

1 **Synergistic retrievals of ice in high clouds from lidar, Ku-band radar and**
2 **submillimeter wave radiometer observations**

3 Mircea Grecu^{a,b} and John E. Yorks ^a

4 ^a *NASA GSFC*

5 ^b *Morgan State University*

6 *Corresponding author:* Mircea Grecu, mircea.grecu-1@nasa.gov

7 ABSTRACT: In this study, we investigate the synergy of elastic backscatter lidar, Ku-band
8 radar, and sub-millimeter-wave radiometer measurements in the retrieval of the ice from satellite
9 observations. The synergy is analyzed through the generation of a large dataset of IWC profile
10 and simulated lidar, radar and radiometer observations. The characteristics of the instruments e.
11 g. frequencies, sensitivities, etc. are set based on the expected characteristics of instruments of
12 the AOS mission. A cross-validation methodology is used to assess the accuracy of the retrieved
13 IWC profiles from various combinations of observations from the three instruments. Specifically,
14 the IWC and associated observations is randomly divided into two datasets, one for the training
15 and the other for the evaluation. The training dataset is used to train the retrieval algorithm, while
16 the evaluation dataset is used to assess the retrieval performance. The dataset of IWC profiles is
17 derived from CloudSat reflectivity observations.

18 The retrieval of the ice water content IWC profiles from the computed observations is achieved in
19 two steps. In the first step, a class, out of 18 potential classes characterized by different vertical
20 distribution of IWC, is estimated from the observations. The 18 classes are predetermined based on
21 k-Means clustering algorithm. In the second step, the IWC profile is estimated using a Ensemble
22 Kalman Smoother (EKS) algorithm that uses the estimated class as a priori information.

23 The results of the study show that the synergy of lidar, radar, and radiometer observations is
24 significant in the retrieval of the IWC profiles. The inclusion of the lidar backscatter observations
25 in the retrieval process has a larger impact on the retrieval performance than the inclusion of the
26 radar observations. As ice clouds have a significant impact on atmospheric radiative processes,
27 this work is relevant to ongoing efforts to reduce uncertainties in climate analyses and projections.

²⁸ **1. Introduction**

²⁹ The future NASA Atmospheric Observing System (AOS) mission (Braun 2022) is expected
³⁰ to feature new combinations of observations that may be used to quantify the amounts of ice in
³¹ high clouds and characterize the microphysical properties of ice particles. These observations
³² include backscatter from an elastic backscatter lidar (Weitkamp 2006), Ku-band radar reflectivity,
³³ and submillimeter wave radiometer brightness temperature measurements. The AOS mission
³⁴ design, while not necessarily optimal for cloud ice estimation, enables characterization of the
³⁵ broader spectrum of cloud and precipitation processes. Nevertheless, these measurements are
³⁶ complimentary for synergistic characterization of ice clouds. That is, despite the fact that lidar
³⁷ observations attenuate quickly in thick ice clouds and the Ku-band radar will not be able to detect
³⁸ clouds characterized by an echo weaker than 8.0 dBZ, the active observations are expected to
³⁹ provide context that may be incorporated into the radiometer retrievals. Herein, the term retrieval
⁴⁰ is defined as the process of estimating geophysical variables from remote sensing observations.
⁴¹ In this study, we investigate the impact of incorporating lidar and radar observations into the
⁴² radiometer retrieval of ice clouds. Because the existing amount of coincident backscatter lidar,
⁴³ Ku-band radar, and submillimeter-wave radiometer observations is rather insufficient to derive
⁴⁴ conclusive results, we employ accurate physical models to simulate lidar, radar and radiometer
⁴⁵ observations and use a cross-validation methodology to characterize the retrieval accuracy. As
⁴⁶ estimates from passive instrument observations strongly depend on "a priori" information (Rodgers
⁴⁷ 2000), for results to be relevant in real applications, it is necessary to base them on realistic vertical
⁴⁸ distributions of ice properties. Such distributions may be derived from cloud-resolving-model
⁴⁹ (CRM) simulations (Pfreundschuh et al. 2020; Liu and Mace 2022) or directly from observations.
⁵⁰ In this study, we employ the latter approach, as CRMs may still be deficient in properly reproducing
⁵¹ the vertical distribution of ice clouds and their associated microphysical properties. Specifically,
⁵² we use observations and products from the CloudSat (CS) mission (Stephens et al. 2002) to derive
⁵³ a database of ice microphysical properties and associated simulated lidar, radar, and radiometer
⁵⁴ observations. This database is used to investigate the accuracy of estimated ice cloud properties
⁵⁵ from the simulated observations. Another major difference relative to previous studies is the
⁵⁶ unique combination of instruments investigated herein. The article is organized as follows. In
⁵⁷ Section 2, we describe the approach used to derive the ice properties and the associated simulated

58 observations, the retrieval and the evaluation methodology. In Section 3, we present the results of
59 the evaluation methodology. We conclude in Section 4.

60 **2. Methodology**

61 As previously mentioned, we use CloudSat (CS) observations (Stephens et al. 2002) to derive
62 the vertical distributions of ice properties needed in the investigation. Although research quality
63 CS cloud ice products exist, to maximize the physical consistency of the approach, we do not use
64 them but derive ice amounts and associated properties directly from CS reflectivity observations.
65 This ensures the consistency between the particle distribution assumptions and the electromagnetic
66 scattering properties used in the CS reflectivity processing and those used the simulation of the
67 lidar, Ku-band radar and radiometer observations. Lidar, Ku-band radar, and submillimeter-
68 wave radiometer observations are simulated from CS observations using accurate physical models
69 and realistic assumptions consistent with the most recent knowledge in the field of ice cloud
70 microphysics, and a non-parametric estimation methodology based on the k-Means clustering
71 algorithm MacKay (2003) is used to investigate the instrument synergy. Details of the methodology
72 are presented below.

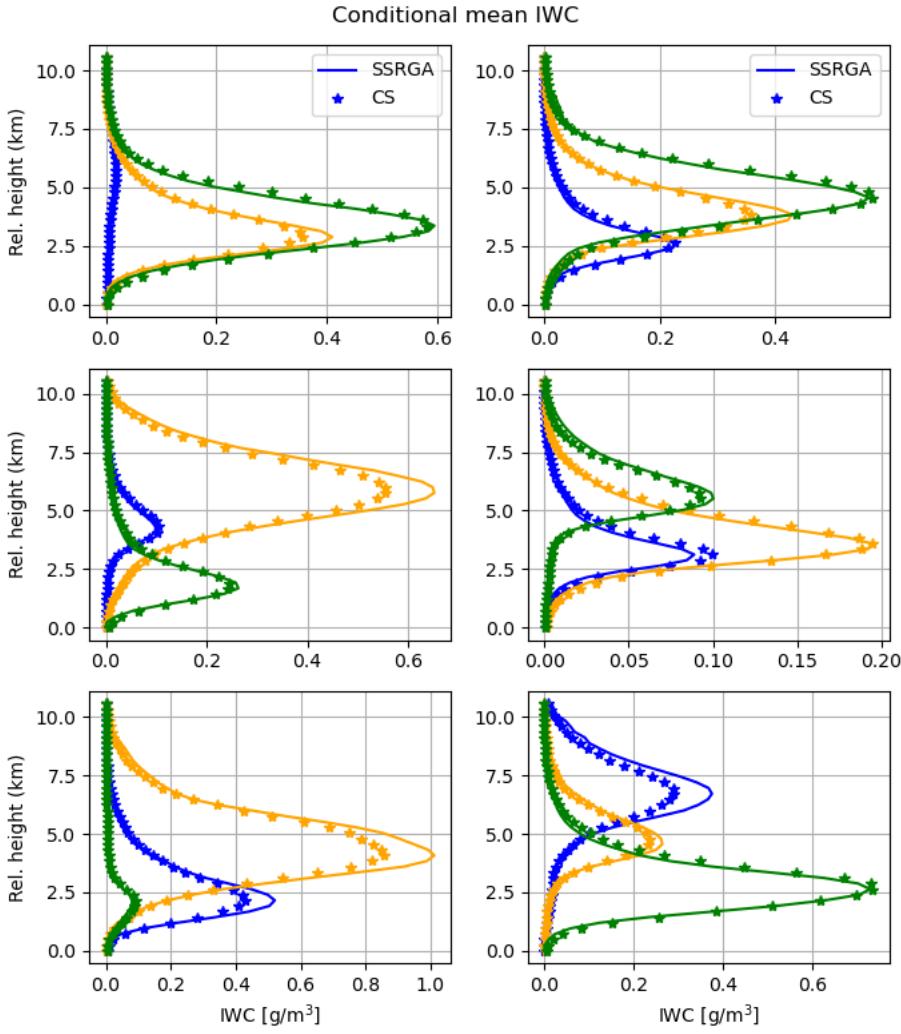
73 *a. Assumptions and forward models*

74 To quantify the number of ice particles in an elementary atmospheric volume as a function of
75 their size, we use normalized gamma functions (Bringi et al. 2003). The benefit of normalized
76 gamma functions is that they encapsulate the variability of Ice Water Content (IWC) - reflectivity
77 relationship into a single parameter, i.e. the normalized Particle Size Distribution (PSD) intercept
78 (Testud et al. 2001; Bringi et al. 2003). The normalized PSD intercept is defined as $N_w = \frac{4^4}{\pi \rho_w} \frac{IWC}{D_m^4}$,
79 where IWC is the ice water content associated with the PSD, and D_m is the mass weighted mean
80 diameter. Testud et al. (2001) showed that the variability in IWC reflectivity (Z) relationships may
81 be fully explained by variability in N_w , and that a formula of the type

$$IWC = N_w^{1-b} a Z^b \quad (1)$$

87 perfectly explains the relationships between IWC and Z calculated from observed PSDs. Equation
88 (1) is not sufficient to derive accurate, unbiased estimates of ice water contents, because N_w

varies considerably in time and space. Nevertheless, multiple studies showed that it is beneficial to parameterize N_w as a function of various variables, such as temperature (Hogan et al. 2006; Delanoë and Hogan 2008; Deng et al. 2010), rather than using N_w independent relations. In this study, we parameterize N_w as a function of temperature based on the CloudSat 2C-ICE product (Deng et al. 2010, 2013). Specifically, we cluster, based on similarity, a large set 2C-ICE profiles into 18 classes using a k-Means procedure. The mean IWC profiles associated with the 18 classes are shown in continuous lines in Figure 1. Alternative estimates, derived using PSD assumptions and electromagnetic scattering calculations that enable accurate and physically consistent simulations of radar observations at Ku-band and radiometer observations of submillimeter-wave frequencies are also shown in Figure 1. These estimates are based on the self-similarity Rayleigh-Gans approximation (SSRGA) of Hogan et al. (2017). Details regarding the estimation process are provided in the following paragraphs. As apparent in Figure 1, the CS and SSRGA estimates are in good agreement. Some discrepancies due to differences between the SSRGA N_w parameterization and the CS 2C-ICE "a priori assumptions" are also apparent, but they are not deemed critical in this study, whose objective is the investigation of synergistic lidar, Ku-band radar and submillimeter-wave radiometer retrievals, because the outcome is not likely to be sensitive to such details. One may notice that the average IWC profiles in Figure 1 are characterized by different peak values and heights. This facilitates a simple way to reverse-engineer to (some extent) the "a priori" assumptions used in the CS 2C-ICE product and use them in formulation of the type described in Equation (1). Specifically, the derivation of relationships of the type $IWC = a_i Z^{b_i}$ for every class i may be used to study a_i as a function of height. Shown in Figure 2 is a representation of the class multiplicative coefficient a_i as a function of relative height scatter plot. As apparent in the figure, and as expected, a_i exhibits a strong variation with the relative height. Coefficient b_i exhibits a height dependency as well (not shown), but the range of variation is significantly smaller, almost zero relative to the mean value of b . Given that any deviation of the multiplicative coefficient in an IWC-Z relation from an average is equivalent to a deviation of the associated N_w from its mean value (Testud et al. 2001), the variation of a as a function of relative-height may be converted into a N_w as a function of relative-height relationship. We, therefore, use the data in Fig. 2 to parameterize N_w as a function of the relative height.



82 FIG. 1. Mean CS IWC profiles for 18 classes derived using the k-Means clustering algorithm. For a
 83 compressed but intelligible representation, three classes are shown in different colors in each panel. The
 84 associated mean profiles derived from CS reflectivity observations using the SSRGA scattering calculations and
 85 N_w parameterization developed in this study are shown using symbol *. The vertical coordinate is defined
 86 relative to the freezing level.

118 For the determination of reference a and b values to be used with Equation (1), we assume that
 119 PSDs are normalized gamma distributions with $N_w = 0.08 \text{ cm}^{-4}$ and $\mu = 2$ and calculate

$$Z = \frac{\lambda^4}{\pi^5 |K_w|^2} \int_0^\infty N(D, D_m) \sigma_b(D) dD \quad (2)$$

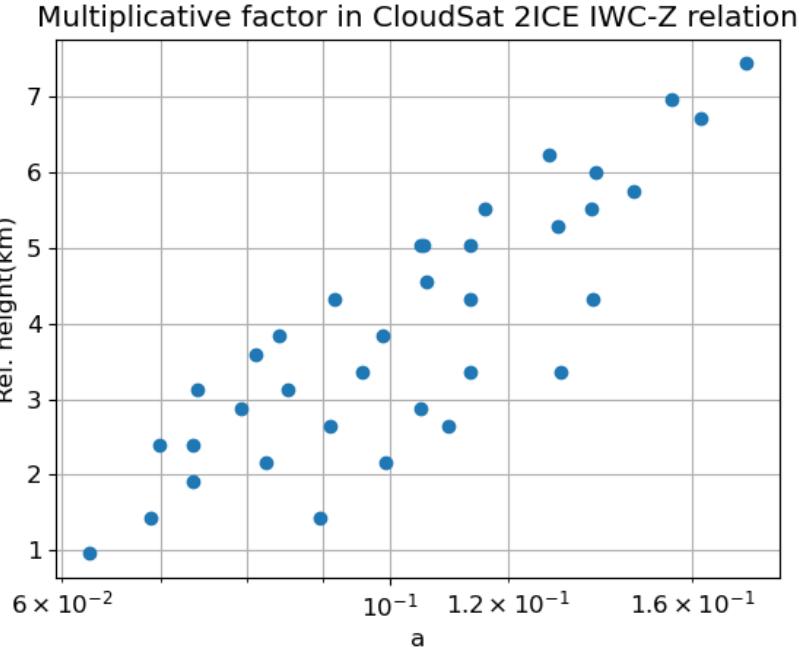
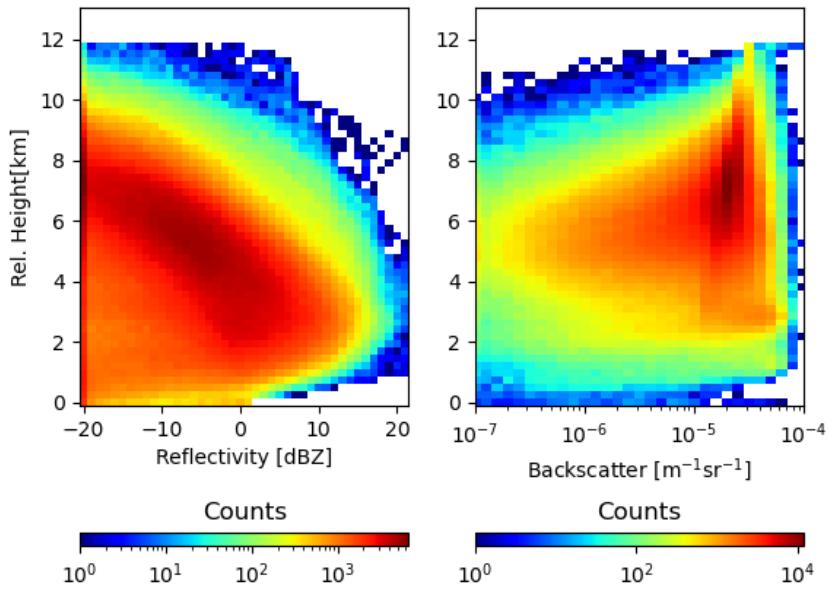


FIG. 2. Multiplicative factor in the 2ICE IWC-relation as a function of the IWC profile peak height.

120 where λ is the radar frequency, $|K_w|$ is the dielectric factor of water, $N(D, D_m)dD$ is the number of
 121 ice particles of diameter with D and $D+dD$ per unit volume, D_m is the mass weighted mean diameter
 122 of the distribution, and $\sigma_b(D)$ is the backscattering cross-section of ice particle of diameter D .
 123 The mass weighted mean diameter is equidistantly sampled to span the entire range of IWC values
 124 in the CS 2C-ICE dataset. The assumed mass-size relation is that of Brown and Francis (1995)
 125 because it works well with the SSRGA scattering calculations (Heymsfield et al. 2022). The open
 126 source software scatter-1.1 of (Hogan 2019) is used to provide the actual scattering properties.
 127 The SSRGA theory was developed for millimeter and submillimeter-wave calculations and may
 128 not be applicable at lidar's wavelength. Therefore, for lidar calculations, we use the Mie solution
 129 included in the scatter-1.1 package. Although more accurate calculations based on more realistic
 130 ice particle shapes exist, they are rather incomplete and not readily available. Moreover, Wagner
 131 and Delene (2022) compared lidar backscatter observations with backscatter calculations based on
 132 coincident PSD observations and the Mie solution and found good agreement, which suggests that
 133 electromagnetic properties derived from Mie calculations are adequate for practical applications.
 134 The lidar molecular backscatter and extinction are calculated at 532 nm using the lidar module of
 135 the CFMIP Observation Simulator Package (COSP; Bodas-Salcedo et al. (2011)). To account for



139 FIG. 3. Simulated distributions of Ku-band radar reflectivity (left) and lidar backscatter (right) as function of
 140 height above the freezing level.

136 multiple-scattering in the lidar observations, we are using the multiscatter-1.2.11 model (Hogan
 137 2015) of Hogan and Battaglia (2008). Shown in Figure 3 are the distributions of simulated Ku-band
 138 radar reflectivity and lidar baescatter as function of height above the freezing level.

141 The radiometer observations are calculated using a one-dimensional efficient, but accurate,
 142 radiative transfer solver based on Eddington's approximation (Kummerow 1993). The Eddington's
 143 approximation has been found to work well in cloud and precipitation retrieval application despite
 144 its simplicity relative to more general (but also computationally intensive) approaches such as the
 145 Monte Carlo radiative transfer solvers (Liu et al. 1996) It should be noted tough that the phase
 146 functions of ice particles tend to be highly asymmetric at sub-millimeter wave frequencies. For
 147 radiative transfer solutions based on the Eddington's approximation to be accurate it is necessary that
 148 the delta-scaling approach (Joseph et al. 1976) be employed. The delta-scaling approach transforms
 149 the initial radiative transfer equation into an equivalent one characterized by a less asymmetric
 150 scattering function and more extinction, which makes the solution Eddington approximation more
 151 stable and accurate. The absorption due to water vapor and other gases is quantified using the

152 Rosenkranz model (Rosenkranz 1998). The water vapor, temperature, and pressure distributions
153 are derived based on a WRF simulation of summer convection over the United States. Specifically,
154 the water vapor, temperature, and pressure profiles associated with times and areas where the
155 model produces anvils are selected and clustered into 40 classes using the k-Means approach. The
156 mean extinction profiles at the radiometer frequencies are calculated for every class and used in
157 process of calculating the brightness temperatures from the estimated ice profiles using a simple
158 Monte Carlo procedure. That is, given a retrieved ice profile and its scattering property, an
159 anvil class and its associated absorption, temperature, and pressure profiles are randomly selected
160 and attached to the ice scattering properties. To make the procedure physically meaningful,
161 temperature rather than height is used in the ice scattering-gas absorption collocation process. The
162 emissivities are randomly chosen between 0.8 and 1.0 and assumed constant for all radiometer
163 frequencies. Brightness temperatures are calculated at 89-, 183.31 ± 1.1 , and 325.15 ± 1.5 GHz,
164 which correspond to three of the 10 channels of the SAPHIR-NG radiometer envisioned to be
165 deployed in the AOS mission (Brogniez et al. 2022). The other channels are centered on the
166 same water vapor absorption lines and are not likely to offer additional information in this rather
167 controlled experiment. Nevertheless, the other channels are expected to be useful in reducing the
168 uncertainties caused by variability in the vertical distribution of water vapor, which may be greater
169 in real life than in the simulated environment.

170 The processing steps used to process the CS reflectivity observations and calculate the lidar,
171 Ku-band and submillimeter-wave radiometer observations may be summarized as follows:

- 172 1. Derivation of physically consistent radar and radiometer lookup tables to relate basic radar
173 and radiometer properties (e.g. reflectivity, attenuation, extinction, scattering-albedo, etc.) to
174 PSD parameters such as IWC and D_m . The tables are derived for a single of N_w , but are
175 usable with any value of N_w using the "normalization" operations described in (Grecu et al.
176 2011).
- 177 2. Derivation of N_w -relative height parameterization using the 2C-ICE product.
- 178 3. Estimation of IWC and related PSD parameters from CS W-band radar observations, using
179 the tables constructed in Step 1 and parameterization derived in Step 2.

180 4. Calculation of lidar, Ku-band radar and radiometer observations from the estimates derived
181 in Step 3 and the tables obtained in Step 1.

182 The application of these steps produces a large dataset of approximately 200,000 cloud ice
183 profiles and associated lidar, radar and radiometer observations that may be used to investigate the
184 synergy of the three sensors. Details are provided in the next section.

185 *b. Estimation and evaluation*

186 Given that the lidar observations may attenuate quickly in thick clouds, while the Ku-band radar
187 will not detect clouds with an echo weaker than 8.0 dBZ, the radiometer is the instrument likely
188 to provide by itself the most complete information about the total amount of ice in its observing
189 volume. However, the vertical distribution of ice is difficult to quantify from radiometer-only
190 observations, because significantly different ice vertical distributions may lead to very similar
191 radiometer observations. This makes radiometer-only retrievals highly dependent on the "a priori"
192 information on the distribution of ice clouds in the atmosphere. As previously mentioned, this
193 is the reason why CS-based IWC retrievals were preferred to CRM simulations, as retrievals are
194 expected to result in more natural and less biased distributions.

195 We employ a two-step estimation methodology similar to that of Grecu et al. (2018). In the
196 first step, we estimate the IWC class, out of the 18 classes of shown in Figure 1, to which the
197 estimated IWC profile is most likely to belong. The class is estimated directly from the synthetic
198 observations. In the second step, we estimate the IWC profile, using a class specific ensemble
199 Kalman Smoother (EKS) methodology similar to that of Grecu et al. (2018). The EKS algorithm
200 updates the estimated IWC relative to the mean IWC of the class to which the profile belongs. The
201 differences between the actual active and passive observations and their mean class values are used
202 in the update. The second step of this procedure is formally identical to the one used in Grecu
203 et al. (2018), but the first step is different. In Grecu et al. (2018), the first step was based on a
204 simple distance-based evaluation. That strategy is likely to be suboptimal in this study, because the
205 joint distribution of IWC profiles and associated observations are significantly more complex. We
206 therefore use a more complex classification methodology based on the TensorFlow library (Abadi
207 et al. 2016). The class estimation model is defined as a TensorFlow Model with two dense layers
208 of 30 neurons each, followed by a softmax layer (Goodfellow et al. 2016). The class estimation

209 model is trained using the 70% of the simulated observations and the corresponding IWC profiles,
210 the remaining 30% of the data being used for evaluation.

$$\mathbf{X} = \bar{\mathbf{X}}_i + \mathbf{Cov}(\mathbf{X}_i, \mathbf{Y}_i) \mathbf{Cov}(\mathbf{Y}_i, \mathbf{Y}_i)^{-1} (\mathbf{Y} - \bar{\mathbf{Y}}_i) \quad (3)$$

211 where \mathbf{X} is the state variable describing the IWC profile, \mathbf{Y} is the vector containing the variation, \mathbf{X}_i is
212 the set of state variables for profiles in class i , and \mathbf{Y}_i is the set of associated observations. Variables
213 $\bar{\mathbf{X}}_i$ and $\bar{\mathbf{Y}}_i$ are the mean values of the state variables and observations in class i , respectively. The
214 covariance matrices between \mathbf{X}_i and \mathbf{Y}_i are denoted by $\mathbf{Cov}(\mathbf{X}_i, \mathbf{Y}_i)$. In step 1, the class is estimated
215 using the TensorFlow model, while in step 2, the IWC profile is estimated using the EKS algorithm
216 summarized in Equation 3.

217 As already mentioned, a cross-validation methodology is used for evaluation, with 70% of the
218 data used for training and the remaining 30% of the data used for validation. The partition of
219 the data into training and evaluation subsets is done randomly. Usually, the partition, training
220 and evaluation steps are repeated several times. However, given the fact that differences in the
221 relationships between the ice property and their associated simulated observations are functions of
222 the meteorological context, and that all regimes are well-sampled in both the training and testing
223 subsets (e.g. out of every 10 pixels in a scene, about 7 end-up in the training dataset, while the others
224 in the testing dataset), the repetition of the partition, training, and evaluation steps multiple times
225 is not necessary. Therefore, in our evaluation, we partition the data into training and evaluation
226 only once and perform all the evaluation for a single partition. The evaluation criteria include the
227 correlation coefficient, the bias, and visual inspections of graphical representations of the estimated
228 properties relative to their references.

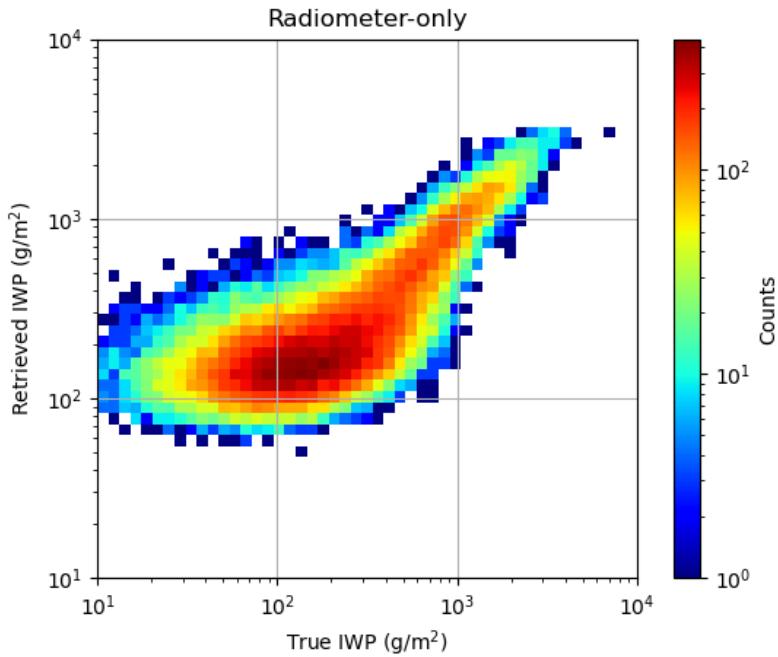
229 3. Results

230 a. Radiometer-only retrievals

233 As previously mentioned, submillimeter-wave radiometers are likely to provide by themselves
234 more complete information about the total amount of ice in their observing volumes than lidars
235 or Ku-band radars with limited sensitivity. However, radiometers observations are an integrated
236 measure of radiative process in clouds that provide little information about the vertical distribution

237 of ice. From this perspective, an evaluation in terms of the ice water path (*IWP*) defined as the
 238 vertical integral of the *IWC*, i.e. $IWP = \int_0^{Z_{top}} IWC(z)dz$ is insightful. Shown in Figure 4 is the
 239 frequency of *IWP* estimated from radiometer-only observations as a function of its true value. As
 240 apparent in the figure, there is good correlation between the retrieved and the true *IWP* values.
 241 The numerical value of the correlation coefficient is 0.92, and there is no-overall bias. That is, the
 242 mean values of retrieved *IWP* and true *IWP* values are equal. However, conditional biases are
 243 apparent, with overestimation of *IWP* for values smaller than 100 g/m^2 and some underestimation
 244 for values larger than 1000 g/m^2 . The biases at the low end of the *IWP* range are not surprising,
 245 given that the impact caused by ice scattering on the total radiometric signal is small for low
 246 values of *IWP* and hard to distinguish from other sources of variability in radiometer observations.
 247 Saturation effects are most likely responsible for underestimation at the high end. It should be
 248 noted that in this evaluation, only atmospheric profiles that exhibit ice detectable by the CS radar
 249 are used. Therefore, a radiometer-only estimation procedure derived from this training dataset
 250 is likely to result in significant overestimation if not used in conjunction with a discrimination
 251 procedure. However, such procedure is not critical in this study, as the lidar observations may be
 252 used to discriminate between clear skies and ice clouds. Although the radiometer-only estimation
 253 procedure is able to estimate the integrated amount of ice in clouds fairly well, its ability to
 254 characterize the vertical distribution of ice in clouds is limited. Figure 5 shows the conditional
 255 vertical distributions of the estimated and true *IWC* for the 18 classes described in Section 2a and
 256 shown in Figure 1. As apparent in the figure, there are significant differences between the estimated
 257 and true *IWC* profiles.

258 Further insight into the radiometer-only estimation performance may be derived by defining the
 259 ice profile gravity center (GC) as $z_{GC} = \frac{\int_0^{Z_{top}} zIWC(z)dz}{\int_0^{Z_{top}} IWC(z)dz}$, where z is the distance relative to the
 260 freezing level, the Z_{top} is the distance from the top of the atmosphere to the freezing level. Shown
 261 in Figure 6 is the frequency of *IWC* gravity center estimated from radiometer-only observations
 262 as a function of its true value. It may be observed in the figure that while the true *IWC* gravity
 263 center exhibits quite a broad distribution, the one retrieved from the radiometer-only observations
 264 exhibits a multimodal narrow distribution. Moreover, there is almost no correlation between the
 265 retrieved and the true *IWC* gravity center. This is another indication that, while the total amount



231 FIG. 4. Frequency plot of estimated IWP derived radiometer-observations as a function of the true IWP used
 232 in observations synthesis.

266 of ice may be reasonably estimated from radiometer-only observations, its vertical distribution can
 267 not be determined from radiometer-only observations.

268 *b. Synergistic retrievals*

269 The synergy of the instrument on the estimates may be investigating by simply incorporating lidar
 270 and radar observations into the retrieval process and comparing the results with the radiometer-
 271 only estimates. Although the lidar observations are subject to attenuation, they are able to provide
 272 information about the vertical distribution of ice in clouds, especially at the top of the clouds. The
 273 radar observations, on the other hand, are able to provide information in the bottom part of the
 274 clouds, where the lidar signal is below the noise level due to attenuation. Therefore, the combined
 275 use of lidar and radar observations is expected to provide a more complete characterization of
 276 the vertical distribution of ice in clouds and enable the derivation of more specific estimates than
 277 those derived from radiometer-only observations.

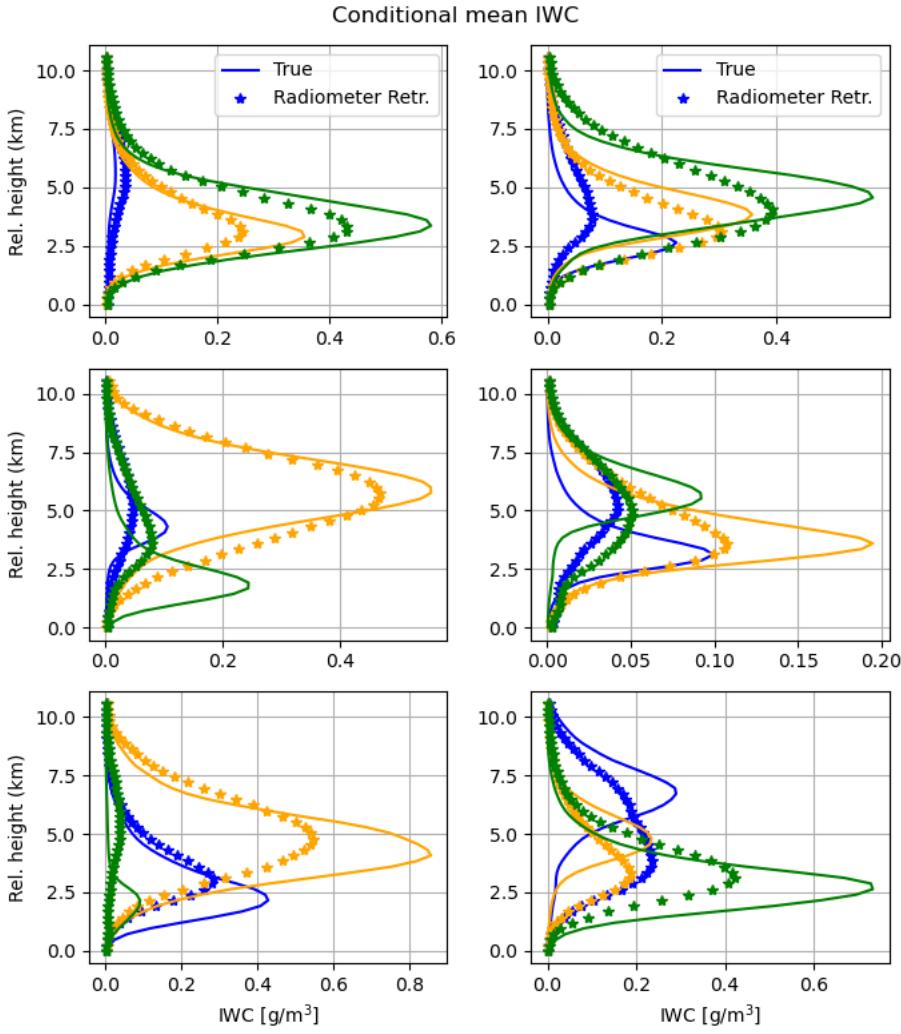


FIG. 5. True and radiometer-only retrieved conditional mean IWC for the 18 classes described in Figure 1.

Shown in Figure 7 is the distribution of the synergistic IWP estimates as a function of their true values. As apparent in the figure, the synergistic IWP estimates are more accurate than the radiometer-only estimates. At the same time, as apparent in Figure 8, the retrieved conditional mean IWC for the 18 classes described section 2a and shown in Figure 1 are in significantly better (almost perfect) agreement with the true IWC profiles than those derived from radiometer-only observations. Furthermore, as seen in Figure 8 the synergistic IWC gravity center estimates are in very good agreement with the true IWC gravity center.

While the estimates based on all instruments are significantly more accurate than those based on radiometer-only observations, it is useful to investigate how the two active instruments (lidar and

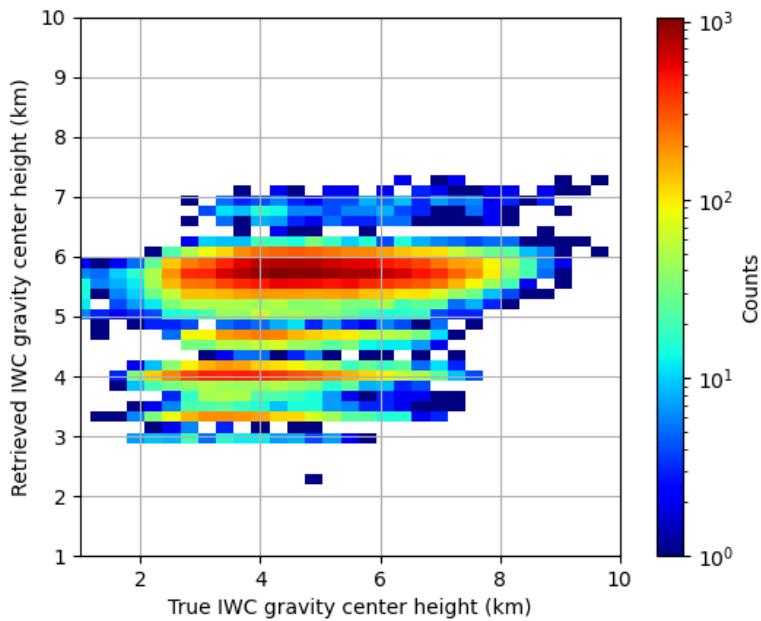


FIG. 6. Same as in Figure 4, but for the *IWC* gravity center.

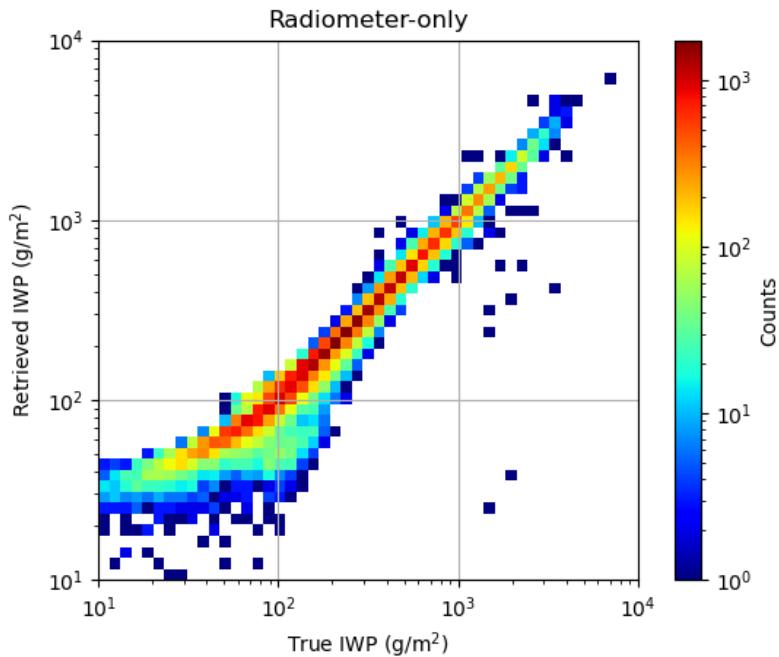


FIG. 7. Same as in Figure 4, but with the lidar and radar observations incorporated in the retrievals.

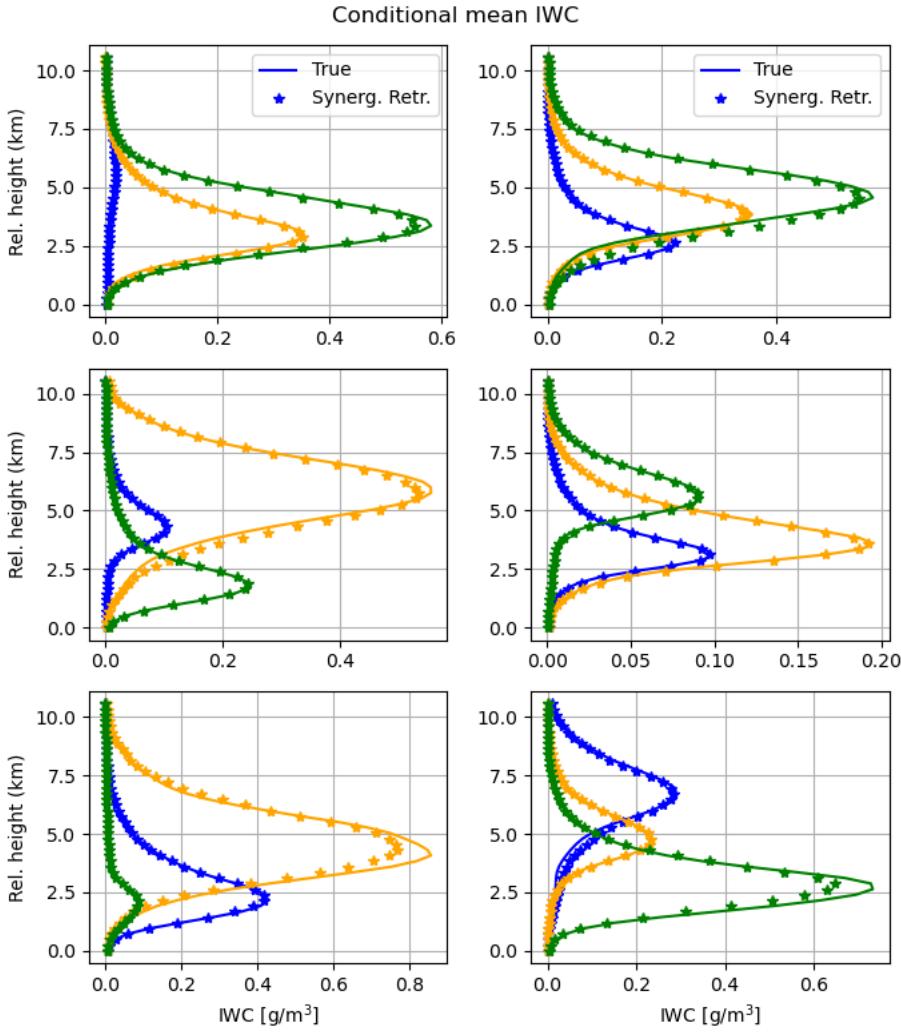


FIG. 8. Same as in Figure 5, but with the lidar and radar observations incorporated in the retrievals.

radar) impact the estimates. For conciseness, we used two statistical scores, namely, the normalized root mean square (NRMS) and the classification accuracy, to summarize the performance of the estimates. The NRMS is defined as

$$NRMS = \frac{\sqrt{\frac{\sum_{i=1}^N (IWC_i - IWC_{true,i})^2}{N}}}{\sqrt{\frac{\sum_{i=1}^N (IWC_{true,i} - \bar{IWC})^2}{N}}} \quad (4)$$

where IWC_i is the estimated IWP for the i -th sample, $IWC_{true,i}$ is the true IWC for the i -th sample, \bar{IWC} is the IWC mean, and N is the size of the estimation dataset. The classification accuracy is

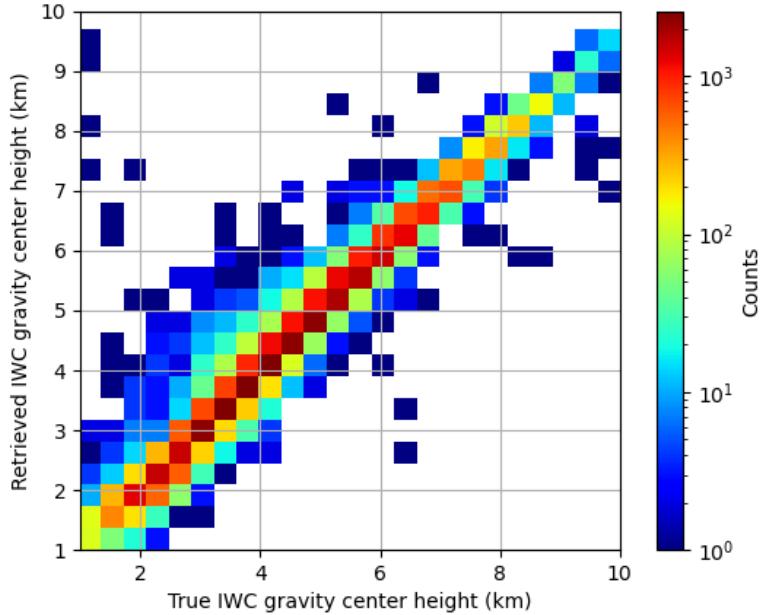


FIG. 9. Same as in Figure 6, but with the lidar and radar observations incorporated in the retrievals.

292 defined as

$$CA = \frac{\sum_{i=1}^N \delta_i}{N} \quad (5)$$

293 where δ_i is a binary variable that is equal to 1 if the estimated IWC class for the i -th sample is
294 equal to the true IWC class for the i -th sample, and 0 otherwise. The performance summary is
295 shown in Table 1 for several combinations of instruments. It may be observed in the table that the
296 performance of the estimates based on all instruments is significantly better than those based on
297 radiometer-only observations. Furthermore, the inclusion of the lidar observations in the retrieval
298 process has a larger impact on the retrieval performance than the inclusion of the radar observations.
299 This is expected since the lidar observations are able to provide information about the top of the
300 clouds, where the radar observations are above the noise level only occasionally. Nevertheless, the
301 inclusion of the radar observations in the retrieval process has a notable impact on the accuracy of
302 the IWC estimates relative to radiometer-only retrievals.

TABLE 1. Performance summary.

Score \ Instruments	Radiometer	Radar-Radiometer	Lidar-Radiometer	Radar-Lidar-Radiometer
NRMS	0.73	0.59	0.32	0.22
Class. Accuracy	0.39	0.48	0.92	0.94

303 4. Conclusions

304 In this study, we investigate the synergy of lidar, Ku-band radar, and sub-millimeter-wave
 305 radiometer measurements in the retrieval of the ice from satellite observations. The synergy is
 306 analyzed through the generation of a large dataset of IWC profile and the calculation of lidar, radar
 307 and radiometer observations using realistic models. The characteristics of the instruments (e.g.
 308 frequencies, sensitivities, etc.) are set based on the expected characteristics of instruments of the
 309 AOS mission. A cross-validation methodology is used to assess the accuracy of the retrieved IWC
 310 profiles from various combinations of observations from the three instruments. Specifically, the
 311 IWC and associated observations is randomly divided into two datasets, one for the training and
 312 the other for the evaluation. The training dataset is used to train the retrieval algorithm, while the
 313 evaluation dataset is used to assess the retrieval performance.

314 To ensure the self-consistency of results and their relevance to practical applications, the dataset
 315 of IWC profiles is derived from CloudSat reflectivity observations. Although subject to potential
 316 biases and uncertainties due to deficiencies in the retrieval models, these profiles are deemed to
 317 be more realistic than those derived from cloud resolving model simulations. Moreover, they
 318 are roughly consistent with the 2C-ICE CloudSat product (Deng et al. 2015), while relying on
 319 assumptions and parameterizations that enable the accurate computation of backscatter lidar, Ku-
 320 band radar, and sub-millimeter-wave radiometer observations.

321 The retrieval of the ice water content (IWC) profiles from the computed observations is achieved
 322 in two steps. In the first step, a class, out of 18 potential classes characterized by different vertical
 323 distribution of IWC, is estimated from the observations. The 18 classes are predetermined based on
 324 k-Means clustering algorithm. In the second step, the IWC profile is estimated using an Ensemble
 325 Kalman Smoother (EKS) algorithm that uses the estimated class as a priori information.

326 The results of the study show that the synergy of lidar, radar, and radiometer observations is
 327 significant in the retrieval of the IWC profiles. The inclusion of the lidar observations in the retrieval

328 process has a larger impact on the retrieval performance than the inclusion of the radar observations.
329 Although results are not directly comparable to those from other studies (Pfreundschuh et al.
330 2020; Liu and Mace 2022), given the differences between the instruments in this study relative
331 to those from other studies, it may be concluded that they are not inconsistent with previous
332 studies. Specifically, previous studies show some skills in radiometer-only retrievals and significant
333 improvements in the retrieval performance when the active observations are incorporated. From
334 this perspective, our findings are consistent with previous studies.

335 Further work is necessary out to assess the impact of sources of uncertainties such as variability
336 in the PSD intercept not captured by the current parameterization, differences in the instruments'
337 footprint sizes, and non-uniform beam filling on the retrievals of the IWC profiles. Other sources of
338 uncertainties that need be considered include the potential existence of supercooled liquid water in
339 the clouds and uncertainties in the electromagnetic scattering properties used in the in instruments'
340 forward models.

341 *Acknowledgments.* This work was supported by the NASA Remote Sensing Theory project
342 through Grant 80NSSC20K1729. The authors thank Dr. Lucia Tsaoussi (NASA Headquarters) for
343 her support of this effort.

344 *Data availability statement.* The CloudSat data can be accessed at:
345 <https://www.cloudsat.cira.colostate.edu/>.

346 References

- 347 Abadi, M., and Coauthors, 2016: Tensorflow: a system for large-scale machine learning. *Osdi*,
348 Savannah, GA, USA, Vol. 16, 265–283.
- 349 Bodas-Salcedo, A., and Coauthors, 2011: Cosp: Satellite simulation software for model assess-
350 ment. *Bulletin of the American Meteorological Society*, **92** (8), 1023–1043.
- 351 Bringi, V., V. Chandrasekar, J. Hubbert, E. Gorgucci, W. Randeu, and M. Schoenhuber, 2003: Raindrop size distribution in different climatic regimes from disdrometer and dual-polarized
352 radar analysis. *Journal of the atmospheric sciences*, **60** (2), 354–365.
- 354 Brogniez, H., and Coauthors, 2022: Time-delayed tandem microwave observations of tropical
355 deep convection: Overview of the c 2 omodo mission. *Frontiers in Remote Sensing*, **3**, 854 735.
- 356 Brown, P. R., and P. N. Francis, 1995: Improved measurements of the ice water content in cirrus
357 using a total-water probe. *Journal of Atmospheric and Oceanic Technology*, **12** (2), 410–414.
- 358 Delanoë, J., and R. J. Hogan, 2008: A variational scheme for retrieving ice cloud properties from
359 combined radar, lidar, and infrared radiometer. *Journal of Geophysical Research: Atmospheres*,
360 **113** (D7).
- 361 Deng, M., G. G. Mace, Z. Wang, and E. Berry, 2015: Cloudsat 2c-ice product update with a new ze
362 parameterization in lidar-only region. *Journal of Geophysical Research: Atmospheres*, **120** (23),
363 12–198.
- 364 Deng, M., G. G. Mace, Z. Wang, and R. P. Lawson, 2013: Evaluation of several a-train ice cloud
365 retrieval products with in situ measurements collected during the sparticus campaign. *Journal*
366 *of applied meteorology and climatology*, **52** (4), 1014–1030.

- 367 Deng, M., G. G. Mace, Z. Wang, and H. Okamoto, 2010: Tropical composition, cloud and climate
368 coupling experiment validation for cirrus cloud profiling retrieval using cloudsat radar and
369 calipso lidar. *Journal of Geophysical Research: Atmospheres*, **115** (D10).
- 370 Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep learning*. MIT press.
- 371 Grecu, M., L. Tian, G. M. Heymsfield, A. Tokay, W. S. Olson, A. J. Heymsfield, and A. Bansemer,
372 2018: Nonparametric methodology to estimate precipitating ice from multiple-frequency radar
373 reflectivity observations. *Journal of Applied Meteorology and Climatology*, **57** (11), 2605–2622.
- 374 Grecu, M., L. Tian, W. S. Olson, and S. Tanelli, 2011: A robust dual-frequency radar profiling
375 algorithm. *Journal of applied meteorology and climatology*, **50** (7), 1543–1557.
- 376 Heymsfield, A., A. Bansemer, G. Heymsfield, D. Noone, M. Grecu, and D. Toohey, 2022: Re-
377 lationship of multiwavelength radar measurements to ice microphysics from the impacts field
378 program. *Journal of Applied Meteorology and Climatology*.
- 379 Hogan, R. J., 2019: scatter-1.1. URL <http://www.met.reading.ac.uk/clouds/ssrga/scatter-1.1.tar.gz>.
- 380 Hogan, R. J., and A. Battaglia, 2008: Fast lidar and radar multiple-scattering models. part ii: Wide-
381 angle scattering using the time-dependent two-stream approximation. *Journal of the Atmospheric
382 Sciences*, **65** (12), 3636–3651.
- 383 Hogan, R. J., R. Honeyager, J. Tyynelä, and S. Kneifel, 2017: Calculating the millimetre-wave
384 scattering phase function of snowflakes using the self-similar rayleigh–gans approximation.
385 *Quarterly Journal of the Royal Meteorological Society*, **143** (703), 834–844.
- 386 Hogan, R. J., M. P. Mittermaier, and A. J. Illingworth, 2006: The retrieval of ice water content from
387 radar reflectivity factor and temperature and its use in evaluating a mesoscale model. *Journal of
388 Applied Meteorology and Climatology*, **45** (2), 301–317.
- 389 Joseph, J. H., W. Wiscombe, and J. Weinman, 1976: The delta-eddington approximation for
390 radiative flux transfer. *Journal of Atmospheric Sciences*, **33** (12), 2452–2459.
- 391 Kummerow, C., 1993: On the accuracy of the eddington approximation for radiative transfer in the
392 microwave frequencies. *Journal of Geophysical Research: Atmospheres*, **98** (D2), 2757–2765.

- 393 Liu, Q., C. Simmer, and E. Ruprecht, 1996: Three-dimensional radiative transfer effects of clouds
394 in the microwave spectral range. *Journal of Geophysical Research: Atmospheres*, **101** (D2),
395 4289–4298.
- 396 Liu, Y., and G. G. Mace, 2022: Assessing synergistic radar and radiometer capability in retrieving
397 ice cloud microphysics based on hybrid bayesian algorithms. *Atmospheric Measurement
398 Techniques*, **15** (4), 927–944.
- 399 MacKay, D. J., 2003: *Information theory, inference and learning algorithms*. Cambridge university
400 press.
- 401 Pfreundschuh, S., P. Eriksson, S. A. Buehler, M. Brath, D. Duncan, R. Larsson, and R. Ekelund,
402 2020: Synergistic radar and radiometer retrievals of ice hydrometeors. *Atmospheric Measure-
403 ment Techniques*, **13** (8), 4219–4245.
- 404 Rodgers, C. D., 2000: *Inverse methods for atmospheric sounding: theory and practice*, Vol. 2.
405 World scientific.
- 406 Rosenkranz, P. W., 1998: Water vapor microwave continuum absorption: A comparison of mea-
407 surements and models. *Radio Science*, **33** (4), 919–928.
- 408 Stephens, G. L., and Coauthors, 2002: The cloudsat mission and the a-train: A new dimension of
409 space-based observations of clouds and precipitation. *Bulletin of the American Meteorological
410 Society*, **83** (12), 1771–1790.
- 411 Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of “normalized”
412 distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing.
413 *Journal of Applied Meteorology and Climatology*, **40** (6), 1118–1140.
- 414 Wagner, S. W., and D. J. Delene, 2022: Technique for comparison of backscatter coefficients
415 derived from in situ cloud probe measurements with concurrent airborne lidar. *Atmospheric
416 Measurement Techniques*, **15** (21), 6447–6466.
- 417 Weitkamp, C., 2006: *Lidar: range-resolved optical remote sensing of the atmosphere*, Vol. 102.
418 Springer Science & Business.