Authors' response on "An evaluation of clouds and radiation in a Large-Scale Atmospheric Model using a Cloud Vertical Structure classification" by Dongmin Lee et al.

We thank Referees and Editor for very insightful and constructive comments on our discussion paper. All the points are well taken, and we have made a number of revisions to the paper to clarify the points raised in the review. The point by point responses and description of changes are included in the authors' response. And a marked-up manuscript version showing the changes made is attached.

Anonymous Referee #1
Received and published: 30 October 2019
Disclaimer: This is my first review of a GMD manuscript. In addition, my background is in atmospheric observations rather than numerical modelling. These circumstances might be reflected in my comments.
Lee at al. describe a method to assess the representation of clouds and their radiative effects in a global model based on A-Train observations and sub-column generators. At least the latter is not within my field of expertise and I had quite some difficulties in understanding how these sub-column generators work. However, I'd leave the assessment of the model-related methodology to an expert reviewer. In terms of presentation, there are a couple of items below that might make it easier for the reader to follow the reasoning of the authors. Finally, the use of jargon and abbreviations quite affect the readability of the text.

Detailed comments:
It has not been clear to me from reading the abstract what this paper is about, particularly in terms of quantitative findings.

A: We have revised the abstract extensively and have now added some specific results. We hope that the revised version makes now clearer what the paper is about.

page 4, line 12/13: Cloudsat and CALIPSO have collected a much longer time series. Wouldn't it be better to base the analysis on 10+ years of observations?

A: We use the same four-year period (2007-2010) of the merged CloudSat/CALIPSO product 2B-CLDCLASS-LIDAR R04 as our previous paper, Oreopoulos et al. (2017), for consistency. This is considered the "golden period" of the merged product as it includes both daytime and nighttime CloudSat observations. CloudSat suffered a battery anomaly in early 2011, and has been conducting only daytime observations since operations resumed.

page 5, lines 4-6: To me the fact that the model produces more than the generally observed four cloud layers sounds like there might be an issue with properly representing clouds in the model. Please comment.

A: First of all, complicated cloud vertical structures are being simplified in both the observation-based product and in the model. Also, the cloud scheme of GEOS-5 is not fundamentally

different than counterpart schemes in similar models. The version of the model used in this study has 72 vertical layers, and clouds can form in about 30 of them. But unlike observations, stratiform modeled clouds form in the vertical layers of a column independently of each other, so that the modeled cloud profiles are typically noisier compared to the smoother vertically coherent profiles of observations. The model is basically forced to produce clouds under the limitation of an atmospheric column that is vertically discretized because the model is built on the 3D grid.

page 5, line 10 - page 7. line 3: I'll leave commenting on this part to modelling experts. Suffice to say that I don't understand from this text how the two sub-column generators work.

A: Subcolumn generators are standard tools in the modeling world to create subgrid cloud variability that the model does not prognosticate or otherwise resolve. They are used for both model parameterizations (radiation, cloud microphysics) and diagnostics such as cloud simulators (the COSP satellite simulator, for example, has a subcolumn generator). Subcolumn generators need rules on how to account for vertical cloud fraction overlap and cloud condensate horizontal inhomogeneity (and how it overlaps in height).

page 7, line 21: not sure maritime continent is used correctly here, but I might have mixed up plots

A: We refer to the oceanic regions around the Maritime Continent. In that region, clear columns are not frequently observed in CloudSat-CALIPSO, but the model seems to be producing them quite frequently.

page 9, line 16: If I understand correctly, this is a comparison of the findings of the two sub-column generators. What is your benchmark for stating that one method produces underestimates?

A: The reviewer's point is well-taken. This refers to just a difference between two subcolumn generators and we have therefore rephrased.

page 10, lines 1-4: This statement should be moved to the figure caption. Also, state at the start of the discussion what kind of averaging has been applied.

A: We have indeed expanded the figure caption using some of this text. The averaging method for the cloudy column has actually been provided before the discussion, see p. 9, lines 25-26.

page 11, discussion of Figure 5: Do I understand correctly that Figure 4 refers to values for all instances at which the respective cloud types are present while Figure 5 presents those numbers normalized by the occurrence rate of the respective cloud types? This information is somewhat hard to extract from the text (but might be due to use of jargon). Also, what is meant with grid-mean? I have really no idea what global grid-mean is supposed to be.

A: Yes, the reviewer's understanding is correct. The term 'grid-mean' is commonly used in climate modeling to indicate averaging over both the clear and cloudy part of a grid cell. Given the confusion of the reviewer, we re-considered our terminology and now use instead "all-sky' CRE for the RFO-weighted CRE of a CVS class to distinguish it from the cloudy column or overcast CRE. This "all-sky" terminology conveys the fact that subcolumns with no cloud or a CVS of another class have zero contribution to the CRE of the CVS class of interest when RFO weighting is used. Furthermore, "Cloudy-column" (CVs-specific) CRE is now referred to as "overcast" CRE throughout the text, and its mean represents the average CRE of a CVS class when it occurs.

page 12: I understand that f and r are expressions for parameters that have been used earlier in the manuscript. Why not introduce them at the first instance they are being used? It would be much less confusing if you would be consistent with the naming parameters throughout the text!

A: Thank you for the suggestion which we have implemented.

Please make use of the sub-figure labelling (a, b, c, etc.) in the discussion of the findings. This would make it much easier for the reader to follow your argument.

A: The text has been revised to include panel labelling.

page 13, lines 23-28: This is a somewhat disappointing conclusion. What's the contribution of this study apart from anything's possible?

A: It may be a somewhat disappointing conclusion, but the existence of compensating errors in GCMs is a well-known fact that is confirmed yet again and documented with substantial detail in this paper. Nevertheless, our results contain specific findings that point to specific clouds processes that need to be fixed in order to improve the overall performance of the model.

page 20, line 19: please update reference.

A: Done, thank you.

Figure 1: I took notice of the dotted lines in the figure only after trying to make sense of the combination of classes from your description myself. The purpose of those lines is not stated anywhere. Please add: dotted lines show which cloud classes have been combined for this study.

A: Thank you again, we added this in the figure 1 caption.

Figures 2 and 3: I would suggest to revise these figures in a more intuitive way. For instance, you could have columns with high clouds on top and low clouds at the bottom. Also have all cases in Figure 2a in one column, then the plots of Figure 2b next to it in another column, and then the findings of Figure 3 in a third column. However, Figure 3 seems pointless to me as it presents the

differences between the two sub-column generators. It would be better to present the difference to the observations for each sub-column generator, i.e. as an individual column added to Figure 2. I would also prefer to see plots from -180 to +180 degree. Finally, please elaborate on the global mean values. Shouldn't they add up to 100%. They don't right now.

A: First of all, we like maps from 0 to 360 (centered at the dateline) better because they accentuate tropical deep convection around the ITCZ. Second, both observed and modeled CVS RFOs (including clear skies) add up to 100% within rounding error. Regarding the re-organization of the panels, we prefer to keep it the way it is, i.e., moving from clear to single layers CVS, to two-layer CVS, and finally to the three-layer CVS (simple to more complex). Subtracting two model results (Fig. 3) allows us to obtain more spatially coherent results (since the various CVS classes tend to occur in the same regions) and to use a more limited range of differences for the colorbar, thus amplifying the difference maps. If we were to subtract each model configuration from observations, we would need double the number of panels.

Figures 4 to 6: Please add a line at zero. See also my comments regarding the discussion on page 11. In Figure 6, you should at a statement to the caption regarding the meaning of the grey bars.

A: Good suggestion to add a zero line, which we have implemented. All bars in Fig. 6, including the gray bar have been defined by the legend and are now defined again in the expanded caption.

Figures 7 to 9: I'd suggest to refer to the two lines as observations and model.

A: Done.

The simulated cloud radiative effects (CRE) are commonly biased in most climate models. This study, with the aid of both CloudSat/CAPLIPSO retrievals and the implemented subcolumn cloud generator in NASA's GEOS-5, explores the CRE biases in association with cloud vertical structure classification. Results show that while the simulation of global CRE is in much agreement with observation, the CRE due to different cloud classification is not. Moreover, by decomposing model's CRE errors into components stemming from biases in RFO and cloudy-column CRE, the relatively good simulations of global grid-mean CRE largely benefit from compensating errors in these two terms. The method introduced in this paper can be used in other models to explore their CRE behavior, thus beneficial for the modeling community.
While I generally find the manuscript suitable for publication in GMD, further improvements are needed before the manuscript is accepted. Below I have included a list of the major comments that I think should be addressed, followed by a list of more specific comments.

Major comments:
1. The study uses 2B-FLXHR-LIDAR product as a guiding reference for CRE, which was obtained by invoking radiative transfer algorithm operating on thermodynamical fields from re-analysis and cloud properties from CloudSat/CAPLIPSO retrievals. As the authors mentioned, the SW CREs in 2B-FLXHR-LIDAR is strongly time dependent. It is worth to add CERES data as a reference as well, since a great many models are commonly tuned to resemble CERES observations.

A: The CVS class occurrences and associated radiative fluxes from 2B-FLXHR-LIDAR (obtained in the way that the reviewer describes) are defined at the scale of individual CloudSat profiles, i.e., at approximately 2 km horizontal resolution. Producing the observed global/zonal average results in our study still requires this flux-CVS type mapping at the CloudSat profile horizontal scale. That is the main reason we cannot use CERES: even at the footprint level, CERES radiative fluxes are coarser than the scale at which a CVS occurrence is defined, so such a mapping cannot be achieved. In the model, it is the subcolumn generator that allows us to define CVS occurrences within a grid cell.

2. Besides cloud overlap assumptions, the cloudiness vertical profile per se is important for the determination of CVS classification. The authors are suggested to replenish the role of layer cloud fraction when revising the paper. The CVS classification may suffer from poor representation of subgrid cloud condensation and/or overlap assumption. In addition, the vertical resolution in GCMs is typically coarser than that in CloudSat/CALIPSO retrievals, which to some extent plays an important role in calculating RFO. The authors need point out this in the paper.

A: The CVS classification of Oreopoulos et al. (2017) used in this paper does not depend on cloud fraction overlap because layer cloud fractions are either zero or one in the 2B-CLDCLASS-LIDAR profiles. In other words, we are considering only cloud *occurrence* overlap at different vertical layers to assign cloud profiles to a CVS class. We just wanted to clarify this point once again. But we believe that what the reviewer refers to here is the quality of the model cloud

profiles passed to the subcolumn generator. These may indeed affect our results. So, what we are really evaluating here is both the quality of the model's cloud scheme in terms of mean profiles, but also the quality of the subcolumn generator to translate these mean profiles into subgrid profiles. The latter issue is partially addressed in this paper by changing the overlap assumption in the subcolumn generator from generalized to maximum-random. One can of course come with other options to create the subgrid columns, and even for generalized overlap, the decorrelation length is a tunable parameter. But both versions of the generator receive as input the same mean profiles, which may be problematic. There is flexibility in changing the generator, but testing different model cloud ("moist") schemes is a bigger effort. Good point about different vertical resolutions. We have added a sentence in the discussion of the concluding section to make the point that "garbage-in" (deficient mean cloud profiles) – "garbage out" (the subcolumn generator will not rectify faulty input profiles).

3. When comparing the two overlap assumptions, the GN assumption yields more clouds than MR in almost all cloud regimes except for isolated low clouds in extratropics. Why does this occur? Is this because clouds in this region are nonadjacent separated by clear skies that have a vertical distance smaller than specified decorrelation length in GN?

A: What the reviewer suggests is certainly possible and makes sense. As shown in Oreopoulos et al. (2017), when multiple distinct cloud layers exist in a standard layer, these are still treated as a single entity in the CVS classification. So, the MR scheme sees them as separate layers (random overlap), while the generalized scheme gives a combined fraction less than random because the separation distance is small.

Specific comments:
1. The abbreviation "RFO" is not fully spelled at the first place in the main text.

A: We fixed this, thanks.

2. P9L1 regions pronounced orography -> regions with pronounced orography
P9L17 exceeds -> exceed 4. P27 As Fig. 3-> As Fig. 4.

A: Corrections applied, thanks.

An evaluation of clouds and radiation in a Large-Scale Atmospheric Model using a Cloud Vertical Structure classification

by

Dongmin Lee[1,2], Lazaros Oreopoulos[2], and Nayeong Cho[3,2]

1. Morgan State University

2. NASA Goddard Space Flight Center

3. University Space Research Association

Confidential manuscript revised for the
Geoscientific Model Development

Deleted: submitted to

Corresponding author address:

Dongmin Lee

NASA-GSFC

Code 613

Greenbelt MD 20771

Dongmin.Lee@nasa.gov

Abstract

We revisit the concept of Cloud Vertical Structure (CVS) classes we have previously employed to classify the planet's cloudiness (Oreopoulos et al., 2017). The CVS classification reflects simple combinations of simultaneous cloud occurrence in the three standard layers traditionally used to separate low, middle, and high clouds and was applied to a dataset derived from active lidar and cloud radar observations. This classification is now introduced in an Atmospheric Global Climate Model, specifically a version of NASA's GEOS-5, in order to evaluate the realism of its cloudiness and of the radiative effects associated with the various CVS classes. Such classes can be defined in GEOS-5 thanks to a subcolumn cloud generator paired with the model's radiative transfer algorithm, and their associated radiative effects can be evaluated against observations. We find that the model produces 50% more clear skies than observations in relative terms, and produces isolated high clouds that are slightly less frequent than in observations, but optically thicker, yielding excessive planetary and surface cooling. Low clouds are also brighter than in observations, but underestimates of the frequency of occurrence (by ~20% in relative terms) help restore radiative agreement with observations. Overall the model reproduces better the longwave radiative effects of the various CVS classes because cloud vertical location is substantially constrained in the CVS framework.

1. Introduction

The large impact of clouds on the Earth's radiation budget and the growing wealth of satellite-based cloud observations are strong motivators for their systematic assessment in climate models. Such evaluation exercises focus on either cloud properties, or metrics of cloud radiative impact, or ideally on both (Pincus et al., 2008; Nam et al., 2012; Klein et al., 2013; Wang and Su, 2013; Dolinar et al., 2015).

Assessments of cloud properties with satellite observations are not always straightforward for a variety of reasons such as inability to define in the model a particular satellite-observed property, or limitations in the satellite observations. For example, the vertically integrated cloud optical depth of the cloudy portion of a model grid cell is an ill-defined quantity that cannot be obtained trivially from the model's optical depth profile since it is intimately associated with a cloud fraction profile, making thus layer optical depths relevant for only the cloudy portions of the grid cell which vary by model height and can conceptually be vertically aligned in various ways. In contrast, vertically-integrated cloud optical depth is quite robustly defined in observations since it is measured with passive imagers at a much higher resolution for which overcast conditions can be more safely assumed. Issues such as these have led to the development of "satellite simulators" that transform Global Climate Model (GCM) cloud fields to forms that are closer analogues to their counterparts observed by satellites (e.g., the COSP simulator – Bodas-Salcedo et al., 2011).

The quality of simulated clouds in GCMs can also be measured in terms of the realism of their radiative impact using quantities such as the Cloud Radiative Effect (CRE), i.e., the difference between all-sky and clear-sky fluxes at the spatial scales of a model grid cell (Wang and Su, 2013). This type of comparison can be performed at a variety of spatiotemporal scales and is often quite illuminating, but the interpretation of findings can suffer from inconsistencies in how the estimates are obtained for satellites and models.

This paper is yet another attempt to evaluate clouds in an atmospheric GCM (AGCM), specifically a version of the Goddard Earth Observing System version 5 (GEOS-5) model (Rienecker et al., 2008; Molod et al., 2012), a multi-purpose global model that is used for a variety of applications. Both approaches of cloud assessment are used, namely comparison of

3

Deleted: is

Deleted: a

Deleted: in the model

Deleted: change at different

Deleted: levels

Deleted: assuming

Deleted: is a far better approximation

Deleted: a

the cloud fields themselves, but also comparison of cloud radiative impacts. Our cloud property

evaluation focuses on a single aspect of cloudiness: Cloud Vertical Structure (CVS). The

comparison is possible because of recent progress in two areas: active cloud remote sensing

which makes resolving cloud vertical profiles possible; and the development of schemes

(subcolumn generators) that create subgrid cloud vertical structures in GCMs. Being able to

categorize clouds in terms of a few CVS categories facilitates the comparison between

observations and models and enables a more rigorous CRE comparison that evaluates the

model's skill with regard to how it simulates the radiative impact of individual CVS classes.

2. Data and methodology

The observational reference dataset of CVS class occurrence and associated radiative fluxes is

essentially the same as Oreopoulos et al. (2017), hereafter O17, and spans four years (2007-

2010.) A schematic illustration of the original CVS classes of O17 is reproduced here as Fig. 1.

The details of how cloud layer boundaries available in the 2B-CLDCLASS-LIDAR R04 dataset

(Sassen and Wang, 2012, see also http://tinyurl.com/2b-cldclass-lidar), a joint product coming

from CloudSat and CALIPSO (hereafter, "CC") active cloud radar and lidar observations, were

interpreted as cloud layer profiles belonging to one of these classes are described exhaustively

in the appendix of O17. The definition of the CVS classes hinges on defining broad categories of

high H, middle M, and low L clouds which are confined to three standard atmospheric layers,

one above 440 hPa, another between 680 and 440 hPa and yet another below 680 hPa,

respectively. The vertical level boundaries defining these standard layers come from the

International Satellite Cloud Climatology Project (ISCCP), (Rossow and Schiffer, 1991). The

reference radiative fluxes come from the 2B-FLXHR-LIDAR R04 CC product (L'Ecuyer et al., 2008;

Henderson et al., 2013; Matus and L'Ecuyer, 2017) and are obtained from a radiative transfer

algorithm operating on observed and re-analysis output which has at its core retrieved CC cloud

properties.

For the purposes of this study, the CVS classes have been reduced to seven by merging

the CVS classes for which clouds occur simultaneously in the same two or three standard

adjacent layers (all multi-layer CVS classes other than "HL"). In other words, we do not

distinguish any longer between CVS classes with clouds occurring in the same adjacent standard

4

Deleted: one

1   layers, even if those were previously discerned based on whether or not a clear layer of

2   substantial vertical extent was present to separate the cloud layers. This means in practice that

3   we do no longer distinguish (cf. Fig. 1) between CVS = "H×M×L" and "HML" (now simply

4   "HML"), CVS = "H×M" and "HM" (now simply "HM"), CVS = "M×L" and "ML" (now simply "ML").

5   The reason for reducing the CVS classes to seven from the original ten is the complexity of the

6   model cloud profiles which can consist of numerous distinct cloud layers and which therefore

7   renders the O17 CVS classification scheme inapplicable. The original scheme was designed for

8   observed cloud profiles from CC that rarely (less than 1% of the time) consisted of more than

9   four distinct cloud layers in which case they were either ignored or processed only in the

10 simplest of cases (such as multiple individual cloud layers residing within a single standard layer

11 – see appendix of O17).

12   A prerequisite for the evaluation of GEOS-5 clouds in terms of their CVS class frequency

13 and the CRE statistics associated with these CVS classes is creating comparable datasets.

14 Assigning CVS classes to grid cell GCM cloud fields is not possible without manipulation of the

15 GCM's cloud profiles. To this end, we use the cloud subcolumn generator that is paired with the

16 RRTMG-LW and RRTMG-SW radiative transfer codes (Mlawer et al., 1997; Iacono et al., 2008) in

17 the model's Monte Carlo Independent Column Approximation (McICA; Pincus et al., 2003)

18 implementation. This subcolumn generator follows Räisänen et al. (2004) and can produce

19 subcolumns that are consistent with specific assumptions about the vertical overlap of both

20 cloud fraction and the horizontal distributions of cloud condensate. While the latter type of

21 overlap is irrelevant to CVS class frequency statistics, it does matter for the radiative transfer

22 calculations producing the radiative fluxes used to estimate CREs. The 140 subcolumns created

23 by the model's generator (which match the number of "g points" in RRTMG-LW's correlated-k

24 scheme) are essentially assumed equivalent to the cloud profiles viewed by the active

25 instruments (CALIPSO's lidar and CloudSat's radar) and whose vertical location information is

26 recorded in the 2B-CLDLASS-LIDAR product. Herein, we will show results from two types of

27 cloud fraction overlap schemes that have been implemented in the cloud subcolumn generator,

28 generalized (GN) overlap, also known as exponential overlap (Hogan and Illingworth, 2000;

5

Oreopoulos and Norris, 2011) and maximum-random overlap (MR overlap, Geleyn and

Hollingsworth, 1979).

The model, GEOS-5 tag "Jason-2_0" was run with fixed sea surface temperatures (SSTs)

for the same period as the reference dataset, 2007-2010. The model integration was driven by

radiative fluxes and heating rates produced by applying generalized overlap in the radiation

calculation. RRTMG-LW and RRTMG-SW were called for an additional set of flux calculations

using this time the MR overlap assumption to produce cloudy subcolumns, but only in

diagnostic mode, i.e., the generated fluxes served only diagnostic purposes and were not

passed back to the model to influence the evolution of its energetics and dynamics. This way,

with one interactive and one diagnostic call to the RRTMG codes, we were able to obtain two

sets of CVS diagnostics and corresponding CREs. In both cases, the subcolumns come from a

common mean cloud fraction and condensate profile. The layer condensates are assumed to

possess horizontal subgrid condensate heterogeneity as prescribed in Oreopoulos et al. (2012).

This subgrid condensate variability affects the model's CRE distribution, but not the CVS fields

and statistics.

In the subcolumn generator, the decorrelation length (e-folding distance) for the

generalized overlap scheme was set to vary zonally as described in Oreopoulos et al. (2012).

The physical meaning of the decorrelation length is that cloud layers separated by a distance

equal to the decorrelation length overlap as a mixture of maximum and random overlap in e

($\approx$0.368) and 1-e ($\approx$0.632) proportions (weights), respectively. At distances greater (smaller)

than the decorrelation length the contribution of random (maximum) overlap contribution

increases (decreases) compared to the above values. In the limit of zero separation cloud

overlap is purely maximum, while in the limit of infinite distance overlap is purely random. The

zonal prescription of decorrelation length by Oreopoulos et al. (2012) is based on CloudSat

observations and is meant to capture a more coherent vertical cloud alignment (i.e., more

maximum overlap and greater decorrelation length) at low latitudes compared to high

latitudes, also seen by Barker (2008). This formulation of overlap is an alternative to maximum-

random overlap which was the standard popular choice in earlier years. The Geleyn and

Hollingsworth (1979) implementation of MR overlap in our Räisänen et al. (2004)-based

Deleted: "

Deleted: the run for which

Deleted: was assumed

Deleted: returned

Deleted: obtained

Deleted: the same

Deleted: at

1 generator allows for random overlap even within a "block" of contiguous clouds: immediately

2 adjacent clouds are maximally overlapped, but non-adjacent clouds within the contiguous block

3 can have portions that are randomly overlapped if there is a local minimum in cloud fraction

4 between them; random overlap applies for those cloudy portions that do not fully overlap with

5 the in-between layers. This type of MR overlap should be contrasted with other

6 implementations (e.g., Chou et al., 1998) where maximum overlap always takes place within

7 the block while the various distinct blocks of the atmospheric column (always separated by

8 clear layers) are themselves randomly overlapped.

9

10 **3. GEOS-5 cloud evaluation with CVS**

11 *a. Climatological CVS occurrence*

12 Figure 2 compares the observed and simulated (from GN overlap) multi-annual maps of

13 the relative frequency of occurrence (RFO) for all seven CVS classes of our study. The observed

14 fields are sampled at rather coarse 4°x4° scales to compensate for the substantial sparseness of

15 the active observations (gridding at higher resolutions would make for relatively noisy maps).

16 Above each panel, we provide the area-weighted RFO global mean of the CVS (equivalent to its

17 global cloud fraction). These fields include nighttime observations and simulations since the

18 former are possible for active sensors and the latter are passed as input for the model's

19 nighttime RRTMG-LW calculations.

20 Before examining consistency (or lack thereof) for cloud fields, we first turn our attention

21 to clear skies. We note that the observations suggest a cloudier world with clear skies occurring

22 only ~25% of the time (or, alternatively, covering 25% of the global area between 82°S and

23 82°N). The GEOS-5 AGCM on the other hand produces clear skies more frequently, ~38% of the

24 time over the entire globe (90°S to 90°N) for GN and ~42% for MR. Despite the model's positive

25 clear-sky fraction bias (negative cloud fraction bias), many patterns of clear-sky occurrence are

26 realistic with peaks occurring in desert areas, western North America and the southern parts of

27 Africa and S. America. Over the ocean, the model seems to be producing clear skies in the

28 Maritime Continent and the far southern oceans more frequently than observations, but these

29 overestimates are still much smaller compared to those in wide subtropical swaths of the

Deleted: .

Formatted: Indent: First line: 0.38"

Deleted: RFO

1 Atlantic and Pacific oceans. The model also exhibits pronounced cloudiness underestimates in

2 the descending branch of the central Pacific's Walker circulation. The only notable model

3 underestimate of clear-sky frequency occurs over western Antarctica. The MR overlap

4 assumption makes the clear sky overestimates worse, with the biggest impact seen in the

5 central and western tropical Pacific (clear subcolumn panel of Fig. 3). Note that the observed

6 global clear sky fraction is lower in 2B-CLDCLASS-LIDAR compared to passive satellite

7 observations such as those from MODIS (King et al., 2003) because of CALIOP's enhanced ability

8 to detect clouds that are optically very thin. Model cloud coverage on the other hand has been

9 traditionally tuned to resemble that seen in cloud climatologies obtained by satellite

10 observations from passive imagers at solar and thermal infrared wavelengths.

11   Moving on to cloudy skies, a quick survey of the remaining panels of Fig. 2 reveals that the

12 model exhibits considerable skill in simulating cloudiness when viewed under the prism of CVS

13 classes. Weaknesses become however apparent upon closer examination. In terms of global

14 values, the only CVS class where the model produces a substantial RFO overestimate is "HM",

15 for both overlap assumptions. For CVS = "HML", global RFOs agree, especially for the GN

16 overlap assumption. The global RFOs of all other CVS classes are underestimated to varying

17 degrees with the underestimates being slightly worse for the MR overlap assumption, except

18 for CVS = "L" for which MR RFO slightly exceeds GN RFO. The total RFO of the four CVS classes

19 containing H clouds is ~40% in observations, and ~36% (GN) or ~32% (MR) in the model. The

20 remaining CVS classes consisting of only L and M clouds add up to a global RFO of ~35% in

21 observations and ~ 26% in the model (both GN and MR). Therefore, most of the 13%

22 discrepancy between GEOS-5/GN in global cloud fraction comes from the three CVS classes

23 containing only L and M clouds, while the larger discrepancy of ~17% for GEOS-5/MR is more

24 evenly split between these three CVS classes and the remaining four containing H clouds.

25   A closer comparison of geographical features is also informative. The bottom part of Fig. 2

26 shows only the GN overlap results and can be directly compared with the top part showing the

27 observed maps. The performance of the MR overlap implementation can be gleaned in terms of

28 its deviation from GN in the Fig. 3 difference maps.

8

Deleted: passive

Deleted: are

Deleted: revealed

Simulating low clouds has been identified as a challenge for large-scale models, but this version of GEOS-5 seems to be simulating the isolated low clouds (CVS = "L") quite well with a global underestimate of ~5% for GN overlap and ~4%, for MR (absolute values), and with characteristic cloud patterns associated with marine stratocumulus being present albeit with less extensive spatial coverage. While GEOS-5 does not produce isolated *M* clouds (CVS class "M") as often as in the observations, the impact is expected to be small as this CVS class is the least frequently observed and with presence exclusively over land and specifically deserts, ice/snow covered surfaces, and regions of pronounced orography. Overall however, there is not such a great paucity of *M* clouds in the model when taking into account the other CVS classes containing this type of cloud. Setting aside deep and multi-layer clouds (the "HML" CVS class), *M* clouds appear only about 11% (for GN – the figure rises to 22% for MR) more frequently (in relative terms) in observations than the model; the combined RFO of "M", "ML" and "HM" is 14.5% in the observations and 13% (11.8%) in the model for the GN (MR) implementation. Finally, *H* over *L* clouds (CVS class "HL") are one of the biggest contributors in the overall cloudiness discrepancy between the real and simulated worlds as they appear twice as often in the observations than in the GN version of the model (and even more relative to the MR implementation of the model). The model seems to be lacking much of the tropical presence of this CVS class.

A closer look at the influence of the overlap assumption on CVS RFOs can be gauged from the Fig. 3 maps. We have previously seen that the MR overlap assumption generally produces less cloudiness than GN. This happens systematically (i.e., virtually all locations) for five out of seven CVS classes. The interesting exception is CVS = "L" (CVS = "M" is absent in GEOS-5 for all practical purposes). The Fig. 3 difference map for CVS = "L" reveals that the GN's reduced cloudiness comes mostly from the extratropics; tropical and subtropical pockets can be found where GN cloud amounts exceed those from MR, as in the other CVS classes. The contrast between CVS = "L" and the other CVS classes illustrates the fact that the specific flavors of these overlap assumptions as implemented in GEOS-5 can produce a variety of outcomes that depend on the total geometrical extent of contiguous or non-contiguous cloud vertical cloud configurations and the detailed shape of the cloud fraction profile.

Deleted:

Deleted: for

Deleted:).

Deleted: miss a lot

Deleted: in general

Deleted: ,

Deleted: GN underestimates come

Deleted: exceeds

Deleted: brings home

Deleted: point

1

b. Global CRE comparison by CVS class

2

3 Figure 4 compares the global mean CRE between model and observations, the latter

4 (r), coming from the aforementioned 2B-FLXHR-LIDAR CC product. It shows the mean values

5 only *when the CVS occurs*, i.e., CRE is weighted by area, but not by global RFO. We call this type

6 of CRE the "cloudy-column" or "overcast" CRE since it is calculated by taking the mean of the

7 CRE values of cloudy subcolumns belonging to the CVS class. CRE values for each cloudy

8 subcolumn also correspond to overcast conditions since there is no partial cloudiness at the

9 subcolumn scale. We show overcast CRE from three perspectives: the Top-of-the-Atmosphere,

10 TOA (Figure 4a), the surface, SFC (Figure 4c), and the atmospheric column, ATM (Figure 4b), the

11 latter derived as the difference between the TOA and SFC CREs. Moreover, we distinguish

12 between shortwave (SW) and longwave (LW) components, and also display their sum which we

13 call "total" CRE (aka "net" CRE). With CRE being defined as the difference between cloudy and

14 clear-sky net (=down-up) fluxes, negative values indicate a radiative cooling effect, while

15 positive values indicate a radiative warming effect. For TOA and SFC, all SW CREs are negative.

16 Note also the magnitudes at TOA and SFC being rather similar for SW, with the slightly larger

17 SFC value resulting from the small positive ATM SW CRE which indicates that clouds slightly

18 enhance atmospheric column absorption. While LW CREs at both TOA and SFC are positive and

19 therefore indicative of warming, the ATM LW CRE can be either positive or negative. Note that

20 all positive global means involve *H* clouds. Again, we show model results for the two overlap

21 assumptions, GN and MR although their CREs are quite close in general. The observed SW CREs

22 depend strongly on the incoming solar flux at the approximate 1:30 pm local overpass time, and

23 are therefore scaled to diurnal fluxes by normalizing with the ratio of the instantaneous to

24 diurnally averaged incoming solar flux at TOA (O17); the LW CREs are simple averages of the

25 daytime and nighttime overpass values. On the other hand, both SW and LW CREs from the

26 model are daily averages of three-hourly mean outputs.

27   For TOA SW CRE, the best agreement between model and observations occurs for CVS =

28 "L", and CVS = "HM". For the remaining CVS classes the model either overestimates (CVS = "H",

29 "M", "HL") or underestimates (CVS = "ML", "HML") overcast TOA SW CRE. The overestimate for

CVS class "H" is very large in relative terms given the small absolute magnitude of the observed CRE. It appears then that H clouds in the model are optically thicker than in observations. Discrepancies are quite smaller for TOA LW CRE, reflecting the lesser dependence of this quantity on cloud properties other than cloud top location (which is constrained because of the CVS class decomposition) once clouds reach a certain value of optical thickness (~ 5). The biggest bias (underestimate) appears for CVS = "HML" CVS, but since it is still smaller than the SW CRE bias it results in an underestimate of net planetary cooling as expressed by total TOA CRE (purple bars). Given the better agreement between LW CREs, total TOA CRE biases largely follow the sign of the SW CRE biases. These findings are very insensitive to the type of chosen overlap, although the differences in magnitudes between the two simulated values are still large enough to be distinguishable in most cases.

When moving to examination of surface (SFC) CREs (bottom panel of Fig. 4) our conclusions about the SW CRE component are the same as before since atmospheric (ATM) SW CREs are small positive values (middle panel). LW CRE values are again simulated quite well since most of the variability is driven by the location of the cloud bottom which is constrained by CVS class. The largest biases occur for CVS = "L" and "HL" (overestimates by the model), and since the TOA CREs have small biases for those cases, errors (excessive cooling) materialize in the ATM LW CRE. Still, the largest ATM LW CRE error occurs for CVS="HM" (excessive warming by the model) because the TOA and SFC CRE errors are in the opposite direction. Given the small magnitude of ATM SW CRE, the total ATM CRE errors track those of the LW component.

Figure 5 compares observed and modeled CRE values that are now weighted by the global mean RFO (f for observations) of the CVS classes in addition to areal weighting. We can call this type of CRE "all-sky" CRE since in the calculation of the mean all subcolumns that do not belong to the CVS class under consideration contribute zero errors. Summing then these CVS-specific values yields the true global CRE of observed and modeled CRE fields. Since the all-sky CRE values and the range of the y-axis are much smaller than in Fig. 4, it makes sense to compare the two figures only with respect to relative biases, essentially focusing on the position of the symbols (simulated values) relative to the bar (observed values). While this will be shown more explicitly in upcoming Fig. 6, comparison of Figs. 4 and 5 basically indicates whether RFO errors

1 suppress (i.e., compensate for) or amplify cloud property only errors. Take CVS = "HL" for

2 example: RFO errors (underestimates) help suppress the TOA and SFC SW (and total)

3 overestimates. In general, we do not see much of the opposite effect, i.e., an amplification of

4 relative error CRE when moving from overcast to all-sky CRE. Of course, a very low RFO also

5 makes an overcast CRE that previously seemed substantial to disappear, with CVS ="M" being a

6 characteristic case in point. The discussion of all-sky CRE error interpretation continues in the

7 next subsection where a more formal error decomposition framework is introduced.

8

9 *c. CRE error decomposition*

10 Figure 6 shows the decomposition of GEOS-5 all-sky CRE global errors ΔCRE of Fig. 5 to

11 overcast CRE and RFO error contributions for the GN case only (conclusions remain the same

12 for MR). The decomposition can be expressed as follows (e.g., Tan et al. 2015):

13 $\Delta CRE = f \times \Delta r + r \times \Delta f + \Delta r \times \Delta f$ (1)

14 This representation of CRE error arises when the model global all-sky CRE of a CVS class

15 (Fig. 5) is expressed as the product of a deviation Δr from the observed mean overcast CRE

16 r, (Fig. 4), and the model global RFO is expressed as a deviation Δf from the observed mean

17 RFO, f:

18 $CRE_{GEOS-5} = (r + \Delta r) \times (f + \Delta f)$ (2)

19 Basically, the model's grid-mean CRE error for a CVS class arises from a combination of

20 overcast CRE bias Δr under the observed RFO f, and the simulated RFO bias Δf under observed

21 overcast CRE r, plus a co-variation term of RFO and CRE errors under observed f and r (Tan et

22 al. 2015). Such a decomposition of CRE error allows us to infer, for example, whether the

23 model's poor simulation of all-sky CRE is mostly due to errors in simulating the occurrence

24 frequency of the CVS class or errors in the optical and physical properties of the CVS class which

25 drive the overcast CRE. Similarly, it potentially reveals cases where good simulations of global

26 all-sky CRE in Fig. 5 benefit from compensating errors in simulated RFO (Fig. 2) and overcast CRE

27 (Fig. 4).

12

Deleted: cloudy-column
Deleted: grid-mean
Deleted: whatever cloudy-column
Deleted: seemed
Deleted: on grid-mean
Deleted: employed
Deleted: grid-mean
Deleted: cloudy column
Deleted: grid-mean
Deleted: 3
Deleted: cloudy-column
Deleted: cloudy-column
Deleted: grid-mean
Deleted: cloudy-column
Deleted: grid-mean
Deleted: cloudy-column

Separate panels are used in Fig. 6 for SW (first column), LW (second column), and total (third column) CRE. The breakdown by TOA, SFC and ATM is also preserved, yielding thus a total of nine panels. In the SW, TOA (Fig. 6a) and SFC (Fig. 6g) results look again very similar, while the ATM CRE errors (Fig. 6d) are too small to merit discussion. For most CVS classes (five out of seven) all-sky SW CRE errors (gray bars) come from overcast CRE errors (red bars), namely errors in CVS optical properties. The excessive planetary cooling of the cloudy columns (negative red bars, four CVS classes) is always dampened by compensating errors, sometimes virtually eliminating the error (as in CVS = "L", "HL") or reducing it slightly (CVS = "H"), or overcorrecting (CVS = "M"). SW TOA overcast CRE (red bars in Fig. 6a) in the opposite direction (cooling underestimates) become bigger all-sky errors due to RFO errors for CVS= "ML" and "HML", while the all-sky errors for CVS = "HM" come almost exclusively from RFO errors (blue bars in Fig. 6a). Finally, three CVS classes have sizable co-variation errors (green bars) in the same direction as RFO errors. The above error interpretation applies virtually intact for surface (SFC) SW errors (Fig. 6g).

Contrary to the SW, the LW CRE errors for all three vantage points (TOA in Fig. 6b, SFC in Fig. 6h, ATM in Fig. 6e) deserve their own discussion as they have different characteristics. At TOA and SFC, the errors are substantially smaller than their SW counterparts. Three of the four CVS classes with H clouds (the exception being CVS = "H") exhibit ~2 Wm^{-2} (absolute) errors, coming from RFO contributions in two out of the three classes. These three classes have smaller all-sky errors at the SFC, in one case (CVS ="HL") because of compensating errors. The largest component errors occur for CVS = "L" which has the largest absolute magnitude of all-sky SFC CRE, but with component errors in the opposite direction, compensation reduces the all-sky CRE error. Because the TOA errors for this CVS class are small, the SFC errors carry to the ATM errors. The other CVS class with large ATM error is "HML" where TOA and SFC errors of the opposite sign conspire to magnify the ATM error.

Errors in total all-sky CRE are driven mainly by SW errors at TOA and SFC (Fig. 6c and Fig. 6i), and LW errors for ATM (Fig. 6f). Errors of the opposite sign reduce the overcast cooling error at the SFC for CVS = "L" and "HL" and the all-sky warming error for CVS="ML". But because the SFC LW CRE errors are in general small, the total CRE SFC errors largely retain the

13

1 characteristics of the SW component. In the atmospheric column, SW and LW overcast (and all-

2 sky) errors are additive for CVS = "HML" and opposing for CVS = "L", the only two classes for

3 which ATM SW CRE registers errors of notable magnitude (cf. middle panels of Fig. 4 and 5).

4   In summary, this decomposition analysis showed the multiple ways relatively good

5 agreement with observed all-sky CRE values from various vantage points can be achieved by

6 GEOS-5 (or any other model evaluated this way). Overcast CRE and RFO errors can compensate,

7 TOA and SFC all-sky CRE errors can compensate (for ATM LW CRE, e.g., CVS = "M", "ML"), SW

8 and LW errors can compensate for total CREs, and finally the errors among various CVS classes

9 can compensate towards decreasing the global CRE error.

10

11 *d. Seasonal CRE comparison*

12 Figures 7-9 compares the multi-year mean annual cycle of TOA, SFC, and ATM total (=SW+LW)

13 all-sky CRE zonal averages between observations and the model (employing the GN overlap

14 assumption) for the four CVS classes with the greatest all-sky SW or LW CREs according to Fig.

15 5.

16   Inspection of the TOA and SFC CRE plots shows that the model has some skill in simulating

17 the seasonal competition between SW and LW CRE, but this should not come as a surprise as it

18 is driven mainly by seasonal changes in insolation. Basically, with everything else staying the

19 same, the SW CRE contribution to total CRE scales with the amount of incoming solar energy.

20 Positive values of total TOA and SFC CRE occur when the solar insolation is weak during the

21 winter allowing thus the positive LW CRE to exceed the negative SW CRE. At TOA, this takes

22 place only for the "HML" CVS class since this is the class with competing SW and LW CREs of

23 relatively large magnitude. Note that the model summer planetary cooling is stronger than in

24 the observations. At the SFC, besides CVS = "HML" the seasonal switch from cooling to warming

25 also takes place for CVS = "L" because the LW CRE is of comparable magnitude to its SW

26 counterpart. The model's CVS = "H" is virtually neutral radiatively at TOA throughout the year,

27 in contrast to the observations, where it provides planetary radiative heating in the tropics and

28 subtropics. It seems then that in the model CVS= "H" consists of optically thicker clouds that

29 reflect more solar radiation to space than in the real world. *H* clouds in GEOS-5 appear to be

Deleted: cloud-column
Deleted: grid-mean
Deleted: for the
Deleted: Cloudy-column
Deleted: grid-mean
Deleted: grid-mean
Deleted: CVS
Deleted: also

also optically thicker when overlapping with *L* clouds (CVS = "HL"), in this case producing planetary cooling in the tropics throughout the year and in the extratropics during the summer months of high insolation, in contrast to the observations where their cooling effect is very weak and occurs only in the austral extratropics during summertime. Evidence for optically thicker *H* clouds in both CVS ="H" and "HL" is also seen at SFC total CREs which are more negative in the model than in the observations. Overall (all CVS classes combined, the rightmost panel of Figs. 7 and 8), the model produces a rather realistic pattern of seasonal variations in zonal mean total CRE.

Total ATM CREs are driven as we have seen earlier by the LW component, and their seasonal cycles are fairly well represented by the model for three of the four most radiatively important CVS classes (Fig. 9). The nature of CVS = "HML" seems however to be different in GEOS-5 compared to observations. At high latitudes, the atmospheric column is cooled by this type of cloudiness, especially during the summer months, as the SFC total CRE (Fig. 8) exceeds the TOA CRE (Fig. 7). Since the SW contribution is relatively small, it then seems that *L* clouds within CVS = "HML" have lower bases or are optically thicker during the summer months in the model compared to observations, making their downward emission towards the surface, and therefore also the contrast between TOA and SFC emission, stronger in the model than the observations. Fig. 9i shows also that the near-zero total ATM CRE for CVS ="HML" in GEOS-5 (Fig. 5) is a result of positive and negative total ATM CRE regional compensations. Overall, the model captures the basic zonal pattern of atmospheric heating and warming (rightmost panel of Fig. 9) with heating prevailing in the tropics and cooling in the extratropics. The tropical heating is however weaker than in the observations while the extratropical atmospheric cooling is stronger.

4. Conclusions

We have introduced a method of cloud evaluation for large-scale atmospheric models that focuses on the vertical structure of cloudiness. Cloud Vertical Structure (CVS) is resolved in a rather simplified way based on the various combinations of cloud presence in three standard layers that have been traditionally used to distinguish between high, middle, and low clouds. A

1   reference dataset for such CVS classification now exists because of CloudSat and CALIPSO active

2   sensor observations (Oreopoulos et al. 2017). For the purposes of model evaluation, the initial

3   dataset of 10 CVS classes was simplified to consist of 7 classes by merging of some of the

4   original classes that had clouds in adjacent standard layers. Beyond comparison of the

5   frequency of occurrence of the CVS classes we also compared their radiative impact in terms of

6   the Cloud Radiative Effect (CRE). While the CVS classes by design constrain cloud vertical

7   location (albeit not in the strictest of ways), they constrain extinction to a lesser extent, and

8   mostly qualitatively (e.g., multi-layer cloud configurations are expected to have a greater total

9   column extinction). This is taken into account when examining the performance of the model in

10   terms of SW and LW CRE. We developed a framework wherein we can compare CRE for only

11   when a CVS class occurs (overcast CRE), or perform a comparison that also accounts for how

12   frequently the CVS class occurs (all-sky CRE). We can then naturally examine to what extent

13   errors in the latter type of CRE come from errors in the overcast CRE of the class and/or biases

14   in the frequency of occurrence.

15   The GEOS-5 model under evaluation produces about 50% more clear skies than

16   observations in relative terms. It produces isolated high clouds (cloud top and base above the

17   440 hPa level) that are slightly less frequent than in observations, but are optically thicker

18   yielding excessive planetary and surface cooling. Low clouds (cloud tops and bases within the

19   lowest layer of the troposphere up to 680hPa) are usually a challenge for global models, but

20   GEOS-5 is doing reasonably well and compensates a lower frequency of occurrence (by ~20% in

21   relative terms) with overestimates in extinction, producing in the end an excellent agreement

22   with observations for SW and LW all-sky CREs at either the TOA, SFC or the atmospheric column

23   vantage points. Overall LW CREs are better simulated since they are mainly driven by vertical

24   cloud location which is substantially constrained when clouds are broken by CVS class. But

25   either component of CRE can be off in terms of contribution to the global CRE if the frequency

26   of occurrence is deficient. The other side of the coin is, of course, that incorrect simulation of

27   the frequency of occurrence can compensate for biased cloud optical and physical properties

28   that determine the overcast CRE of the CVS class. Needless to say, CRE biases among different

29   CVS classes can also cancel out to various degrees when global or regional CREs encompassing

Deleted: natural

Deleted: result in

Deleted: ,

Deleted: .

Deleted: are because of

Deleted: inherent

Deleted: inherent

1   all clouds represented by the CVS classes are calculated. In such a holistic view, the model

2   appears able, for example, to reproduce the aggregate planetary feature of atmospheric

3   radiative warming in the tropics and cooling in the extratropics driven by cloud configurations

4   dominated by high and low clouds, respectively, albeit with magnitudes that differ from those

5   observed.

6   The evaluation we conducted requires that the model has the capability to produce

7   cloudy subcolumns which are then considered equivalent to the atmospheric column profiles

8   seen by the active observations. There is no unique way to go from mean cloud fraction profiles

9   to subcolumns having layer cloud fractions that are either one or zero. We tried two ways to

10   produce subcolumns that assume different cloud fraction overlaps and obtained rather close

11   results. By adopting our framework of cloud evaluation, which, incidentally, should be used in

12   conjunction with other cloud evaluation methodologies (e.g., cloud regimes as in Jin et al.

13   2017a, b), one can assess whether other large scale models are more sensitive (i.e., produce a

14   greater diversity of CVS climatologies) to different overlap assumptions applied to the same

15   original mean cloud fraction profiles. What one should always keep in mind however is that no

16   matter how good the cloud subcolumn generator is, observed CVS class global frequencies and

17   patterns cannot be reproduced if the model's underlying mean cloud profiles used as input to

18   the generator are deficient.

19

20   **Code availability**

21   The GEOS-5 source code is available under the NASA Open-Source Agreement at

22   http://opensource.gsfc.nasa.gov/projects/GEOS-5/.

23

24   **Author contribution**: D. Lee and L. Oreopoulos designed the metrics and experiments. D. Lee

25   adapted the model code for the new metric and performed the simulations. N. Cho processed

26   the observational dataset. D. Lee and N. Cho created the graphics and figures. D. Lee and L.

27   Oreopoulos authored the text with contributions from N. Cho.

28

29   **Competing interests**: The authors declare that they have no conflict of interest.

Deleted: is

1

2 **List of Figures**

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: cloudy-column

Formatted: Font: Bold

Deleted: W/m²) SW, LW, total=SW+LW CREs at

Deleted: ,

Deleted: within

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: 3

Deleted: grid-mean

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: grid-mean

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: grid-mean

1   greatest all-sky CREs according to Fig. 5. The rightmost set of panels displays the scaled (half)

2   total CRE of all CVS classes combined.

3   **Figure 8.** As Fig.7, but for SFC total all-sky CRE.

4   **Figure 9.** As Fig. 7, but for ATM total all-sky CRE.

5

6

1 **References**

2 Barker, H. W. (2008), Overlap of fractional cloud for radiation calculations in GCMs: A global

3   analysis using CloudSat and CALIPSO data, *J. Geophys. Res.*, **113**, D00A01,

4   doi:10.1029/2007JD009677.

5 Bodas-Salcedo, A., et al. (2011), COSP: Satellite simulation software for model assessment, *Bull.*

6   *Am. Meteorol. Soc.*, **92**(8), 1023–1043.

7 Dolinar, E.K., X. Dong, B. Xi, J.H. Jiang, H. Su (2015), Evaluation of CMIP5 simulated clouds and

8   TOA radiation budgets using NASA satellite observations, *Clim. Dyn.*, **44**, 2229–2247,

9   https://doi.org/10.1007/s00382-014-2158-9

10 Chou, M. D., M. J. Suarez, C. H. Ho, M. M. H. Yan, and K. T. Lee (1998), Parameterizations for

11   cloud overlapping and shortwave single-scattering properties for use in general circulation

12   and cloud ensemble models, *J. Clim.*, **11**, 202–214.

13 Geleyn, J. F., and A. Hollingsworth (1979), An economical analytical method for the

14   computation of the interaction between scattering and line absorption of radiation, *Contrib.*

15   *Atmos. Phys.*, **52**, 1–16.

16 Henderson, D. S., T. L'Ecuyer, G. Stephens, P. Partain, and M. Sekiguchi (2013), A multi-sensor

17   perspective on the radiative impacts of clouds and aerosols, *J. Appl. Meteorol.*

18   *Climatol.*, **52**, 853–871, doi:10.1175/JAMC-D-12-025.1.

19 Hogan, R. J., and A. J. Illingworth (2000), Deriving cloud statistics from radar, *Q. J. R. Meteorol.*

20   *Soc.*, **126**, 2903–2909.

21 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W.

22   D. (2008), Radiative forcing by long-lived greenhouse gases: Calculations with the AER

23   radiative transfer models, *J. Geophys. Res.*, 113, D13103, doi:10.1029/2008JD009944.

24 King, M.D., W.P. Menzel, Y.J. Kaufman, D. Tanre, Bo-Cai Gao, S. Platnick, S.A. Ackerman, L.A.

25   Remer, R. Pincus, P.A. Hubanks (2003), Cloud and aerosol properties, precipitable water,

26   and profiles of temperature and water vapor from MODIS, *IEEE Trans. Geosci. Remote*

27   *Sensing*, **41**, 442–458, https://doi.org/10.1109/TGRS.2002.808226

Klein, S.A., Y. Zhang, M.D. Zelinka, R. Pincus, J. Boyle, P.J. Gleckler (2013), Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, *J. Geophys. Res. Atmos.*, **118**, 1329–1342. https://doi.org/10.1002/jgrd.50141

Jin, D., Oreopoulos, L., & Lee, D. (2017a). Regime-based evaluation of cloudiness in CMIP5 models. *Climate Dynamics*, **48**(1), 89–112. https://doi.org/10.1007/s00382-016-3064-0.

Jin, D., Oreopoulos, L., & Lee, D. (2017b). Simplified ISCCP cloud regimes for evaluating cloudiness in CMIP5 models. *Climate Dynamics*, **48**(1-2), 113–130. https://doi.org/10.1007/s00382-016-3107-6.

L'Ecuyer, T. S., N. B. Wood, T. Haladay, G. L. Stephens, and P. W. Stackhouse Jr. (2008), Impact of clouds on atmospheric heating based on the R04 CloudSat fluxes and heating rates data set, *J. Geophys. Res.*, **113**, D00A15, doi:10.1029/2008JD009951.

Matus, A. V., and T. S. L'Ecuyer (2017), The role of cloud phase in Earth's radiation budget, *J. Geophys. Res. Atmos.*, **122**, 2559–2578, doi:10.1002/2016JD025951.

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, (2012) The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna. NASA Tech. Rep. NASA TM-2012-104606, Vol. 28, 117 pp.

Nam, C., S. Bony, J.-L. Dufresne, H. Chepfer (2012), The 'too few, too bright' tropical low-cloud problem in CMIP5 models, *Geophys. Res. Lett.,* **39**, https://doi.org/10.1029/2012GL053421

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

Oreopoulos, L., N. Cho, and D. Lee (2017), New insights about cloud vertical structure from CloudSat and CALIPSO observations, *J. Geophys. Res. Atmos.*, 122, 9280–9300, doi:10.1002/2017JD026629.

Deleted: .,

Deleted: P. M.

Deleted: (2011),

Formatted: Font: Not Italic

Deleted: *Disc.,* **11**, 597–625.

1   Pincus, R., Barker, H. W., and Morcrette, J.-J. (2003), A fast, flexible, approximate technique

2   for computing radiative transfer in inhomogeneous cloud fields, *J. Geophys. Res.*, 108,

3   4376, doi:10.1029/2002JD003322, D13.

4   Pincus, R., C.P. Batstone, R.J.P. Hofmann, K.E. Taylor, P.J. Glecker (2008), Evaluating the

5   present-day simulation of clouds, precipitation, and radiation in climate models, *J. Geophys.*

6   *Res.*, **113**, D14209. https://doi.org/10.1029/2007JD009334

7   Räisänen, P., H. W. Barker, M. F. Khairoutdinov, J. Li, and D. A. Randall (2004), Stochastic

8   generation of subgrid-scale cloudy columns for large-scale models, *Q. J. R. Meteorol.*

9   *Soc.*, **130**, 2047–2067.

10 Rienecker, M. M., and Coauthors, (2008) The GEOS-5 data assimilation system – documentation

11   of versions 5.0.1 and 5.1.0, and 5.2.0. NASA Tech. Rep. NASA/TM-2008-104606, Vol. 27, 92

12   pp.

13 Rossow, W. B., and R. A. Schiffer (1991), ISCCP cloud data products. Bulletin of the American

14   Meteorological Society, 72, 2– 20, https://doi.org/10.1175/1520-

15   0477(1991)072<0002:ICDP>2.0.CO;2.

16 Sassen, K., and Z. Wang (2012), The clouds of the middle troposphere: Composition, radiative

17   impact, and global distribution, *Surv. Geophys.*, **33**(3–4), 677–691, doi:10.1007/s10712-011-

18   9163-x.

19 Tan, J., C. Jakob, W. B. Rossow, and G. Tselioudis (2015), Increases in tropical rainfall driven by

20   changes in frequency of organized deep convection, *Nature*, **519**, 451–454.

21 Tselioudis, G., W. Rossow, Y.-C. Zhang, and D. Konsta (2013), Global weather states and their

22   properties from passive and active satellite cloud retrievals, *J. Clim.*, **26**, 7734–7746,

23   doi:10.1175/JCLI-D-13-00024.1.

24 Wang, H. and W. Su (2013), Evaluating and understanding top of the atmosphere cloud

25   radiative effects in Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment

26   Report (AR5) Coupled Model Intercomparison Project Phase 5 (CMIP5) models using

27   satellite observations, *J. Geophys. Res. Atmos.*, **118**, 683–699.

28   https://doi.org/10.1029/2012JD018619

29

1

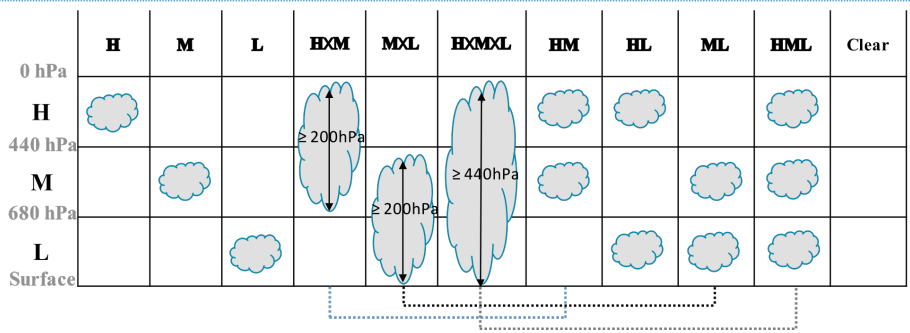|	H	M	L	HXM	MXL	HXMXL	HM	HL	ML	HML	Clear
0 hPa											
H 440 hPa				≥ 200hPa		≥ 440hPa					
M 680 hPa					≥ 200hPa						
L Surface											

2
3
4 **Figure 1**. The original ten CVS classes of Oreopoulos et al. (2017) used as reference for the
5 comparison of this paper. The multi-layer CVS classes other than "HL" are merged in this paper
6 thus reducing the total number of CVS classes to seven. We essentially do no distinguish
7 between contiguous and non-contiguous clouds in adjacent standard layers. Dotted lines show
8 which pairs of CVS classes have been combined for this study.
9

a) Relative Frequency Occurrences CloudSat/CALIPSO
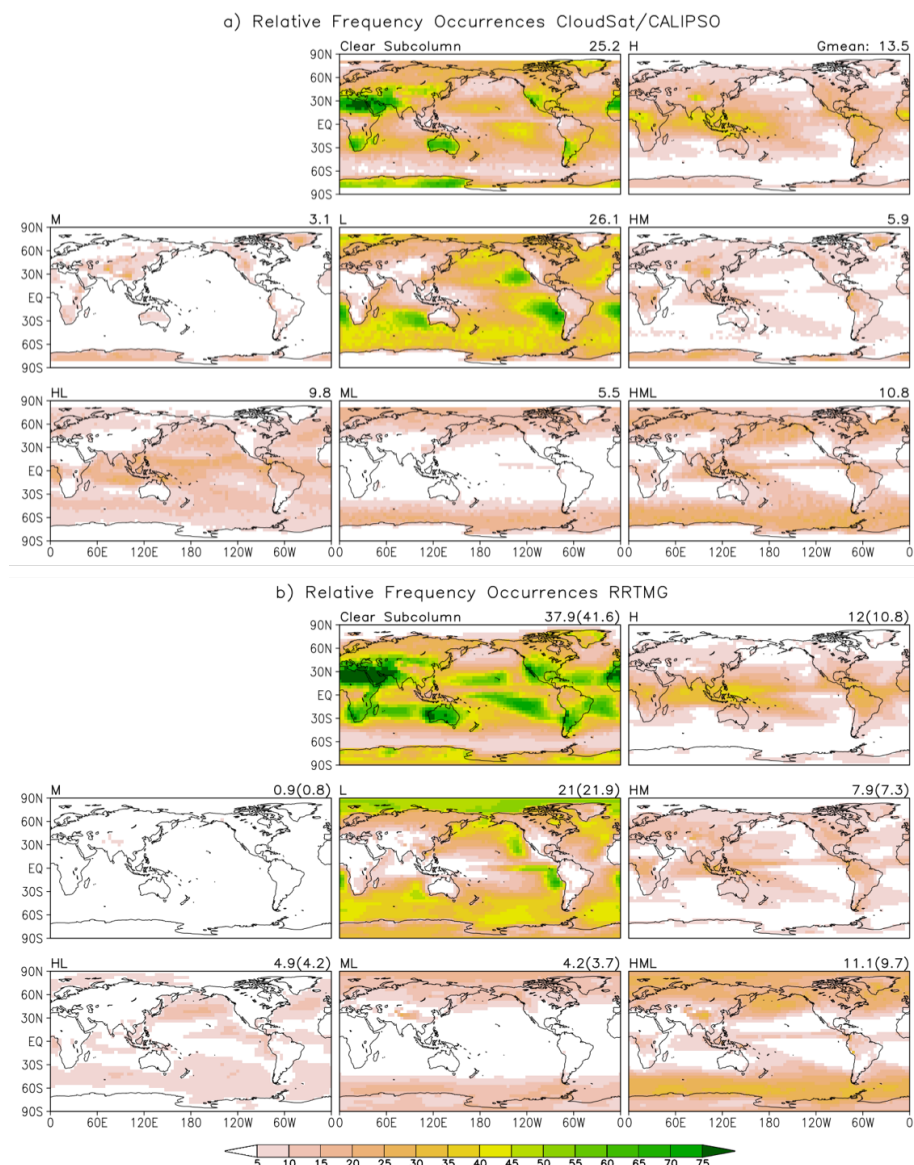
1
Figure 2. Geographical RFO distribution (%) for cloudless skies and the seven CVS classes

according to CloudSat/CALIPSO observations (top 8 panels), and for GEOS-5 (GN overlap

25
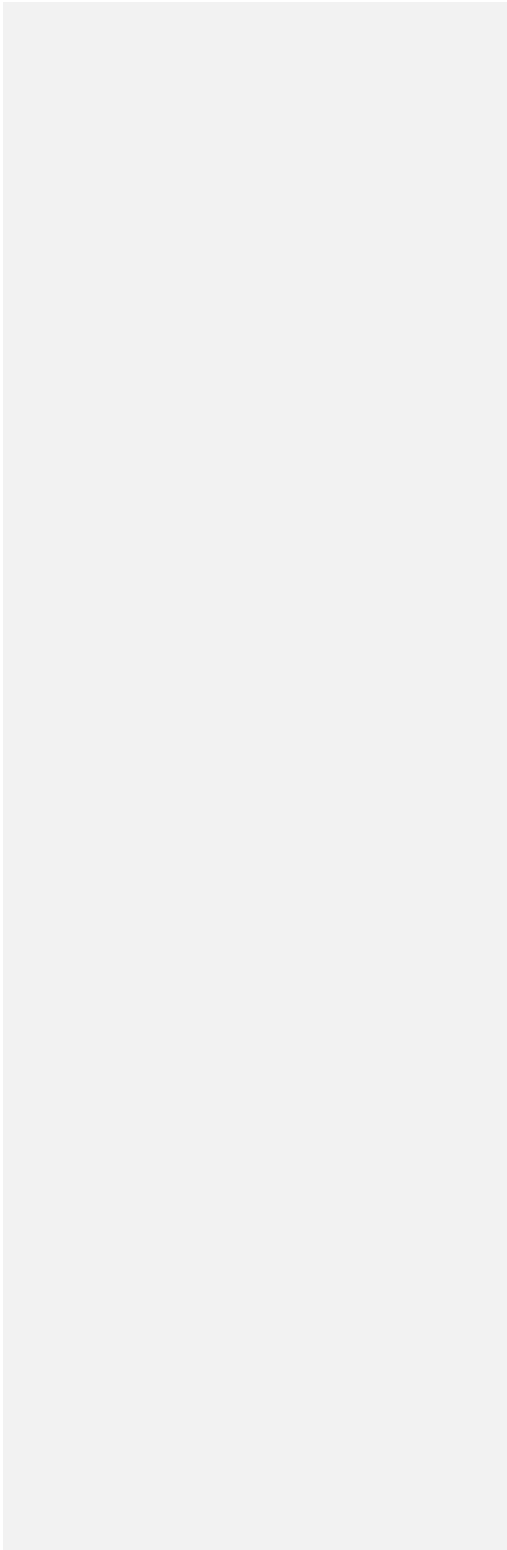
1   assumption, bottom 8 panels). Global mean values are shown above each panel, in the case of

2   GEOS-5 we provide the global values for both the GN and MR overlap (in parentheses).

3

4

1
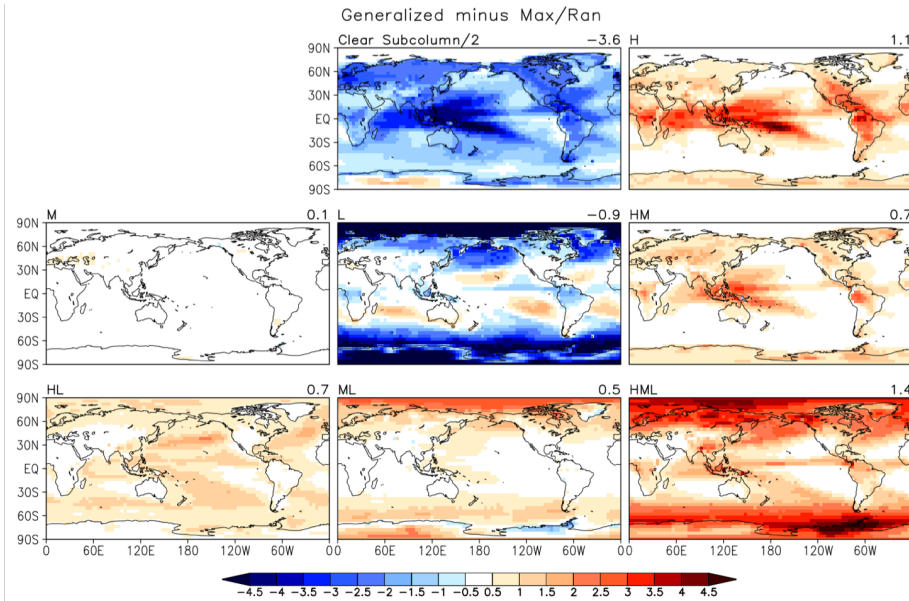
Figure 3. RFO difference (%) maps for clear skies (divided by two to use a common color scale)

and the seven CVS classes as simulated by GEOS-5 using the GN and MR overlap assumptions in

the cloudy subcolumn generator.

6
7

27

1

2
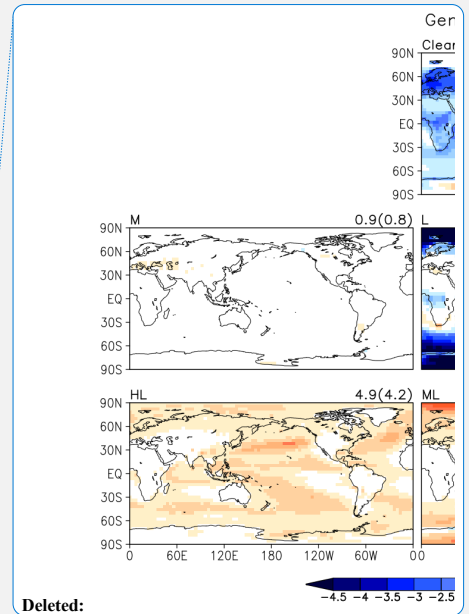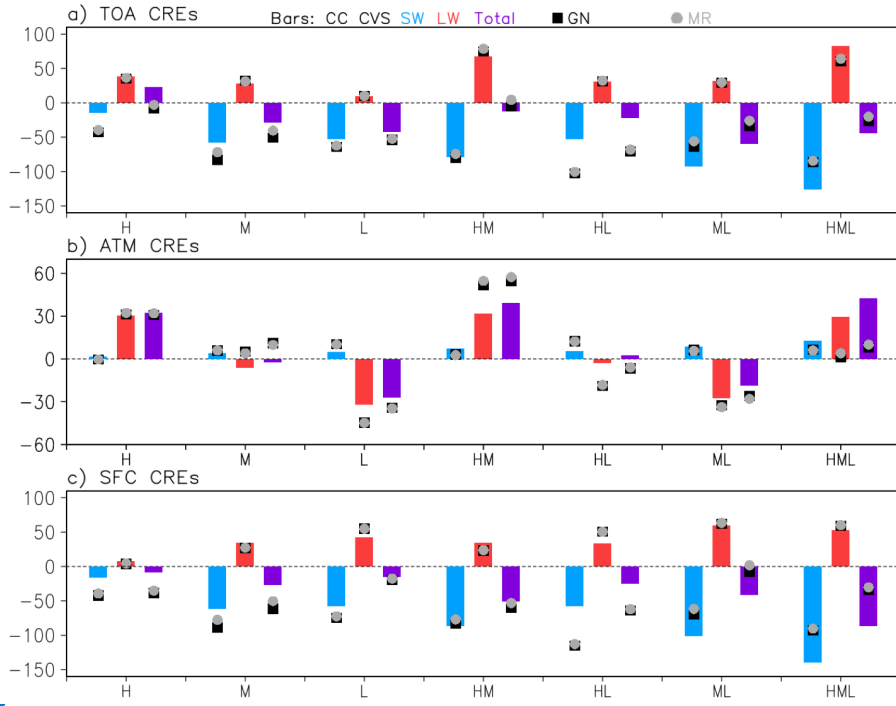
Figure 4. Comparison between observations and model (GN and MR) of global overcast CREs

(Wm⁻²): Top-of-the-Atmosphere, TOA (top), surface, SFC (bottom), and atmospheric column,

ATM (middle) derived as the difference between the TOA and SFC CREs. CREs are distinguished

into shortwave (SW) and longwave (LW) components, and their sum, "total" CREs for each CVS

class are also shown. Note that the y-axis range is the same for TOA and SFC CRE, but it is

substantially more compressed for ATM CRE.

1

2
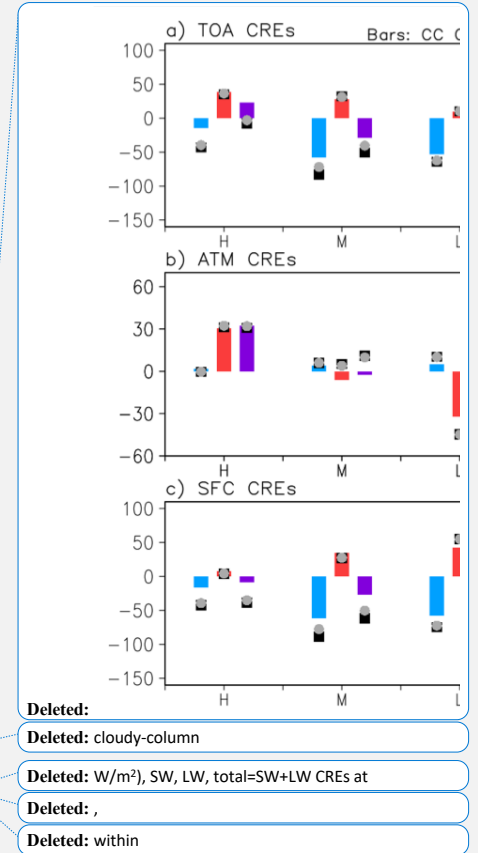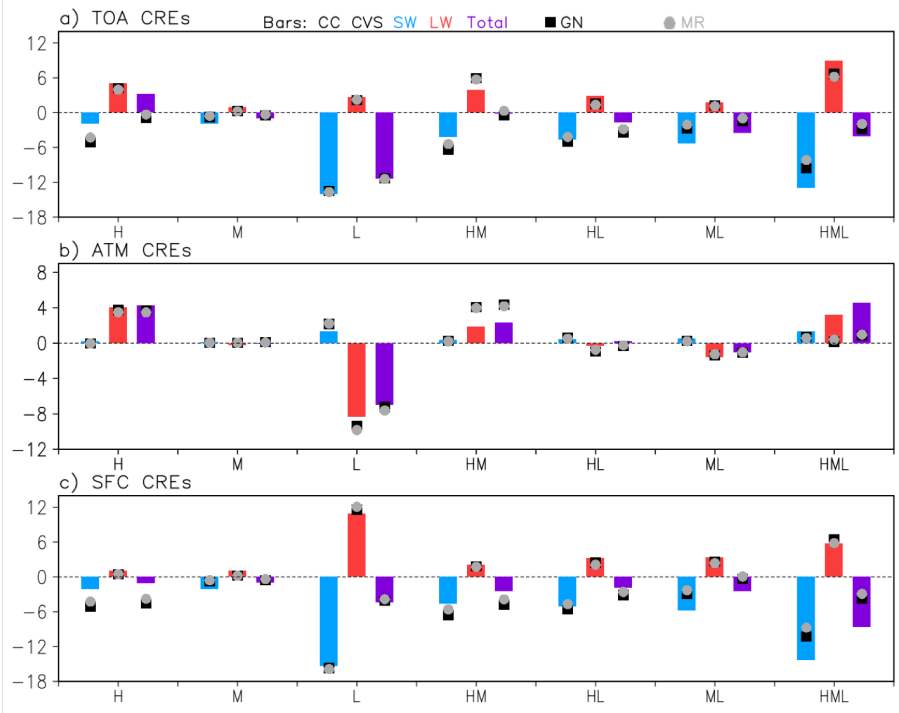
Figure 5. As Fig. 4, but for all-sky (RFO-weighted) CREs.

4
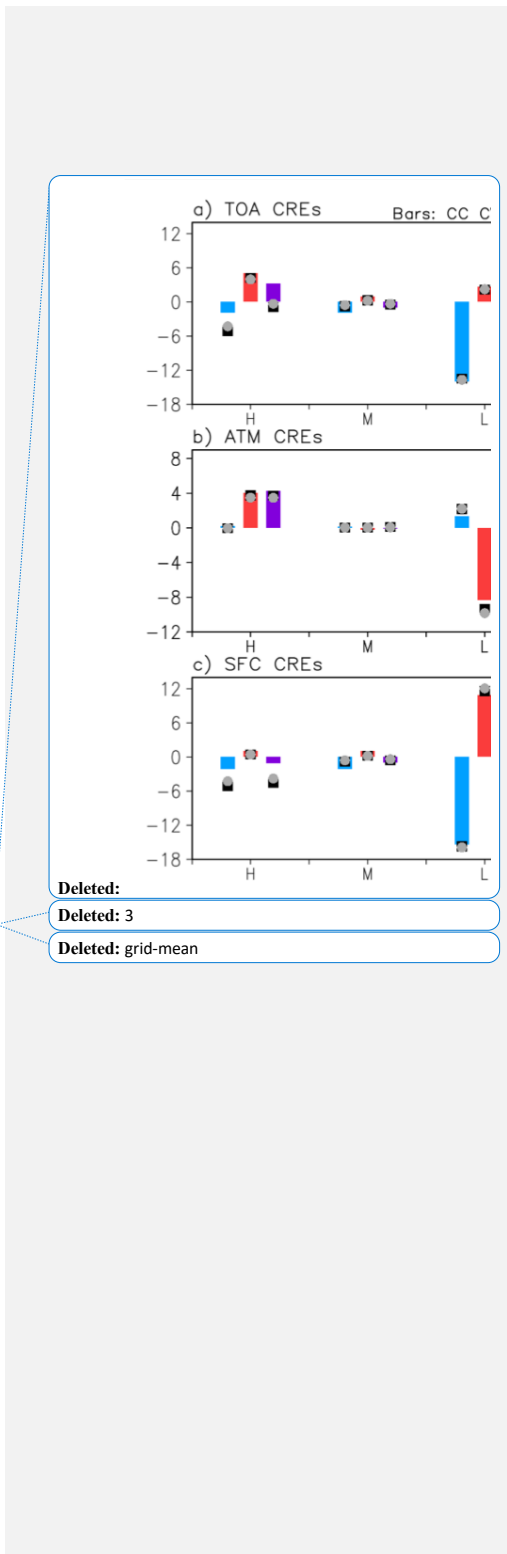
1
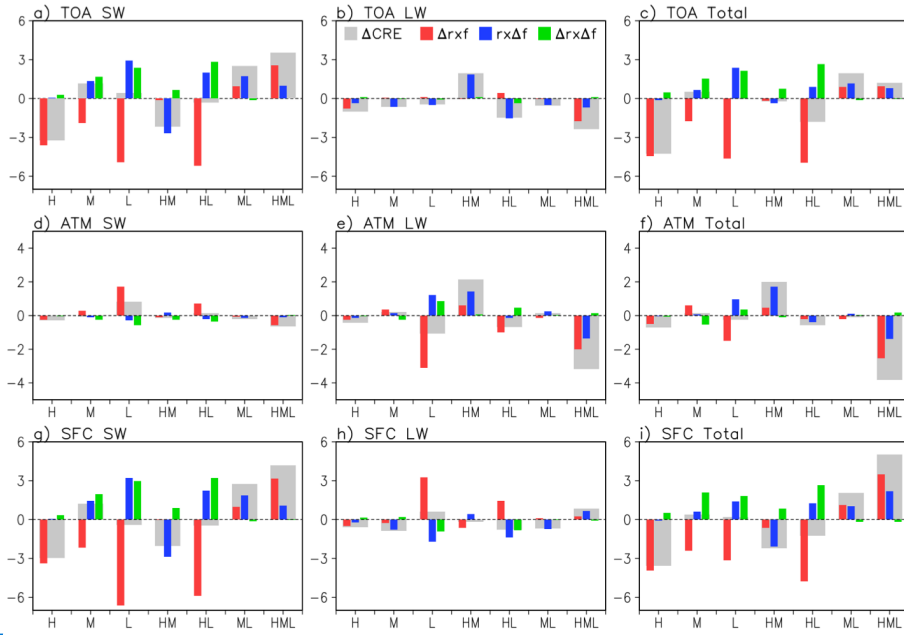
2
3
4   **Figure 6.** Decomposition of all-sky CRE error (Eq. 1) for GEOS-5 CVS classes when the GN

5   overlap assumption is used. Gray bars represent the overall all-sky CRE error, and the remaining

6   bars contributions to that error as follows: red bars represent overcast CRE errors, blue bars

7   RFO errors, and green bars co-variation errors. The nine panels represent all combinations of

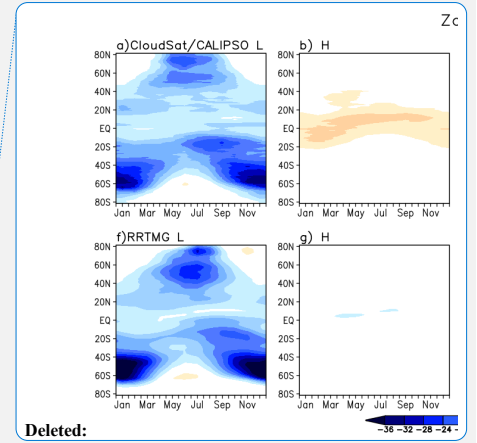8   CRE, namely SW, LW, total at TOA, SFC and within ATM.

9

Deleted:

Deleted: grid-mean

1

2
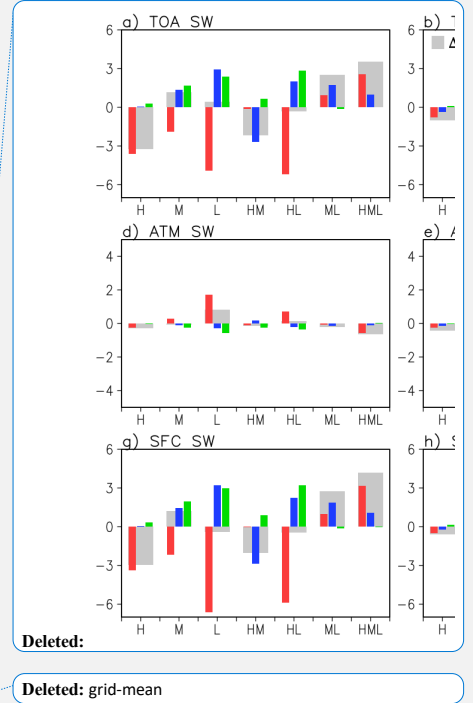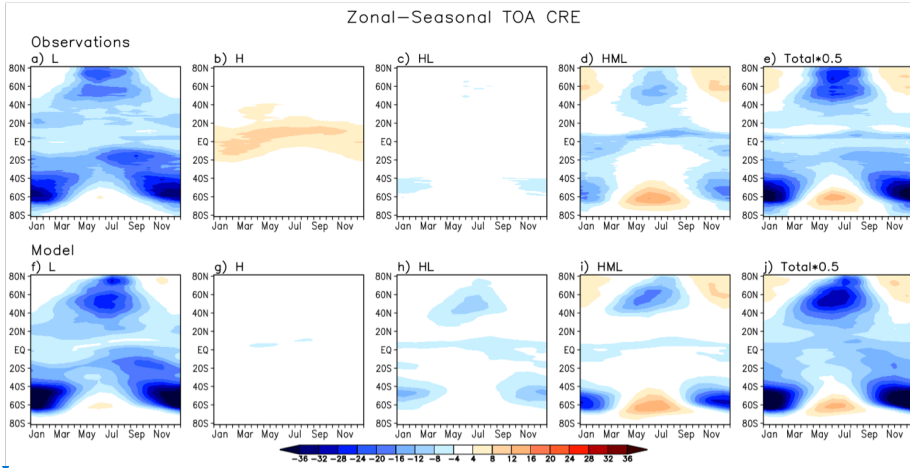
Figure 7. Comparison of the multi-year annual cycle of TOA, total (=SW+LW) all-sky CRE zonal

averages (W/m^2) between observations (top row, panels a to e) and the model (bottom row,

panels f to j) when employing the GN overlap assumption for the four CVS classes with the

greatest all-sky CREs according to Fig. 5. The rightmost set of panels displays the scaled (half)

total CRE of all CVS classes combined.

8

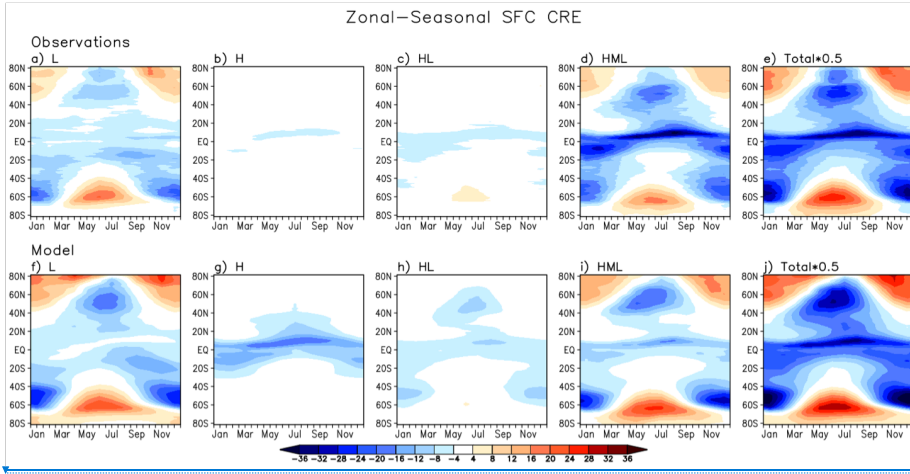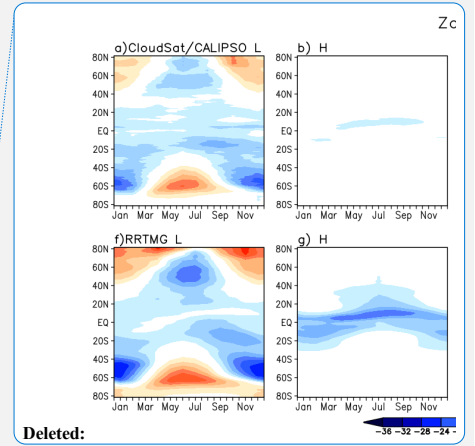**Zonal−Seasonal SFC CRE**

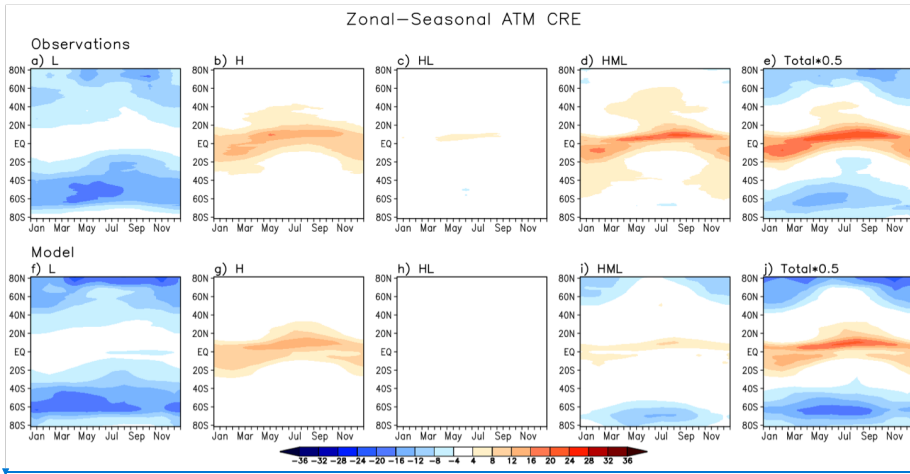**Figure 8.** As Fig.7, but for SFC total all-sky CRE.

1

2

3

Deleted:

Figure 9. As Fig. 7, but for ATM total <u>all-sky</u> CRE.