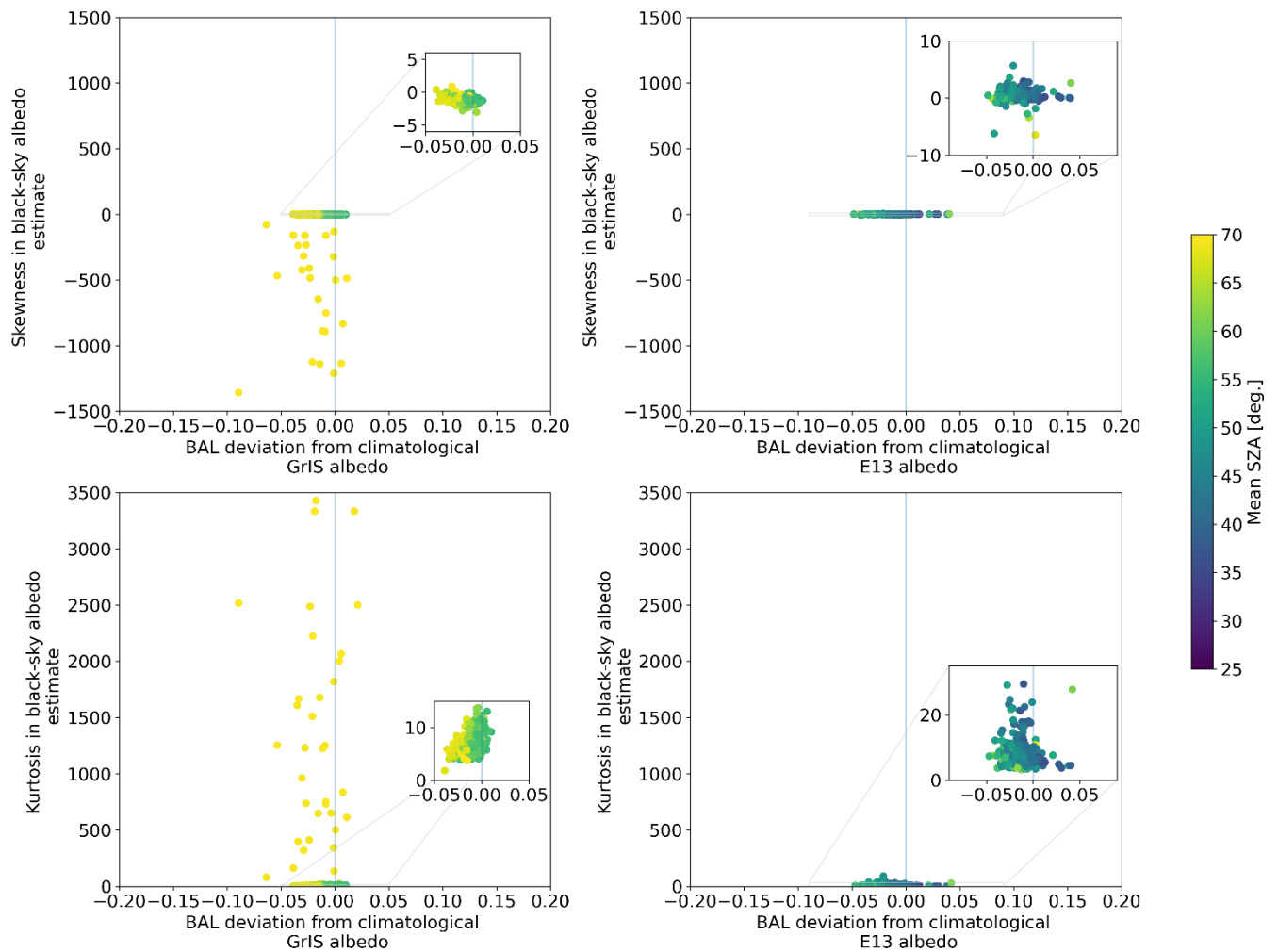
Figure S1: Deviation of monthly mean BAL from climatologically expected surface albedos over central GrIS (left column) and the Kansas plains containing the E13 BSRN site (right column) as a function of skewness (top row) and kurtosis (bottom row) of SAL. Marker color indicates mean SZA of clear-sky observations for each month.

Supplementary Table 1: Climatological blue-sky albedos used to derive albedo maps from Dynamic World LU data

Land cover	Assigned climatological blue-sky albedo	Source
Water	0.076	Consistent with CLARA-A1/A2
Forest	0.14	He et al. (2014)
Flooded vegt.	0.16	He et al. (2014)
Cropland	0.19	He et al. (2014)
Shrub	0.2	He et al. (2014)
Built area	0.15	Trlica et al. (2017)
Bare ground	0.25	He et al. (2014)
Snow and ice	0.8	He et al. (2014)
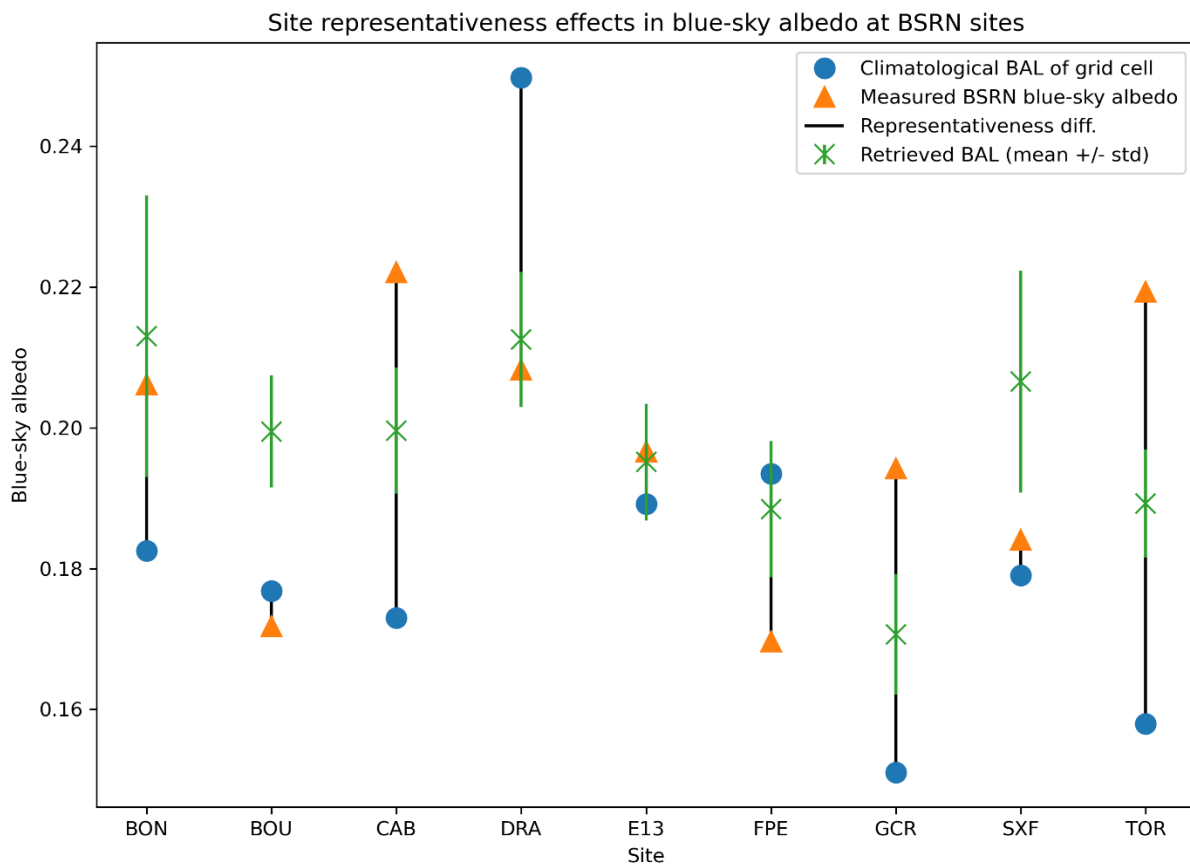
Figure S2: Spatial representativeness analysis for BSRN sites. Blue circles indicate expected grid cell-mean albedos from Google Earth Engine's Dynamic World land cover data combined with climatological surface albedos for each land cover class. Orange triangles indicate measured mean albedo at the BSRN site. The lengths of the black lines connecting the two are indicative of spatial unrepresentativeness of the station measurement for its surroundings; retrieved mean CLARA BAL of the grid cell in question is marked with green cross-with-whiskers, whiskers denoting standard deviation of BAL. Dynamic World data is from June-August 2018; BAL and BSRN observations are means of all data available at each site in question.

CAB, DRA, GCR, and TOR display the largest differences (>0.04) between (expected) grid cell and *in situ* albedo. However, a close inspection of the classifications in Dynamic World reveals that the area around DRA in Southwestern US is classified as "bare ground", whose climatological surface albedo is markedly high at 0.25 (He et al., 2014). Conversely, the ESA CCI land cover data (ESA, 2017) classifies the region as "shrubland", whose climatological albedo of 0.2 is in close agreement with both the *in situ* measurement and the mean

retrieved BAL. We therefore consider DRA as representative despite the apparent disagreement. At CAB, CCI land cover reports predominance of grassland, whereas DW assigns the area to a grassland-cropland mix. The fractions of built area also disagree somewhat at CAB, leading to some ambiguity in whether the site should be considered representative or not. We elected to retain CAB in the summary statistics. However, at GCR and TOR the disagreement clearly results from a large proportion of forest cover in the CLARA grid cell, whose effects on albedo are captured by the BAL retrievals but not the near-surface *in situ* measurement. GCR and TOR are therefore excluded from further retrieval performance analysis.

Interestingly, albeit expectedly, BAL retrievals with their standard deviation fall within the 'representativeness range' at all sites excepting two: BOU and SXF. Both sites are characterized by the presence of large metropolitan areas in the CLARA grid cell though the site itself is rural. The higher-than-expected BAL may therefore indicate suboptimal atmospheric correction accuracy close to urban centers, which would be in accordance with the design of SMAC, and the potentially missed small-scale aerosol variability over such areas.

For 3 of the 4 polar BSRN sites (ALE, GVN, SYO), the principal representativeness challenge rises from the variable delineation between sea ice and open water in the CLARA grid cells containing them. We calculated and extracted time series (not shown) of observed open water fraction from Sentinel-2 in DW at these sites, noting that open water fractions were at times quite large around GVN and ALE (>0.2), though cloudiness often limited available sampling markedly. Therefore, we elected to also exclude GVN and ALE from the results.
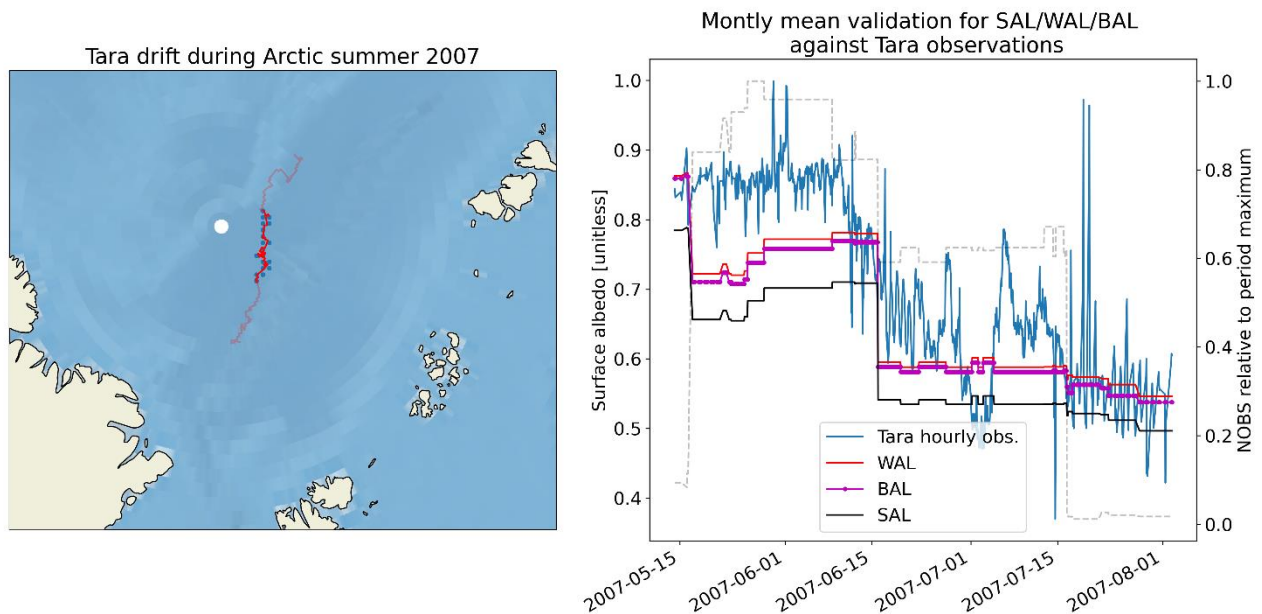
Figure S3: Left: Tara drift across the Arctic Ocean during summer 2007. Right: Hourly in situ observed surface albedo at Tara (blue), and the retrieved grid cell-scale BAL (purple), WAL (red), and SAL (black). Identificaiton of the CLARA grid cell containing the Tara ice camp is updated for each hourly observation, explaining the at-times rapid shifts in satellite-based albedo estimates. Dashed line on the right y-axis shows the number of valid AVHRR-GAC observations in each monthly mean, normalized to the maximum observed during the summer in question.
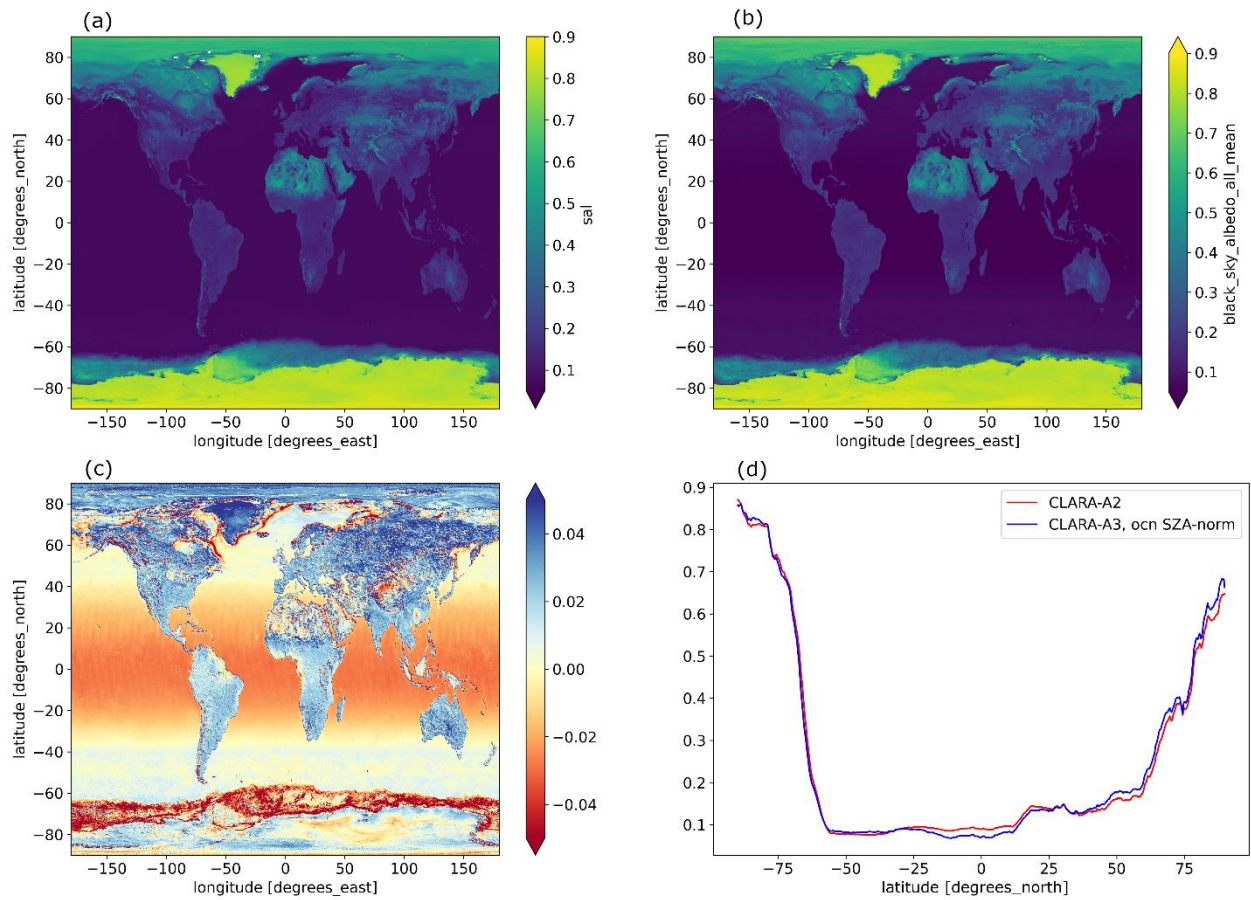
Figure S4: Intercomparison of CLARA-A3 and CLARA-A2 black-sky albedo estimates. Data shown corresponds to the mean of April-September 2015. (a) CLARA-A2, (b) CLARA-A3, (c) difference (CLARA-A3 - CLARA-A2), (d) zonal means of CLARA-A2 and CLARA-A3.

Additional references:

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf