Assessment of two stochastic cloud subcolumn generators using observed fields of vertically resolved cloud extinction

Lazaros Oreopoulos,[a] Nayeong Cho,[a,b] Dongmin Lee,[a,c] Matthew Lebsock,[d] Zhibo Zhang[e,b]

[a] *Earth Sciences Division, NASA's Goddard Space Flight Center, Greenbelt MD*

[b] *Goddard Earth Sciences Technology and Research II, UMBC, Baltimore, MD*

[c] *Morgan State University, Baltimore, MD*

[d] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA*

[e] *Department of Physics, University of Maryland Baltimore County, Baltimore, MD*

Corresponding author: Lazaros Oreopoulos, Lazaros.Oreopoulos@nasa.gov

ABSTRACT

We evaluate two stochastic subcolumn generators used in GCMs to emulate subgrid cloud variability enabling comparisons with satellite observations and simulations of certain physical processes. Our evaluation necessitated the creation of a reference observational dataset that resolves horizontal and vertical cloud variability. The dataset combines two CloudSat cloud products that resolve two-dimensional cloud optical depth variability of liquid, ice, and mixed phase clouds when blended at ~200 m vertical and ~ 2 km horizontal scales. Upon segmenting the dataset to individual "scenes", mean profiles of the cloud fields are passed as input to generators that produce scene-level cloud subgrid variability. The assessment of generator performance at the scale of individual scenes and in a mean sense is largely based on inferred joint histograms that partition cloud fraction within predetermined combinations of cloud top pressure – cloud optical thickness ranges. Our main finding is that both generators tend to underestimate optically thin clouds, while one of them also tends to overestimate some cloud types of moderate and high optical thickness. Associated radiative flux errors are also calculated by applying a simple transformation to the cloud fraction histogram errors, and are found to approach values almost as high as 3 W m^{-2} for the cloud radiative effect in the shortwave part of the spectrum.

SIGNIFICANCE STATEMENT

The purpose of the paper is to assess the realism of relatively simple ways of producing fine-scale cloud variability in global models from coarsely-resolved cloud properties. The assessment is achieved via comparisons to observed cloud fields where the fine-scale variability is known in both the horizontal and vertical directions. Our results show that while the generators have considerable skill, they still suffer from consistent deficiencies that need to be addressed with further development guided by appropriate observations.

1. Introduction

Satellite simulators are important tools of modern GCMs. Subcolumn cloud generators are in turn critical components of satellite simulators striving to emulate subgrid-scale cloud properties in order to bridge the coarse resolution of these models and the scales at which satellite retrievals are performed. Emulation of subgrid variability is not only required for mimicking satellite observations, but also contributes to improved calculations of quantities with non-linear dependences on cloud properties, such as radiative flux and precipitation rate (Song et al. 2018). For some radiation schemes such as the Monte Carlo Independent Column Approximation (McICA, Pincus et al. 2003), the coupling of the radiative transfer algorithm with cloud fields resolved at subgrid scales is actually an inherent feature of their design.

Since simple profiles of averaged quantities at larger scales do not by themselves fully constrain the distribution of profiles at smaller scales, there is considerable amount of freedom (and uncertainty) in simulating the more detailed satellite views of clouds. The emulated subgrid variability depends on how cloud microphysical properties are distributed both horizontally and in height, and how cloud occurrence and the distributions of microphysical properties overlap vertically. Within the satellite simulator framework, the tool that handles subgrid variability is a cloud subcolumn generator. This tool produces stochastic samples of subcolumn cloud profiles, which preserve the average profiles of the gridbox in the limit of a large number of samples. An essential ingredient of subcolumn generators is a set of rules on how the cloudy parts of a gridbox overlap vertically (Jakob and Klein, 1999). The combination of cloud horizontal variability and how it correlates between various atmospheric levels produces the subgrid variability of total (vertically integrated) cloud water path and column cloud optical depth TAU. Note that subgrid variability of vertically integrated (column) cloud properties can exist even if the properties themselves at individual levels are distributed homogeneously; this is because the rules of cloud occurrence overlap create variability in the number of layers that are cloudy in each subcolumn.

The subcolumn generator SCOPS (Subgrid Cloud Overlap Profile Sampler) of the CFMIP (Cloud Feedback Modeling Intercomparison Project) Operational Satellite Simulator Package (COSP, Bodas-Salcedo et al. 2011) assumes that vertical cloud occurrence follows a combination of maximum and random overlap (Hillman et al. 2018) and that a cloudy layer's condensate is homogeneous at the scale of model (GCM) gridboxes. While SCOPS subcolumns are generally not passed outside of COSP, its underlying overlap assumptions have been proven inadequate for the simulation of accurate cloud-sky radiative fluxes when implemented in radiation schemes (Barker et al., 1999; Oreopoulos et al., 2012). A first step then towards improvement of radiation from subcolumn generators would therefore be direct statistical comparison of their cloud fields with observed subgrid cloud variability. Note that horizontal coherence, i.e., the spatial arrangement of subcolumns does not need to be part of the subgrid variability description and comparison since horizontal interactions within a model gridbox are typically neglected.

Because subcolumn generators also emulate vertical variability, reference data appropriate for generator validation must provide such information as well. An approach embraced previously was to use simulated cloud fields more highly resolved than in GCMs as proxy for observations, for example cloud fields from Multiscale Modeling Framework (MMF) simulations (Hillman et al. 2018). In this treatment, model gridcolumns are treated as observed subcolumns and serve as “truth” in comparisons with subcolumns produced by SCOPS or other generators. The usefulness of such an evaluation is obviously limited by the degree of realism of MMF clouds. Moving beyond previous work, here we make the case that simulated cloud fields are not the only recourse, but rather that observed cloud fields exist as a reference source for subgrid cloud variability for the purposes of subcolumn generator validation and development. Such cloud fields come from combined observations of the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) lidar aboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite and its near-synchronous CloudSat satellite carrying the Cloud Profiling Radar (CPR). The main feature of the cloud fields retrieved from these two instruments is that the variability is resolved horizontally not only for the atmospheric column as a whole, but also vertically within the column.

Our paper thus uses such a reference dataset in an attempt to evaluate SCOPS and another established subcolumn generator by Räisänen et al. (2004). It consists of two main parts: The first part is dedicated to describing the construction of the observational reference dataset, while the second part presents the approach and findings of our effort to assess the performance of these two generators.

2. Quality assessment and improvement of reference dataset

a. Key Cloud Products

Our construction of 2D cloud fields resolving horizontal (along the path of the two observing satellites) and vertical (height) variability of visible cloud optical depth (COD) is achieved by combining two CloudSat “release 5” (R05) products: 2B-CWC-RVOD (Leinonen et al. 2016, hereafter 2BCWC) and the CALIPSO-enhanced 2C-ICE product

(Deng et al. 2015, hereafter 2CICE). To begin with, some quick clarifications on terminology are in order before we proceed. Herein we use COD for "cell" optical depth and TAU for the integrated/column optical depth, i.e., the vertical sum of CODs. In CloudSat terminology CPR measurements are resolved vertically in 240 m "bins" and horizontally in 1.7 km (along track) "rays". A bin therefore represents a vertical layer, a ray essentially represents a subcolumn, and a "cell" is a bin within a subcolumn. We refrain from using the term "bin" for vertical layers because it is also used for histogram discretization.

1) LIQUID PHASE: 2B-CWC-RVOD (2BCWC)

This product focuses on clouds of liquid phase, with key inputs to the algorithm being the CloudSat 2B-GEOPROF radar reflectivity profile (Marchand et al., 2008) and the column cloud optical depth TAU from the Collection 6 level 2 Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) cloud product (MYD06) (Platnick et al., 2017), which has been collocated with CloudSat CPR measurements and is available in CloudSat's MOD06-1KM-AUX product. The incorporation of MODIS TAU (TAU_{MODIS}) as a constraint in the retrievals makes this a purely daytime product. The algorithm is based on the optimal estimation framework (Rodgers 2000), with the measurement vector consisting of the logarithm of the total column liquid optical depth TAU_{liq} and the profile of all valid CPR reflectivities. To infer TAU_{liq}, the algorithm uses ice cloud optical depth provided by the 2CICE product (Deng et al. 2015, see below) which is subtracted from MODIS total optical depth, i.e., $TAU_{liq} = TAU_{MODIS} – TAU_{2CICE}$. Ancillary temperature estimates from the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis provided as CloudSat's ECMWF-AUX product are used to delineate layers where ice, mixed, and liquid clouds are expected. The state vector retrieved by the algorithm consists of the particle number concentration N_T, assumed constant throughout the column, and the profile of the logarithm of geometric mean particle radius $r_g = \exp(\langle \ln r \rangle)$ (brackets denote the expected value). By assuming a lognormal distribution of liquid particle size with geometric standard deviation σ_{log}, the Liquid Water Content *LWC* profile can be derived one CloudSat subcolumn/ray at a time. A preliminary version of the algorithm was previously outlined in Leinonen et al. (2016) and has been revisited to better handle confounding factors such as precipitation, and overlying ice and mixed-phase clouds. See https://www.cloudsat.cira.colostate.edu/cloudsat-static/info/dl/2b-cwc-rvod/2B-CWC-RVOD_PDICD.P1_R05.rev0_.pdf for additional details.

2) ICE PHASE: 2C-ICE (2CICE)

The Cloudsat and CALIPSO Ice Cloud Property Product (2C-ICE, hereafter 2CICE) contains retrieved estimates of profiles of ice cloud water content, effective radius, and extinction coefficient for each active measurement ray that contains ice particles as indicated by CloudSat's CPR and/or CALIPSO's CALIOP. Specifically, the 2CICE cloud product uses combined inputs of measured radar reflectivity profiles from CloudSat's 2B-GEOPROF product and measured attenuated backscattering coefficient profiles at 532 nm from CALIOP (Rogers et al. 2011) to constrain the ice cloud retrieval more tightly than the radar-only product, and to yield improved retrievals (Deng et al. 2015). As in 2BCWC, an optimal estimation framework is used with the measurement vector consisting of the CALIOP-measured backscattering coefficient profile and the CPR-measured reflectivity profile. The state vector is initialized with a priori estimates from extensive in situ measurements and/or literature-supported empirical relations and algorithms.

3) CLOUD MASK: 2B-CLDCLASS-LIDAR (2BCL)

The 2B-CLDCLASS-LIDAR product, hereafter 2BCL, combines CPR and CALIOP measurements for cloud phase determination and cloud scenario classification (Sassen and Wang 2008; Sassen and Wang 2012). Cloud classification is achieved by synthesizing information about the horizontal and vertical variability of cloud properties, the precipitating state of the cloud field, cloud temperature, and coincident MODIS radiances. For this work, we use 2BCL as a 2D (along track-height) cloud mask on the same grid as the previous two products. For each ray in the dataset, the height of the cloud top and base, as well as the cloud thermodynamic phase (liquid, ice, mixed) is extracted for vertically distinct cloud "objects" consisting of contiguous vertical layers. With each distinct cloud object assigned one of the three thermodynamic phases, all the vertical layers it encompasses have the same phase.

b. Construction of base 2D cloud optical depth fields

2CICE provides vertically (240m) and horizontally (~1.7 km) resolved (i.e., "cell") ice water content (*IWC*) and particle effective radius (r_e) from which cell visible ice cloud optical depth COD_{ice} can be calculated according to (e.g., Stein et al. 2011):

$$COD_{ice} = \frac{3Q_{ext}IWC}{4\rho_{ice}r_e}\Delta z \tag{1}$$

where ρ_{ice} = 0.92 g cm^{-3} is the density of ice, Q_{ext} = 2 is the extinction efficiency at visible wavelengths, and Δz = 240 m is the cell's physical thickness.

On the other hand, the 2D field of visible cloud optical depth for liquid phase clouds COD_{liq} is estimated from 2BCWC using the retrieved optical extinction coefficient. Per Leinonen et al. (2016), the cell's extinction coefficient is calculated by combining the cell's N_T, r_g, and the droplet distribution's geometric standard deviation σ_{log}:

$$\sigma_{ext} = 2\pi N_T r_g^2 \exp\left(2\sigma_{log}^2\right) \qquad (2)$$

Cell cloud optical depth COD_{liq} is then obtained by multiplying extinction with the cell's physical thickness Δz, i.e., $COD_{liq} = \sigma_{ext}\,\Delta z$.

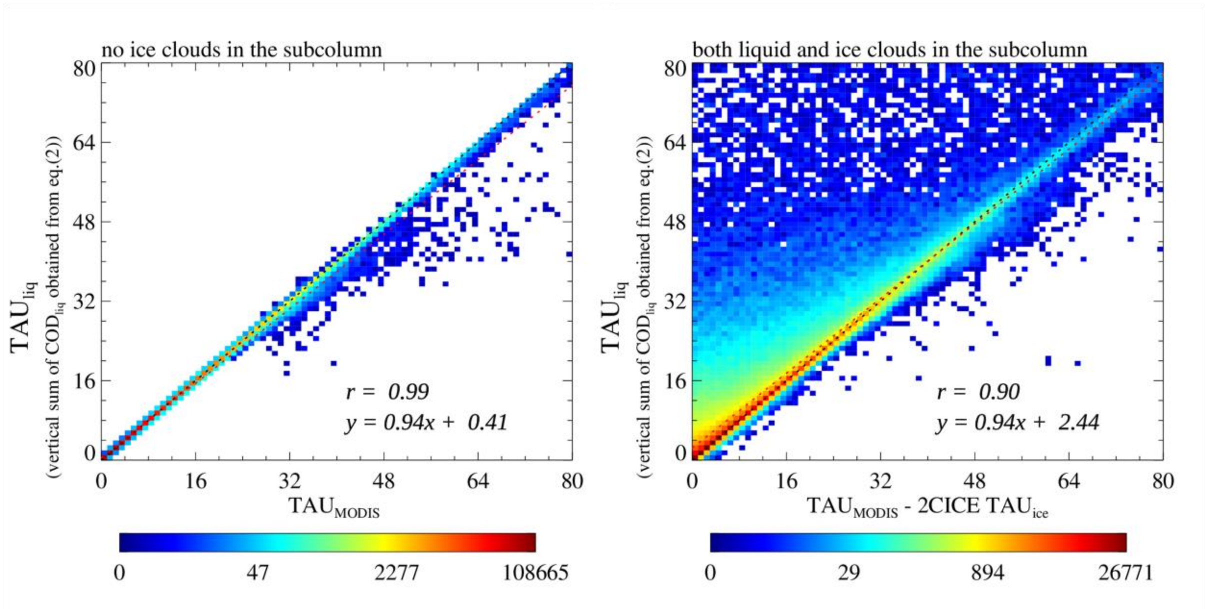

Fig. 1. Comparison of either directly observed (subcolumns with liquid only clouds, left panel) or inferred (subcolumns with both liquid and ice clouds, right panel) $TAU_{MODIS,liq}$ against TAU_{liq} from 2BCWC (eq. 2). One month of data (January 2007) with solar zenith angle below 45° were used, as in Leinonen et al. (2016).

We evaluate COD_{liq} obtained from eq. (2) by comparing its vertical integral TAU_{liq} with the collocated MODIS TAU_{liq}, inferred or directly observed. TAU_{MODIS} in MOD06-1KM-AUX corresponds to the TAU_{liq} used as constraint in the 2BCWC retrieval when there is no ice in the subcolumn; TAU_{liq} can also be inferred by subtracting 2CICE TAU_{ice} from total MODIS TAU_{MODIS} for subcolumns that have both liquid and ice clouds, as indicated above. A comparison for the first scenario is shown in the left panel of Fig. 1 for a sample month (January 2007), while a comparison for the second scenario is shown in the right panel. The general agreement seen in the figure serves primarily as a sanity check rather than an independent validation given that the 2BCWC retrievals are algorithmically constrained by

TAU_{MODIS}. Since MODIS is not sensitive to precipitating particles, the MODIS optical depth constraint signifies that the combination of retrieved N_T, r_g in eq. (2) represents the extinction of non-precipitating particles. However, occasional substantial deviations from the 1:1 to line are indicative of imperfect filtering of rays containing precipitation (drizzle) in which case the a priori constraints of the optimal estimation algorithm are inappropriate and may introduce biases.

The blended 2D COD field from 2BCWC and 2CICE consists then of cells whose COD comes from either 2CICE-alone (COD_{ice}), 2BCWC-alone (COD_{liq}), or combined CODs from the two datasets ("mixed" phase cells with $COD_{mixed} = COD_{liq}+COD_{ice}$).

1) Missing Retrievals

Comparison of our 2D COD fields, constructed as described above, with CloudSat's 2D 2BCL cloud mask product reveals that numerous cells identified as cloudy in that product do not have COD > 0 in our combined product. This turns out to be a problem mostly confined to liquid phase clouds whose CODs come from 2BCWC. It appears that most missing liquid phase retrievals correspond to optically thin (and therefore of low CPR reflectivity) clouds that are however still detected by the CALIOP lidar and therefore present in 2BCL. Since upper-level clouds with substantial extinction (TAU ~ 5 and above) fully attenuate the lidar beam, many of these low clouds seen by the lidar are either completely unobscured by clouds above, or co-occur with upper level clouds still thin enough to allow the lidar beam to reach the lower troposphere. Such clouds, present in 2BCL, but not in 2BCWC, are assigned COD values according to a filling scheme described below.

Figure 2 shows two examples, each corresponding to a CALIOP-CloudSat "curtain" within a 1° gridbox. In the first example (upper row) liquid cloud retrievals seem to be missing (no yellow) in segment *A*. More missing liquid cloud retrievals apparently exist in segment *B*, which is classified as being of mixed phase by virtue of being part of a distinct cloud object in 2BCL assigned in its entirety to the mixed phase because of ice presence in its upper part. The second example (bottom row) has far more missing liquid cloud retrievals: indeed, very few yellow cells of successful 2BCWC retrievals exist compared to the large liquid cloud cell population (orange cells) in 2BCL. Missing liquid retrievals can be found both in parts of the scene with (segment *D*) and without (segment *E*) overlying high clouds. Below we describe our simple approach to restore some of the missing retrievals for liquid clouds that are apparently present (orange cells in the left panels of Fig. 2).

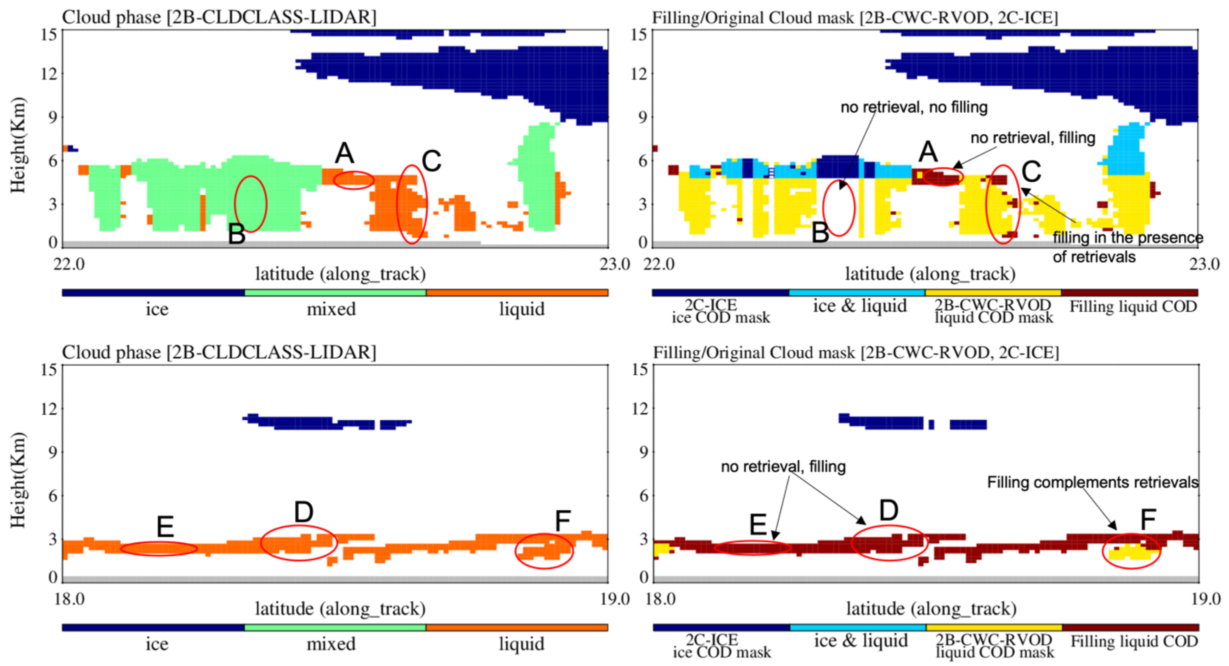

Fig. 2. Two examples 2BCL 2D cloud mask scenes (left panels) and reconstructed 2D COD scenes. Highlighted segments designated with letters are used as examples to explain major aspects of our filling scheme.

2) Filling Missing Liquid Phase Retrievals

A simple scheme was devised to assign COD_{liq} values to cells with no retrievals in 2BCWC but flagged as having liquid clouds in 2BCL. We follow the logic of the 2BCWC algorithm and calculate the subcolumn's TAU_{liq} by either simply assigning total TAU_{MODIS} to subcolumns with only liquid clouds, or assigning TAU_{MODIS} minus TAU_{ice} from 2CICE to mixed subcolumns. TAU_{liq} for the missing cells can then be calculated by subtracting the sum of existing 2BCWC COD_{liq} values in the subcolumn (if applicable, as in segment *C* of Fig. 2). This adjusted TAU_{liq} is then partitioned into equal COD_{liq} values among all missing cells. The COD_{liq} profile of a subcolumn can therefore contain a mixture of COD_{liq} values from the original 2BCWC retrievals, and equal COD_{liq} values for all cells with missing retrievals from the partitioning of the adjusted TAU_{liq}. Assigning COD_{liq} values to cells with unavailable retrievals through a more sophisticated scheme would require information that is not currently at our disposal. Actually, application of even this procedure (constituting "Pass 1" of our filling scheme) is not always possible either because of unavailability of TAU_{MODIS} or because the adjusted TAU_{liq} after subtracting TAU_{ice} (and in some cases also the sum of 2BCWC-provided COD_{liq}) from TAU_{MODIS} turns out to be negative. Missing COD_{liq} values for such cases are filled with available neighboring COD$_{liq}$ values (“Pass 2”), as described below.

The possible Pass 1 scenarios to obtain profiles of COD$_{liq}$ can then be summarized as follows:

(1) Pure liquid subcolumn according to 2BCL, no 2BCWC retrievals in the subcolumn (segment *E* in Fig. 2): Apportion equally TAU$_{MODIS}$ ≡ TAU$_{MODIS,liq}$ to all cells with liquid phase in 2BCL. The COD$_{liq}$ profile consists of equal values.

(2) Pure liquid subcolumn according to 2BCL with available 2BCWC retrievals (segment *F* in Fig. 2): Subtract TAU$_{2BCWC,liq}$ from TAU$_{MODIS}$ ≡ TAU$_{MODIS,liq}$ and if the result is positive apportion equally to all missing cells with liquid phase in 2BCL. The COD$_{liq}$ profile is a mixture of 2BCWC retrievals and the equal values from the apportionment of this adjusted TAU$_{liq}$ corresponding to all cells of the subcolumn with missing retrievals.

(3) Overlying ice clouds and no retrievals in 2BCWC (segments *A* and *D* in Fig. 2): Subtract TAU$_{2CICE}$ from TAU$_{MODIS}$ and if the result is positive apportion the resulting TAU$_{liq}$ equally to all missing cells with liquid phase in 2BCL. The COD$_{liq}$ profile consists of equal values.

(4) Both overlying ice clouds with 2CICE retrievals and liquid clouds with 2BCWC retrievals (segment *C* in Fig. 2): Subtract the combined ice and liquid column optical depths (TAU$_{2CICE}$ + TAU$_{2BCWC}$) from TAU$_{MODIS}$ and if the result is positive apportion the resulting TAU$_{liq}$ equally to all cells of liquid phase in 2BCL without 2BCWC retrievals. The COD$_{liq}$ profile is a mixture of 2BCWC retrievals and the equal values from the apportionment of adjusted TAU$_{liq}$ corresponding to all cells of the subcolumn with missing retrievals.

As previously mentioned, our scheme has also a second part (“Pass 2”) to fill 2BCL cells with liquid clouds that remain unfilled after Pass 1 because TAU$_{MODIS}$ is either unavailable or inconsistent with available 2CICE and 2BCWC CODs in the subcolumn. Pass 2 applies a “nearest-neighbor” (NN) scheme using cells with available COD$_{liq}$, from 2BCWC retrievals or Pass 1. The scheme works as follows: We move from “left” to “right” (from start to end of the data granule). We form a 3×3 domain centered around the missing value and calculate the missing value as the weighted average of the available COD$_{liq}$ values in the domain either from 2BCWC, Pass 1, or Pass 2 on preceding subcolumns. We then multiply by the volume liquid cloud fraction of the domain (fraction of cells in the domain with liquid clouds, i.e., number of cloudy cells divided by 9). Unity weights are used when averaging immediate vertical and horizontal neighbors of the cell being filled, and $1/\sqrt{2}$ (pythagorean distance) for diagonal neighbors. If the 3×3 domain contains no COD_{liq} values, we expand the domain centered around the missing value to be a rectangle domain of size 21×7 (horizontal×vertical) and calculate the missing value as the median of all available COD_{liq} values in the domain. Further expansion to 51×7 and 101×7 domains is built into the scheme, but is rarely invoked.

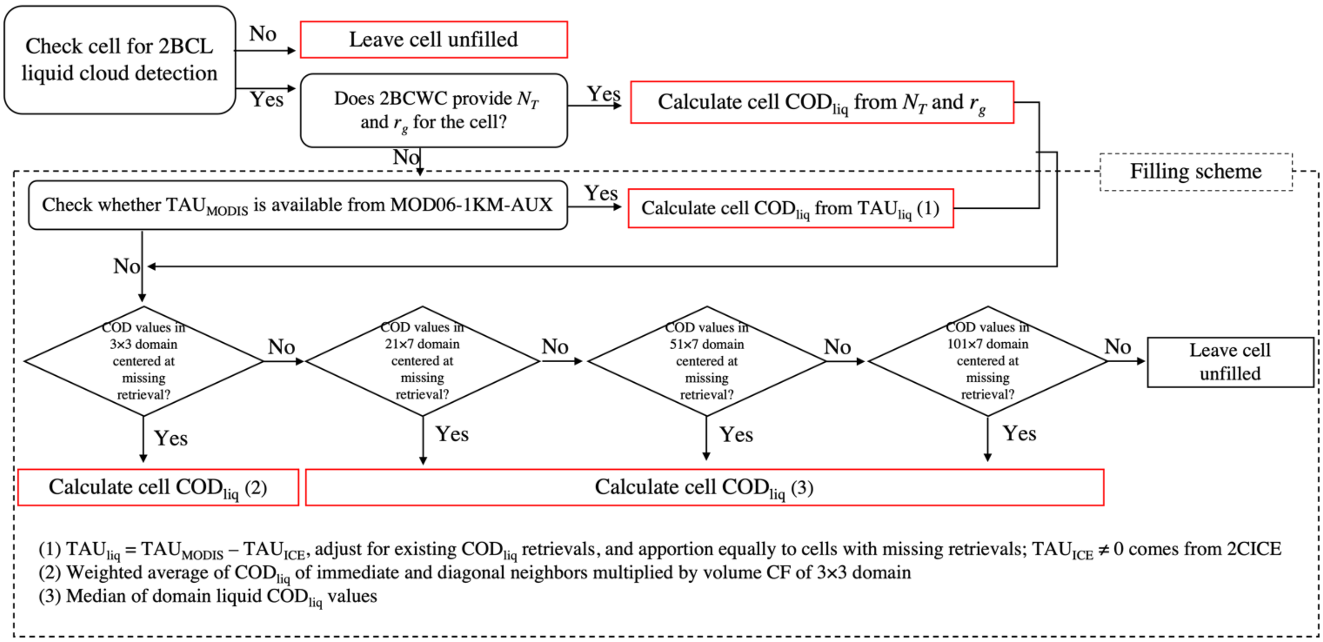

Fig. 3. Flowchart summarizing the construction of our 2D COD_{liq} reference fields.

Provenance of COD_{liq} cells	**Number of cells**	**Percentage of cells**
Derived from r_g, N_T	289,197,600	80.7%
Filled using MODIS TAU_{liq}	45,438,111	12.7%
Filled using NN, 3×3 domain	22,188,733	6.2%
Filled using NN, 21×7 domain	1,270,571	0.4%
Filled using NN, 51×7 domain	184,610	0.1%
Filled using NN, 101×7 domain	82,989	0.0%
All COD_{liq} cells	358,362,614	100%

Table 1. Population information on the provenance of COD_{liq} cells in our 2D COD field constructed by combining 2BCWC and 2CICE retrievals.

The flowchart of Fig. 3 encapsulates the full scheme for constructing the 2D COD_{liq} fields. Table 1 summarizes the provenance of COD_{liq} cells for the 2007 data we processed. About 81% of COD_{liq} values come from 2BCWC; from the remaining 19%, about two-thirds are filled via TAU_{MODIS} (explicit or adjusted MODIS TAU_{liq}, or inferred where appropriate using TAU_{2CICE}), i.e., Pass 1, and the remaining one-third by NN filling (Pass 2), the vast majority of which comes from very close neighbors (within the 3×3 domain).

3) ASSESSMENT OF FILLING

A major application of the type of generators assessed in this work is to provide subcolumns to COSP's International Satellite Cloud Climatology Project (ISCCP) and MODIS simulators whose main diagnostic is joint cloud fraction (CF) histograms in cloud top pressure (CTP) – TAU space. We therefore opt to use such histograms as the cornerstone for evaluating not only the generators, but also the performance of the filling scheme. Specifically, we examine whether average joint CTP-TAU histograms corresponding to the modified COD fields are more alike to their counterparts from coincident Aqua cloud observations than average joint histograms coming from the original (unfilled) COD fields.

The central role of CTP-TAU histograms in this paper merits some additional elaboration. Given that we use in these histograms the subcolumn TAU and the CTP of its topmost cloudy layer, one would initially think that the vertical distribution of COD does not matter, but rather only its vertical integral, in accordance with the simplicity of our filling scheme's Pass 1. But the irrelevance of the COD profiles for the TAU of individual subcolumns casts doubt at the same time on the appropriateness of CTP-TAU histograms as a rigorous evaluation metric of the filling scheme. While it is true that using only TAU for individual observed subcolumns considered in isolation makes the COD profile irrelevant, the details of the vertical COD profile matter *for the ensemble of subcolumns* forming a scene. This is because in addition to the observed scene CF profile (calculated as the fraction of the scene's subcolumns with valid clouds in that layer), the generators also use the observed mean COD profile as input. This layer mean COD across subcolumns *does depend* on how TAU is vertically apportioned within individual subcolumns in the observations, making thus CTP-TAU histograms coming from generators sensitive to the observed COD profile. Nonetheless, the lack of better alternatives compels us to stay with the simple equal apportionment of adjusted TAU_{liq} for Pass 1 of our filling scheme, and also CTP-TAU joint histograms as the primary metric for evaluating both the filling scheme and the performance of the subcolumn generators.

Aqua joint histograms were obtained from the equal area 3-hour histogram dataset used in Cho et al. (2021) to derive MODIS Cloud Regimes (CRs) on the ISCCP grid. Because for the year of our analysis (2007) Aqua and CloudSat-CALIPSO (CC) were part of the A-Train constellation, temporal matching is already built into the dataset. We simply identify the segment of the 2D COD field that falls within the 110 km gridbox on that day and create a 2D CTP-TAU histogram for that segment. To construct the 2D CTP-TAU histograms we convert COD profiles from height to pressure coordinates using CloudSat's ECMWF-AUX product. We then eliminate all cloudy subcolumns with TAU < 0.3 (about 1% of all subcolumns), by setting TAU = 0. This is done because MODIS detection and retrieval of clouds with such low optical thickness is of low confidence, something accounted for in the MODIS simulator (Pincus et al. 2012) used in this paper for generator evaluation. All coincident joint histograms from the active CC and the passive MODIS observations are then averaged. Figure 4 shows the comparison of global joint histograms resolving CF into 42-bins, using ISCCP's CTP-TAU bin discretization (Jakob and Tselioudis 2003), also used for a CTP-TAU joint histogram version found in the MODIS cloud products.

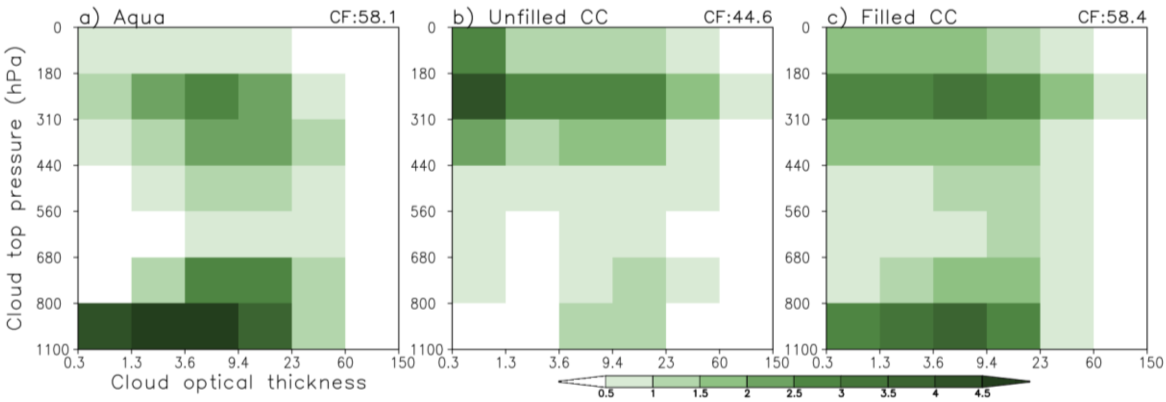

Fig. 4. Globally-averaged joint CTP-TAU histograms from 2007 coincident observations of (a) Aqua (left); (b) "Unfilled" COD fields from CC; (c) "Filled" COD fields from CC.

While MODIS clouds cannot be considered as "truth", a certain degree of consistency between passive and active retrievals is expected, with large discrepancies potentially being a cause for concern. Our comparison is therefore highly instructive and clearly shows the improvements brought by the filling scheme with respect to low clouds. The unfilled joint histogram has a far smaller overall CF than the MODIS global histogram (44.6 vs 58.1, in %, henceforth implied for all CF values) with the difference in low clouds, CTP > 680 hPa, being even larger: low CF values are 28.5 for MODIS and 7.4 for unfilled CC 2D COD fields. The reconstructed COD dataset compensates some of the low cloud difference with more high clouds, especially in the bin of smallest TAU due to CALIOP's sensitivity to even the optically thinnest clouds. But filling is essential in raising the low CF of CC to 19.7, still not as high as MODIS (which may be overestimating low CF because MODIS pixels are assumed overcast when cloud is detected), but good enough to bring down the overall CF discrepancy to (absolute) 0.3%.

The evaluation of the filling scheme via such a CTP-TAU joint histogram comparison can also be broken down by the MODIS CRs of Cho et al. (2021). These CRs represent the most common cloud mixtures observed by MODIS at daily ~100 km scales as represented by the mean of all CTP-TAU histograms deemed alike by a k-means clustering algorithm. The CTP-TAU histograms of the CC COD segments collocated with MODIS-Aqua equal area gridboxes are assigned to the MODIS CRs identified therein. For this comparison, histogram averaging is therefore performed as before, but now separately for each MODIS CR. The relative proximity of joint histograms coming from the unfilled and filled cloud fields to the Aqua reference can be summarily captured by Euclidean Distances (EDs) between members of histogram pairs. EDs derived for unfilled and filled CC COD fields are very similar for CRs with relatively small populations of low clouds (CR1-CR6). But for the other CRs, CR7-CR11, with plentiful low clouds, the lower EDs of the filled cloud fields indicate better histogram resemblance, particularly for CR7-CR10. These results instill confidence in the beneficial effects of our simple approach to fill cells with missing COD_{liq} values.

3. Performance of subcolumn generators

a. Description of subcolumn generators and implementation specifics

1) SCOPS GENERATOR

The SCOPS generator can produce subcolumns obeying random, maximum, or (our choice for this paper) maximum-random overlap, and can treat separately convective and stratiform clouds if such a distinction is known in the source dataset and of interest (neither applies for this work). SCOPS takes as input the gridbox's or (in our case) the scene's mean CF profile (the fraction of the scene at each layer with COD above the threshold that indicates cloud presence) and then outputs a set of subcolumn cloud occurrence profiles where the individual subcolumn cells are either overcast or cloud free (i.e., binary 0 or 100 cell CF). Condensate amounts are then assigned to the overcast cells according to the standard implementation in COSP which assumes constant in-cloud condensate mixing ratio across each layer equal to the gridbox layer mean provided by the host GCM. In our study each SCOPS subcolumn is multiplied by the scene's mean COD rather than condensate profile. The output is a 2D scene of COD that obeys the cloud occurrence overlap rules of SCOPS and has horizontally uniform layer COD, but horizontally variable subcolumn TAU. The scene COD field is then passed to the MODIS simulator to generate a CTP-TAU joint histogram.

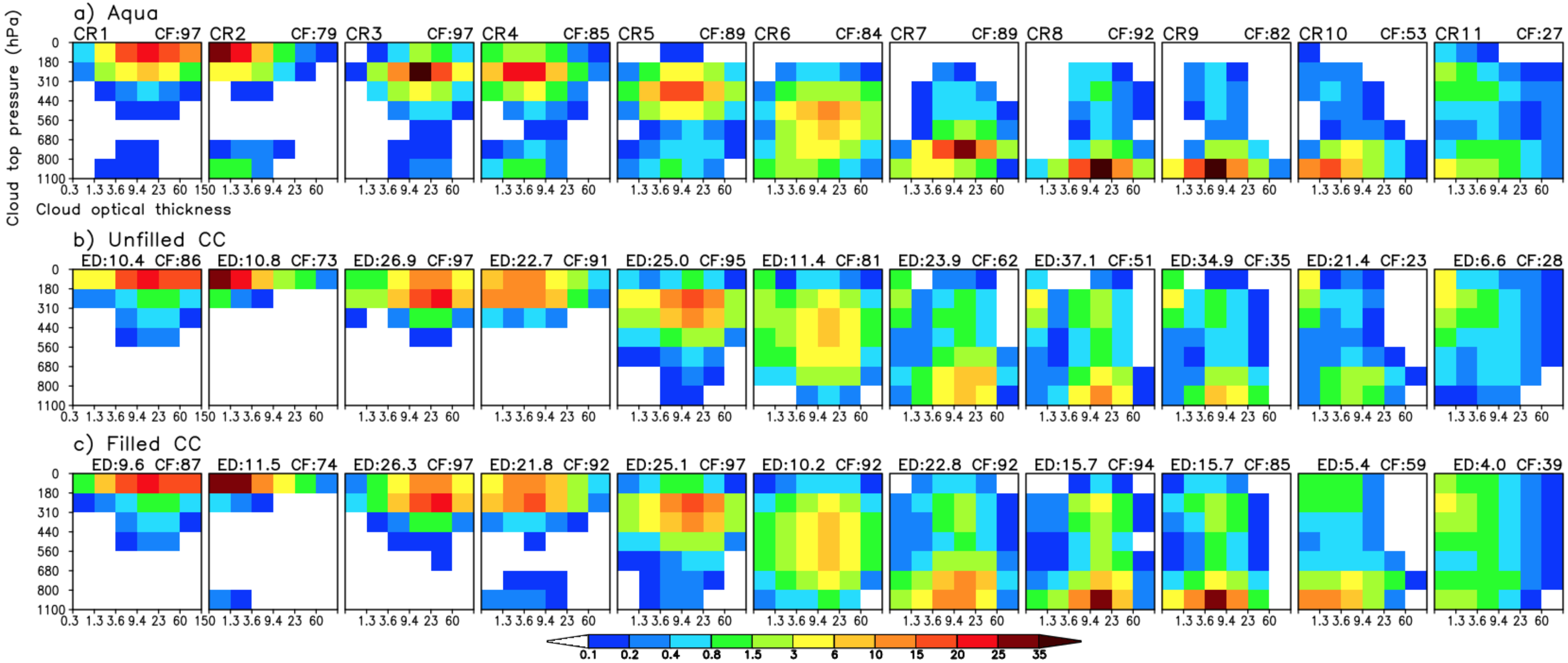

Fig. 5. Globally-averaged joint CTP-TAU histograms from coincident 2007 observations broken down by MODIS CR. (a) Aqua (top); (b) “Unfilled” COD fields from CC (middle); (c) “Filled” COD fields from CC (bottom). Above the CC panels, we also provide in addition to CF the CR-specific EDs from observations.

2) Raisanen Generator

The "Raisanen" generator also yields an ensemble of stochastically generated cloudy subcolumns (Räisänen et al. 2004). Each layer within each subcolumn (cell) is assumed homogeneous, with a CF of either 0 or 100, but condensate amount can vary across a vertical layer consisting of the cells of the subcolumns at the same altitude, forming thus the layer's condensate probability density function (PDF) which can be described by an analytical function such as gamma, beta or lognormal. In our case, we pass to the generator the CF, mean COD, and COD variance profiles of the 2D scene, the latter coming from the CF profile using the variance parameterization in Oreopoulos et al. (2012). The COD variance profile is then used to create a profile of beta distribution PDFs.

The generator allows for a continuous range of cloud occurrence overlap rates between maximum and random overlap according to the generalized overlap paradigm of Hogan and Illingworth (2000). In this paradigm a weighting factor controls the relative contribution of maximum (dominating for values of weighting factor close to one) and random overlap (dominating for values of the weighting factor close to zero) to the combined CF of two cloudy layers. The weighting factor is parameterized as an exponentially decaying function with an e-folding distance or "decorrelation length" describing its rate of decrease as a function of the separation distance between cloud layer pairs. Small values of decorrelation length denote rapid decline of the weighting factor with separation distance (near-random overlap) while large values denote a slow decline (near-maximum overlap). We use a parameterization that captures the day-to-day latitude dependence of decorrelation length with a Gaussian function fit to CloudSat observations of cloud occurrence overlap (Oreopoulos et al. 2012). Similarly, vertical correlations of COD PDFs are captured by correlations of COD ranks (i.e., Spearman rank correlations), also assumed to decay exponentially with layer separation distance, according to a second decorrelation length. This decorrelation length is also parameterized with a latitude and day-of-the-year-varying Gaussian function which fits CloudSat reflectivity observations (Oreopoulos et al. 2012) and represents a more rapid decay with vertical separation distance of COD rank than cloud occurrence overlap.

b. Results

We use the one-year (2007) dataset of reconstructed CC COD fields to assess the skill of the two subcolumn generators described above in simulating cloud subgrid variability of scenes consisting of 100 subcolumns (~110 km given CC ray sampling). We only use oceanic scenes (about 586 thousand scenes) because height in CC datasets is referenced relative to local surface elevation, making the averaging of vertically resolved subcolumns ambiguous when surface elevation varies.

Our preferred (but not only) method of assessing skill is comparison of CTP-TAU joint histograms coming from the MODIS simulator. These capture how clouds and their condensate align vertically in real and simulated overlap scenarios to generate subcolumn TAUs, and at the same time provide a rough cloud type discretization according to cloud top height (pressure). Because in constructing these joint histograms the MODIS simulator rejects subcolumns with TAU < 0.3, the reconstructed COD fields and the statistics derived from them treat these subcolumns as cloud-free. All in all, the generators are evaluated using comparisons between: (1) Mean CTP-TAU histograms obtained by extensive averaging of individual scene histograms and their corresponding bin-resolved cloud radiative effects (CREs); (2) Quantities derived from individual scene CTP-TAU histograms; (3) Profiles of cumulative CF and (the closely related) CF exposed to space.

1) GRAND AVERAGE HISTOGRAMS

Figure 6 shows the one-year mean global ocean of 42-bin CTP-TAU joint histograms from observations (left) and from the subcolumn generator reconstruction (middle and right panels). The leftmost column of observed globally (ocean-only) averaged joint histograms includes the numerical values of bin CF. The middle column shows the globally-averaged histograms coming from SCOPS-generated subcolumns; the numbers stand for the bin CF values while color displays differences from the reference observed histograms. The rightmost column similarly conveys results from the Raisanen generator.

Once again, the overall resemblance of the reconstructed histogram to the true histogram can be captured by the ED between the two histograms. According to that metric, the Raisanen generator performs better overall (smaller ED) despite its slightly worse than SCOPS underestimation of total CF . Both generators have reasonably good skill in reproducing the total vertically projected CF. This is expected to some extent since they are supplied observed CF profiles. Nevertheless, how the CFs of individual layers are overlapped still matters: the underestimation by the generators suggest that they overlap clouds slightly more maximally than in observations. A common deficiency of the two generators is the underestimation of optically thin (TAU < 3.6) clouds, which is less severe and mostly concentrated to TAU < 1.3 for the Raisanen generator. These underestimates are possibly related at least in part to the greater than observed tendency for maximum overlap and the resulting greater vertical cloud alignment that reduces the probability of optically thin clouds. SCOPS seems to compensate for the deficit of optically thin clouds with overestimates of clouds of moderate and large optically thickness, something that the Raisanen generator is much less prone to. Note that while absolute biases for the optically thickest TAU class (60-150) are mild for both generators, the small differences from observations correspond to small CFs to begin with (i.e., such clouds are rare). The CF of the second largest TAU class (23-60) is overestimated similarly by both generators (7.4 vs 6.7 in observations), but SCOPS suffers more error compensation. On the whole, the Raisanen generator produces fewer extreme biases (both overestimates and underestimates) than SCOPS, and most of the overall CF underestimate comes from optically thin clouds. SCOPS on the other hand would have suffered a much greater total CF underestimate due to thin clouds were it not for compensatory overestimates for clouds of moderate and high optical thickness.

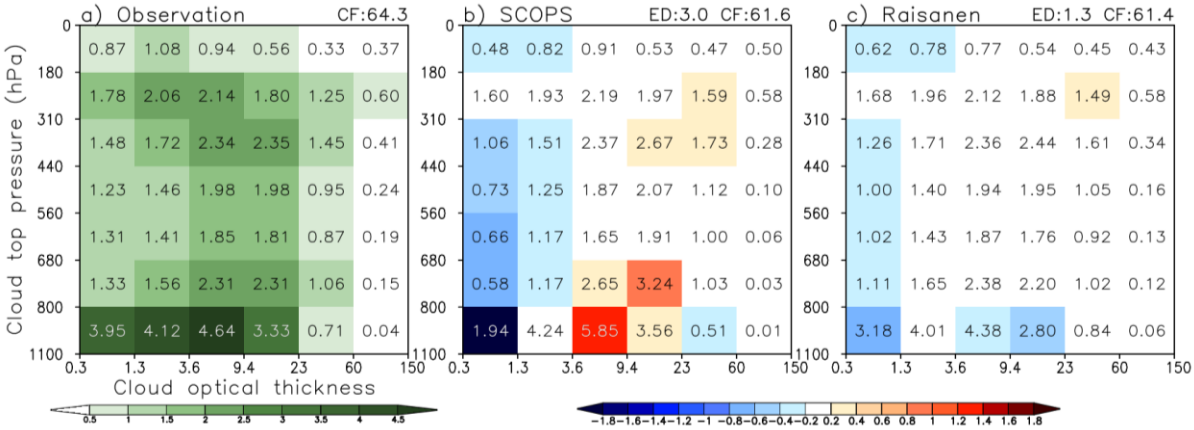

Fig. 6. One-year (2007) mean global ocean CTP-TAU joint histograms of observations (left panels) and from the subcolumn generator reconstruction (SCOPS-middle and Raisanen-right panels). For the middle and right panels the numbers represent the bin CF values themselves while the colors represent differences from the reference observed histograms.

The comparison of average histograms can be performed at a greater level of detail using again the Cho et al. (2021) MODIS CRs in the manner previously employed in assessing the quality of COD_{liq} filling. Mean joint CTP-TAU histograms by CR are compared between our reference COD fields and those produced by the two generators in Fig. 7. The top row shows the observed mean joint histograms by CR from CC COD fields for the year 2007 and for ocean only. The middle and bottom rows show joint histogram differences obtained by subtracting the observed mean CR histogram from its reconstructed counterpart produced by the two generators. Both generators are capable of closely reproducing the mean CF corresponding to each CR, but again with a systematic underestimation. The Raisanen generator performs overall better when performance is measured in terms of ED: for 8 out of 11 CRs, Raisanen EDs are smaller than SCOPS EDs. Raisanen is notably inferior for CR8 even though it reproduces the mean CF of this CR quite well. This CR along with CR9 appear to go against Raisanen's tendency of optical thin cloud underestimation; on the other hand, SCOPS's underestimation of optically thin cloud is persistent across all CRs.

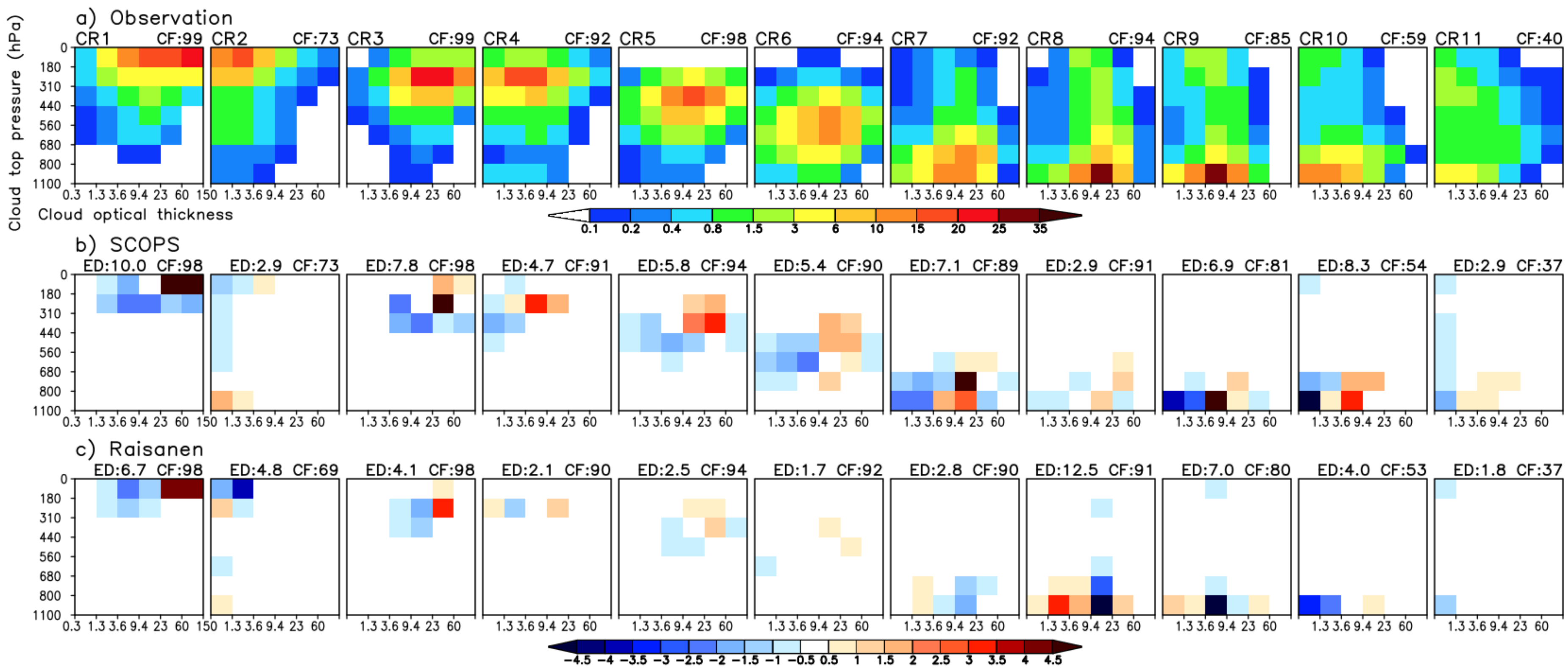

Fig. 7. Top row: Mean 42-bin joint CTP-TAU histograms from CC for 2007 aggregated by MODIS CR over oceans only. Middle row: difference between SCOPS-reconstructed and observed mean joint histograms by CR (negative values indicate underestimate by SCOPS). Bottom row: As middle row, but for the Raisanen generator.

2) Individual Scenes

The performance of the two subcolumn generators can also be assessed at the scene level by comparing the statistics of individual scene EDs and other metrics. Figure 8 depicts two examples of how the EDs between simulated and observed 42-bin joint histograms for individual 100-subcolumn scenes can be used to compare the performance of the two generators. The left panel shows how the mean of scene EDs varies as a function of their CF. The mean EDs of the two generators start to diverge at CF ≈ 20%, with the average EDs of the Raisanen generator remaining consistently below those from the SCOPS generator, indicating greater resemblance to observations on average. Two factors contribute to the monotonic increases of ED with CF: larger CF values for joint histogram bins that are already populated creating larger squared differences, and greater number of populated bins contributing more terms to the sum of squared differences.

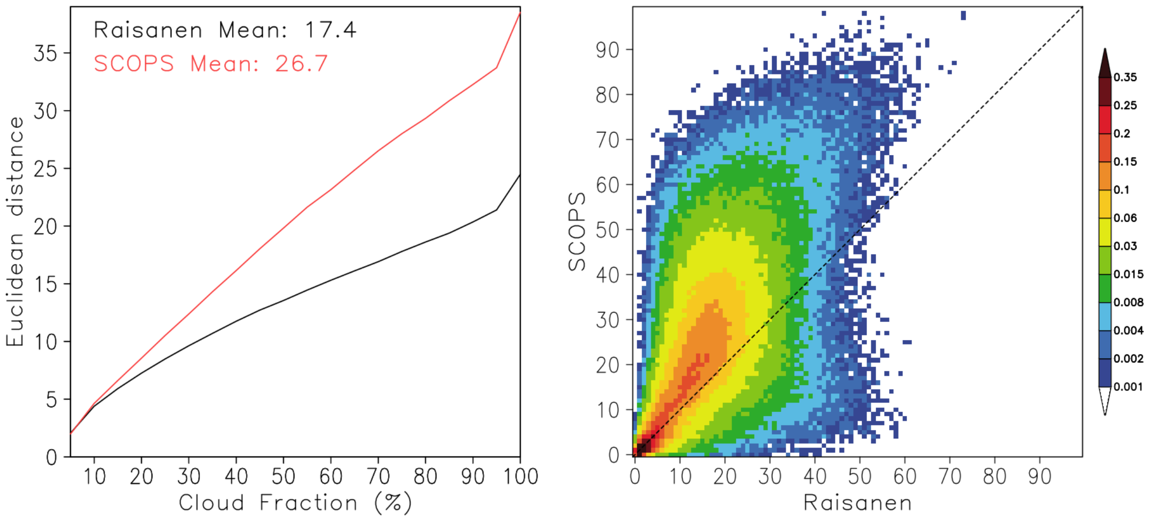

Fig. 8. Left panel: Average of individual 100-subcolumn scene EDs for the SCOPS and Raisanen generators from scene 42-bin histograms, discretized by observed scene CF. Right panel: Density plot of ED pairs from the two generators for individual scenes.

The right panel of Fig. 8 provides another glimpse of relative generator performance using the same scene ED dataset. This time we create a density plot of ED pairs from the two generators. The population of pairs above the diagonal containing scenes where SCOPS ED exceeds Raisanen ED is much larger. There is a hint that SCOPS is doing better than Raisanen at very small EDs (and thus likely small scene CFs), as indicated by the larger density below the diagonal up to ED ≈ 5, but the density asymmetry reverses quickly, with far more scenes having greater ED for SCOPS than Raisanen above ED ≈ 10.

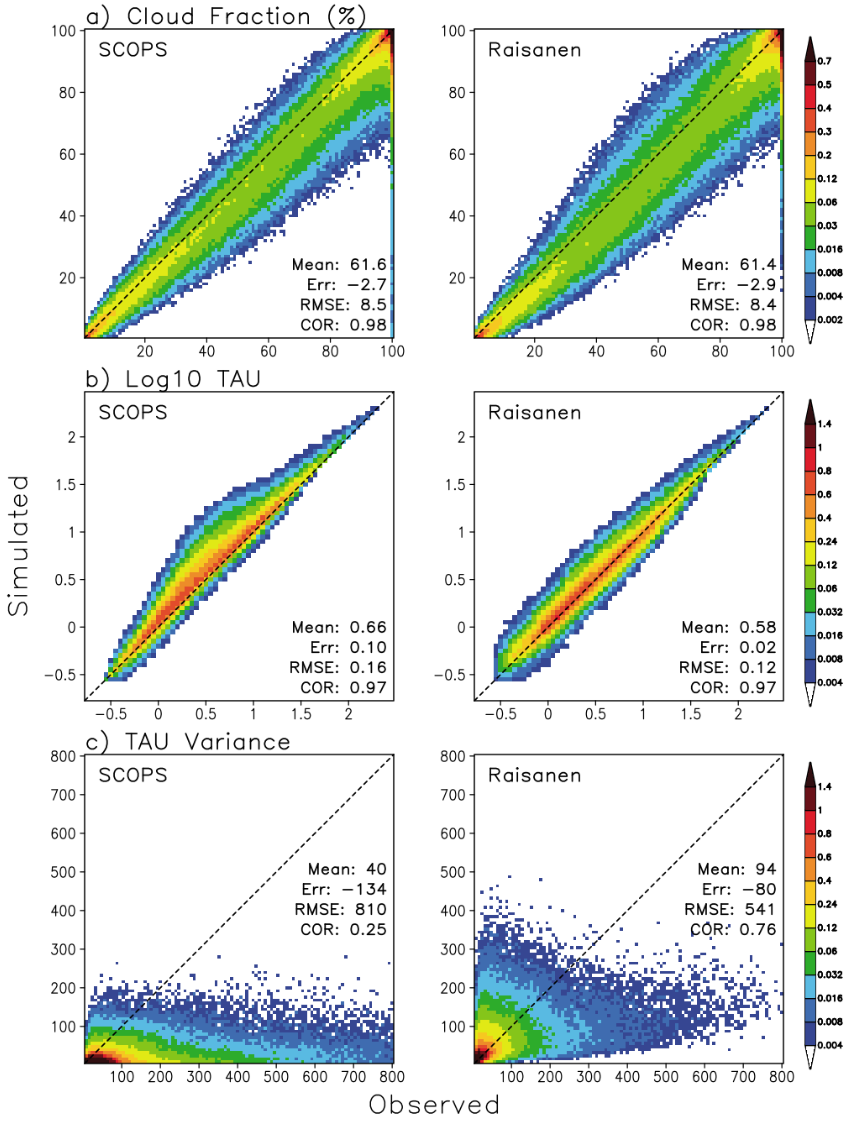

Fig. 9. Top row: Density plot of pairs of observed and simulated (SCOPS: left column; Raisanen: right column) vertically projected scene CFs. Middle row: As top row, but for mean logarithmic TAU. Bottom row: As top row but for scene TAU variance.

Additional comparisons using density plots are shown in Fig. 9. This time we compare observed and simulated vertically projected scene-level CFs (top row), mean TAUs (middle row) and TAU variances (bottom row). The density plots for CFs indicate similar performance for the two generators and a preponderance of scene CF underestimates (i.e., fewer points above the diagonal) over the full range of CFs, a result consistent with the overall underestimate of CF seen in Fig. 6. The two generators produce a broad overestimate of mean logarithmic TAU which is however less pronounced for Raisanen, consistent with the underestimate of optical thin and overestimate of optically moderate and thick cloud populations seen previously in averaged joint histogram results. One has to keep in mind though when comparing to previous results that what is being assessed here is the mean (logarithmic) TAU of individual scenes which corresponds to the TAU distribution of individual scene joint histograms. The Raisanen generator appears to overestimate scene mean TAU even when the scene is optically thin on average, something that does not occur for SCOPS which is performing better for such scenes. But for intermediate TAUs Raisanen is clearly better, ultimately yielding a smaller overall bias and RMSE. Where the two generators diverge greatly is with respect to the variance of TAU: SCOPS's variance is far below observations for the vast majority of scenes. This is hardly surprising given that SCOPS distributes COD homogeneously across vertical layers and all variance of column TAU comes from cloud occurrence overlap. Raisanen on the other hand tends to overestimate variance of low and moderate magnitude. But once observed variance becomes very pronounced (> 200), which is though quite rare, Raisanen typically underestimates it.

Even with this information, the radiative implications of generator performance at the level of individual scenes are not easily predictable outside of an actual model implementation. Yet, one can hypothesize on the potential impact of competing effects in the shortwave part of the spectrum where cloud heterogeneity (subgrid variability) matters more: reflected solar radiation would be underestimated when scene total vertically projected CF is underestimated (both generators), but this would be compensated to some degree by overestimates in mean TAU (both generators). Once variance of TAU is taken into account, its smaller underestimate by Raisanen would contribute a smaller overestimate than SCOPS. In other words, SCOPS can potentially provide greater compensation for its CF underestimate through its mean TAU overestimate and TAU variance underestimate. Radiative flux errors implied by generator deficiencies in producing correct subgrid variability are discussed later, but only in the context of the grand-average joint histograms of Fig. 6, and not for individual scenes where a more involved setup is required.

3) SENSITIVITY EXPERIMENTS WITH THE RAISANEN GENERATOR

As elaborated previously, the overlap of cloud occurrence and COD PDF in the Raisanen generator is regulated by two decorrelation lengths which control the proportion with which maximum and random overlap mix. In the results shown previously we used decorrelation lengths values obtained by the parameterizations of Oreopoulos et al. (2012), specifically their equations 10 and 11, which express decorrelation lengths as a function of latitude and day of the year. This means that we have shown results from only a single realization of the

Raisanen generator. Yet, other decorrelation length combinations may give better agreement with observations. In this subsection, we therefore show results from a limited number of experiments applying simple scaling on the default values of the two decorrelation lengths. Specifically, we show results from eight experiments corresponding to all possible combinations of halving and doubling the default values of the two decorrelation lengths.

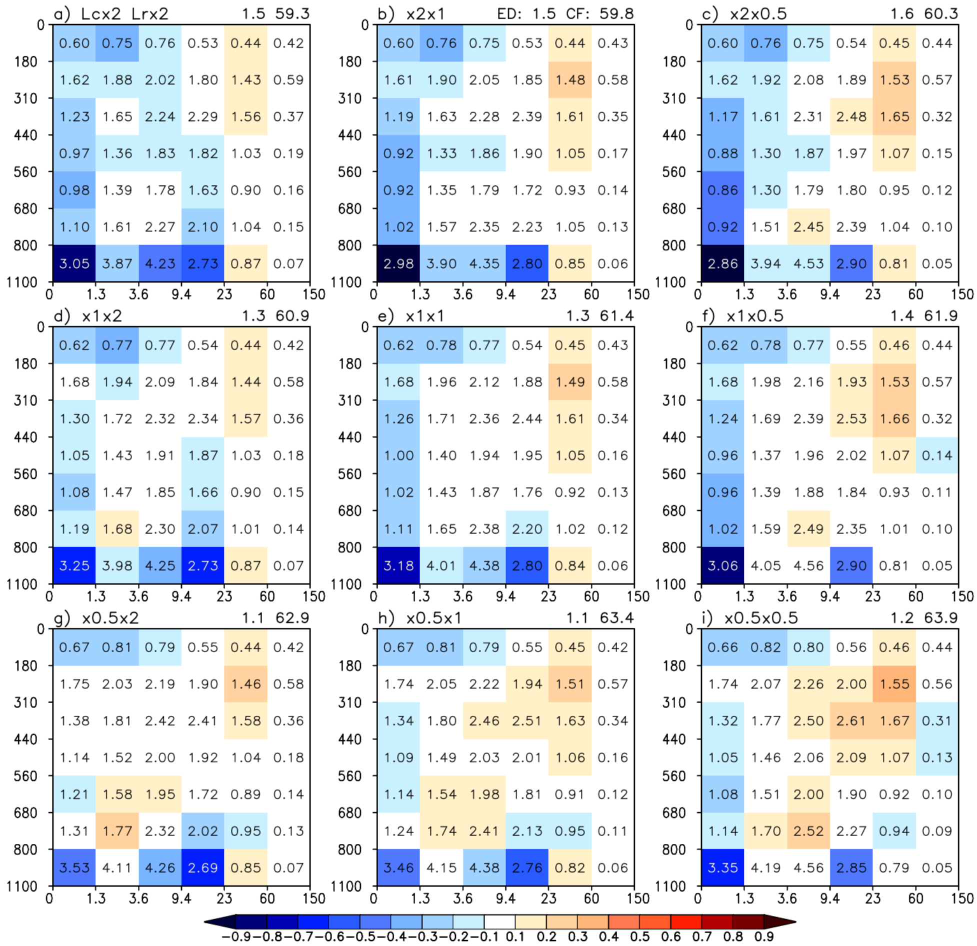

Fig. 10. Mean (2007 global average over ocean) CTP-TAU joint histograms for various experiments with the Raisanen generator using different decorrelation lengths, as indicated above each panel: “×0.5” indicates halving, “×1” same as default (Oreopoulos et al. 2012), and “×2” doubling the decorrelation length; the scaling factor for cloud occurrence decorrelation length is given first and that for COD PDF overlap second. Also, above each panel the ED of the mean histogram from observations, and the histogram CF are provided. The center plot is the default experiment, i.e., Fig. 6c.

The center plot corresponds to the default experiment previously shown in Fig. 6c. Total CF and ED shown above each panel facilitate a quick assessment of performance. It can be seen that the observed CF (64.3) can be further approached by halving the cloud occurrence overlap decorrelation length (bottom row). This makes sense, because a smaller decorrelation length makes cloud overlap more random which favors larger total CF. Doubling this decorrelation length (top row) has the opposite than desired effect: overlap becomes more maximum which decreases total CF compared to the already lower than observed CF of the default experiment. Note that CF across rows can change even if the decorrelation length for cloud occurrence overlap remains constant because of our previously discussed rejection of TAU < 0.3 subcolumns which do not count towards cloudy skies; the number of such subcolumns depends on the decorrelation length regulating COD PDF overlap.

Matching total CF better does not guarantee superior (smaller) ED, as seen by the ED value of the lower right panel which is (slightly) larger than that two panels of the same row with worse CF. The best performing experiment is probably the one where decorrelation length for COD PDF remains the same while decorrelation length for cloud occurrence is halved (middle panel of bottom row). The fact that halving the default value for cloud occurrence overlap improves results (compare middle and bottom row) is somewhat surprising because previous results suggest that cloud occurrence overlap randomizes substantially slower with cloud layer separation distance than COD PDF overlap (Räisänen et al. 2004; Pincus et al. 2005; Oreopoulos and Norris 2011; Oreopoulos et al. 2012).

4) PROFILES OF OVERLAPPED CLOUD FRACTION

The generators by design should reproduce the observed mean CF profiles. However, the cloud occurrence is imperfectly overlapped vertically and this can be captured by comparing CFs of combinations of vertical layers between observations and simulations. Relevant profiles conveying such information are those for cumulative CF and CF exposed to space as in Barker (2008), the latter actually being the profile of differences between adjacent cumulative CF values.

Figure 11 compares the profiles of mean downward and upward cumulative CF and CF exposed to space between observations and simulations as well as the profiles of root mean square differences in these quantities from the scene level data. Fig. 11 shows that in a mean sense both generators handle cloud overlap quite well. The downward cumulative cloud fraction profile (Fig. 11a) shows that the two generators handle the overlap of small CF in the upper troposphere similarly and start to diverge only at a height of about 8 km, with the cumulative CF of the Raisanen generator remaining closer to observations until abrupt CF increases around 3 km (seen as a change in the slope of the curve) bring the two generator curves closer together and with observations. This suggests errors in the CF overlap of the generators at those heights which compensate for excessive overlap aloft causing underestimates of cumulative CF. At lower levels, the generators once again overlap too much, restoring the underestimates in cumulative CF. When CF is accumulated in the other direction (panel b) the excessive overlap of the generators starts appearing at around 1 km and once established continues unabated since getting closer to the observed curve would now require a severe underestimate of overlap. The RMSE curves from scene values suggests slightly more compensating error for Raisanen.

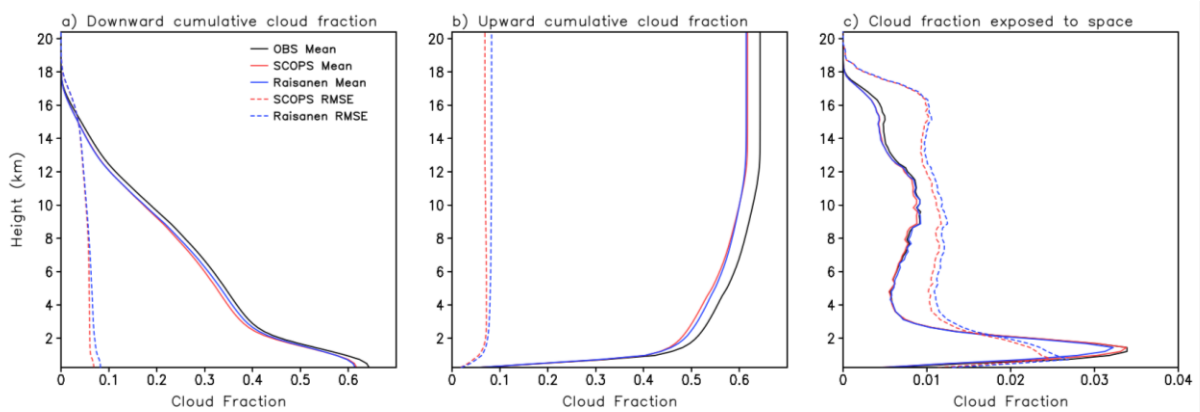

Figure 11. Downward (a) and upward (b) global (ocean) average cumulative CF for the observations and two generators. The rightmost panel (c) shows profiles of CF exposed to space. For all three panels, the profile of root mean square errors is also provided.

The rightmost panel shows the profile of CF exposed to space which peaks where the difference between two successive cumulative CF values is maximum. SCOPS outperforms Raisanen at that height, but the curves from the two generators are otherwise close and their deviations from observations are very small, until the highest levels of the troposphere where deviations re-emerge. Keep in mind that one can get good CF exposed to space even if the cumulative CF profile is biased because it is the shape of the profile that primarily matters.

5) First Order Radiative Flux Errors

Here we present a simple method for translating the observed and simulated grand-averaged CF distribution resolved in CTP-TAU space shown in Fig. 6 to a cloud radiative effect distribution (CRE, i.e., the difference between all-sky and clear-sky fluxes) in the same phase space. To accomplish this, we use the concept of cloud radiative kernels (CRKs) introduced by Zelinka et al. (2012) in which the impact of cloud on the shortwave (SW), longwave (LW) and total (combined SW+LW) radiative flux is modeled for each of the 42 bins of the CTP-TAU joint histogram. The CRKs give the change in radiative flux due to clouds (i.e., compared to clear-sky conditions) per unit CF across all bins of the joint CTP-TAU histogram. When this normalized difference between clear and overcast radiative fluxes, i.e., the overcast CRE is multiplied with the distribution of CF shown in Fig. 6, a distribution of CREs in the CTP-TAU phase space can be obtained. In our own implementation of this CRE calculation, rather than using the model-derived Zelinka et al. (2012) CRKs, we use roughly equivalent observational counterparts coming from the Clouds and the Earth Radiant System (CERES) FluxByCldTyp product (Sun et al. 2022). Specifically, from the monthly version of the product, we calculate globally-averaged ocean-only pseudo-CRKs for the year 2007. These are shown in Fig. S1. When the pseudo-CRKs are multiplied with the CF histograms of Fig. 6, the results of Fig. 12 are obtained. Above each panel of Fig. 12, we provide the global (ocean only) CRE value.

As expected, LW CRE errors are less widespread than SW errors, based on the extent of white space (indicating small errors) in the middle and right panels of the first and second row. The Raisanen generator performs remarkably well on a global basis, but is aided in the SW by non-negligible compensation of errors in individual bins; the bin errors are larger and more extensive for SCOPS and result in a substantial 2.7 Wm^{-2} global SW CRE error. Because SCOPS's LW CRE error is small, most of the SW CRE error propagates to total CRE, which as a global value is actually perfect for Raisanen. The overestimates of SW CRE by SCOPS are consistent with the previous discussion of Fig. 6, identifying underestimates of optically thin clouds and overestimates of clouds with moderate and large optical thickness. The distribution of SW CRE errors in Fig. 12 (colors) tracks the distribution of CF errors in Fig. 6 (colors) to some extent, but the one-to-one mapping is imperfect because CF errors for optically thin clouds are radiatively inconsequential. Binned LW CRE errors outside the ±0.1 Wm^{-2} range are more frequent for SCOPS, but generally rare besides the more sensitive to LW radiation high clouds. Errors in total CRE outside this range are limited to very few bins for Raisanen, but quite more frequent for SCOPS when TAU > 3.6.

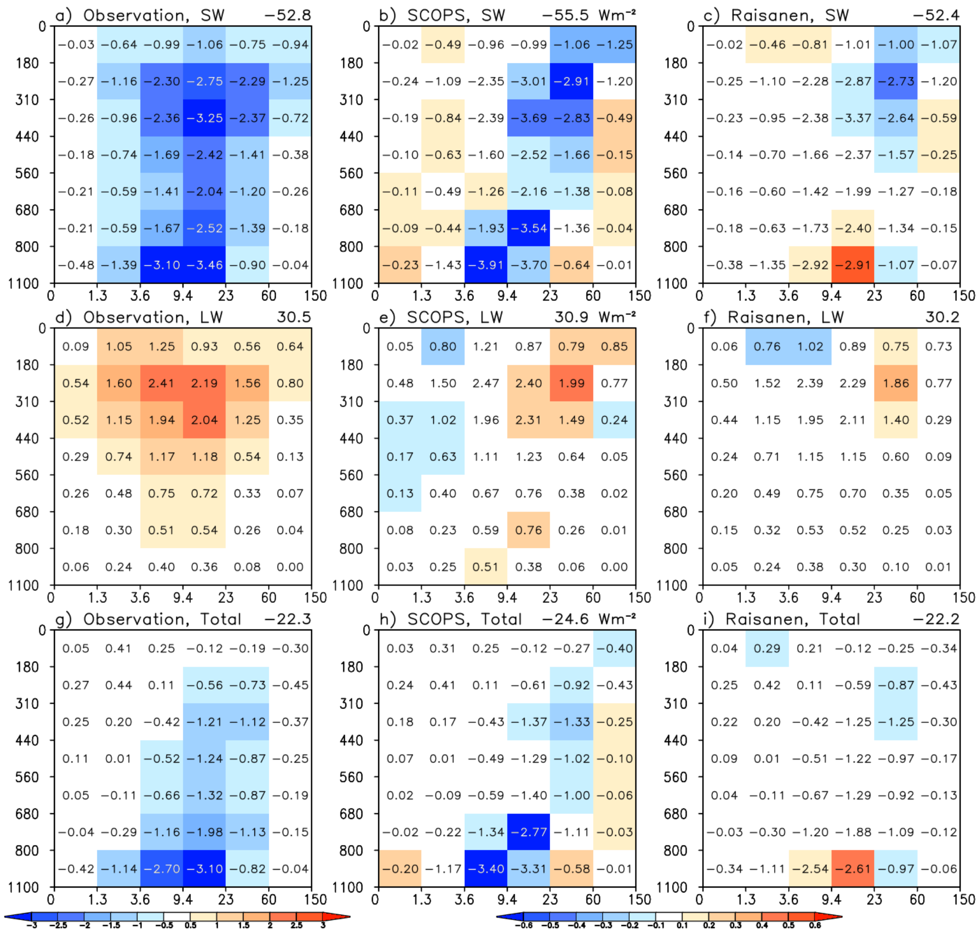

Fig. 12. Distributions of shortwave (SW), longwave (LW) and total = (SW+LW) CREs in the CTP-TAU phase space (numbers in all panels) and errors (colors in the middle and right panels) obtained by multiplying the "pseudo-CRKs" of Fig. S1 with the CF histograms of Fig. 6. Above each panel the global (ocean only) CRE value is provided.

4. Discussion and conclusions

We have created a dataset of 2D cloud optical depth fields that is extensive enough for statistical evaluation of cloud subcolumn generators used by satellite simulators in GCMs for emulating real-world subgrid variability. The dataset is based on active CloudSat-CALIPSO global observations for the year 2007. The appropriateness of such observations compared to those obtained from passive instruments stems mainly from their ability to resolve the vertical (height-dependent) variability of clouds; passive observations typically provide only column-integrated quantities. Our dataset was constructed by combining two CloudSat products. Upon doing this, it quickly became apparent that the portion of the cloud field contributed by liquid clouds required corrections as it was found to lack retrievals for a non-negligible fraction of cells identified to be cloudy and of liquid phase in the CALIPSO-enhanced CloudSat 2B-CLDCLASS-LIDAR product. We came up with a relatively simple filling scheme for the missing clouds whose performance was assessed via comparisons with MODIS-Aqua cloud retrievals. The comparisons were based on discretized cloud fraction (CF) distributions in the form of joint histograms of cloud top pressure (CTP)-cloud optical thickness, (TAU), and even encompassed a MODIS cloud regime (CR) segregation in order to gauge possible dependences on cloud types.

The reconstructed 2D cloud optical depth fields including our improvements were segmented into scenes comprising 100 subcolumns (~110 km) and whose mean CF and TAU profiles were passed to the two generators to produce their own set of 100 subcolumns as a simulated version of the scene. All three sets of subcolumns were passed to COSP's MODIS simulator which transformed them to joint CTP-TAU scene histograms for potential averaging across spatiotemporal scales. Note that there is no fundamental reason the generators should be configured to also produce 100 subcolumns, since individual subcolumns are never compared, but rather scene-level cloud field properties. When we experimented with different numbers of subcolumns, we noticed a slow progressive improvement in performance as the subcolumn number grew, but without much benefit above 100 subcolumns. Actually, on the opposite side of fewer subcolumns, even as few as 20 would not have affected the results of this study substantiatively.

A sensible way of assessing the performance of the two generators in terms of joint histograms is to compare the Euclidean distances of their grand averages against their observational counterpart (which again can be made more detailed by applying a CR breakdown), or statistics of EDs coming from individual scene histograms. Simpler comparisons bypassing joint histograms altogether are of course also possible using vertically-integrated quantities directly derived from the scene's subcolumns, such as vertically-projected cloud fraction, (logarithimic) mean scene TAU, and variance of TAU.

Both types of comparisons described above reveal a clear superiority of the Raisanen generator in our default implementation, i.e., using the parameterization of decorrelation lengths for cloud occurrence and cloud optical depth PDFs proposed by Oreopoulos et al. (2012). Sensitivity experiments allowing the decorrelation lengths to vary unveil that simple modifications to this parameterization may yield even better results. Nevertheless, the main deficiencies of both generators, namely an overestimation of overlap and an underestimation of the occurrence of optically thin clouds remains a persistent theme. For SCOPS, this error combines with overestimates of optically thicker clouds yielding an overestimate of shortwave cloud radiative effect that is close to 3 Wm^{-2} on global scales, but for Raisanen where optically thicker clouds are better simulated, the shortwave CRE error remains contained.

We are fully aware that vertically integrated quantities and joint histograms describing how integrated extinction and location of the highest cloud co-vary cannot be viewed as the only way to evaluate subgrid variability simulated by generators. Even in a two-dimensional world, the subgrid-scale profiles of cloud occurrence and cloud (liquid and ice) condensate and particle size should be well-reproduced since they may matter for radiative heating rate profiles and the physical parameterizations developed for GCMs. We have taken a first step by evaluating subgrid cloud occurrence profiles via grand-average profiles of cumulative cloud fraction and cloud fraction exposed to space, which revealed satisfactory skill of similar parity for both generators. Nevertheless, one should keep in mind that knowing and being able to simulate the subgrid variability of a wider range cloud properties is imperative for an assessment of subgrid realism based on radiative flux and heating rate profiles. We anticipate to be in a position to confront the simulators with such stricter tests in future endeavors. Efforts of this kind would also benefit by a wider range of choices of empirical and easy to use cloud subcolumn generators that operate on the (more readily available from observations) condensed part of total water content, criteria that are unfortunately not met by other existing generators (Norris et al. 2008; Larson and Schanen 2013).

Acknowledgments.

All authors gratefully acknowledge support by NASA's CloudSat and CALIPSO Science Team program. The research by ML was carried out at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, under contract with the National Aeronautics and Space Administration. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

Data Availability Statement.

The CloudSat products used in this study are available at https://www.cloudsat.cira.colostate.edu. The CERES FluxByCldType product is available at https://ceres.larc.nasa.gov/data/. The COSP simulator which includes the MODIS simulator can be downloaded from https://github.com/CFMIP/COSPv2.0.

REFERENCES

Barker, H. W., Stephens, G. L., and Fu, Q., 1999: The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry, Q. J. Roy. Meteor. Soc., 125, 2127–2152.

Barker, H. W., 2008: Overlap of fractional cloud for radiation calculations in GCMs: A global analysis using CloudSat and CALIPSO data, J. Geophys. Res., 113, D00A01, doi:10.1029/2007JD009677.

Bodas-Salcedo, A., et al., 2011: COSP: Satellite simulation software for model assessment. Bull. Amer. Meteor. Soc., 92 (8), doi:10.1175/2011BAMS2856.1.

Cho, N., J. Tan, and L. Oreopoulos, 2021: Classifying Planetary Cloudiness with an Updated Set of MODIS Cloud Regimes, J. of Appl. Meteor. and Clim., 60(7), 981-997, https://doi.org/10.1175/JAMC-D-20-0247.1.

Deng, M., G. Mace, Z. Wang, and E. Berry (2015), CloudSat 2C-ICE product update with a new Ze parameterization in lidar-only region, J. Geophys. Res. Atmos., 120, 12,198–12,208, doi:10.1002/2015JD023600.

Hillman, B. R., R. T. Marchand, and T. P. Ackerman, 2018: Sensitivities of Simulated Satellite Views of Clouds to Subgrid-Scale Overlap and Condensate Heterogeneity. J. Geophys. Res. Atmos., 123, 7506-7529, doi:10.1029/2017JD027680.

Hogan, R. J., and A. J. Illingworth, 2000: Deriving cloud overlap statistics from radar. Quart. J. Roy. Meteor. Soc., 126, 2903–2909.

Jakob, C. and Klein, S. A. 1999 The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ECMWF model. Q. J. R. Meteorol. Soc., 125, 941–965.

Jakob, C. and Tselioudis, G., 2003. Objective identification of cloud regimes in the tropical western Pacific. Geophys. Res. Lett., 30(21), https://doi.org/10.1029/2003GL018367.

Larson, V. E., and D. P. Schanen, 2013: The Subgrid Importance Latin Hypercube Sampler (SILHS): A multivariate subcolumn generator. Geosci. Model Dev., 6, 1813–1829, https://doi.org/10.5194/gmd-6-1813-2013.

Leinonen, J., Lebsock, M. D., Stephens, G. L., & Suzuki, K. (2016). Improved retrieval of cloud liquid water from CloudSat and MODIS. Journal of Applied Meteorology and Climatology, 55, 1831–1844. https://doi.org/10.1175/JAMC-D-16-0077.1.

Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens (2008), Hydrometeor detection using Cloudsat — an Earth-orbiting 94-GHz cloud radar, J. Atmos. Oceanic Technol., 25, 519–533, doi:10.1175/2007JTECHA1006.1.

Norris, P. M., L. Oreopoulos, , A. Y. Hou, , W.-K. Tao, , and X. Zeng, 2008: Representation of 3D heterogeneous cloud fields using copulas: Theory for water clouds. Quart. J. Roy. Meteor. Soc., 134, 1843–1864.

Oreopoulos, L., and P. Norris, 2011: An analysis of cloud overlap at a midlatitude atmospheric observation facility. Atmos. Chem. Phys., 11, 5557–5567, doi:10.5194/acp-11-5557-2011.

Oreopoulos, L., D. Lee, Y. C. Sud, and M. J. Suarez, 2012: Radiative impacts of cloud heterogeneity and overlap in an atmospheric General Circulation Model. Atmos. Chem. Phys., 12, 9097–9111, doi:10.5194/acp-12-9097-2012.

Pincus, R., H. W. Barker, and J. J. Morcrette, 2003: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. J. Geophys. Res., 108 .4376, doi:10.1029/2002JD0x03322.

Pincus, R., C. Hannay, , S. A. Klein, , K.-M. Xu, , and R. S. Hemler, 2005: Overlap assumptions for assumed probability distribution function cloud schemes in large-scale models. J. Geophys. Res., 110, D15S09, doi:10.1029/2004JD005100.

Pincus, R., S. Platnick, S. A. Ackerman, R. S. Hemler, and R. J. P. Hofmann, 2012: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and and the limits of instrument simulators. J. Climate, 25, 4699–4720, doi:10.1175/JCLI-D-11-00267.1.

Platnick, S., and et al. , 2017: The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua. IEEE Trans. Geosci. Remote Sens., 55, 502–505, doi:10.1109/TGRS.2016.2610522.

Räisänen, P., Barker, H. W., Khairoutdinov, M., Li, J., & Randall, D. (2004). Stochastic generation of subgrid-scale cloudy columns for large-scale models. Quarterly Journal of the Royal Meteorological Society, 130, 2047–2068.

Rodgers, C. D. (2000), Inverse Methods for Atmospheric Sounding: Theory and Practice. Series on Atmospheric and Oceanic and Planetary Physics, Vol. 2, World Scientific, 256.

Rogers, R. R., Hostetler, C. A., Hair, J. W., Ferrare, R. A., Liu, Z., Obland, M. D., Harper, D. B., Cook, A. L., Powell, K. A., Vaughan, M. A., and Winker, D. M.: Assessment of the CALIPSO Lidar 532 nm attenuated backscatter calibration using the NASA LaRC airborne High Spectral Resolution Lidar, Atmos. Chem. Phys., 11, 1295–1311, https://doi.org/10.5194/acp-11-1295-2011, 2011.

Sassen, K., and Z. Wang, 2008: Classifying clouds around the globe with the CloudSat radar: 1 year of results. Geophys. Res. Lett., 35, L04805, https://doi.org/10.1029/2007GL032591.

Sassen, K., and Z. Wang, 2012: The clouds of the middle troposphere: Composition, radiative impact, and global distribution. Surv. Geophys., 33, 677–691, https://doi.org/10.1007/s10712-011-9163-x.

Song, H., Z. Zhang, P. L. Ma, S. J. Ghan, and M. Wang, 2018: An evaluation of marine boundary layer cloud property simulations in the Community Atmosphere Model using satellite observations: Conventional subgrid parameterization versus CLUBB. J. Climate, 31, 2299–2320, https://doi.org/10.1175/JCLI-D-17-0277.1.

Stein, T. H. M., J. Delanoë, and R. J. Hogan, 2011: A Comparison among Four Different Retrieval Methods for Ice-Cloud Properties Using Data from CloudSat, CALIPSO, and MODIS, J. Appl. Meteor. Clim., 50(9), 1952-1969, https://doi.org/10.1175/2011JAMC2646.1.

Sun, M, D. R. Doelling, N. G. Loeb, R. C. Scott, J. Wilkins, L. Nguyen, and P. Mlynczak, 2022: Clouds and the Earth's Radiant Energy System (CERES) FluxByCldTyp Edition 4 Data Product. J. Atm. Ocean. Tech., https://doi.org/10.1175/JTECH-D-21-0029.1.

Zelinka, M., , S. A. Klein, , and D. L. Hartmann, 2012: Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. J. Climate, 25, 3715–3735.