

A monthly 1-degree resolution dataset of daytime cloud fraction over the Arctic during 2000–2020 based on multiple satellite products

Xinyan Liu¹, Tao He¹, Shunlin Liang², Ruibo Li³, Xiongxin Xiao⁴, Rui Ma¹, Yichuan Ma¹

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

² Department of Geography, University of Hong Kong, Hong Kong 999077, China

³ State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

⁴ Institute of Geography and Oeschger Center for Climate Change Research, University of Bern, Bern 3012, Switzerland

Correspondence to: Tao He (taohers@whu.edu.cn)

Abstract. The low accuracy of satellite cloud fraction (CF) data over the Arctic seriously restricts the accurate assessment of the regional and global radiative energy balance under a changing climate. Previous studies have reported that no individual satellite CF product could satisfy the needs of accuracy and spatio-temporal coverage simultaneously for long-term applications over the Arctic. Merging multiple CF products with complementary properties can provide an effective way to produce a spatiotemporally complete CF data record with higher accuracy. This study proposed a spatiotemporal statistical data fusion framework based on cumulative distribution function (CDF) matching and the Bayesian maximum entropy (BME) method to produce a synthetic 1°×1° CF dataset in the Arctic during 2000–2020. The CDF matching was employed to remove the systematic biases among multiple passive sensor datasets through the constraint of using CF from an active sensor. The BME method was employed to combine adjusted satellite CF products to produce a spatiotemporally complete and accurate CF product. The advantages of the presented fusing framework are that it not only uses the spatiotemporal autocorrelations but also explicitly incorporates the uncertainties of passive sensor products benchmarked with reference data, i.e., active sensor product and ground-based observations. The inconsistencies of Arctic CF between passive sensor products and the reference data were reduced by about 10–20% after fusing, with particularly noticeable improvements in the vicinity of Greenland. Compared with ground-based observations, R^2 increased by about 0.20–0.48, the root mean square error (RMSE) and bias reductions averaged about 6.09% and 4.04% for land regions, respectively; these metrics for ocean regions were about 0.05–0.31, 2.85%, and 3.15%, respectively. Compared with active sensor data, R^2 increased by nearly 0.16, and RMSE and bias declined by about 3.77% and 4.31%, respectively, in land; meanwhile, improvements in ocean regions were about 0.3 for R^2 , 4.46% for RMSE and, 3.92% for bias. The results of the comparison with the ERA5 and the MRI-AGCM3-2-S climate model suggest an obvious improvement in the consistency between the satellite-observed CF and the reanalysis and model data after fusion. This serves as a promising indication that the fused CF results hold the potential to deliver reliable satellite observations for modeling and reanalysis data. Moreover, the fused product effectively supplements the temporal gaps of AVHRR-based products caused by satellite faults and the data missing from MODIS-based products prior to the launch of Aqua, and extends

40 the temporal range better than the active product; it addresses the spatial insufficiency of the active sensor
41 data and the AVHRR-based products acquired at latitudes greater than 82.5 °N. A continuous monthly 1-
42 degree CF product covering the entire Arctic during 2000–2020 was generated and is freely available to
43 the public at <https://doi.org/10.5281/zenodo.7624605> (Liu et al., 2022). This is of great importance for
44 reducing the uncertainty in the estimation of surface radiation parameters and thus helps researchers to
45 better understand the earth’s energy imbalance.

46 **1 Introduction**

47 Clouds substantially affect Earth’s energy budget by reflecting solar radiation back to space and by
48 restricting emissions of thermal radiation into space (Ramanathan et al., 1989; Van Tricht et al., 2016;
49 Danso et al., 2020). Clouds are also an essential variable in the climate system because they are directly
50 associated with precipitation and aerosol loading (Toll et al., 2019; Poulsen et al., 2016). The cloud
51 fraction (CF), which represents the amount of sky estimated to be covered by a specified cloud type or
52 level (partial CF) or by all cloud types and levels (total CF), has long been recognized as a major source
53 of uncertainty when estimating radiation flux and future climate change (Xie et al., 2010; Liu et al., 2011a;
54 Qian et al., 2012; Danso et al., 2020). An accurate representation of CF is essential for the evaluation of
55 regional and global energy budgets as well as for predicting future climatic conditions. However,
56 variances in CF definitions and system differences commonly exist among different sources of data. As
57 a solution, the fused product provides a higher level of definition consistency and accuracy in comparison
58 to alternative datasets.

59 By making spatially continuous observations, satellites provided us with an unprecedented
60 advantage in assessing regional and global cloud effects. In the last few decades, increased effort has
61 been made to develop, analyze, and validate global or regional cloud property datasets that are based on
62 long-term satellite observations (Heidinger et al., 2014; Hollmann et al., 2013; Karlsson and Devasthale,
63 2018; Marchant et al., 2016; Rossow and Schiffer, 1999; Stubenrauch et al., 2013; Enriquez-Alonso et
64 al., 2016; Sun et al., 2015; Tzallas et al., 2019; Wu et al., 2014). Studies have also shown that although
65 different cloud datasets were derived from different observation instruments and algorithms, most of
66 them provide quite consistent CF observations in middle and lower-latitude regions (Karlsson and
67 Devasthale, 2018; Stengel et al., 2017; Claudia, 2012). However, systematic errors and artifacts exist in
68 CF data, so some inconsistencies inevitably occur among different datasets (Sun et al., 2015; Tzallas et
69 al., 2019; Wu et al., 2014), especially in the polar regions (Liu et al., 2022). Perennial snow/ice coverage
70 coupled with frequent moisture inversions in Arctic has limited the cloud detection capabilities of passive
71 sensor datasets, where the differences between these various datasets tend to be about two-fold in
72 magnitude when compared with datasets acquired at other latitudes (Karlsson and Devasthale, 2018; Liu
73 et al., 2022; Stubenrauch et al., 2013). The uncertainties of the annual global surface downward longwave
74 (LW) and shortwave (SW) fluxes caused by satellite-derived cloud properties were calculated at about
75 2% (7 Wm^{-2} and 4 Wm^{-2} , respectively) and those for global surface upward LW and SW were about
76 0.8% (about 3 Wm^{-2}) and 13% (also 3 Wm^{-2}), respectively (Kato et al., 2011; Kato et al., 2012; Kim
77 and Ramanathan, 2008). It should be noted that the differences in CF may have a more obvious impact
78 on the surface radiation budget in high-latitude polar regions. Kennedy et al. (2012) found that the CF
79 bias might cause monthly biases in Arctic surface SW and LW fluxes over 90 and 60 Wm^{-2} for some
80 reanalyses, respectively (Kennedy et al., 2012). Walsh et al. (2009) proposed that the bias of summer
81 low-level CF would create deviations of about 160 Wm^{-2} in estimated SW radiation (Walsh et al., 2009).

82 Some other related studies have also found that the variances of annual Arctic surface radiation
83 estimation caused by CF uncertainty were higher than 10 Wm^{-2} (Hakuba et al., 2017; Kato et al., 2018b;
84 Huang et al., 2017). Therefore, relying on a single CF dataset may introduce large uncertainty when
85 analyzing the cloud dynamics over the Arctic, further affecting the estimated energy budget and related
86 climate applications.

87 Each cloud dataset has its own advantages and disadvantages in Arctic CF detection. The Advanced
88 Very High Resolution Radiometer (AVHRR) offers the longest continuous satellite observation records
89 extending from 1978 to the present and provides daily global coverage based on data from several
90 AVHRRs. With the successful operation of new generations of satellites, the frequency of global view
91 has increased to more than eight each day, which provides richer angular information for CF observations
92 (Heidinger et al., 2014; Karlsson et al., 2017). Many cloud products exist that are based on AVHRR
93 sensors. The International Satellite Cloud Climatology Project (ISCCP) H-series product relies on newer
94 passive imagers with higher spectral, spatial, radiometric, and temporal resolutions; it provides revised
95 daytime cloud detection over snow and ice in polar regions (Young et al., 2018). Moreover, the ISCCP
96 is largely unaffected by the AVHRR orbital drifts (Loyola R et al., 2010; Liu et al., 2022). The CM SAF
97 cLoud, Albedo, and RAdition datasets (CLARA-A1/A2) systematically use CALIPSO-CALIOP cloud
98 information for development and validation purposes, and it optimizes the detection conditions during
99 the polar day over snow- and ice-covered surfaces (Karlsson et al., 2017; Karlsson and Hakansson, 2018).
100 The AVHRR Pathfinder Atmospheres - Extended (PATMOS-x) product is the first multi-parameter
101 dataset that is making use of all AVHRR channels. This product has a relatively finer spatial resolution
102 than other AVHRR-based records, and it also improves cloud detection based on active sensor data
103 (Heidinger et al., 2012; Heidinger et al., 2014). However, the AVHRR-based products are often reported
104 to underestimate Arctic CF because of the limitations in radiation correction and spatial bands (Stengel
105 et al., 2017; Kotarba, 2015). In addition, the United States National Oceanic and Atmospheric
106 Administration's (NOAA's) archiving of data has its own problems with intermittent occurrences of gaps,
107 duplications, and corrupt data as well as the orbit drifts of satellites (Karlsson et al., 2017). Beginning in
108 2000, the higher resolution, higher calibration accuracy, and larger number of spectral bands used in the
109 Moderate Resolution Imaging SpectroRadiometer (MODIS) cloud products resulted in more robust, but
110 shorter-length products than AVHRR (Kennedy et al., 2012; Claudia, 2012; Stengel et al., 2017),
111 including MOD08/MYD08 (Marchant et al., 2016) and the Clouds and the Earth's Radiant Energy
112 System (CERES) (Kato et al., 2018b; Minnis et al., 2011). Meanwhile, the MODIS-based products are
113 usually reported to overestimate the CF in the Arctic (Trepte et al., 2019; Liu et al., 2022). Although
114 passive sensor data provide a long time series of continuous CF data covering the entire Arctic region,
115 the limitations of visible and thermal channels in distinguishing clouds from snow and ice cause the cloud
116 results of passive sensor data in the high-latitude bright cold polar regions to have questionable accuracy
117 (Eastman and Warren, 2010; Liu et al., 2010; Liu et al., 2012a; Philipp et al., 2020). Active instruments,
118 such as CALIOP, do not rely on thermal or visible contrasts in detecting clouds, so they are regarded as
119 an excellent reference for passive data collection in transient and zonal scenarios (Stubenrauch et al.,
120 2013; Stengel et al., 2017). However, the number of CALIPSO spatial samplings is too low to overlap
121 large areas repeatedly in a short time, and the CALIPSO imagers only cover the regions within 82.5°N
122 latitudes, which greatly reduced spatial and temporal coverages when compared with passive sensor
123 sensors (Liu et al., 2022; Claudia, 2012; Stubenrauch et al., 2013). Moreover, differences in
124 instrumentation impose these different cloud definitions, which further increase the biases between the
125 passive sensor data and the active sensor data. Therefore, an effective method for blending the advantages

126 of multiple satellite products should yield more accurate Arctic CF products based on a variety of
127 observations and algorithms.

128 Several studies have been dedicated to correcting passive sensor data based on active sensor data
129 with the goal of improving the accuracy of CF products. Philipp et al. (2020) corrected passive sensor
130 CF data by constructing a function of the sea ice concentration in different seasons and the CF bias in
131 data acquired from active and passive sensors, which showed reliable results for low-level cloud cover
132 identification where the sea ice concentration was known (Philipp et al., 2020). Kotarba (2020) matched
133 the CALIPSO profile data and the MODIS instantaneous field of view to correct passive sensor data
134 (Kotarba, 2020). This method can be used as an important reference for short-term research that focused
135 a small area, while the efficiency of the algorithm is also important for the correction of long-time series
136 and large-scale data. Given that passive sensor CFs exhibit seasonal fluctuations similar to those of active
137 sensor data (peaking in September and minimizing in April in the Arctic), an approach based on
138 cumulative distribution function (CDF) matching using time series data may be able to improve both the
139 accuracy and efficiency of CF detection. Using CDF matching can reduce the systematic bias and root
140 mean square errors (RMSEs) between target and reference datasets while maintaining the relative
141 relationship, which has been successfully applied in the study of soil moisture, surface emissivity spectra,
142 precipitation, and land surface temperature (Drusch, 2005; Brocca et al., 2011; Liu et al., 2011b; Zhang
143 et al., 2018; Nie et al., 2016; Xu and Cheng, 2021).

144 In the field of meteorology, to obtain more accurate cloud coverage information, multi-source data
145 fusion is usually performed based on spectral bands and scale geometry information of instantaneous
146 satellite images. Examples include various transforms including the contourlet (Miao and Wang, 2006;
147 Jin et al., 2011), curvelet (Li and Yang, 2008; Liu et al., 2015), NSCT (Wang et al., 2012), and tetrolet
148 transforms (Zhang et al., 2014). Alternatively, based on the field of view of different observation
149 instruments used to acquire satellite images and of ground-based stations, methods such as the stepwise
150 revision method (Kenyon et al., 2016) and data assimilation technology (Hu and Xue, 2007) have been
151 used. However, in the climate domain, the estimation of a radiative energy budget on a large scale over
152 a long time series usually requires monthly climate model grid data (Kato et al., 2018a; Sledd and
153 L'ecuyer, 2021). Using fused instantaneous data to extrapolate climate-scale data may result in a large
154 accumulation of errors. In recent decades, the fusion of multi-sensor thematic products in climate-scale
155 studies has been widely used and developed. Two main types of methods exist for merging multiple
156 satellite thematic products based on the principle of calculation. One type of fusing approach provides
157 spatiotemporal data fusion by spectral correlation, which is more suitable for the regions where the spatial
158 information of objects has no obvious change, such as the Spatial and Temporal Adaptive Reflectance
159 Fusion Model (STARFM) and the improved STARFM (Gao et al., 2006; Hilker et al., 2009; Zhu et al.,
160 2010; Zhang et al., 2014). The other type of spatiotemporal data fusing method is data-driven, which
161 involves developing geostatistical models to solve the problem created when the same parameter is
162 inconsistent among different satellite products. This method includes the Kriging family of techniques
163 (Chatterjee et al., 2010; Li et al., 2014; Savelyeva et al., 2008), the spatiotemporal interpolation method
164 (Yang and Hu, 2018), and the Bayesian melding framework (Fuentes and Raftery, 2005; Christakos,
165 2010). However, these methods rely on Gaussian assumptions and linear models, which limits their
166 estimation accuracy (Nazelle et al., 2010; He and Kolovos, 2017). A nonlinear spatiotemporal
167 geostatistical method, Bayesian maximum entropy (BME), has been proposed to fuse the parameters that
168 have apparent spatiotemporal variations (Nazelle et al., 2010). The BME method can integrate
169 information from different sources and then consider the data uncertainties in achieving improved

170 prediction accuracy. The most important advantage of BME is that it does not restrict the complex
171 stochastic relationship between predictions/observations and ‘true’ values to the Gaussian linearized
172 model; this is an obvious breakthrough over approaches restricted to using normal distributions (Nazelle
173 et al., 2010; Li et al., 2013; Xu et al., 2019). The BME method has broad application in the assessment
174 of many different atmosphere parameters, such as ozone concentration (Nazelle et al., 2010; Bogaert et
175 al., 2009; Christakos et al., 2004), $PM_{2.5}$, PM_{10} (Yu and Wang, 2010; Beckerman et al., 2013), and aerosol
176 optical depth (Xia et al., 2022; Tang et al., 2016). These parameters have similar spatiotemporal
177 properties to CF, i.e., they vary rapidly in both time and space. Therefore, BME has the potential for use
178 in merging multiple satellite CF products to produce spatiotemporally complete, accurate, and coherent
179 Arctic CF products.

180 In this paper, we present a spatiotemporal data fusion framework based on a CDF matching
181 approach and BME methodology to generate a fused monthly daytime CF product with $1^\circ \times 1^\circ$ resolution
182 in the Arctic region from 2000 to 2020. The CDF matching approach is used to correct the bias of passive
183 sensor data based on active sensor data, thereby improving the quality of the passive data. The BME
184 method is used to produce spatiotemporally complete monthly CF data from corrected multiple-satellite
185 CF products. The uncertainties of passive sensor CF products benchmarked with active sensor data and
186 ground-based data are all considered in the fusing process. The study area was in the Arctic region above
187 $60^\circ N$, including land and marine areas. The structure of this paper is as follows. Section 2 describes the
188 data, while Section 3 introduces the data preprocessing and methods. The results and discussion are
189 presented in Sections 4 and 5, respectively. Finally, the conclusions are provided in Section 6.

190 **2 Data**

191 **2.1 Satellite Data**

192 In view of the complementarity among the AVHRR-based, MODIS-based, and active sensor
193 products, this study involved ten passive-satellite-derived products from MODIS and AVHRR, with the
194 time period spanning from 2000 to 2020 along with an active-satellite-derived product from CALIPSO,
195 with the time period spanning from 2006 to 2016. The experimental period only included the sunlit
196 months from April to September because of the darkness of the Arctic winter. All the data are briefly
197 described in Table 1. Our study aimed to provide accurate and reliable measurements of cloud fraction
198 during the daytime in the Arctic region. To achieve this objective, we utilized cloud fraction data labeled
199 as "daytime" from multiple satellite datasets.

200 The AVHRR sensors are onboard sun-synchronous orbit satellites collecting data in the morning or
201 afternoon (NOAA, Metop-A/B). The morning (afternoon) orbits cross the equator on their descending
202 (ascending) node at approximately 0730 (1330) local time (LT). Starting with NOAA-17 and all MetOp
203 satellites, AVHRR data are available from a midmorning orbit with the equator crossing time at
204 approximately 0930 LT. However, complications arose from changes in the equatorial crossing times of
205 individual AVHRR sensors due to satellite drift (Heidinger et al., 2014; Karlsson et al., 2013). The
206 AVHRR has a nominal spatial resolution of 1.1 km at the nadir point, facilitating full global coverage
207 twice daily (daytime and nighttime), but the products this study employed provide global area coverage
208 data with a nadir footprint size of $1.1 \text{ km} \times 4.4 \text{ km}$ (Stengel et al., 2017). Cloud detection algorithms of
209 these latest satellite data have improved greatly in polar regions. However, some data gaps exist as a
210 result of AVHRR scan motor errors (e.g., the NOAA-15 orbits were blacklisted in 2000 and 2001) and

211 limitations of observation conditions (e.g., CLARA-A2 could not cover the central Arctic Sea in
212 September).

213 The MODIS sensor is onboard both the morning satellite Terra and the afternoon satellite Aqua,
214 with overpass times at the equator of approximately 1030 LT and 1330 LT, respectively. The MODIS
215 produces complete near-global coverage in less than 2 days. The 36 channels from the visible to thermal
216 infrared spectrum provide abundant spectral information for cloud parameter retrieval. The new version
217 datasets have improved the cloud detection algorithms in polar regions, whereas some researchers found
218 overestimated CF in snow/ice surface in the new datasets when compared with active sensor data
219 (Marchant et al., 2020; Marchant et al., 2016; Paul, 2017; Trepte et al., 2019). Although some differences
220 exist between Terra and Aqua, the consistency between these two satellites cannot be ignored (Trepte et
221 al., 2019).

222 The CALIPSO satellite combines an active light detection and ranging (lidar) instrument (Cloud-
223 Aerosol Lidar with Orthogonal Polarization - CALIOP Lidar) with passive infrared (Imaging Infrared
224 Radiometer) and visible imagers (Wide Field Camera) to probe the vertical structure and properties of
225 thin clouds and aerosols worldwide (Winker et al., 2007; Vaughan et al., 2004; Hunt et al., 2009; Vaughan
226 et al., 2009; Winker et al., 2009). As the most accurate currently active space-borne instrument for
227 detecting clouds, CALIPSO has a 16-day repeat cycle with equatorial overpass time at 1:30 PM. The
228 CAL_LID_L3_GEWEX_Cloud-Standard-V1-00 is a widely used grid cloud product with a spatial
229 resolution of an equal angle grid $1^\circ \times 1^\circ$ (Claudia, 2012).

230 Table 1. Satellite cloud fraction products used in this research.

Products	Cloud detect method	Satellite	Sensor	Overpass time	Time range	Temporal resolution	Spatial resolution
MOD08-M3 Terra	MOD 35	Terra	MODIS	1030am	2000.2-2020.12	daily	$1^\circ \times 1^\circ$
MYD08-M3 Aqua	MYD 35	Aqua	MODIS	1330pm	2002.7-2020.12	daily	$1^\circ \times 1^\circ$
CERES-SSF Terra	CERES Edition 4	Terra	MODIS	1030am	2000.3-2020.12	daily	$1^\circ \times 1^\circ$
CERES-SSF Aqua	CERES Edition 4	Aqua	MODIS	1330pm	2002.7-2020.12	daily	$1^\circ \times 1^\circ$
CLARA-A2 AM	EUMETSAT NWC SAF PPS	NOAA-15	AVHRR3	0730am	2000.1-2000.7	daily	$0.25^\circ \times 0.25^\circ$
		NOAA-17	AVHRR3	0930am	2001.3-2002.7		
		METOPA	AVHRR3	0930am	2002.8-2007.6		
		NOAA-14	AVHRR2	1330pm	2007.7-2019.6		
CLARA-A2 PM	EUMETSAT NWC SAF PPS	NOAA-16	AVHRR3	1400pm	2000.1-2000.12	daily	$0.25^\circ \times 0.25^\circ$
		NOAA-18	AVHRR3	1330pm	2001.1-2003.5		
		NOAA-19	AVHRR3	1330pm	2003.6-2005.7		
		NOAA-18	AVHRR3	1330pm	2005.8-2009.5		
PATMOS-x AM	Naive Bayesian	NOAA-15	AVHRR3	0730am	2000.1-2000.7	daily	$0.1^\circ \times 0.1^\circ$
		NOAA-17	AVHRR3	0930am	2001.3-2002.8		
		METOPA	AVHRR3	0930am	2002.9-2007.6		
		NOAA-14	AVHRR2	1330pm	2007.7-2020.12		
PATMOS-x PM	Naive Bayesian	NOAA-16	AVHRR3	1400pm	2000.1-2001.3	daily	$0.1^\circ \times 0.1^\circ$
		NOAA-18	AVHRR3	1330pm	2001.4-2003.5		
		NOAA-19	AVHRR3	1330pm	2003.6-2005.7		
		NOAA-19	AVHRR3	1330pm	2005.8-2009.5		
					2009.6-2020.12		

ISCCP-H AM	IR and VIS threshold	NOAA-14- NOAA-19; METOPA	AVHRR2 / AVHRR3	9000am <hr/> 1500pm	2000.1-2017.6	daily	1°×1°
ISCCP-H PM							
CALIPSO- GEWEX	5km merged layer product level 2	CALIPSO	CALIOP	1330pm	2006.6-2016.12	Monthly	1°×1°

231 **2.2 Ground Observation Data**

232 **2.2.1 Climatic Research Unit Gridded Time Series**

233 The Climatic Research Unit gridded Time Series (CRU TS) is a widely used climate dataset
234 covering all land surfaces except Antarctica, which uses angular distance weighting to interpolate
235 monthly climate anomalies from extensive networks of weather station observations onto a 0.5° grid
236 (Harris et al., 2020; Harris et al., 2014). This dataset was first published in 2000, and the latest version,
237 CRU TS4.05, contains ten variables including cloud cover for the period 1901–2020 (Harris et al., 2020).
238 The percentage of cloud cover was derived from observations of sunlit hours, and CRU TS4.05 output
239 files are actual values, not anomalies.

240 **2.2.2 International Comprehensive Ocean-Atmosphere Data Set**

241 The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is the most extensive
242 freely available archive of global surface marine data, which has been assimilated into all major
243 atmospheric, oceanic, and coupled reanalysis (Freeman et al., 2017). The ICOADS report is derived
244 from synthetical observations of ships, buoys, coastal platforms, or oceanographic instruments. This
245 dataset offers a gridded monthly summary for 2° latitude × 2° longitude boxes dating back to 1800 (and
246 1°×1° boxes since 1960) (Woodruff et al., 2005). The available climatic variables include cloud cover
247 and other atmospheric parameters (Bojinski et al., 2014). In this study, we used the 1°×1° cloud cover
248 data in sunlight months (April to September) spanning 2000 to 2020. In particular, we obtained the
249 "fraction of observations in daylight" data from the ICOADS dataset, which allowed us to select only
250 the data points corresponding to daytime observations. During our analysis, we imposed a threshold of
251 0.8 for the fraction of observations in daylight, ensuring that we only included the data with high
252 confidence in our study.

253 **2.3 Reanalysis Data and Model Data**

254 In recent decades, atmospheric reanalysis datasets have emerged as a valuable resource for studying
255 climate processes and predictability, offering a long-term, gridded depiction of atmospheric conditions.
256 These datasets rely on state-of-the-art data assimilation systems, which integrate observational data and
257 underlying models to create a continuous record of historical weather patterns. Through the use of
258 various atmospheric variables, they provide insight into past weather phenomena. The utilization of
259 these datasets could prove imperative in conducting research within areas that are limited in data
260 availability, such as the Arctic. Several studies have investigated the performance of reanalyses over
261 the Arctic for a variety of fields including CF (Yeo et al., 2022; Kennedy et al., 2012; Huang et al.,
262 2017). However, the systematic errors of climatological reanalysis CF are substantial for Arctic clouds
263 because of the complexity of cloud microphysical processes and lack of good observation. In-depth
264 comparisons, as conducted by Walsh, have identified difficulties in adequately depicting persistent low-

265 level CF in summer via reanalysis models (Walsh et al., 2009).

266 ERA5 is an advanced atmospheric reanalysis product developed by the European Centre for
267 Medium-Range Weather Forecasts (ECMWF). It provides information on cloud properties, including
268 cloud fraction, cloud ice, cloud liquid, rain, and snow water content, which are estimated using the
269 prognostic equations developed by Tiedtke in 1993(Tiedtke, 1993). This method accounts for physical
270 processes that act as sources or sinks of clouds, such as convection and condensation. In addition, the
271 outdated diagnostic temperature-dependent approach for phase partitioning in mixed-phase clouds has
272 been replaced with a more sophisticated, prognostic method developed by Forbes and Ahlgrimm in
273 2014 (Forbes and Ahlgrimm, 2014). The updated radiation scheme in ERA5 employs the Monte Carlo
274 independent column approximation with generalized overlap for sub-grid cloud representation,
275 enhancing the accuracy of the product.

276 This study uses the CF of ‘ERA5 hourly data on single levels from 1959 to present,’ and the CF
277 parameter has been regridded to a regular lat-long grid of 0.25° and calculated by making assumptions
278 about the degree of overlap/randomness between clouds at different heights.

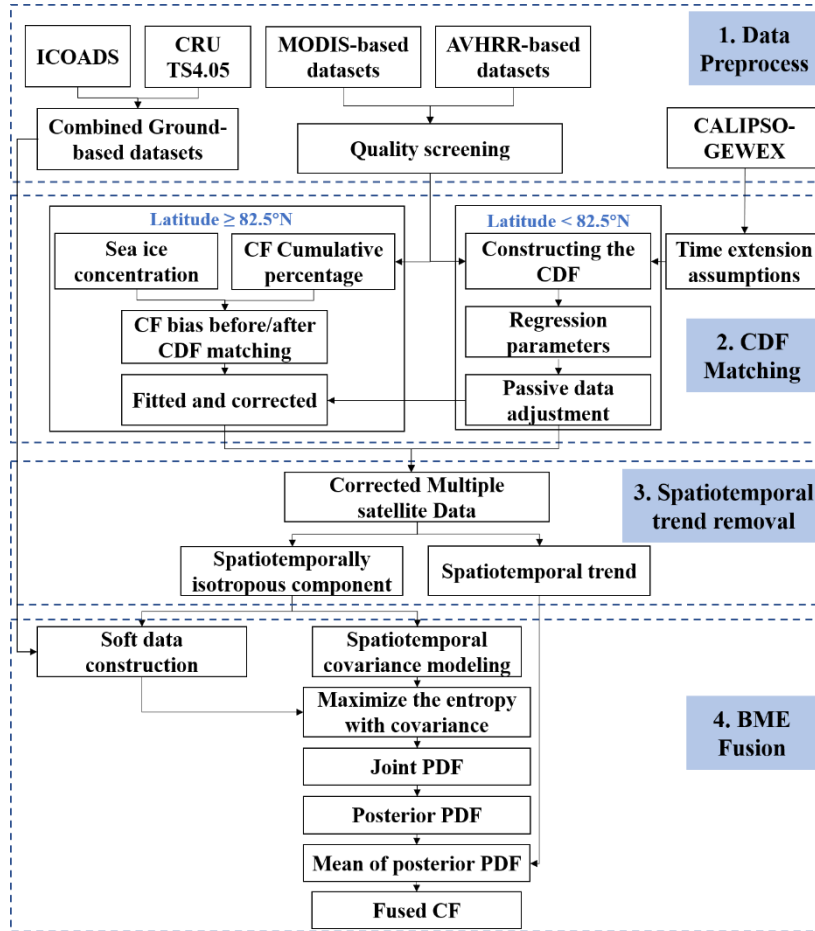
279 The climate model is also a valuable tool for climate studying. However, comparisons of climate
280 models to Arctic observations over the past three decades have revealed persistent challenges
281 simulating Arctic climate that partially attribute to imprecise cloud fraction (English et al. 2014). The
282 sixth phase Coupled Model Intercomparison Project (CMIP6) models have been used in many research
283 papers about climate. Among them the simulation data of the MRI-AGCM3-2-S climate model
284 provides a basis for climate research designed to answer fundamental scientific questions and serves as
285 a resource for authors of the Sixth Assessment Report of the Intergovernmental Panel on Climate
286 Change (IPCC-AR6). The model employed in this study is derived from the operational weather
287 prediction model of the Japan Meteorological Agency (JMA). It integrates quasiconservative semi-
288 Lagrangian dynamics, a radiation scheme, and a land surface scheme that was initially designed for a
289 climate model. Utilizing observed sea surface temperature (SST) as well as SST alterations forecasted
290 by atmosphere-ocean coupled models, we carried out simulations of both present-day and future
291 climate conditions. This model was released in 2017 and provided CF parameters at native nominal
292 resolutions of 25 km. This resolution employed in the model is as fine as that employed by regional
293 climate models (RCMs) in recent studies. Smallscale phenomena are realistically simulated in the high-
294 resolution model, with keeping the same quality of global-scale climate representation as the lower-
295 resolution models.

296 The study involved a comparison of pre- and post-fusion CF data with reanalysis and model data.
297 The aim was to underscore the significant role of fused data in improving the consistency of CF
298 between satellite observations, reanalysis data, and model data.

299 **3 Data Preprocessing and Methodology**

300 In this study, we propose a fusion algorithm framework that combines data from multiple satellites
301 to provide CF datasets with high spatiotemporal coverage and improved accuracy. Figure 1 shows a
302 flowchart of the general process, which includes four parts. First, the original data were preprocessed
303 before data fusion, a process that included data quality control and data resampling. Second, bias
304 correction of passive sensor data was conducted using active data with the CDF matching method. Third,
305 to comply with BME's stationarity prerequisite that assumes constancy of mean and variance, we
306 removed the spatiotemporal trend of the original satellite CF data over the study area using the

307 spatiotemporal moving window filter method. Fourth, the spatiotemporal covariance function was
 308 modeled based on the isotropic residual data, and then the entropy was maximized with covariance
 309 constraint. All the satellite-based CF data were treated as soft data so that the associated uncertainties
 310 were incorporated into the fusing process.



311
 312 Figure 1. Flowchart for merging the multiple satellite cloud fraction products based on cumulative
 313 distribution function matching and the Bayesian maximum entropy method.
 314

315 3.1 Data Preprocessing

316 Over the Arctic, the cloud detection capabilities of passive sensors are always limited by spectral
 317 channels, while active sensors are not susceptible to these effects (Liu et al., 2010; Liu et al., 2012b;
 318 Kotarba, 2020; Shupe et al., 2013). To obtain more accurate fused CF results, it is necessary to correct
 319 these passive sensor products using active sensor data before merging.

320 For satellite datasets, statistics always have the Scientific Data Set (SDS) name suffix
 321 “_Standard_Deviation” and which are computed by calculating an unweighted standard deviation of all
 322 pixels or samples within a given 1° grid cell. The large CF standard deviations (STDs) of satellite datasets
 323 represent the large uncertainties of CF detection (Ackerman et al., 2008; Stengel et al., 2017). In this
 324 study, we calculated the relationship between differences in STDs and CFs of passive/active sensor
 325 datasets and found that the larger the standard deviation, the more serious the underestimation of passive
 326 sensors. For the products with standard deviation flags, including MOD08 Terra/Aqua, CLARA-A2
 327 AM/PM, and the PATMOS-x AM/PM, we used the 90 percentile of the daily standard deviation as scene-

328 based dynamic thresholds to screen CF data.

329 However, no standard deviation information was available for CERES-SSF Terra/Aqua and the
330 ISCCP-H AM/PM datasets. Based on research that shows ignoring optically very thin clouds could
331 increase the agreement between passive sensor data and the CALIPSO data, the 0.15 COT dataset was
332 selected as the quality threshold in this study.

333 3.2 CDF Matching

334 A widely used scaling strategy known as CDF matching can be used to adjust the distribution of the
335 target dataset to the range of reference data under the constant relative relationship. Several studies have
336 proved that the process of adjusting this distribution does not change the variation of original satellite-
337 based products, but rather aligns the value range with that of the reference data (Liu et al., 2011b; Brocca
338 et al., 2011; Xu and Cheng, 2021). Based on similar seasonal fluctuations of the passive sensor CFs and
339 active sensor data, the time series of passive sensor data from each grid box in the Arctic region were
340 adjusted to the values of the paired CALIPSO-GEWEX latitude and longitude grid. However, the
341 CALIPSO-GEWEX data could not cover regions with a latitude greater than 82.5°N and the temporal
342 range only covers 2006–2016. To correct the CF bias over the entire Arctic region, two strategies were
343 considered.

344 First, for the regions with enough reference data, the CF data of all passive sensors were directly
345 adjusted by CDF matching. The matching approach includes three steps: (1) constructing the cumulative
346 distribution function, (2) deriving regression parameters, and (3) adjusting the original data with
347 regression parameters. In our study, we use a three-month moving mean to eliminate the uncertainties in
348 CALIPSO-GEWEX data caused by the limitation of sampling quantities and frequencies. The filtered
349 daily passive sensor datasets were resampled as monthly mean data, and then the CDFs were constructed
350 for every dataset based on the same method used for the active data. A least-square fit was used to derive
351 the relationship between the reference and the target datasets. Based on the analysis of Liu et al.(2022),
352 the seasonal variation of CF for multiple satellites was greater than the interannual changes in CF (Liu
353 et al., 2022). We propose an additional assumption that the CDF ratio between active and passive sensor
354 data remains constant over the years in a 1°×1° grid cell.

355 Second, it was difficult to implement the CDF matching strategy for areas beyond the coverage of
356 active sensor data. Considering the relationship among the CF bias before and after CDF correction, the
357 cumulative percentage of CF (CPCF, the average CF over an interval of SIC), and the sea ice
358 concentration (SIC), a fitting function is proposed to correct the CF data.

359 After executing the abovementioned steps, we obtained the corrected multiple satellite data.

360 3.3 Spatiotemporal Trend Analysis and Removal

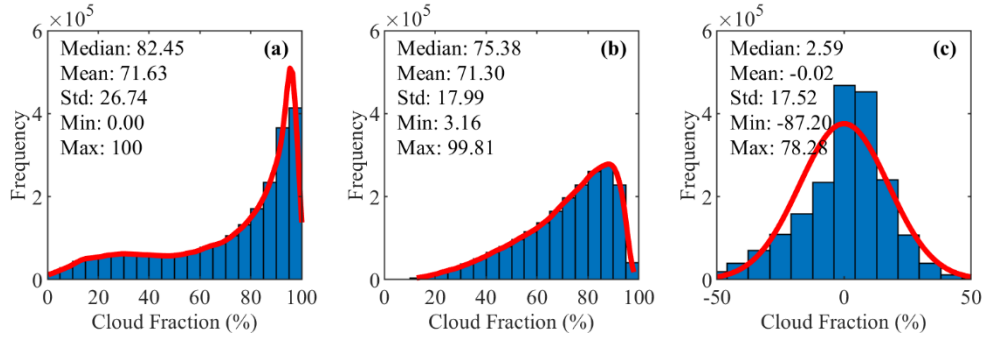
361 The BME theory was constructed based on the hypothesis of spatiotemporal random field (S/TRF)
362 (Nazelle et al., 2010; Christakos, 2000; He and Kolovos, 2017), which means that all the variables used
363 for this process are homogeneous and isotropous. However, a natural process that evolves in space–time,
364 such as the distribution of CF, can be divided into a heterogenetic global spatiotemporal trend and a
365 spatiotemporally isotropous residual, following Eq. (1):

$$366 \quad CF_{(s,t)} = \overline{CF}_{(s,t)} + CF_{res(s,t)}, \quad (1)$$

367 where (s, t) represents the space and time, $\overline{CF}_{(s,t)}$ represents the global spatiotemporal trend, and $CF_{res(s,t)}$

368 represents the stochastic anomalies of the variable. To meet the second-order stationarity assumption
 369 (constant mean and variance), it is necessary to remove the global spatiotemporal trend before estimating
 370 the spatiotemporally autocorrelated structure of the data (Spadavecchia and Williams, 2009; Tang et al.,
 371 2016). In this study, the global spatiotemporal trend was calculated using a spatiotemporal filter window
 372 with a size of 5° (longitude) \times 5° (latitude) \times 3 (months).

373 Figure 2 shows a histogram of the original combined satellite CF data, the global spatiotemporal
 374 trend, and the residual spatiotemporally isotropous component. From these distributions of the histogram,
 375 the residual is approximately normally distributed, which meets the requirement for modeling the
 376 structure of the spatiotemporal autocovariance.



377
 378 Figure 2. Histograms of (a) original combined satellite cloud fraction, (b) global spatiotemporal trend, and (c)
 379 spatiotemporally isotropous component, for the entire Arctic area (Example using 2010 data).

380

381 3.4 BME Fusion

382 3.4.1 Spatiotemporal Covariance Modelling

383 In spatiotemporal geostatistics, a covariance function indicates the spatial and temporal dependency
 384 of the data, which decreases as distance/time increases (Griffith, 1993). The spatiotemporal variation of
 385 the CF also can be expressed by a spatiotemporal covariance function. In the BME method, the
 386 experimental covariance can be calculated from the point pairs at specific distances and then modeled by
 387 the commonly used covariance model (Cressie, 2015). This study uses a nested covariance model with
 388 two spatiotemporal exponential models to model the spatiotemporal covariance of the detrended
 389 combined CF data, following Eq. (2):

$$390 \quad \text{cov}(d, \tau) = c_1 \exp\left(-\frac{3d}{a_{s1}}\right) \exp\left(-\frac{3\tau}{a_{t1}}\right) + c_2 \exp\left(-\frac{3d}{a_{s2}}\right) \exp\left(-\frac{3\tau}{a_{t2}}\right), \quad (2)$$

391 where d is the spatial lag and τ is the temporal lag between point pairs at coordinates $p(s, t)$ and
 392 coordinates $p'(s', t')$; c_1 and c_2 are the partial sill variances of the two exponential models; a_{s1} and a_{s2}
 393 are the spatial ranges of the two exponential models; a_{t1} and a_{t2} are the temporal ranges of the two exponential
 394 models. When the S/TRF is characterized by spatial and temporal stationarity, it is only the relative
 395 distance between any couple of locations that affects the covariance function. Specifically then, the
 396 covariance function has the same value $\text{cx}(p, p') = \text{cx}(r, t)$ for any location pair (p, p') separated by the
 397 same spatial distance vector $r = s' - s$ and same temporal distance lag $\tau = t' - t$ (Christakos and Serre 2000).
 398 In this study, the parameters for spatiotemporal covariance are modeled separately for each year. The
 399 modelled results shown that the model has a spatial range of 2° , a temporal range of 3 months, and a
 400 partial sill variance of 0.85 for local scale CF (the first nested covariance model). And for the large range

401 CF the model has a spatial range of 30°, a temporal range of 6 months, and a partial sill variance of 0.15
 402 (the second nested covariance model).
 403

404 3.4.2 Construction of Soft Data

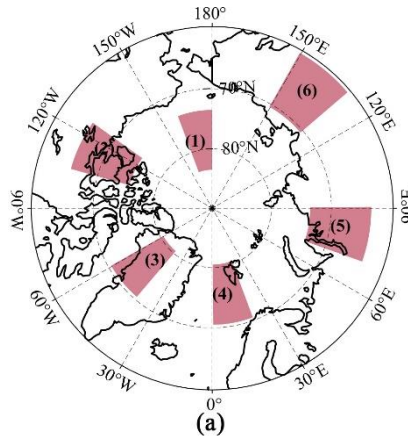
405 BME treated the informative content with uncertainty from different sources as soft data (He and
 406 Kolovos, 2017). For example, the observed data that accompanied by obvious sources of uncertainty
 407 such as inaccuracy in measuring devices, modeling uncertainties, and human error. In this study, the CF
 408 data of passive sensor products are viewed as soft data. For the BME method, a key conceptual aspect is
 409 that the framework does not impose any restrictive assumptions about the PDFs of soft data. Hence, a
 410 parameterized statistical distribution of different sources of information can be used to replace the real
 411 PDFs (Nazelle et al., 2010). Soft data could be probabilistic or interval soft data (Christakos, 2000). In
 412 this study, the differences between satellite data and ground observations followed normal distributions
 413 approximately. Therefore, the passive sensor data used for fusion were all treated as soft data with a
 414 Gaussian distribution, following Eq. (3):

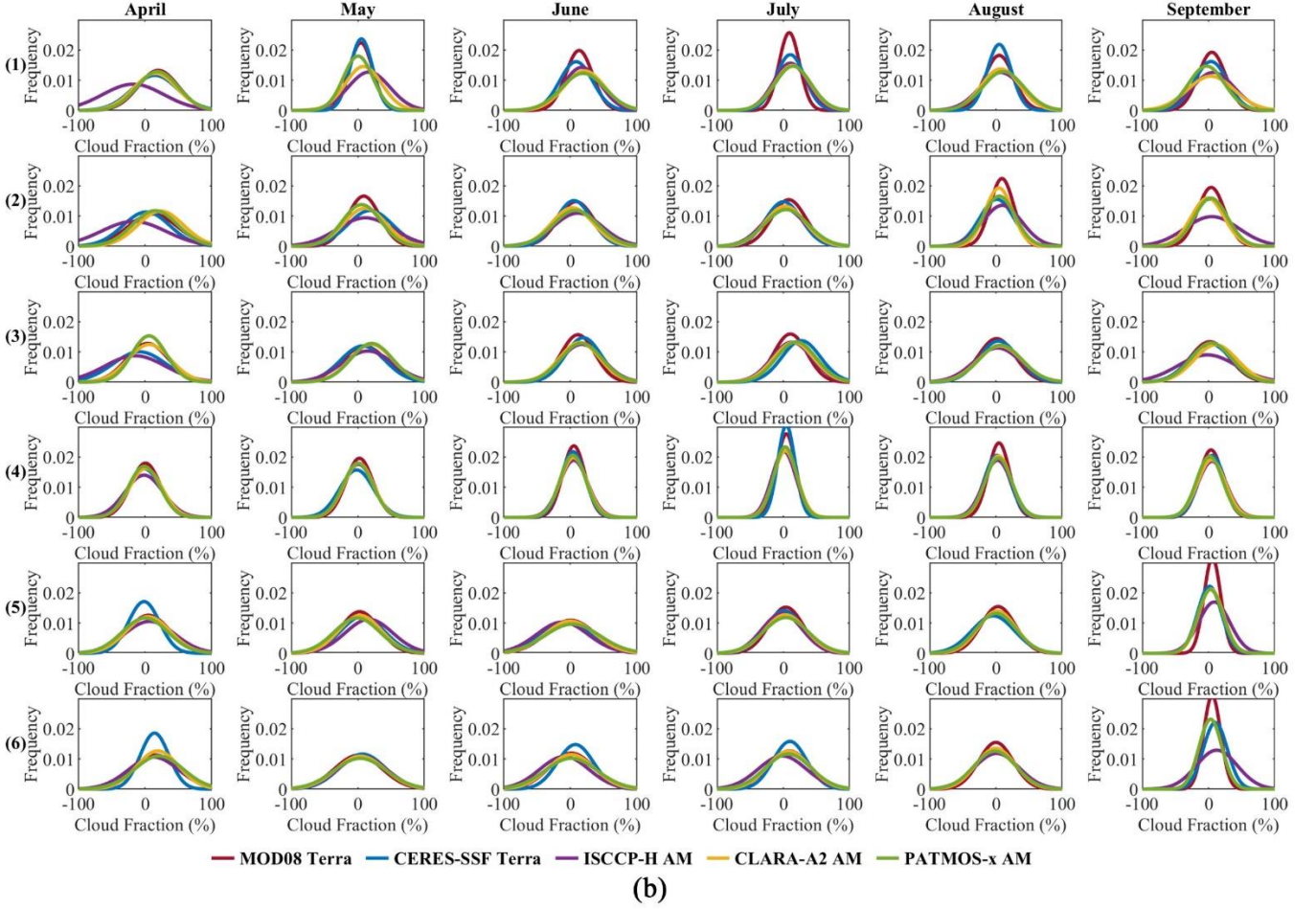
$$415 \quad CF_{sate,x} = CF_{ground,x} + \varepsilon_x, \quad (3)$$

416 where $CF_{sate,x}$ and $CF_{ground,x}$ are the satellite CF data and the corresponding ground observation,
 417 respectively, and ε_x is an independent random error, following Eq. (4):

$$418 \quad \varepsilon \sim N(\mu_\varepsilon, \sigma_\varepsilon^2), \quad (4)$$

419 where μ_ε represents the mean of random error and σ_ε^2 represents the variance (Tang et al., 2016).





420 Figure 3. Gaussian probability density functions of the random errors between each type of satellite data and
 421 ground observations at six randomly selected regions of interest from April to September.

422

423 Because the uncertainties in each satellite CF data vary at different spatial and temporal scales, using
 424 the average uncertainty of the entire dataset to construct soft data over the entire study area will
 425 undoubtedly neglect the spatiotemporal variation of uncertainties. In this study, six regions were
 426 randomly selected to analyze the probability density functions (PDF) of random errors (Fig. 3). Large
 427 inconsistencies were observed for the PDF in land and ocean regions, and the temporal variation was
 428 also an important factor in inconsistencies. We constructed the soft data for CF data over land and ocean
 429 regions in every month separately. Considering the large errors in the Greenland Ice Sheet (GrIS), we
 430 calculated the PDF of random error separately for that region.

431 For each grid box, the CFs of different satellite data were converted into a Gaussian distribution
 432 probability soft data, individually (Tang et al., 2016). The soft data were expressed as:

433
$$CF_{soft,sate} \sim N(CF_{sate} + \mu_{\epsilon}, \sigma_{\epsilon}^2), \quad (5)$$

434 where CF_{sate} is the detrended CF value of multiple satellite datasets; the mean and variance of the
 435 Gaussian distribution probability soft data were expressed by $CF_{sate} + \mu_{\epsilon}$ and σ_{ϵ}^2 , respectively.

436 3.4.3 Using the BME Method for Multiple CF Data Fusion

437 The BME method can be used to merge continuous variables of satellite data for some atmospheric

438 parameters. To simplify the heterogeneity and anisotropic variability, the residuals were considered only
 439 in the fusion process. Assuming that various adjacent observations from satellites were available with
 440 irregular spatial and temporal gaps, the nonlinear mean estimation \bar{x}_k of CF at the location (s_x, s_y) at
 441 time t was estimated as:

$$442 \quad \bar{x}_k = \int x_k f(x_k | x_{soft,1}, x_{soft,2} \dots x_{soft,n}) dx_k, \quad (6)$$

443 where $f(x_k | x_{soft,1}, x_{soft,2} \dots x_{soft,n})$ is a posterior PDF over the spatiotemporal adjacent grid observations, and
 444 $x_{soft,1}, x_{soft,2} \dots x_{soft,n}$ are the probabilistic Gaussian soft data derived from multiple satellite data. The
 445 posterior PDF at the estimation point updates from the prior PDF in the Bayesian rule when soft data are
 446 involved, so the relationship can be expressed as:

$$447 \quad f(x_k | x_{soft,1}, x_{soft,2} \dots x_{soft,n}) = \frac{f(x_{soft,1}, x_{soft,2} \dots x_{soft,n}, x_k)}{f(x_{soft,1}, x_{soft,2} \dots x_{soft,n})}, \quad (7)$$

448 where $f(x_{soft,1}, x_{soft,2} \dots x_{soft,n})$ represents the prior PDF of the spatiotemporally isotropous CF at the adjacent
 449 grid, $f(x_{soft,1}, x_{soft,2} \dots x_{soft,n}, x_k)$ is the joint PDF without specific information. Generally, the joint PDF is
 450 represented by $f_g(x_{map})$, which can be calculated by maximizing the entropy under the constraint of the
 451 general knowledge g (Jaynes, 1957). When predicting the probability distribution of a random event, the
 452 larger the information entropy, the larger the amount of information obtained, and the result is closer to
 453 the actual situation under a most uniform probability distribution. In this study, general knowledge is the
 454 spatiotemporal covariance model, and to maximize the entropy, we introduce a Lagrange multiplier λ
 455 (Xia et al., 2022).

$$456 \quad f_g(x_{map}) = \frac{\exp\left(\sum_{\alpha=1}^n \lambda_{\alpha} g_{\alpha}(x_{map})\right)}{\int \exp\left(\sum_{\alpha=1}^n \lambda_{\alpha} g_{\alpha}(x_{map})\right) dx_{map}}, \quad (8)$$

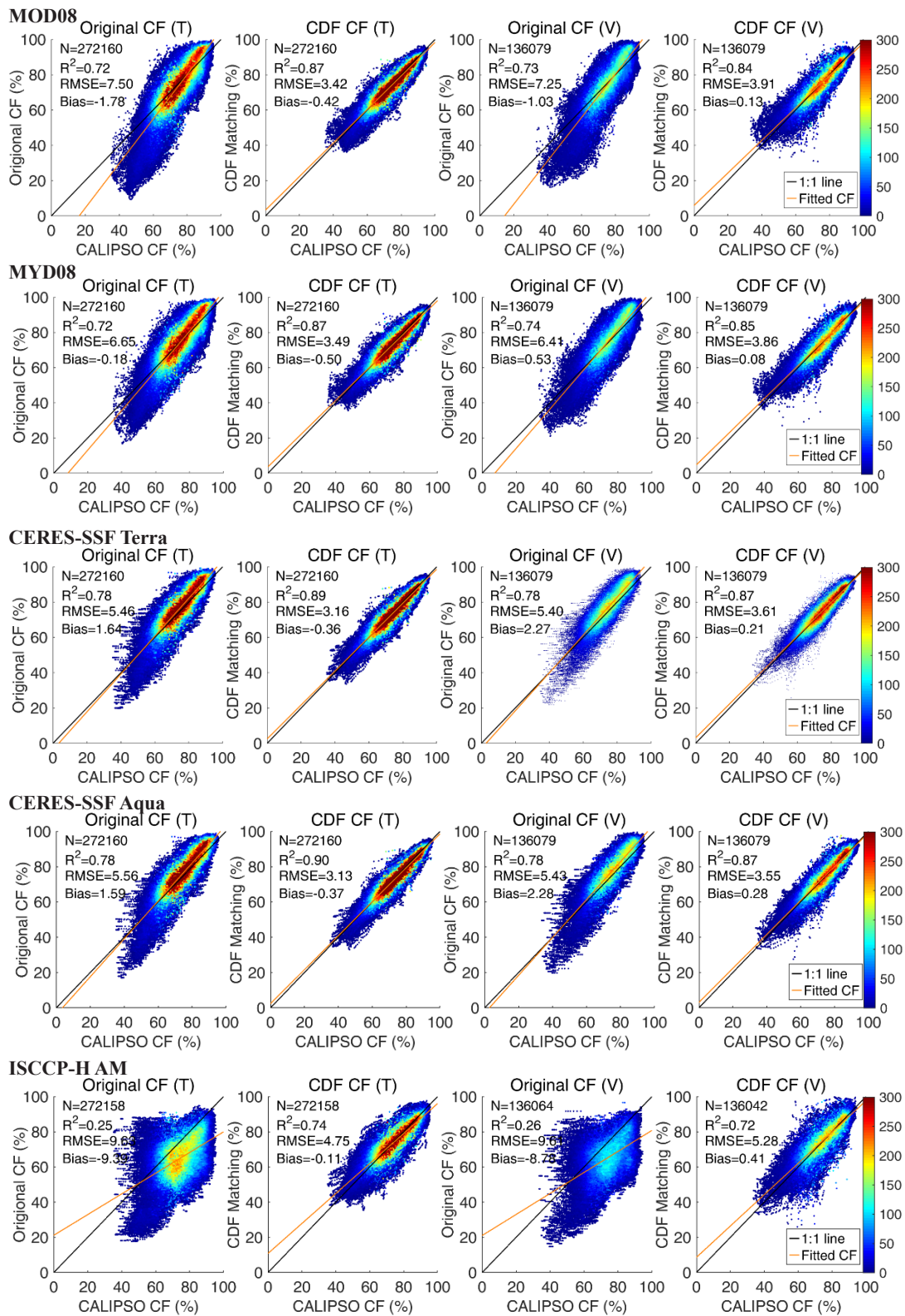
457 Finally, the expectation of spatiotemporally CF isotropous component can be calculated by solving
 458 these equations. Then the anisotropic spatiotemporal trend component of each grid was added to the
 459 expectation at the corresponding point to obtain the merged CF product.

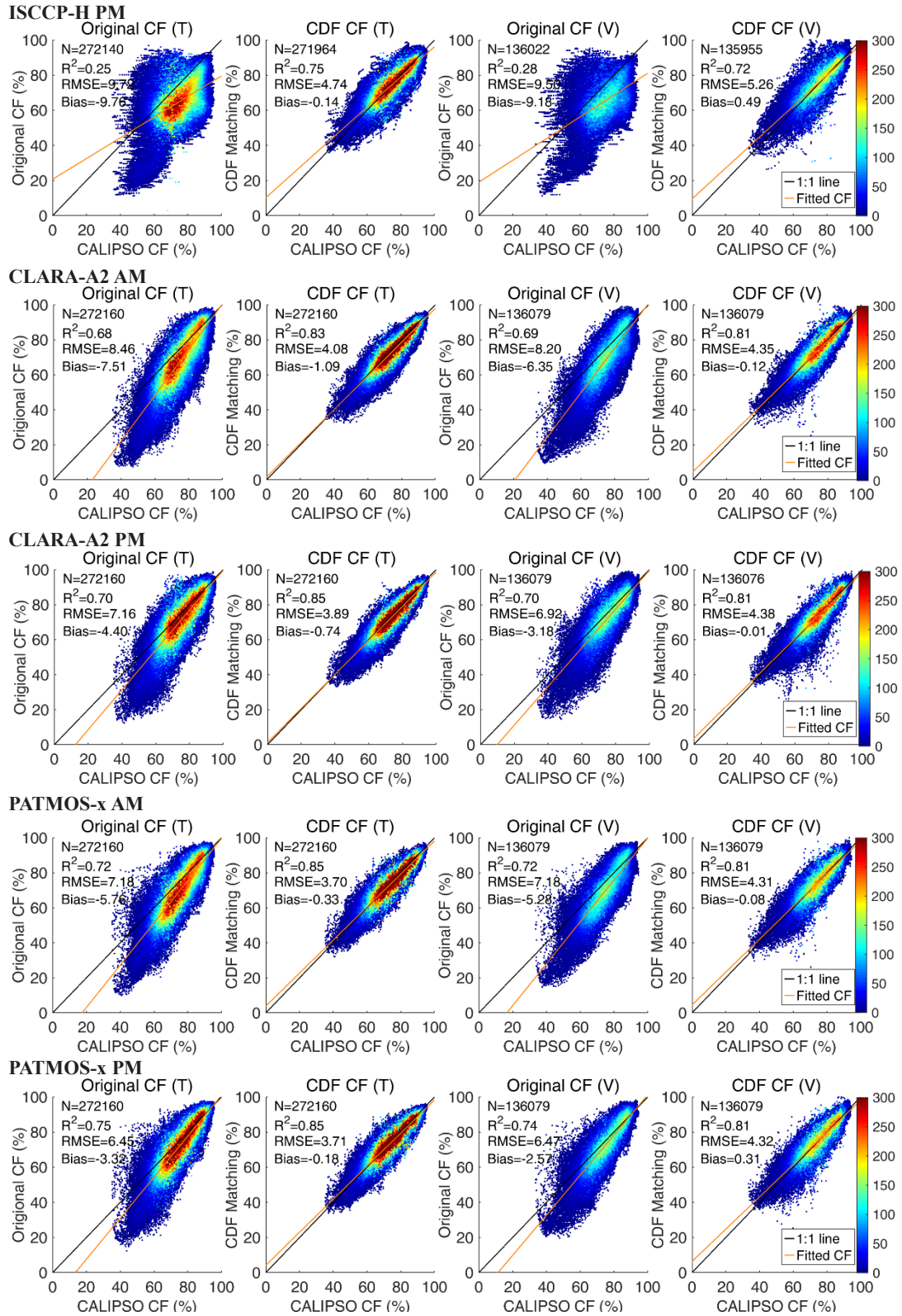
460 **4 Results**

461 **4.1 Result of CDF Matching**

462 Figure 4 shows the scatter plots of the CF distribution before and after CDF matching from multiple
 463 passive and active sensors at the valid grid boxes with a latitude of less than 82.5°N. Based on the fact
 464 that the assumption that the correction coefficient does not vary over time, the training datasets (T) were
 465 processed from 2008 to 2014 and the validation datasets (V) were processed in 2006, 2007, 2015, and
 466 2016. In Fig. 4, the ‘Original CF (T)’ and ‘Original CF (V)’ indicate the comparison of CALIPSO-
 467 GEWEX CF and that of the original passive sensor data, so that the ‘CDF CF (T)’ and ‘CDF CF (V)’
 468 represent the comparison between CALIPSO-GEWEX CF and the corrected CF. In general, for all the
 469 passive sensor datasets, the CFs after CDF matching were closer to the 1:1 line than before CDF matching.
 470 R^2 increased by about 0.07–0.15, while that for ISCCP-H products was over 0.45. The RMSEs decreased
 471 to one-third to one-half of what they were, and the biases decreased to approximately zero, which means
 472 that the CDF matching obviously corrected outliers and eliminated the average differences between the

473 passive and active sensor CFs. From these scatter plots, we also understand that CDF matching plays an
 474 important role in low CFs (less than 60%), which was always seen in April or on the GrIS(Liu et al.,
 475 2022).





476

477

478

479

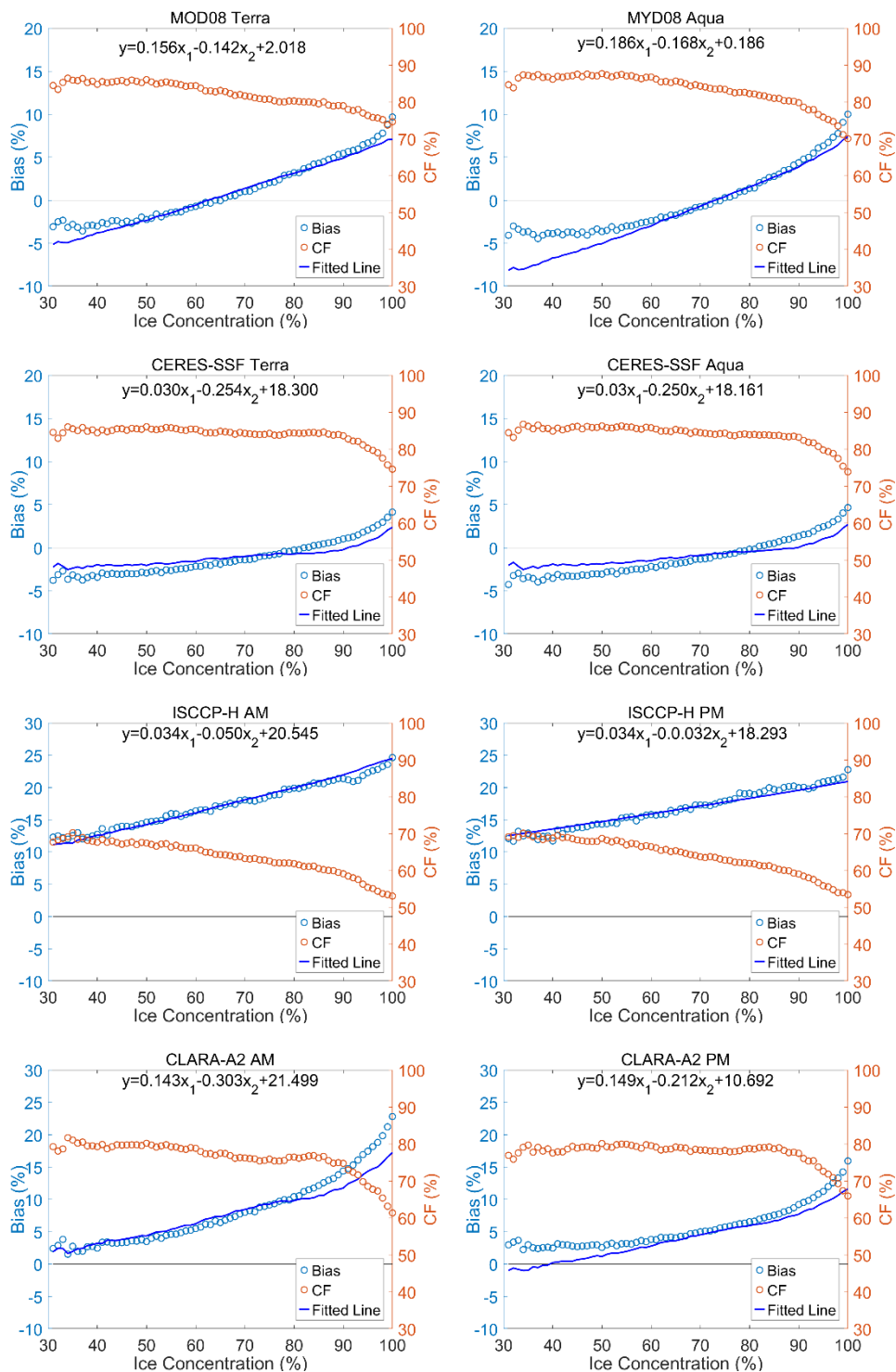
480

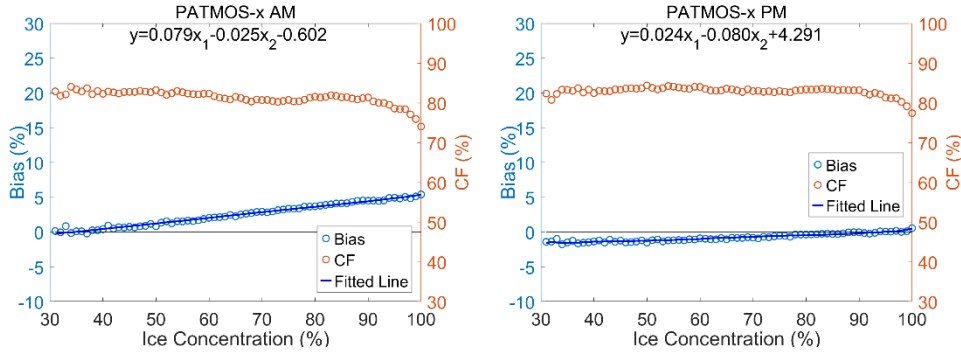
481

Figure 4. The scatter plots of the cloud fraction comparison between the passive and active sensor datasets at regions with latitudes less than 82.5°N before and after cumulative distribution function matching: (T) means training data with time ranges from 2008 to 2014 and (V) means validation data from 2006, 2007, 2015, and 2016.

In the sea ice regions, the relationships between CF bias of passive sensor data after and before CDF

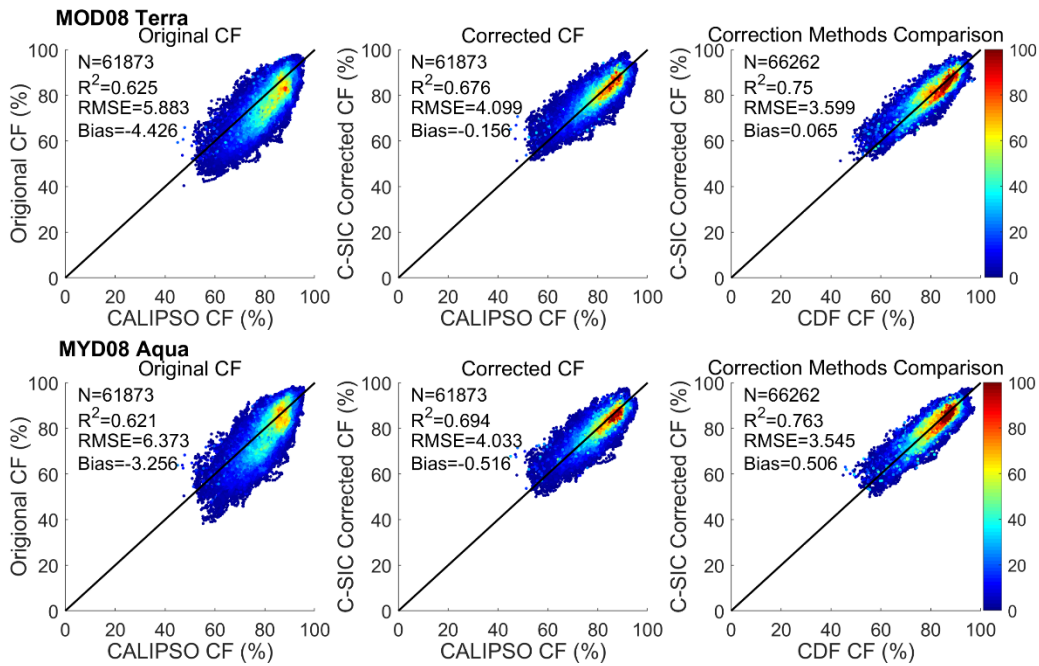
482 matching, CPCF, and SIC are shown in Fig. 5. The results indicated that the mean of bias increased with
 483 the SIC. Moreover, the CPCF appeared to decrease with increasing SIC, a negative correlation between
 484 CPCF and bias was also evident.

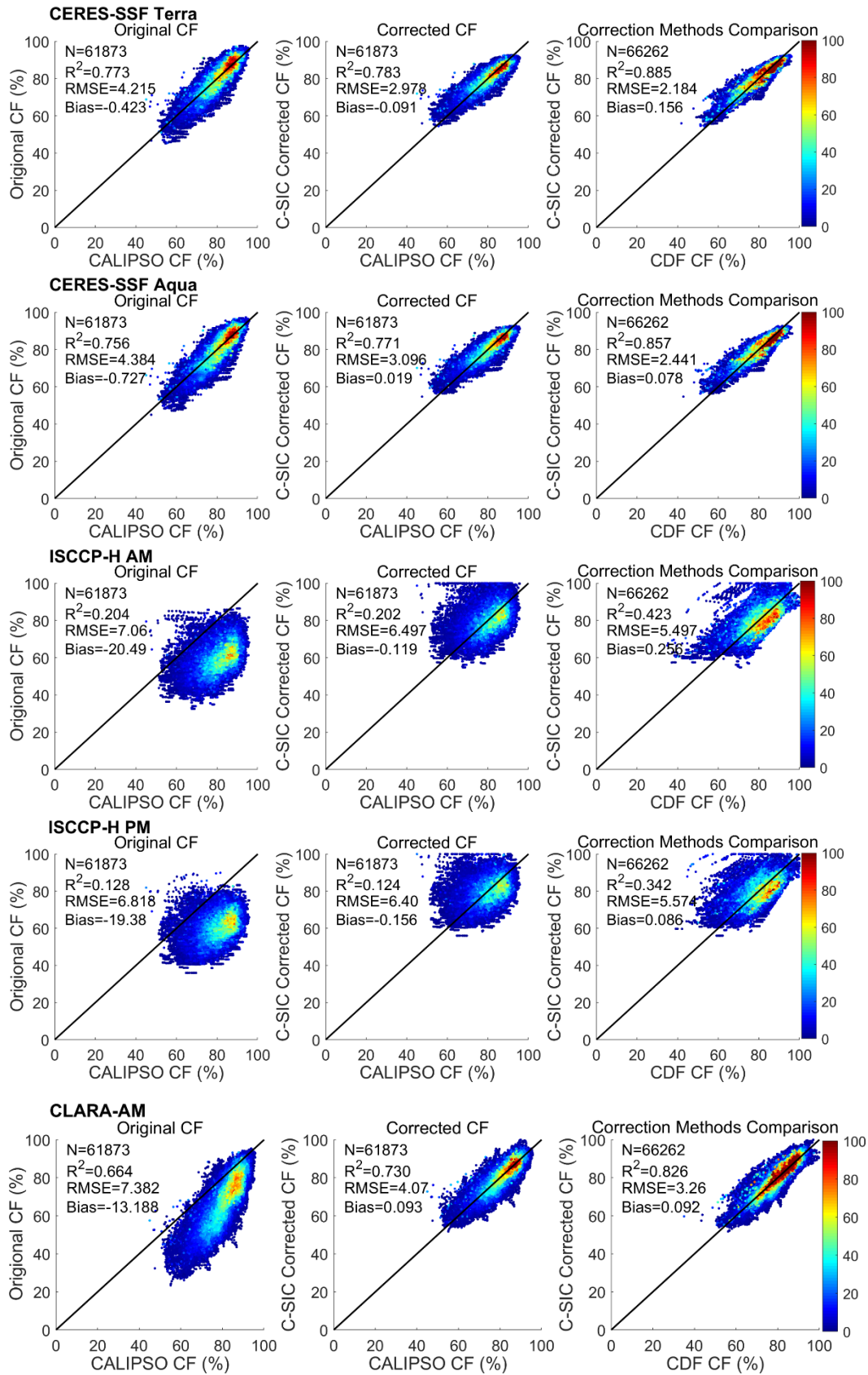


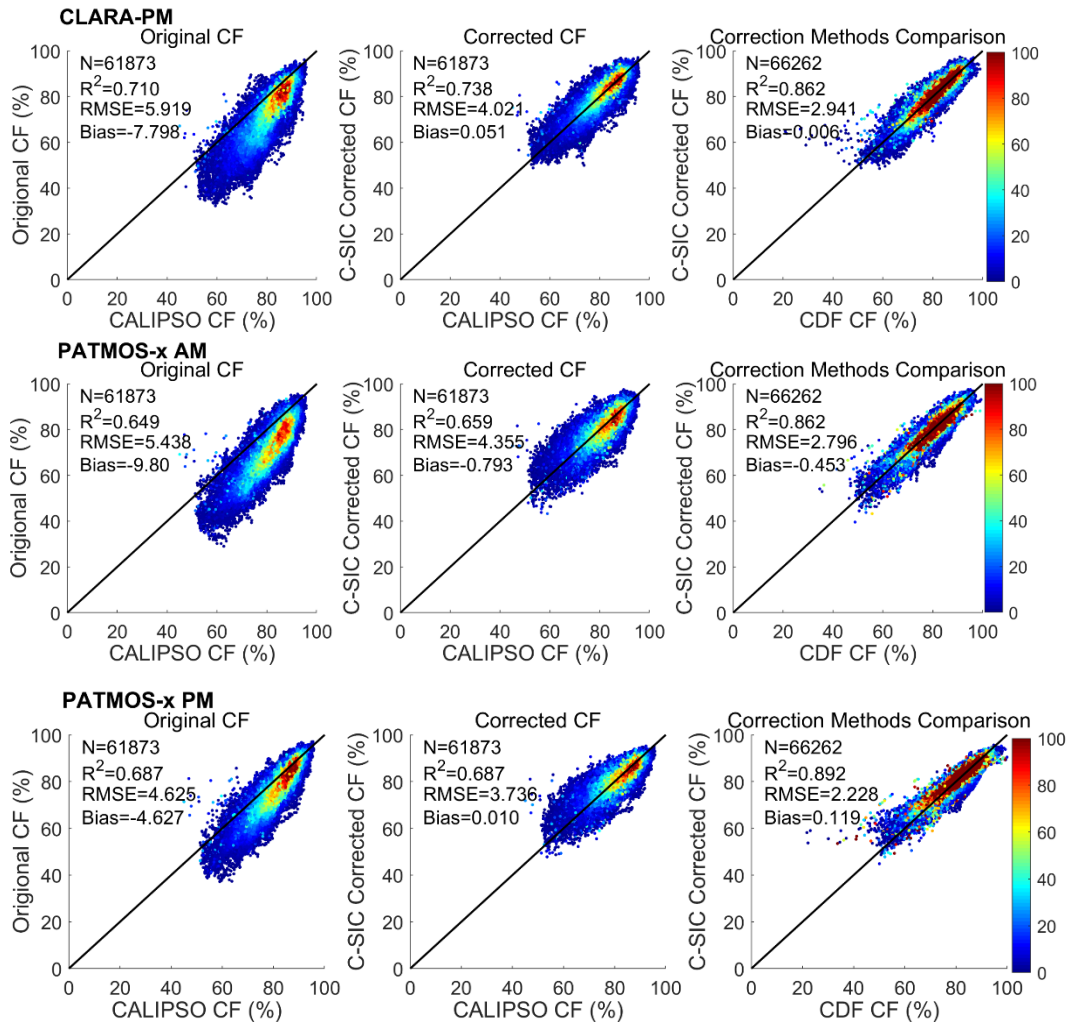


485 Figure 5. The relationship between cloud fraction bias of passive sensor data after and before cumulative
 486 distribution function matching, the cumulative percentage of cloud fraction, and the sea ice concentration in sea ice
 487 regions with latitude less than 82.5°N.
 488

489 By virtue of this association, SIC and CPCF are modeled as dependent variables of the bias. Due to
 490 the predominant presence of sea ice over the domain located above 82.5N, we employ this functional
 491 association to remediate CF inaccuracies in the region, called C-SIC Corrected CF. Figure 6's initial two
 492 panels depict a comparison between the CF of active data and passive data before and after correction by
 493 C-SIC in sea ice regions below 82.5°N. The results indicate that R^2 of the corrected scatter plots increased
 494 slightly, but the RMSEs and bias were greatly reduced. In particular, the CF underestimated by passive
 495 sensors was similar to that of active sensors after correction. In our previous study, we have proven that
 496 this type of underestimation is very common(Liu et al., 2022). The third panel of Figure 6 shows the
 497 comparison of C-SIC Corrected CF and the CDF matching CF in sea ice regions with latitude less than
 498 82.5°N. The results also showed that the C-SIC Corrected CFs have high degree of consistency with the
 499 CFs corrected by the CDF matching, with R^2 over 0.75, RMSE less than 3.6, and bias less than 0.5.
 500 However, although the correction has improved the ISCCP-H CFs, they also showed large
 501 inconsistencies with the passive sensor data and the CDF matching data. Therefore, the ISCCP-H CFs in
 502 regions north of 82.5°N were not included in the following fusion process.



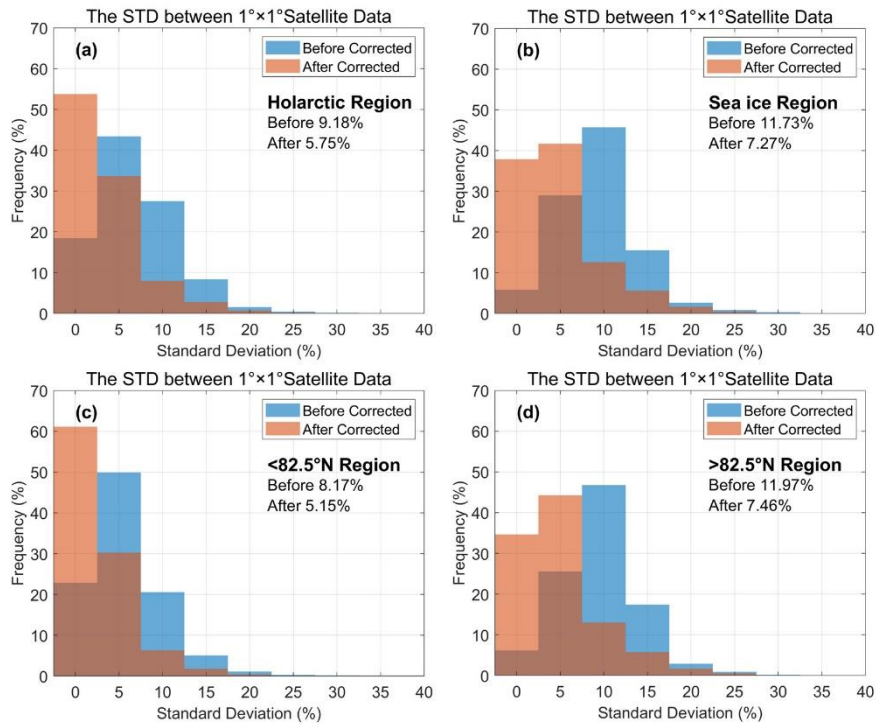




503 Figure 6. The scatter plots of the cloud fraction (CF) comparison between the passive sensor datasets and the
 504 active sensor dataset before (the first panel) and after (the second panel) using the method of CF corrected by the
 505 cumulative percentage of CF and SIC (C-SIC). And the scatter plots of the results comparison between C-SIC and
 506 cumulative distribution function matching (the third panel).

507
 508 Accompanying the decreases in the CF differences of the active and passive sensor data, the
 509 accuracy of individual passive sensor datasets for the entire Arctic during the experimental period was
 510 also generally improved. Moreover, the consistency of multiple satellite data has improved greatly.
 511 Figure 7 displays the standard deviation between $1^\circ \times 1^\circ$ passive sensor CF data before and after the
 512 application of cumulative distribution function matching (latitude $\leq 82.5^\circ\text{N}$) and C-SIC correction
 513 (latitude $> 82.5^\circ\text{N}$). The results obtained from different regions indicate an obvious decrease in the
 514 inconsistency between multiple passive sensor data after the correction with the aforementioned methods.
 515 In the Holarctic region, multiple passive sensor CFs saw a decrease in mean STD from 9.18% to 5.75%,
 516 with more than 50% of the corrected data displaying a standard deviation within 5%. The sea ice region
 517 saw the largest reduction rate of the mean STD, approximately 4.5%. This reduction was mainly derived
 518 from a STD value range of 10–15%, due to the limited detection capacity of passive sensor data in sea
 519 ice areas. Regions with latitude less than 82.5°N saw a decrease in mean STD of only 3.02%. In contrast
 520 to the sea ice region, these land regions saw a smaller standard deviation between multiple satellite data.
 521 The distribution of STD frequency in regions over 82.5°N and the entire sea ice area appeared similar,

522 indicating that the C-SIC correction method was highly effective in 82.5°N regions. Although the relative
 523 values showed improvement, the absolute change appeared inconspicuous.

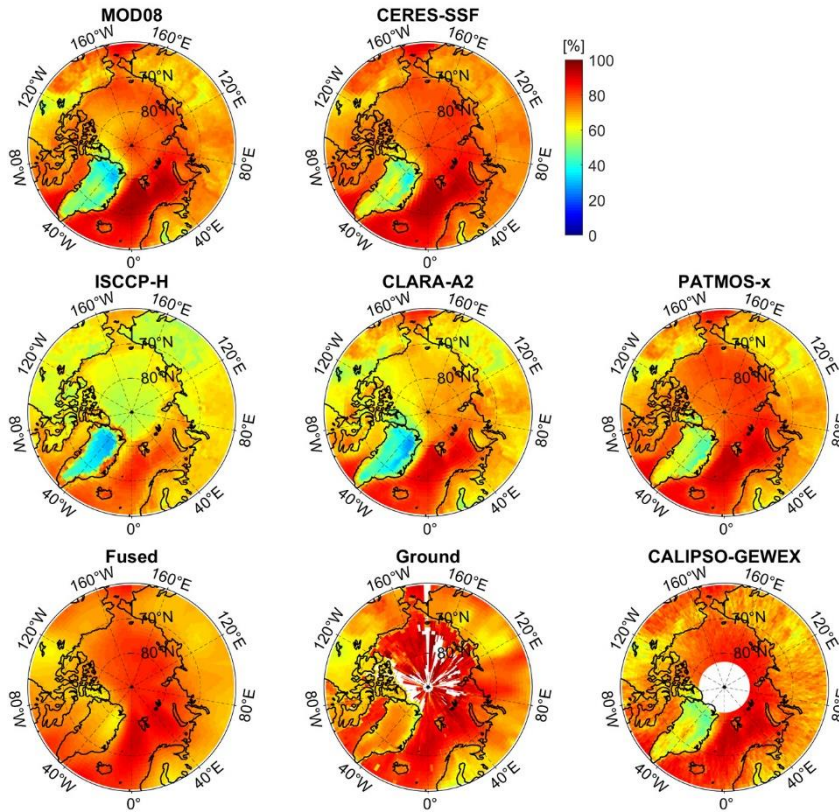


524
 525 Figure 7. Standard deviation between $1^{\circ} \times 1^{\circ}$ passive sensor cloud fraction before and after cumulative distribution
 526 function matching (latitude < 82.5°N) and C-SIC Corrected (latitude > 82.5°N).
 527

528 4.2 Result of BME Fusing

529 4.2.1 Spatial and Temporal Distribution of the Fused CF

530 Figure 8 shows the spatial distribution of Arctic CF from the fused product, multiple satellite data,
 531 and ground observations. The results indicate that although most satellite-based products agreed
 532 relatively well with the ground-based observations in both the geographical distribution and the zonal
 533 average of Arctic CF at first glance, large disparities also appeared in some specific regions, whereas the
 534 fused product we proposed reduced these disparities apparently. For instance, nearly all the passive and
 535 active sensor products show the CFs over the GrIS were less than 60%. However, CFs of ground-based
 536 observations over this region were reported as nearly 70%, which is closer to that of the fused product.
 537 The sea regions of the central Arctic, which are covered by perennial sea ice/snow, are another area where
 538 the passive sensor products always underestimate CF. From these figures, some passive sensor products,
 539 especially for the AVHRR-based datasets, have CFs that are about 10–20% lower than those of active
 540 sensor data and ground-based observations. However, the fused CF has a similar magnitude to these two
 541 referred datasets.



542

543 Figure 8. Distribution of the average cloud fraction of different datasets over the Arctic from 2000 to 2020. The
 544 time ranges for ISCCP-H and CALIPSO-GEWEX were from 2000 to 2017 and from 2006 to 2016, respectively.

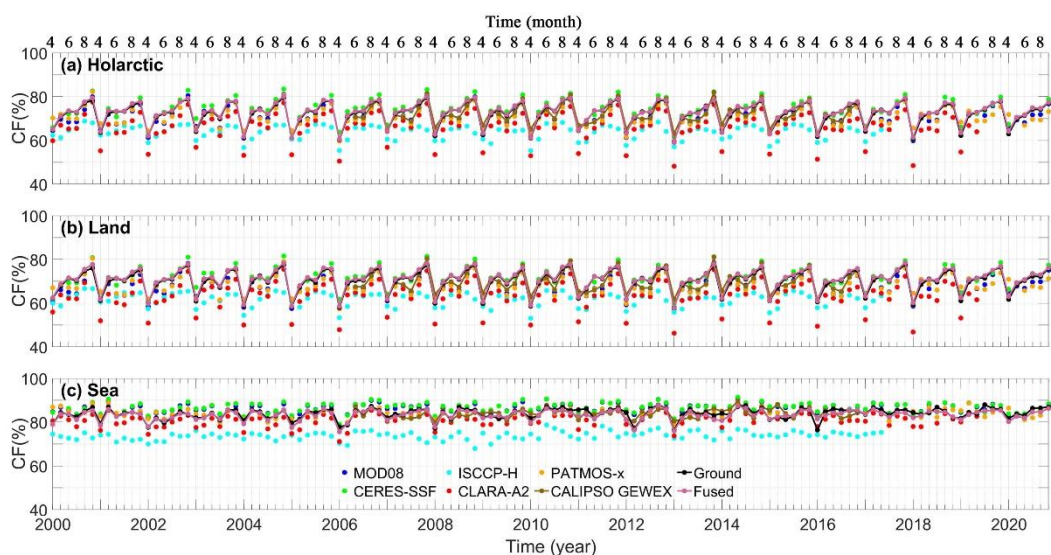
545

546 By contrast, the ground-based CF products have a large data gap because ground weather stations
 547 are sparsely distributed in the Arctic, so the limitation of sampling quantities and frequencies had the
 548 effect of limiting the spatial and temporal ranges of active sensor data. Moreover, the AVHRR-based
 549 products often suffer from missing data as a result of satellite failures or band switching (Hollmann,
 550 2018); in addition, some passive sensor products such as CLARA-A2 have some spatial gaps over the
 551 Arctic Sea during autumn (Karlsson et al., 2017). Although we have eliminated a large number of low-
 552 precision daily data in preprocessing, the completeness of the merged multiple-satellite CF products is
 553 obviously higher than those of the original satellite-based data and ground-based observations in both
 554 spatiality and temporality, especially in regions of the Arctic Ocean. The spatial completeness (the ratio
 555 of available data to the CF grids of the entire Arctic) of the fused CF product was nearly 100%, which is
 556 much larger than 54.09% of ground-based products and 73.15% of the active sensor product. Therefore,
 557 the fusion algorithm proposed by this study can not only obviously reduce the inconsistencies of Arctic
 558 CF between multiple satellite products and reference datasets but also effectively compensate for the data
 559 gaps caused by the lack of reference data.

560

561 It is well known that the CF in the Arctic regions fluctuates apparently with the change in seasons.
 562 To show the temporal accuracy of the fusion products, we analyzed the long time series area-weighted
 563 mean of the CF. Figure 9 depicts the fluctuation of the mean value on a monthly basis for all data
 564 during sunshine periods (April to September) before and after fusion, as demonstrated by the time
 565 series. It is clear that the CF peaks in September and reaches a minimum in April. However, only the
 fused product always maintains a high level of consistency with the reference data, with the monthly

566 mean CF varying from 62% to 79%. The overall area-weighted mean of the differences between fused
 567 CF and CALIPSO-GEWEX CF and between fused CF and ground-based CF was about 0.91% and
 568 0.40%, respectively, which are about one-third of the differences for MODIS-based products and
 569 reference products and about one-fifth to one-twentieth of the differences for AVHRR-based products
 570 and reference products. In land and ocean areas, the fusion algorithm clearly corrects the outliers with
 571 large deviations, such as the CF from CLARA-A2, PATMOS-x products, and the CERES-SSF
 572 products. The first two datasets are well-known for underestimating the Arctic CF dramatically
 573 (Karlsson et al., 2017; Karlsson and Dybbroe, 2010). In this study, the underestimation mainly occurred
 574 in April, with approximately 8% and 3% for those two datasets, respectively. The latter has often been
 575 reported to overestimate CF (Doelling et al., 2016; Trepte et al., 2019), and in this study the CERES-
 576 SSF products nearly overmeasure CF all year long from April to September. However, the fusion
 577 framework proposed by this study scales these underestimated values or overestimated values to a
 578 range similar to that of active sensor data by CDF matching; meanwhile, it takes into account the
 579 deviation from ground observations in the BME fusion process. The fused CFs can not only reduce the
 580 overestimation of CF by MODIS-based products but also decrease the underestimation of CF for
 581 AVHRR-based products, which obviously improves the consistency of CF between the active sensor,
 582 passive sensor, and ground observation dataset compared with the original data.

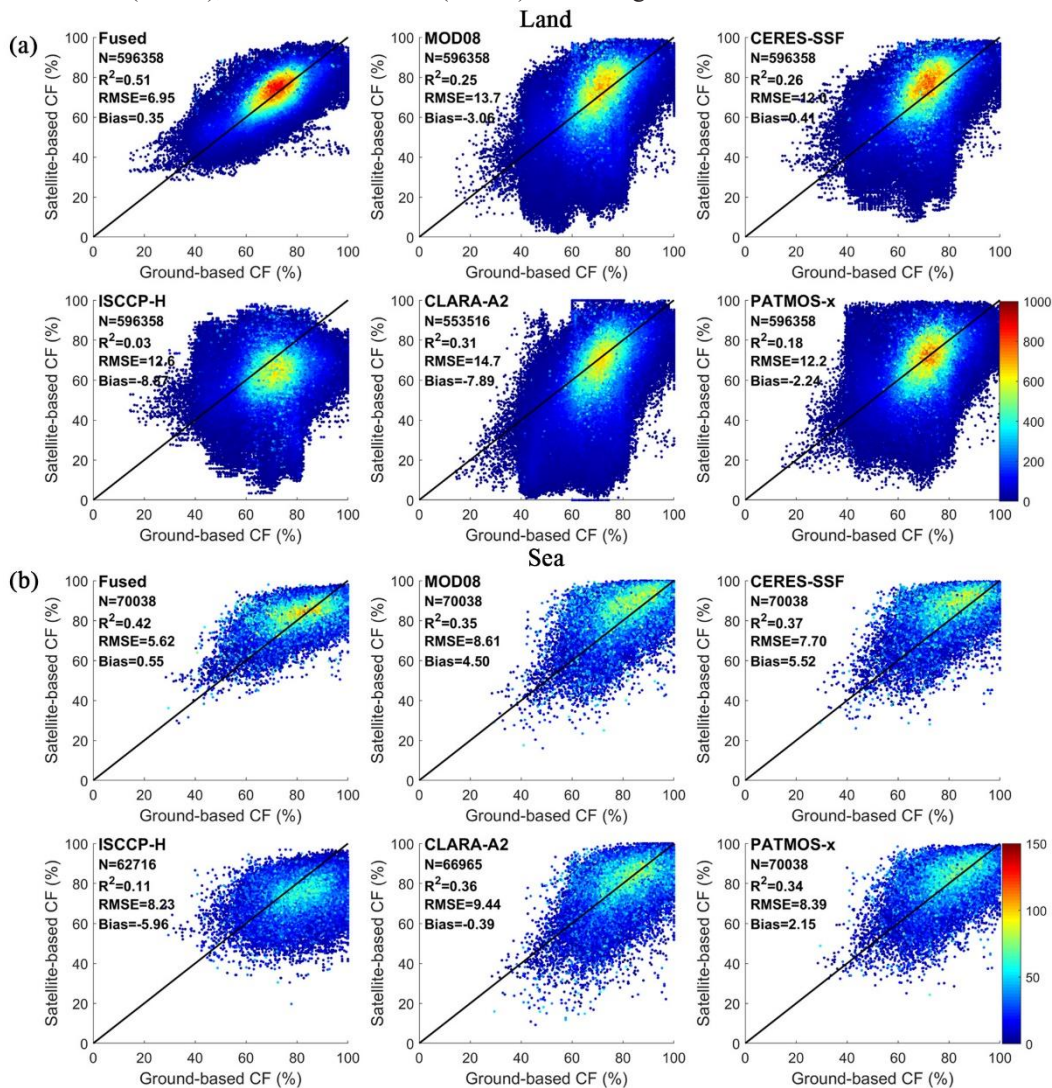


583 Figure 9. The area-weighted means of cloud fraction over (a) Holarctic, (b) Land, and (c) Sea for different
 584 products in the Arctic from April to September during 2000 to 2020. The time ranges for ISCCP-H and CALIPSO-
 585 GEWEX were from 2000 to 2017 and from 2006 to 2016, respectively.

587 4.2.2 Quantitative Assessment of Fused CF

588 To validate the fused CF and compare the accuracy of the fused results to that of several original
 589 satellite CFs, all the passive sensor CF products and the merged CF product were spatiotemporally
 590 compared with the CRU TS4.05 in land regions and ICOADS measurements in sea regions. The
 591 correlation coefficient (R^2), root-mean-square error (RMSE), and mean bias (bias) were used to
 592 quantitatively evaluate the accuracies of the original and merged CF products. As Fig. 10 indicates, the
 593 scatters of the fused CF product and ground-based observations were closer to the 1:1 line than that of
 594 the original satellite data. In this case, the fused data had the largest R^2 (0.51), lowest RMSE (6.95%),

595 and the lowest bias (0.35%) for land regions. In addition, the fused data had the largest R^2 (0.42), the
 596 lowest RMSE (5.62%), and the lowest bias (0.55%) for sea regions.

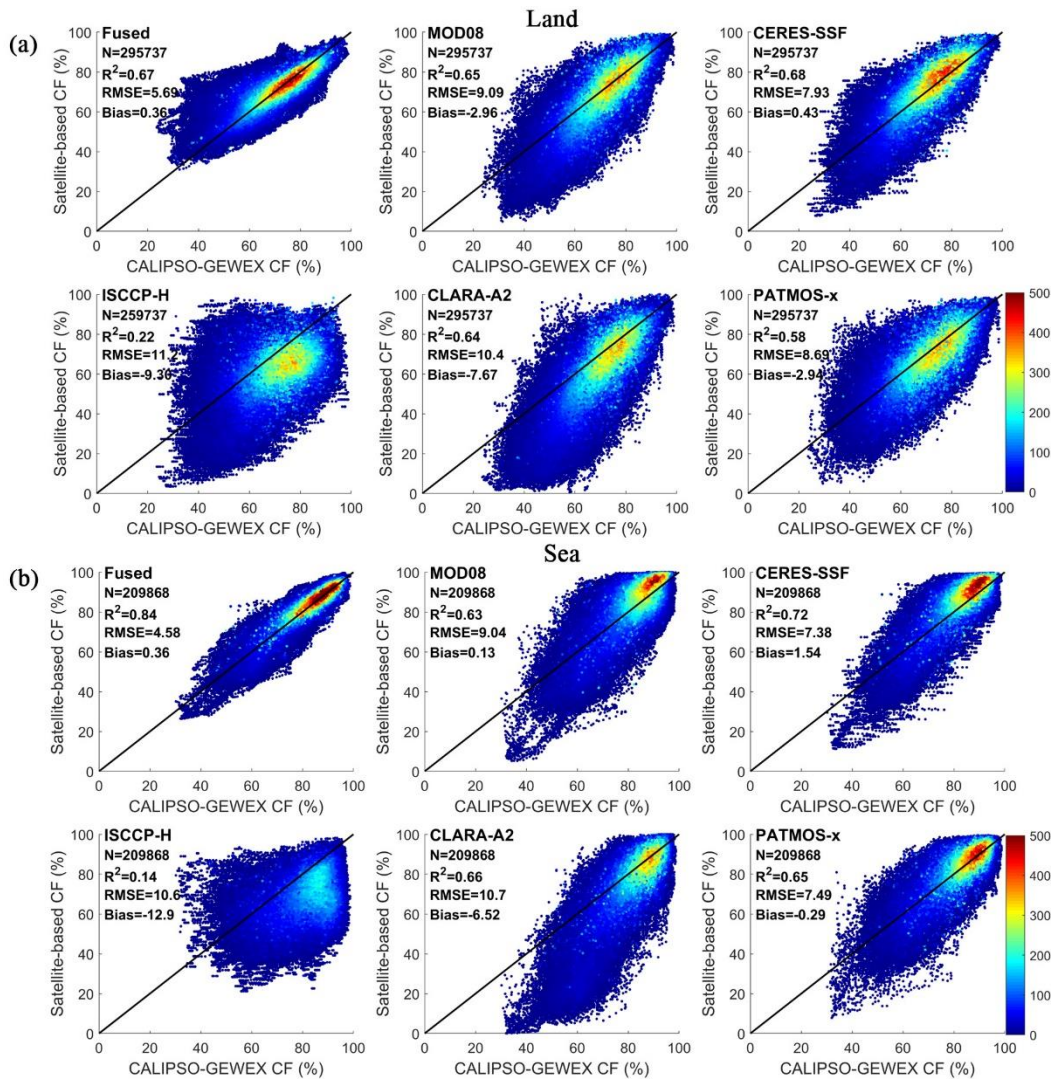


597
 598 Figure 10. Validation of the fused cloud fraction and the original passive sensor datasets against the (a) CRU
 599 TS4.05 and (b) ICOADS datasets.
 600

601 For land, it can be also seen that the fusion results have a strong ability to correct the satellite CF
 602 that is less than 30%. These values were mainly found on the GrIS, in the Canadian Islands, and on the
 603 central Eurasian continent. In addition, the RMSE of CF after fusion was only one-half of the original
 604 satellite data, which means that the overall distribution of the fused CF is better fitted to the reference
 605 data, and most of the CFs with differences over 30% were well-corrected.

606 The observations of ICOADS come from multiple observation platforms, and most of these
 607 platforms operate in open waters. The open water regions varied mostly with the growth and decline of
 608 the SIC, which brings great spatiotemporal heterogeneity for the sampling of ICOADS. Therefore, in the
 609 verification process, the first step was to spatiotemporally collocate the satellite data with ocean site.
 610 Figure 10 (b) shows that R^2 of the fused CF only improved by about 0.05–0.08 when compared with most
 611 satellite data. However, the fusion algorithm reduces the RMSEs and bias obviously. The RMSEs of the
 612 fusion CF were about one-fourth to one-third of the original MODIS-based products and one-third to
 613 three-fifths of the original AVHRR-based products. The reductions of bias were about 4–5% for MODIS-

614 based products and about 2–5.4% for AVHRR-based products.

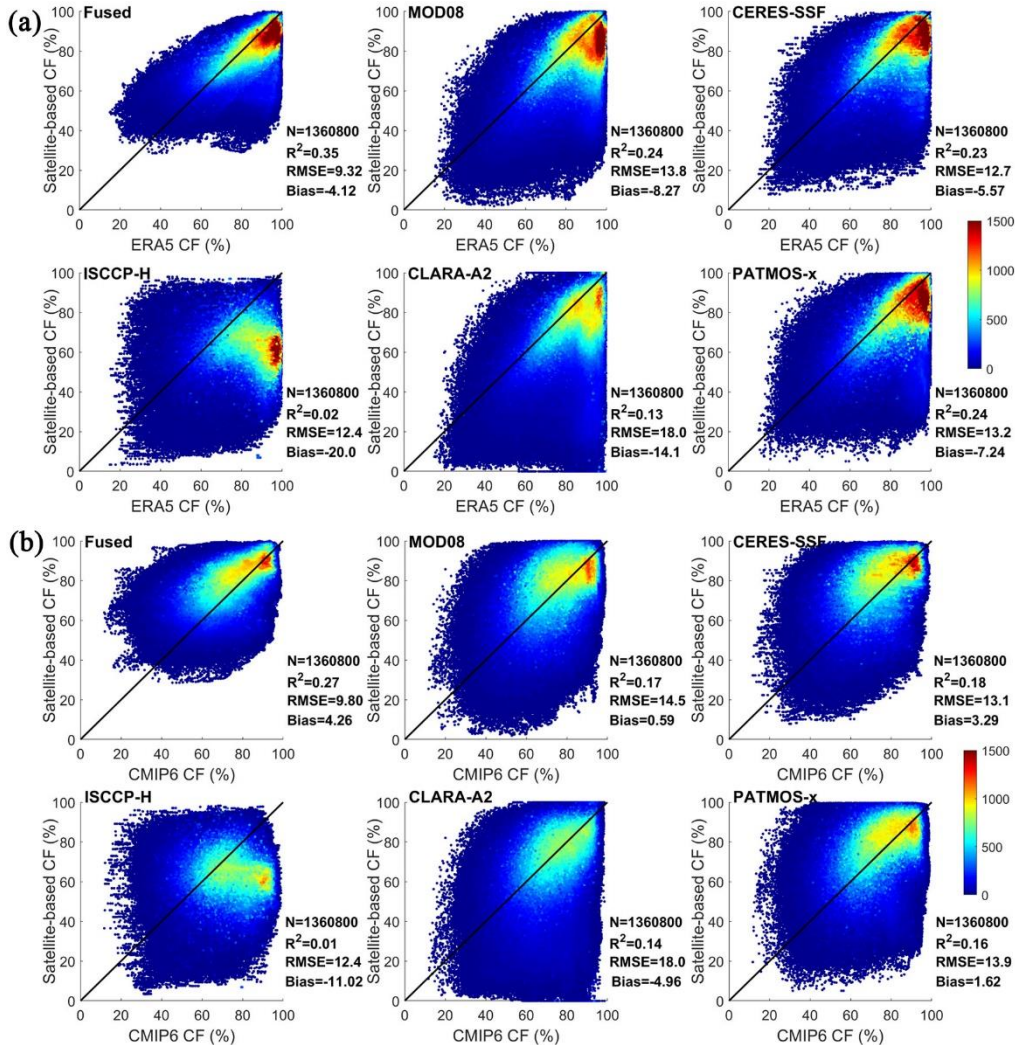


615
616 Figure 11. Validation of the fused cloud fraction and the original passive sensor datasets against the CALIPSO-
617 GEWEX dataset over (a) land and (b) sea regions, with a temporal range from 2006 to 2016.

618

619 As the accepted reference for passive sensor products, CALIPSO-based products are
620 considered to provide excellent data and are always used to validate the accuracy of cloud datasets.
621 In Fig. 11, we compare the CFs of passive sensor products before and after fusion with that of the
622 CALIPSO-GEWEX product. The results show that when compared with the original satellite data,
623 the consistencies between the fused product and the active sensor product were further improved in
624 both land and sea regions. The RMSEs were reduced to about one-third to one-half of the original
625 values, or approximately 5.69% and 4.58% for land and sea regions, respectively. Actually, the
626 consistency of CFs between passive and active sensor datasets was higher than that between satellite
627 data and ground observations. Except for the ISCCP-H products, R^2 of original satellite data was
628 over 0.63; that of fused CF only improved obviously in sea regions (about 0.12–0.21), while it
629 improved slightly but in inconspicuously in land regions (about -0.01–0.1). This can be explained
630 by the fact that the fusion algorithm greatly improves the low-value CFs in the land areas (especially
631 on the GrIS) to levels similar to that of ground-based observations, while the CF of the active sensor
632 data was no more than 60%. Therefore, some overestimations for the fused CF existed when

633 compared with the CALIPSO-GEWEX CF data. From the bias of Fig. 11 (a), we also see that the
 634 fusion algorithm can obviously improve the CF underestimated by the original satellite data.
 635 However, in the sea regions, the MODIS-based datasets seem to overidentify CF, especially when
 636 the CF was over 80%. Meanwhile, the AVHRR-based datasets show underestimation when CF was
 637 less than 80%. Obviously, the fused product corrected these CFs to a more suitable range.



638
 639 Figure 12. Validation of the fused cloud fraction (CF) and the original passive sensor datasets against (a) ERA5 CF
 640 dataset and (b) CMIP6 CF dataset over the Holarctic.

641 Reanalysis data and the climate model data are commonly used to provide a consistent and
 642 continuous dataset for long-term climate trends and variability studies. These datasets can provide
 643 insights into the behavior of the climate system that would be difficult to obtain from direct observations
 644 alone. To further show the advantages of the fusion results, we analyzed the difference in CFs between
 645 different satellite data, ERA5 reanalysis datasets and the MRI-AGCM3-2-S climate model. As can be
 646 seen from Fig. 12, the fusion product greatly reduced the deviation in CF between the satellite data and
 647 the reanalysis dataset and the model data. When compared with the ERA5 CF dataset, the scatters of
 648 fused CFs were more concentrated around the 1:1 line than those of the original satellite data. R^2 of the
 649 fusion product was about 1.5 times higher (improved about 0.18) than that of the original data, and the
 650 RMSEs and bias decreased to one-third of their original values (decreases of about 3.08–8.68% and 1.45–
 651 15.88%, respectively). This means that the distributions of the CFs over the entire Arctic of the fusion

652 product were more consistent with those of the reanalysis CF dataset than the original satellite. However,
653 the low absolute values also indicated that there were inescapable inconsistencies in some grids. The
654 ERA5 dataset has usually been reported to overestimate CF in some regions of the Arctic, especially in
655 the ocean regions (Yeo et al., 2022). In these regions the fused CF has lightly higher values than that of
656 the ERA5 data.

657 The comparison results with the MRI-AGCM3-2-S CF show that when compared with the original
658 satellite data, the fusion method reduced the CF underestimation partly; these underestimations were
659 often seen in April or over the central and western GrIS. In addition, R^2 was improved by about 0.14, and
660 the RMSEs were reduced to one-fourth of their values of original satellite data (about 2.60–8.20%
661 reduction). However, although the fusion data relieve some CF overestimations that occurred in original
662 passive sensor datasets, the scatter plot in Fig. 12 shows that the fusion CFs in some grids were
663 significantly higher than the CF of model data (with bias by 4.26%). These grids are usually found in the
664 open water areas of Arctic Ocean, central Alaska, central Eurasia, and along the eastern margin of
665 Greenland. Several studies have shown that the climate models underestimate the CF over these regions
666 (Vignesh et al., 2020).

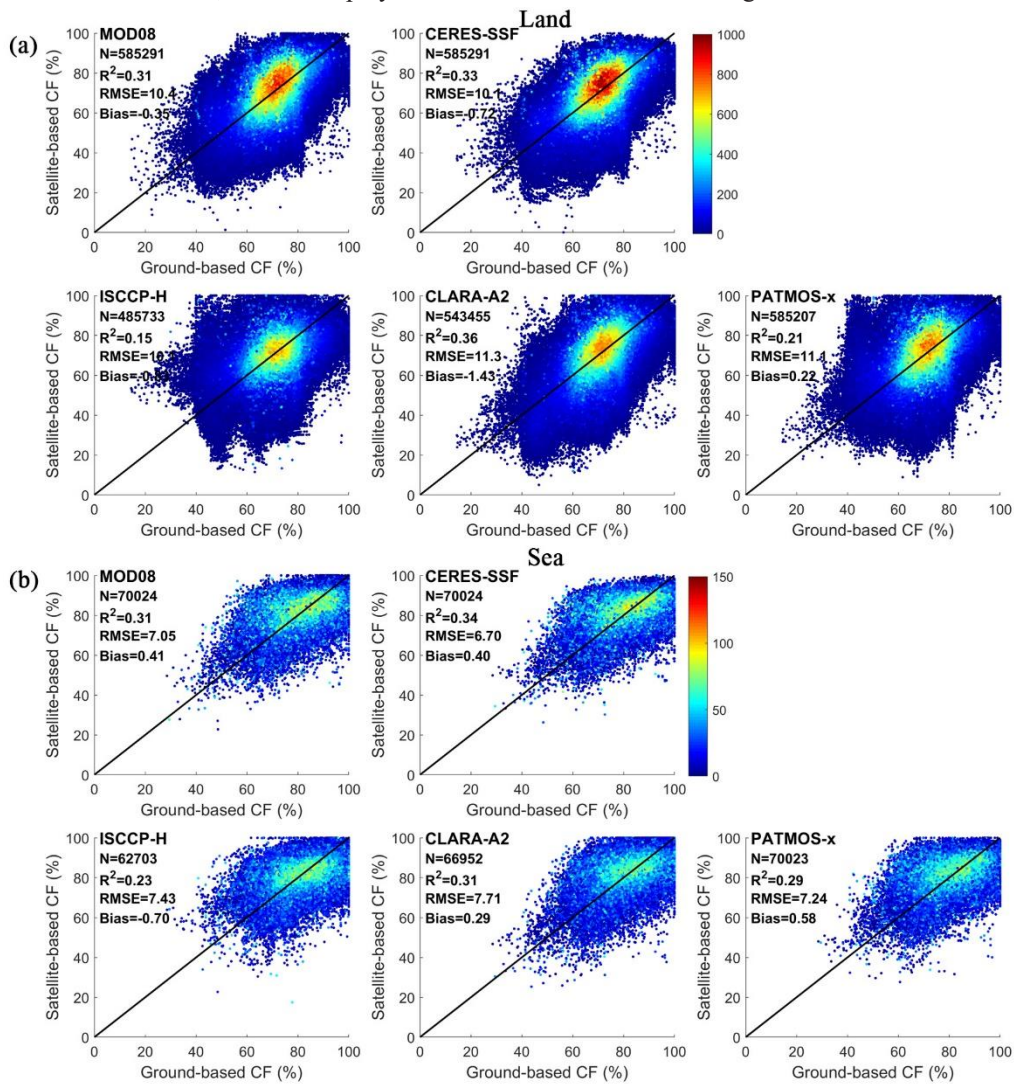
667 **5 Discussion**

668 **5.1 The Efficacy of CDF Matching in CF Fusion**

669 The CDF matching approach was operated based on a time series CF considering the time-varying
670 process of CF products at a specific longitude–latitude grid box. Compared with the metrics for the
671 traditional approach, the CF of multiple passive sensor products was scaled to a level similar to the active
672 sensor CF after CDF matching, so that the inconsistencies among multiple passive sensor CF datasets
673 were reduced. To further evaluate the efficacy of CDF matching in the CF fusion process, we
674 quantitatively evaluated the deviation between satellite data before and after CDF correction with ground
675 observation data.

676 By comparing Fig. 10 and Fig. 13, we can infer that CDF matching can obviously improve the low
677 value of CFs typical of satellite data, making such data more similar to that observed by ground-based
678 sites. These improvements were more obvious for CFs over land regions. Among them, the largest bias
679 correction was seen for the ISCCP-H products (about 7.9% improvement) and the CLARA-A2 products
680 (about 6.5% improvement); the former always underestimated CF in the Arctic (Kotarba, 2015; Liu et
681 al., 2022) and the latter have often been reported to under-identify CF over northern Canada, northern
682 Russia, and across the entire GrIS in land regions and over the entire Arctic Ocean in April (Karlsson
683 and Dybbroe, 2010). Note that the bias of CERES-SSF changes from 0.4% to -0.72% after CDF
684 matching, because CERES-SSF products are usually reported to overestimate CF and these
685 overestimations were corrected reasonably. For the ocean regions, the ground references used in this
686 paper were derived from multiple platform observations, which have great spatio-temporal heterogeneity.
687 Therefore, a large CF discrepancy existed between satellite data and ocean observations. Almost all the
688 passive sensor data have RMSEs and bias that would decrease after CDF correction by about 0.8–1.7%
689 and 0.68–5.26%, respectively. The CDF matching mainly improves the CF in the high-value grid boxes
690 of MODIS-based data and PATMOS-x data as well as in the CF in low-value grid boxes of ISCCP-H and
691 CLARA-A2. Satellite observation covering open sea areas typically presents a higher CF compared to
692 station observation. Consequently, partial overestimation may persist despite correction by the CDF

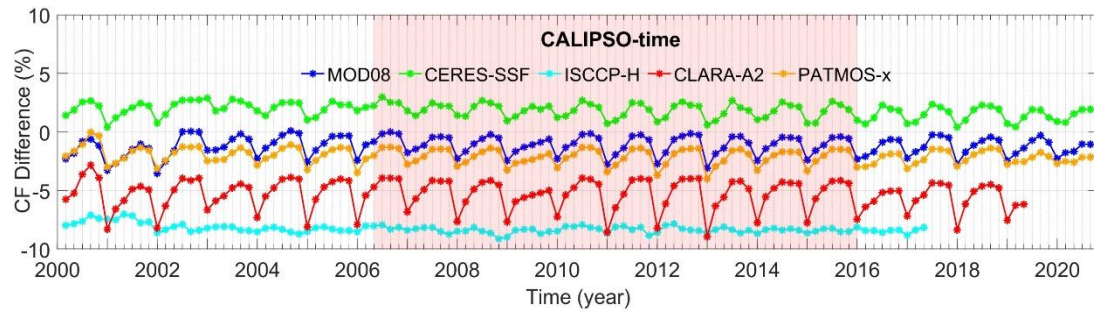
693 matching approach. In the subsequent fusion process, the difference between satellite CF and ground CF
 694 was taken into account, which can play a certain extent role in overfitting correction.



695
 696 Figure 13. Validation of the corrected cloud fraction of passive sensor datasets after cumulative distribution
 697 function matching against (a) CRU TS4.05 dataset over land regions and against (b) ICOADS dataset over sea
 698 regions.
 699

700 In addition, in the land area, CDF matching was directly carried out grid by grid. However, the short
 701 temporal range (2006–2016) of the reference data limits the production of long time-series CF products.
 702 In this study, we proposed a hypothesis that the matching parameter in a specific grid box does not change
 703 over time. To prove the validity of this hypothesis, we conducted sensitivity analysis on the matching
 704 parameters from the fifth to the eleventh year at one-year intervals. The findings indicate that any
 705 deviations in matching parameters were under 0.05% when the time horizon exceeded 8 years. This
 706 demonstrates a level of stability in the correction coefficient when utilizing data for a period exceeding
 707 11 years (Figure A1). Figure 14 displays the variation in differences between satellite data and ground
 708 observations before and after conducting CDF matching throughout the duration of the study. These
 709 differences are calculated by subtracting the deviation between satellite data and ground observations
 710 subsequent to CDF matching from that prior to CDF matching. Clearly, the differences remained steady
 711 over time, and the maximum average annual difference was no more than 1.56%, while part of it was

712 derived from the orbit drift of satellite and variations in the spectral channel.



713

714 Figure 14. The difference in results between satellite data and ground observations before and after cumulative
715 distribution function matching over the Arctic from 2000 to 2020.
716

717 5.2 The Uncertainties of the original Satellite Data

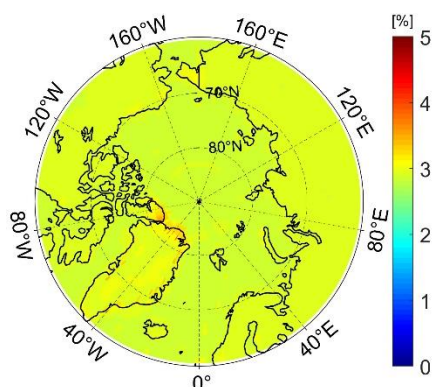
718 CF products from different sensors have different degree of uncertainties. As a knowledge-centered
719 approach, Bayesian maximum entropy approach could integrate informative content with uncertainty
720 from different sources based on a rigorous theoretical support of considerable generality to achieve
721 improved prediction accuracy. For example, the observed data that accompanied by obvious sources of
722 uncertainty such as inaccuracy in measuring devices, modeling uncertainties, and human error are treated
723 as soft data in BME strategy. For the CF datasets of multiple satellite, the uncertainties come from
724 calibration error, orbit drift, signal degradation as well as the errors of cloud detection algorithms (Liu et
725 al., 2022). To achieve optimum estimation of CFs by combining data from multiple sensors, it is
726 imperative to explicitly consider the uncertainties associated with the CF data that is being merged. In
727 our study, the CF data of passive sensor products are viewed as soft data, and the uncertainty associated
728 with different error sources can be expressed explicitly by probability distribution.

729 Specifically, the soft data of multiple satellite CF datasets were constructed by comparing the
730 spatiotemporally collocated satellite CFs and the ground-based records from CRU TS4.05 over land and
731 from ICOADS over sea. Traditionally, the deviations between each satellite dataset and ground site
732 observations at different times and different regions have been averaged to the entire datasets, and then
733 used to calculate the average uncertainty of these data. In this way, the spatial variation of uncertainty in
734 each satellite dataset was ignored. Because the conditions that cause uncertainty are variable in time and
735 space, the uncertainties in each satellite dataset were definitely not the same everywhere (Tang et al.,
736 2016). This is especially true for the ICOADS data, which come from different platforms and introduce
737 large inconsistencies in results. In this study, we constructed soft data for CF over land, ocean, and GrIS
738 regions every month separately by analyzing the PDF differences for different regions and different
739 months, which realized more consistent results with the ground observations. However, despite concerted
740 efforts, determining the uncertainty for each grid remains challenging in light of the substantial temporal
741 and spatial gaps of the reference data, particularly that which pertains to the marine domain.

742 5.3 The Uncertainties of the Fusion CF

743 To assess the fusion algorithm's reliability, we used the standard deviation of error within each grid value
744 in the fusion process to quantify the uncertainties. Specifically, we determined the standard deviation of
745 the predicted posterior probability density function on each grid point. Our findings demonstrate that,

746 with the exception of the northern region of Greenland and part of the margin error, the standard deviation
747 of error in other areas was within 3% (Fig 15). We attribute these discrepancies primarily to the
748 underestimation of satellite observations, particularly the ISCCP-H data, by around 10-30% in the central
749 zone of Greenland. Moreover, the CF of ISCCP-H was significantly overestimated beyond the Greenland
750 margin. Such significant inconsistencies can adversely affect the fusion results. In addition, because the
751 CF of satellite data, particularly satellite data based on AVHRR, was significantly lower than that of
752 ground observation data and active sensor data in April, and a significant difference existed between
753 different datasets, the standard deviation of error after fusion marginally increased in April, with some
754 areas at approximately 4%. It should be noted that our fused data shows an overestimation in the
755 Greenland region. This is mainly because the fusion process prioritizes consistency between the fused
756 data and ground observations. In specific applications, users can make corresponding adjustments based
757 on active sensor data for calibration purposes.



758
759

Figure 15. The mean error standard deviation of the fusion results

760 **6 Data Availability**

761 The fused CF product is available on the Zenodo repository at [https://doi.org/](https://doi.org/10.5281/zenodo.7624605)
762 10.5281/zenodo.7624605 (Liu et al., 2022). The gridded CF data are provided both in *.mat format
763 (Fused_CF_Arctic_MAT, with file size 9.91 MB) and netCDF format (Fused_CF_Arctic_netCDF, with
764 file size 10.7MB) at 1° spatial resolution and monthly temporal resolution during 2000–2020 in
765 percentages. The results in these two folders are exactly the same, someone can download either format
766 as needed.

767 **7 Conclusions**

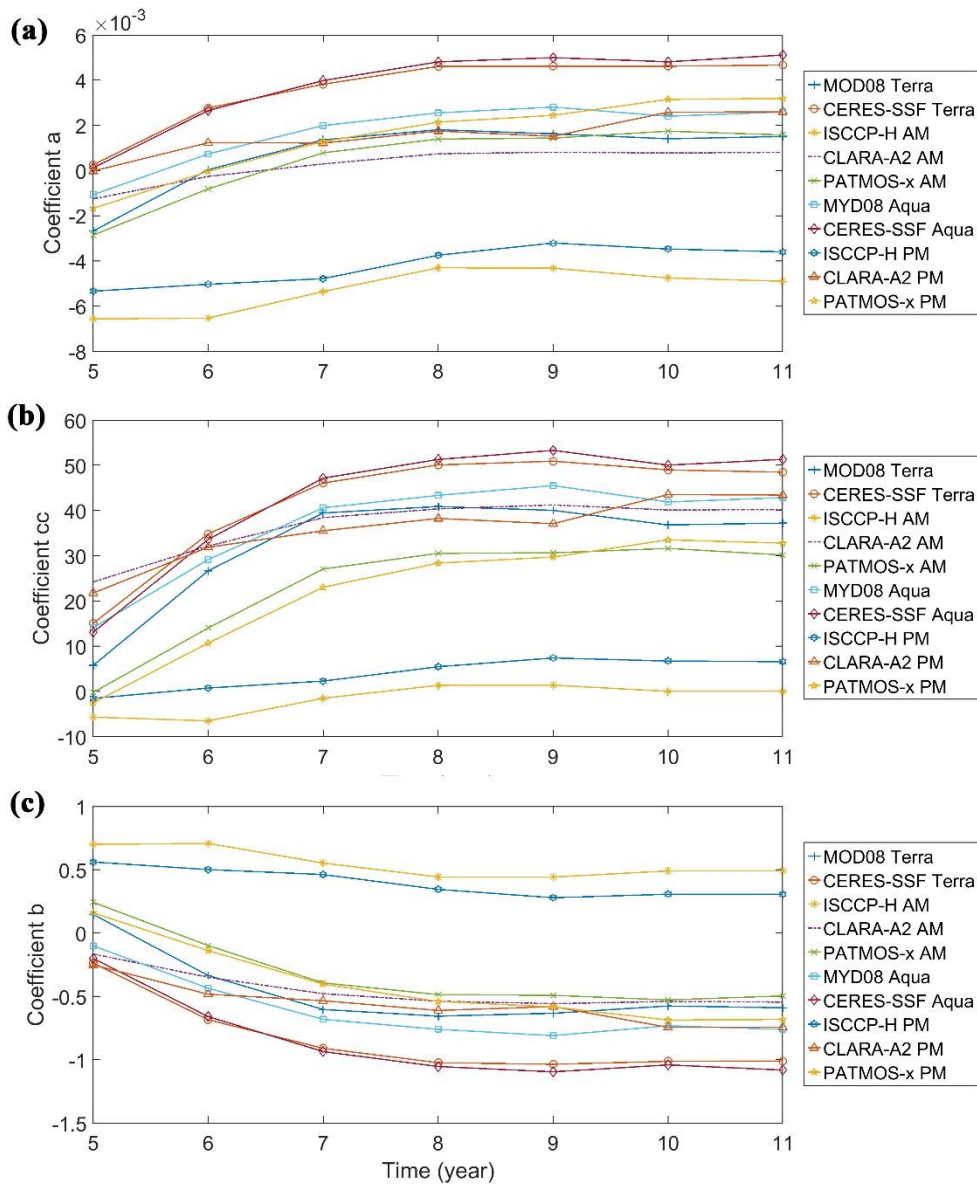
768 The spatiotemporal inconsistency in existing satellite CF products would inhibit their application in
769 climatological and energy budget studies. Over the Arctic region, the special climatic conditions and
770 underlying surface characteristics limit the cloud detection abilities of passive/optical satellite sensors.
771 The complementary features of the CF products derived from multiple satellite sensors in spatial
772 completeness and accuracy make it possible to produce an improved CF product by merging data from
773 multi-sensor satellite CF products.

774 In this paper, we propose a data fusion strategy for producing high-quality monthly CF data over
775 the entire Arctic with a latitude larger than 60°N during sunlit months from 2000 to 2020. Four key steps
776 were involved in the proposed strategy: (1) data quality control; (2) correct the bias of passive sensor

777 data using CDF matching; (3) obtain the spatiotemporally isotropous component by removing the
778 spatiotemporal trends; and (4) produce very accurate CF data by fusing multiple satellite products and
779 considering the uncertainty between satellite data and ground observations with the BME approach.

780 The fusion algorithm proposed by this study apparently reduced inconsistencies in the Arctic CF
781 data acquired by multiple satellite products and the reference products spatiotemporally, resulting in 10–
782 20% reductions of CF differences between fused satellite products and the reference data, and an obvious
783 improvement was seen across the GrIS and in the central Arctic Ocean. The results from 21-year data
784 sets in the study areas demonstrate that the monthly mean CF of the fusion product varied from 62%
785 (April) to 79% (September) during the study period, which is similar to that of the two reference datasets.
786 After CDF matching, the inconsistencies of multiple satellite CF products were reduced by about 3.43%
787 for the entire Arctic, with a larger reduction (4.46%) for sea ice regions. The overestimation of MODIS-
788 based products and the underestimation of AVHRR-based products have been effectively corrected, with
789 the CERES-SSF bias changing from 0.4% to -0.72% and the bias of ISCCP-H and CLARA-A2
790 decreasing by about 7.9% and 6.5%, respectively. After BME fusing, comparisons with the ground-based
791 observations (CRU TS4.05 in land and ICOADS in marine areas) and the active sensor data CALIPSO-
792 GEWEX show that R^2 improved by about 0.05–0.48 for different products; meanwhile, the overall
793 RMSEs and bias of fusion product were reduced by about 2.08–7.75% and 1.6–12.54%, with reductions
794 of nearly 50% and 67% when compared with that of the original passive sensor data, respectively. When
795 compared with the reanalysis CF dataset ERA5 and the model dataset MRI-AGCM3-2-S, R^2 increased
796 by about 0.18 and 0.14, RMSE and bias for reanalysis data decreased by about one-third of that for the
797 original data, with reductions about 3.08–8.68% and 1.45–15.88% for different data, respectively. The
798 RMSEs for model data dropped to one-fourth of their original values (about a 2.60–8.20% reduction).
799 These mean that the proposed fusion algorithm effectively removed CF data with differences greater than
800 30% and made the fused Arctic CF estimation more robust than those data from a single satellite.
801 Nevertheless, the fused product could completely cover the entire Arctic, especially the ocean regions,
802 where the active sensor data and the ground-based data have large data gaps. Temporally, the fused data
803 can complement the missing data caused by the faults of satellites carrying AVHRR sensors and the
804 absence of Aqua data before 2002 as well as the temporal limitation of passive sensors.

805 In general, the proposed fusion algorithm combines the complementary features of multiple satellite
806 CF datasets; it not only takes full advantage of the spatiotemporal autocorrelation among neighboring
807 grids but also incorporates uncertainty estimates of multi-sensor CFs, such as the uncertainties of each
808 passive sensor dataset, the uncertainties between passive and active sensor datasets, as well as the
809 uncertainties between satellite data and ground-based observations. Through temporal and spatial
810 expansion schemes, this fusion framework makes up for the disadvantages in spatiotemporal ranges of
811 reference data. Finally, the fusion algorithm can generate monthly $1^\circ \times 1^\circ$ CF product covering the entire
812 Arctic region during 2000 to 2020, which has positive significance for reducing the uncertainties of
813 assessment of surface radiation flux and improving the accuracy of research related to climate change
814 and energy budgets both regionally and globally. However, some overestimations were observed,
815 especially in ocean regions. This may be attributed to the fact that the ocean stations are too sparse to
816 play a certain role in correcting the overfitting of CDF. Although ICOADS is a widely used ocean
817 validation dataset, it has great spatiotemporal heterogeneity because it comes from a variety of different
818 observation platforms and the sampling is affected by the extent of sea ice. Better reference data should
819 be explored to further improve the uncertainty involved in the assessment of the fused product.



821

822 Figure A1. The sensitivity analysis on the CDF matching parameters from the fifth (2011) to the eleventh year (2017)
 823 of CALIPSO time at one-year intervals. The Coefficient a, b and c are calculated by the least-square fit method. And
 824 the time period only contains sunlight month from April to September.

825 **Author contributions**

826 XL performed the method, validation, and writing the original draft of the paper. TH was responsible
 827 for conceptualization, supported and supervised the study and reviewed the paper. SL was responsible
 828 for conceptualization and reviewed the paper. RL provided guidance on data processing. XX, RM and
 829 YM contributed to the editing and revising of the manuscript. XL prepared the manuscript with
 830 contributions from all co-authors.

831 **Competing interests**

832 The contact author has declared that none of the authors has any competing interests.

833 **Disclaimer**

834 Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in
835 published maps and institutional affiliations.

836 **Acknowledgments**

837 We thank the relevant teams and organizations for providing the data sets used in this study. We
838 thank the NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive
839 Center (LAADS DAAC) for providing the MOD08_M3/MYD08_M3 products, the NASA Langley
840 Research Center Atmospheric Science Data Center (ASDC) for providing CERES_SSF and CALIPSO-
841 GEWEX data, the Satellite Application Facility on Climate Monitoring (CM SAF) for providing the
842 CLARA-A2 product, the NOAA National Centers for Environmental Information (NCEI) for providing
843 PATMOS-x and ISCCP-H products. We also thank the University of East Anglia Climatic Research
844 Unit for their providing the CRU TS4.05 data, the NOAA Physical Sciences Laboratory for their
845 providing the ICOADS marine data, and the European Centre for Medium-Range Weather Forecasts
846 (ECMWF) for their archiving the ERA5 data. Many thanks to the LetPub (<http://letpub.com.cn/>) for its
847 linguistic assistance during the preparation of this manuscript.

848 **Financial support**

849 This work was supported by the National Natural Science Foundation of China Grant (42090012),
850 Hubei Natural Science Foundation Grant (2021CFA082) and National Key Research and Development
851 Program of China (2020YFA0608704).

852 **References**

- 853 Ackerman, S. A., Holz, R. E., Frey, R., Eloranta, E. W., Maddux, B. C., and McGill, M.: Cloud detection
854 with MODIS. Part II: Validation, *Journal of Atmospheric and Oceanic Technology*, 25, 1073-1086,
855 10.1175/2007jtecha1053.1, 2008.
- 856 Beckerman, B. S., Jerrett, M., Serre, M., Martin, R. V., Lee, S.-J., van Donkelaar, A., Ross, Z., Su, J., and
857 Burnett, R. T.: A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in
858 the Contiguous United States, *Environmental Science & Technology*, 47, 7233-7241, 10.1021/es400039u,
859 2013.
- 860 Bogaert, P., Christakos, G., Jerrett, M., and Yu, H. L.: Spatiotemporal modelling of ozone distribution in
861 the State of California, *Atmospheric Environment*, 43, 2471-2480, 10.1016/j.atmosenv.2009.01.049,
862 2009.

863 Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: THE CONCEPT
864 OF ESSENTIAL CLIMATE VARIABLES IN SUPPORT OF CLIMATE RESEARCH, APPLICATIONS,
865 AND POLICY, *B Am Meteorol Soc*, 95, 1431-1443, 10.1175/bams-d-13-00047.1, 2014.

866 Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P.,
867 Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., and Bittelli, M.: Soil moisture estimation
868 through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, *Remote
869 Sensing of Environment*, 115, 3390-3408, 10.1016/j.rse.2011.08.003, 2011.

870 Chatterjee, A., Michalak, A. M., Kahn, R. A., Paradise, S. R., Braverman, A. J., and Miller, C. E.: A
871 geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol
872 optical thickness, *Journal of Geophysical Research-Atmospheres*, 115, 10.1029/2009jd013765, 2010.

873 Christakos, G.: *Modern Spatiotemporal Geostatistics*, Modern spatiotemporal geostatistics2000.

874 Christakos, G.: *Integrative problem-solving in a time of decadence*, Springer Science & Business Media
875 2010.

876 Christakos, G., Kolovos, A., Serre, M. L., and Vukovich, F.: Total ozone mapping by integrating
877 databases from remote sensing instruments and empirical models, *Ieee Transactions on Geoscience and
878 Remote Sensing*, 42, 991-1008, 10.1109/Tgrs.2003.822751, 2004.

879 Christakos G, Serre ML.: BME analysis of spatiotemporal particulate matter distributions in North
880 Carolina. *Atmospheric Environment* 34:3393–3406,2000.

881 Claudia, S., William R., Stefan K.: *Assessment of Global Cloud Data Sets from Satellites A Project of
882 the World Climate Research Programme Global Energy and Water Cycle Experiment (GEWEX)
883 Radiation Panel*, World Climate Research Program Proport, 2012.

884 Cressie, N.: *Statistics for spatial data*, John Wiley & Sons2015.

885 Danso, D. K., Anquetin, S., Diedhiou, A., Kouadio, K., and Koba, A. T.: Daytime low-level clouds in
886 West Africa - occurrence, associated drivers, and shortwave radiation attenuation, *Earth Syst Dynam*, 11,
887 1133-1152, 10.5194/esd-11-1133-2020, 2020.

888 Doelling, D. R., Sun, M., Nguyen, L. T., Nordeen, M. L., Haney, C. O., Keyes, D. F., and Mlynyczak, P.
889 E.: Advances in Geostationary-Derived Longwave Fluxes for the CERES Synoptic (SYN1deg) Product,
890 *Journal of Atmospheric and Oceanic Technology*, 33, 503-521, 10.1175/Jtech-D-15-0147.1, 2016.

891 Drusch, M.: Observation operators for the direct assimilation of TRMM microwave imager retrieved soil
892 moisture, *Geophysical Research Letters*, 32, 10.1029/2005gl023623, 2005.

893 Eastman, R. and Warren, S. G.: Arctic Cloud Changes from Surface and Satellite Observations, *Journal
894 of Climate*, 23, 4233-4242, 10.1175/2010jcli3544.1, 2010.

895 Enriquez-Alonso, A., Sanchez-Lorenzo, A., Calbo, J., Gonzalez, J. A., and Norris, J. R.: Cloud cover
896 climatologies in the Mediterranean obtained from satellites, surface observations, reanalyses, and CMIP5
897 simulations: validation and future scenarios, *Climate Dynamics*, 47, 249-269, 10.1007/s00382-015-
898 2834-4, 2016.

899 Forbes, R. M. and Ahlgrim, M.: On the Representation of High-Latitude Boundary Layer Mixed-Phase
900 Cloud in the ECMWF Global Model %J *Monthly Weather Review*, 142, 3425-3445,
901 <https://doi.org/10.1175/MWR-D-13-00325.1>, 2014.

902 Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan,
903 P., Eastman, R., Gates, L., Gloeden, W., Ji, Z., Lawrimore, J., Rayner, N. A., Rosenhagen, G., and Smith,
904 S. R.: ICOADS Release 3.0: a major update to the historical marine climate record, *International Journal*
905 *of Climatology*, 37, 2211-2232, 10.1002/joc.4775, 2017.

906 Fuentes, M. and Raftery, A. E.: Model evaluation and spatial interpolation by Bayesian combination of
907 observations with outputs from numerical models, *Biometrics*, 61, 36-45, 10.1111/j.0006-
908 341X.2005.030821.x, 2005.

909 Gao, F., Masek, J., Schwaller, M., Hall, F. J. I. T. o. G., and sensing, R.: On the blending of the Landsat
910 and MODIS surface reflectance: Predicting daily Landsat surface reflectance, 44, 2207-2218, 2006.

911 Griffith, D. A.: STATISTICS FOR SPATIAL DATA - CRESSIE,N, *Geographical Analysis*, 25, 271-275,
912 1993.

913 Hakuba, M. Z., Folini, D., Wild, M., Long, C. N., Schaepman-Strub, G., and Stephens, G. L.: Cloud
914 effects on atmospheric solar absorption in light of most recent surface and satellite measurements,
915 10.1063/1.4975543, 2017.

916 Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic
917 observations - the CRU TS3.10 Dataset, *International Journal of Climatology*, 34, 623-642,
918 10.1002/joc.3711, 2014.

919 Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution
920 gridded multivariate climate dataset, *Sci Data*, 7, 109, 10.1038/s41597-020-0453-3, 2020.

921 He, J. and Kolovos, A.: Bayesian maximum entropy approach and its applications: a review, *Stochastic*
922 *Environmental Research and Risk Assessment*, 32, 859-877, 10.1007/s00477-017-1419-7, 2017.

923 Heidinger, A. K., Evan, A. T., Foster, M. J., and Walther, A.: A Naive Bayesian Cloud-Detection Scheme
924 Derived from CALIPSO and Applied within PATMOS-x, *Journal of Applied Meteorology and*
925 *Climatology*, 51, 1129-1144, 10.1175/Jamc-D-11-02.1, 2012.

926 Heidinger, A. K., Foster, M. J., Walther, A., and Zhao, X. P.: The Pathfinder Atmospheres-Extended Avhrr
927 Climate Dataset, *Bulletin of the American Meteorological Society*, 95, 909-+, 10.1175/Bams-D-12-
928 00246.1, 2014.

929 Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., Gao, F., and White, J.
930 C.: A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance
931 based on Landsat and MODIS, *Remote Sensing of Environment*, 113, 1613-1627,
932 10.1016/j.rse.2009.03.007, 2009.

933 Hollmann, R.: ESA Cloud_cci Product Validation and Intercomparison Report(PVIR),
934 10.5676/DWD/ESA_Cloud_cci/AVHRR-PM/V002, 2018.

935 Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E.,
936 Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van
937 Roozendaal, M., and Wagner, W.: THE ESA CLIMATE CHANGE INITIATIVE Satellite Data Records
938 for Essential Climate Variables, *B Am Meteorol Soc*, 94, 1541-1552, 10.1175/bams-d-11-00254.1, 2013.

939 Hu, M. and Xue, M. J. G. r. l.: Implementation and evaluation of cloud analysis with WSR-88D
940 reflectivity data for GSI and WRF-ARW, 34, 2007.

941 Huang, Y. Y., Dong, X. Q., Xi, B. K., Dolinar, E. K., Stanfield, R. E., and Qiu, S. Y.: Quantifying the
942 Uncertainties of Reanalyzed Arctic Cloud and Radiation Properties Using Satellite Surface Observations,
943 *Journal of Climate*, 30, 8007-8029, 10.1175/Jcli-D-16-0722.1, 2017.

944 Hunt, W. H., Winker, D. M., Vaughan, M. A., Powell, K. A., Lucker, P. L., and Weimer, C.: CALIPSO
945 Lidar Description and Performance Assessment, *Journal of Atmospheric and Oceanic Technology*, 26,
946 1214-1228, 10.1175/2009jtech1223.1, 2009.

947 Jaynes, E. T. J. P. r.: *Information theory and statistical mechanics*, 106, 620, 1957.

948 Jin, W., Fu, R.-d., Ye, M., and Li, J.-x.: Meteorological Cloud Image Fusion Using Contourlet Transform
949 and Compressed Sensing, *International Conference on Ecological Protection of Lakes-Wetlands-
950 Watershed and Application of 3S Technology (EPLWW3S 2011)*, Nanchang, PEOPLES R CHINA, 2011
951 Jun 25-26, WOS:000391516000097, 413-416, 2011.

952 Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Meirink, J. F., Devasthale, A., Hanschmann, T.,
953 Kothe, S., Jaaskelainen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D.,
954 Finkensieper, S., Hakansson, N., and Hollmann, R.: CLARA-A2: the second edition of the CM SAF
955 cloud and radiation data record from 34 years of global AVHRR data, *Atmospheric Chemistry and
956 Physics*, 17, 5809-5828, 10.5194/acp-17-5809-2017, 2017.

957 Karlsson, K. G. and Devasthale, A.: Inter-Comparison and Evaluation of the Four Longest Satellite-
958 Derived Cloud Climate Data Records: CLARA-A2, ESA Cloud CCI V3, ISCCP-HGM, and PATMOS-
959 x, *Remote Sens-Basel*, 10, 10.3390/rs10101567, 2018.

960 Karlsson, K. G. and Dybbroe, A.: Evaluation of Arctic cloud products from the EUMETSAT Climate
961 Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations, *Atmospheric
962 Chemistry and Physics*, 10, 1789-1807, DOI 10.5194/acp-10-1789-2010, 2010.

963 Karlsson, K. G. and Hakansson, N.: Characterization of AVHRR global cloud detection sensitivity based
964 on CALIPSO-CALIOP cloud optical thickness information: demonstration of results based on the CM
965 SAF CLARA-A2 climate data record, *Atmos Meas Tech*, 11, 633-649, 10.5194/amt-11-633-2018, 2018.

966 Karlsson, K. G., Riihela, A., Mueller, R., Meirink, J. F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann,
967 J., Kaspar, F., Hollmann, R., and Wolters, E.: CLARA-A1: a cloud, albedo, and radiation dataset from
968 28 yr of global AVHRR data, *Atmospheric Chemistry and Physics*, 13, 5351-5367, 10.5194/acp-13-5351-
969 2013, 2013.

970 Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., and Chen, Y.: Uncertainty
971 Estimate of Surface Irradiances Computed with MODIS-, CALIPSO-, and CloudSat-Derived Cloud and
972 Aerosol Properties, *Surveys in Geophysics*, 33, 395-412, 10.1007/s10712-012-9179-x, 2012.

973 Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., Huang, X., Smith, W. L.,
974 Su, W., and Ham, S.-H.: Surface Irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy System
975 (CERES) Energy Balanced and Filled (EBAF) Data Product, *J Climate*, 31, 4501-4527, 10.1175/jcli-d-
976 17-0523.1, 2018a.

977 Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., Huang, X. L., Smith, W.
978 L., Su, W. Y., and Ham, S. H.: Surface Irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy
979 System (CERES) Energy Balanced and Filled (EBAF) Data Product, *Journal of Climate*, 31, 4501-4527,
980 10.1175/Jcli-D-17-0523.1, 2018b.

981 Kato, S., Rose, F. G., Sun-Mack, S., Miller, W. F., Chen, Y., Rutan, D. A., Stephens, G. L., Loeb, N. G.,
982 Minnis, P., Wielicki, B. A., Winker, D. M., Charlock, T. P., Stackhouse, P. W., Xu, K.-M., and Collins,
983 W. D.: Improvements of top-of-atmosphere and surface irradiance computations with CALIPSO-,
984 CloudSat-, and MODIS-derived cloud and aerosol properties, *Journal of Geophysical Research*, 116,
985 10.1029/2011jd016050, 2011.

986 Kennedy, A., Xi, B., Dong, X., and Zib, B. J.: Evaluation and Intercomparison of Cloud Fraction and
987 Radiative Fluxes in Recent Reanalyses over the Arctic Using BSRN Surface Observations, *J Climate*, 25,
988 2291-2305, 10.1175/jcli-d-11-00147.1, 2012.

989 Kenyon, J. S., Moninger, W. R., Smith, T. L., Peckham, S. E., Lin, H., Grell, G. A., Dowell, D. C., James,
990 E. P., Olson, J. B., Smirnova, T. G., Alexander, C. R., Hu, M., Brown, J. M., Weygandt, S. S., Benjamin,
991 S. G., and Manikin, G. S.: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid
992 Refresh, *Mon Weather Rev*, 144, 1669-1694, 10.1175/mwr-d-15-0242.1, 2016.

993 Kim, D. and Ramanathan, V.: Solar radiation budget and radiative forcing due to aerosols and clouds,
994 *Journal of Geophysical Research*, 113, 10.1029/2007jd008434, 2008.

995 Kotarba, A. Z.: Evaluation of ISCCP cloud amount with MODIS observations, *Atmos Res*, 153, 310-317,
996 10.1016/j.atmosres.2014.09.006, 2015.

997 Kotarba, A. Z.: Calibration of global MODIS cloud amount using CALIOP cloud profiles, *Atmospheric*
998 *Measurement Techniques*, 13, 4995-5012, 10.5194/amt-13-4995-2020, 2020.

999 Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., and He, Y.: Blending multi-resolution satellite sea surface
1000 temperature (SST) products using Bayesian maximum entropy method, *Remote Sensing of Environment*,
1001 135, 52-63, 10.1016/j.rse.2013.03.021, 2013.

1002 Li, L., Shi, R., Zhang, L., Zhang, J., and Gao, W.: The data fusion of aerosol optical thickness using
1003 universal kriging and stepwise regression in East China, *Conference on Remote Sensing and Modeling*
1004 *of Ecosystems for Sustainability XI*, San Diego, CA, 2014

1005 Aug 18-20, WOS:000344548600027, 10.1117/12.2061764, 2014.

1006 Li, S. and Yang, B.: Multifocus image fusion by combining curvelet and wavelet transform, *Pattern*
1007 *Recognition Letters*, 29, 1295-1301, 10.1016/j.patrec.2008.02.002, 2008.

1008 Liu, X., He, T., Sun, L., Xiao, X., Liang, S., and Li, S.: Analysis of Daytime Cloud Fraction
1009 Spatiotemporal Variation over the Arctic from 2000 to 2019 from Multiple Satellite Products, *Journal of*
1010 *Climate*, 35, 3995-4023, 10.1175/jcli-d-22-0007.1, 2022.

1011 Liu, Y., Liu, S., and Wang, Z.: A general framework for image fusion based on multi-scale transform and
1012 sparse representation, *Information Fusion*, 24, 147-164, 10.1016/j.inffus.2014.09.004, 2015.

1013 Liu, Y., Wu, W., Jensen, M. P., and Toto, T.: Relationship between cloud radiative forcing, cloud fraction
1014 and cloud albedo, and new surface-based approach for determining cloud albedo, *Atmospheric Chemistry*
1015 *and Physics*, 11, 7155-7170, 10.5194/acp-11-7155-2011, 2011a.

1016 Liu, Y., Ackerman, S. A., Maddux, B. C., Key, J. R., and Frey, R. A.: Errors in Cloud Detection over the
1017 Arctic Using a Satellite Imager and Implications for Observing Feedback Mechanisms, *Journal of*
1018 *Climate*, 23, 1894-1907, 10.1175/2009jcli3386.1, 2010.

1019 Liu, Y., Key, J. R., Liu, Z., Wang, X., and Vavrus, S. J.: A cloudier Arctic expected with diminishing sea

1020 ice, *Geophysical Research Letters*, 39, n/a-n/a, 10.1029/2012gl051251, 2012a.

1021 Liu, Y. H., Key, J. R., Ackerman, S. A., Mace, G. G., and Zhang, Q. Q.: Arctic cloud macrophysical
1022 characteristics from CloudSat and CALIPSO, *Remote Sensing of Environment*, 124, 159-173,
1023 10.1016/j.rse.2012.05.006, 2012b.

1024 Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe,
1025 M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active
1026 microwave satellite-based retrievals, *Hydrology and Earth System Sciences*, 15, 425-436, 10.5194/hess-
1027 15-425-2011, 2011b.

1028 Loyola R, D. G., Thomas, W., Spurr, R., and Mayer, B.: Global patterns in daytime cloud properties
1029 derived from GOME backscatter UV-VIS measurements, *Int J Remote Sens*, 31, 4295-4318,
1030 10.1080/01431160903246741, 2010.

1031 Marchant, B., Platnick, S., Meyer, K., and Wind, G.: Evaluation of the MODIS Collection 6 multilayer
1032 cloud detection algorithm through comparisons with CloudSat Cloud Profiling Radar and CALIPSO
1033 CALIOP products, *Atmospheric Measurement Techniques*, 13, 3263-3275, 10.5194/amt-13-3263-2020,
1034 2020.

1035 Marchant, B., Platnick, S., Meyer, K., Arnold, G. T., and Riedi, J.: MODIS Collection 6 shortwave-
1036 derived cloud phase classification algorithm and comparisons with CALIOP, *Atmos Meas Tech*, 9, 1587-
1037 1599, 10.5194/amt-9-1587-2016, 2016.

1038 Miao, Q. and Wang, B.: A Novel Image Fusion Method Using Contourlet Transform, *International*
1039 *Conference on Communications*,

1040 Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A., Arduini,
1041 R. F., Trepte, Q. Z., Smith, W. L., Ayers, J. K., Gibson, S. C., Miller, W. F., Hong, G., Chakrapani, V.,
1042 Takano, Y., Liou, K. N., Xie, Y., and Yang, P.: CERES Edition-2 Cloud Property Retrievals Using TRMM
1043 VIRS and Terra and Aqua MODIS Data-Part I: Algorithms, *Ieee Transactions on Geoscience and Remote*
1044 *Sensing*, 49, 4374-4400, 10.1109/tgrs.2011.2144601, 2011.

1045 Nazelle, A. D., Arunachalam, S., and Serre, M. L.: Bayesian maximum entropy integration of ozone
1046 observations and model predictions: an application for attainment demonstration in North Carolina,
1047 *Environ Sci Technol*, 44, 5707-5713, 10.1021/es100228w, 2010.

1048 Nie, S., Wu, T., Luo, Y., Deng, X., Shi, X., Wang, Z., Liu, X., and Huang, J.: A strategy for merging
1049 objective estimates of global daily precipitation from gauge observations, satellite estimates, and
1050 numerical predictions, *Advances in Atmospheric Sciences*, 33, 889-904, 10.1007/s00376-016-5223-y,
1051 2016.

1052 Paul, A. H.: Collection 6.1 Change Summary Document MODIS Atmosphere Level-3 Algorithm and
1053 Global Products, 2017.

1054 Philipp, D., Stengel, M., and Ahrens, B.: Analyzing the Arctic Feedback Mechanism between Sea Ice
1055 and Low-Level Clouds Using 34 Years of Satellite Observations, *Journal of Climate*, 33, 7479-7501,
1056 10.1175/jcli-d-19-0895.1, 2020.

1057 Poulsen, C. J., Tabor, C., and White, J.: Response to Comment on "Long-term climate forcing by
1058 atmospheric oxygen concentrations", *Science*, 353, 10.1126/science.aad8550, 2016.

1059 Qian, Y., Long, C. N., Wang, H., Comstock, J. M., McFarlane, S. A., and Xie, S.: Evaluation of cloud
1060 fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations,
1061 *Atmospheric Chemistry and Physics*, 12, 1785-1810, 10.5194/acp-12-1785-2012, 2012.

1062 Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., and Hartmann,
1063 D.: Cloud-Radiative Forcing and Climate - Results from the Earth Radiation Budget Experiment, *Science*,
1064 243, 57-63, DOI 10.1126/science.243.4887.57, 1989.

1065 Rossow, W. B. and Schiffer, R. A.: Advances in understanding clouds from ISCCP, *Bulletin of the*
1066 *American Meteorological Society*, 80, 2261-2287, 10.1175/1520-0477(1999)080<2261:Aiucfi>2.0.Co;2,
1067 1999.

1068 Savelyeva, E., Utkin, S., Kazakov, S., and Demyanov, V.: Modeling Spatial Uncertainty for Locally
1069 Uncertain Data, 7th International Conference on Geostatistics for Environmental Applications,
1070 Southampton, ENGLAND, 2010

1071 Sep, WOS:000288481100026, 295-+, 10.1007/978-90-481-2322-3_26, 2010.

1072 Shupe, M. D., Turner, D. D., Walden, V. P., Bennartz, R., Cadetdu, M. P., Castellani, B. B., Cox, C. J.,
1073 Hudak, D. R., Kulie, M. S., Miller, N. B., Neely, R. R., Neff, W. D., and Rowe, P. M.: HIGH AND DRY
1074 New Observations of Tropospheric and Cloud Properties above the Greenland Ice Sheet, *B Am Meteorol*
1075 *Soc*, 94, 169-+, 10.1175/Bams-D-11-00249.1, 2013.

1076 Sledd, A. and L'Ecuyer, T. S.: Emerging Trends in Arctic Solar Absorption, *Geophys Res Lett*, 48,
1077 10.1029/2021gl095813, 2021.

1078 Spadavecchia, L. and Williams, M.: Can spatio-temporal geostatistical methods improve high resolution
1079 regionalisation of meteorological variables?, *Agricultural and Forest Meteorology*, 149, 1105-1117,
1080 10.1016/j.agrformet.2009.01.008, 2009.

1081 Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., Christensen, M., Carbajal
1082 Henken, C., Preusker, R., Fischer, J., Devasthale, A., Willén, U., Karlsson, K.-G., McGarragh, G. R.,
1083 Proud, S., Povey, A. C., Grainger, R. G., Meirink, J. F., Feofilov, A., Bennartz, R., Bojanowski, J. S., and
1084 Hollmann, R.: Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the
1085 framework of the Cloud_cci project, *Earth Syst Sci Data*, 9, 881-904, 10.5194/essd-9-881-2017, 2017.

1086 Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L.,
1087 Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick,
1088 S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G.: Assessment of
1089 Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel,
1090 *Bulletin of the American Meteorological Society*, 94, 1031-1049, 10.1175/bams-d-12-00117.1, 2013.

1091 Sun, B. M., Free, M., Yoo, H. L., Foster, M. J., Heidinger, A., and Karlsson, K. G.: Variability and Trends
1092 in U.S. Cloud Cover: ISCCP, PATMOS-x, and CLARA-A1 Compared to Homogeneity-Adjusted
1093 Weather Observations, *Journal of Climate*, 28, 4373-4389, 10.1175/jcli-d-14-00805.1, 2015.

1094 Tang, Q., Bo, Y., and Zhu, Y.: Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD)
1095 products using Bayesian maximum entropy method, *Journal of Geophysical Research: Atmospheres*, 121,
1096 4034-4048, 10.1002/2015jd024571, 2016.

1097 Tiedtke, M.: Representation of Clouds in Large-Scale Models, *Monthly Weather Review*, 121, 3040-
1098 3061, Doi 10.1175/1520-0493(1993)121<3040:Rocils>2.0.Co;2, 1993.

1099 Toll, V., Christensen, M., Quaas, J., and Bellouin, N.: Weak average liquid-cloud-water response to
1100 anthropogenic aerosols, *Nature*, 572, 51-55, 10.1038/s41586-019-1423-9, 2019.

1101 Trepte, Q. Z., Bedka, K. M., Chee, T. L., Minnis, P., Sun-Mack, S., Yost, C. R., Chen, Y., Jin, Z., Hong,
1102 G., Chang, F.-L., and Smith, W. L.: Global Cloud Detection for CERES Edition 4 Using Terra and Aqua
1103 MODIS Data, *IEEE Transactions on Geoscience and Remote Sensing*, 57, 9410-9449,
1104 10.1109/tgrs.2019.2926620, 2019.

1105 Tzallas, V., Hatzianastassiou, N., Benas, N., Meirink, J. F., Matsoukas, C., Stackhouse, P., and Vardavas,
1106 I.: Evaluation of CLARA-A2 and ISCCP-H Cloud Cover Climate Data Records over Europe with
1107 ECA&D Ground-Based Measurements, *Remote Sens-Basel*, 11, 10.3390/rs11020212, 2019.

1108 Van Tricht, K., Lhermitte, S., Lenaerts, J. T. M., Gorodetskaya, I. V., L'Ecuyer, T. S., Noel, B., van den
1109 Broeke, M. R., Turner, D. D., and van Lipzig, N. P. M.: Clouds enhance Greenland ice sheet meltwater
1110 runoff, *Nature Communications*, 7, ARTN 10266
1111 10.1038/ncomms10266, 2016.

1112 Vaughan, M., Young, S., Winker, D., Powell, K., Omar, A., Liu, Z. Y., Hu, Y. X., and Hostetler, C.: Fully
1113 automated analysis of space-based lidar data: an overview of the CALIPSO retrieval algorithms and data
1114 products, *Bba Lib*, 5575, 16-30, 10.1117/12.572024, 2004.

1115 Vaughan, M. A., Powell, K. A., Kuehn, R. E., Young, S. A., Winker, D. M., Hostetler, C. A., Hunt, W. H.,
1116 Liu, Z. Y., McGill, M. J., and Getzewich, B. J.: Fully Automated Detection of Cloud and Aerosol Layers
1117 in the CALIPSO Lidar Measurements, *Journal of Atmospheric and Oceanic Technology*, 26, 2034-2050,
1118 10.1175/2009jtecha1228.1, 2009.

1119 Vignesh, P. P., Jiang, J. H., Kishore, P., Su, H., Smay, T., Brighton, N., and Velicogna, I.: Assessment of
1120 CMIP6 Cloud Fraction and Comparison with Satellite Observations, *Earth and Space Science*, 7,
1121 10.1029/2019ea000975, 2020.

1122 Walsh, J. E., Chapman, W. L., and Portis, D. H.: Arctic Cloud Fraction and Radiative Fluxes in
1123 Atmospheric Reanalyses, *Journal of Climate*, 22, 2316-2334, 10.1175/2008jcli2213.1, 2009.

1124 Wang, D., Bi, S., Wang, B., and Yan, J.: Satellite cloud image fusion based on regional feature with
1125 nonsubsamped contourlet transform, *Journal of Computer Applications*, 32, 2585-2587, 2012.

1126 Winker, D. M., Hunt, W. H., and McGill, M. J.: Initial performance assessment of CALIOP, *Geophysical*
1127 *Research Letters*, 34, Artn L19803
1128 10.1029/2007gl030135, 2007.

1129 Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y. X., Powell, K. A., Liu, Z. Y., Hunt, W. H., and Young,
1130 S. A.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, *Journal of*
1131 *Atmospheric and Oceanic Technology*, 26, 2310-2323, 10.1175/2009jtecha1281.1, 2009.

1132 Woodruff, S. D., Diaz, H. F., Worley, S. J., Reynolds, R. W., and Lubker, S. J.: Early ship observational
1133 data and ICOADS, *Climatic Change*, 73, 169-194, 10.1007/s10584-005-3456-3, 2005.

1134 Wu, W., Liu, Y. G., Jensen, M. P., Toto, T., Foster, M. J., and Long, C. N.: A comparison of multiscale
1135 variations of decade-long cloud fractions from six different platforms over the Southern Great Plains in
1136 the United States, *J Geophys Res-Atmos*, 119, 3438-3459, 10.1002/2013jd019813, 2014.

1137 Xia, X., Zhao, B., Zhang, T., Wang, L., Gu, Y., Liou, K.-N., Mao, F., Liu, B., Bo, Y., Huang, Y., Dong,
1138 J., Gong, W., and Zhu, Z.: Satellite-Derived Aerosol Optical Depth Fusion Combining Active and Passive
1139 Remote Sensing Based on Bayesian Maximum Entropy, *IEEE Transactions on Geoscience and Remote*
1140 *Sensing*, 60, 1-13, 10.1109/tgrs.2021.3051799, 2022.

1141 Xie, S. C., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Clothiaux, E. E., Gaustad, K.
1142 L., Golaz, J. C., Hall, S. D., Jensen, M. P., Johnson, K. L., Lin, Y. L., Long, C. N., Mather, J. H., McCord,
1143 R. A., McFarlane, S. A., Palanisamy, G., Shi, Y., and Turner, D. D. D.: ARM CLIMATE MODELING
1144 BEST ESTIMATE DATA A New Data Product for Climate Studies, *Bulletin of the American*
1145 *Meteorological Society*, 91, 13-+, 10.1175/2009bams2891.1, 2010.

1146 Xu, S. and Cheng, J.: A new land surface temperature fusion strategy based on cumulative distribution
1147 function matching and multiresolution Kalman filtering, *Remote Sensing of Environment*, 254,
1148 10.1016/j.rse.2020.112256, 2021.

1149 Xu, S., Cheng, J., and Zhang, Q.: Reconstructing All-Weather Land Surface Temperature Using the
1150 Bayesian Maximum Entropy Method Over the Tibetan Plateau and Heihe River Basin, *IEEE Journal of*
1151 *Selected Topics in Applied Earth Observations and Remote Sensing*, 12, 3307-3316,
1152 10.1109/jstars.2019.2921924, 2019.

1153 Yang, J. and Hu, M.: Filling the missing data gaps of daily MODIS AOD using spatiotemporal
1154 interpolation, *Science of the Total Environment*, 633, 677-683, 10.1016/j.scitotenv.2018.03.202, 2018.

1155 Yeo, H., Kim, M.-H., Son, S.-W., Jeong, J.-H., Yoon, J.-H., Kim, B.-M., and Kim, S.-W.: Arctic cloud
1156 properties and associated radiative effects in the three newer reanalysis datasets (ERA5, MERRA-2,
1157 JRA-55): Discrepancies and possible causes, *Atmospheric Research*, 270,
1158 10.1016/j.atmosres.2022.106080, 2022.

1159 Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., and Rossow, W. B.: The International Satellite
1160 Cloud Climatology Project H-Series climate data record product, *Earth Syst Sci Data*, 10, 583-593,
1161 10.5194/essd-10-583-2018, 2018.

1162 Yu, H.-L. and Wang, C.-H.: Retrospective prediction of intraurban spatiotemporal distribution of PM2.5
1163 in Taipei, *Atmospheric Environment*, 44, 3053-3065, 10.1016/j.atmosenv.2010.04.030, 2010.

1164 Zhang, C.-J., Chen, Y., Duanmu, C., and Feng, H.-J.: Multi-channel satellite cloud image fusion in the
1165 tetrolet transform domain, *Int J Remote Sens*, 35, 8138-8168, 10.1080/01431161.2014.980918, 2014.

1166 Zhang, Q., Cheng, J., and Liang, S.: Deriving high-quality surface emissivity spectra from atmospheric
1167 infrared sounder data using cumulative distribution function matching and principal component analysis
1168 regression, *Remote Sensing of Environment*, 211, 388-399, 10.1016/j.rse.2018.04.033, 2018.

1169 Zhu, X., Chen, J., Gao, F., Chen, X., and Masek, J. G.: An enhanced spatial and temporal adaptive
1170 reflectance fusion model for complex heterogeneous regions, *Remote Sensing of Environment*, 114,
1171 2610-2623, 10.1016/j.rse.2010.05.032, 2010.

1172

1173