Dear referee#1,

Thank you for taking the time to review our manuscript. We are especially happy that you agree that the core objective of the paper is clear and highly relevant for the community. Thank you also for the suggested minor revisions and following is a point by point response to each question/suggestion:
Dear referee#2,

Thank you for taking the time to review our paper for your overall supportive comments and useful suggestions about how to improve the article, and in particular, so that it may be more useful as a reference paper for the simulator. Following is a point by point response to each question/suggestion:

General comments

Reviewer 2: Given that this paper will become the main documentation reference for this simulator, we think it would benefit from some discussion and results on the impact of the different methods (mainly #1 vs #3) on other variables listed in Table 1, not only cloud fraction.

Answer: We agree that, since this will be the reference paper for the CLARA-A2 simulator, we should also describe all the simulated variables and not just the cloud fraction. We have expanded the section describing the simulator to also describe every variable in more detail, and decided that this is sufficient for the sake of a reference paper. In this paper, we want to focus on cloud fraction since it is absolutely central to this simulator.

Reviewer 2: The message regarding benefits of method #3 with respect to previous analyses needs to be more specific (e.g. in L385-390). The largest differences between methods occur in the polar regions, with much smaller differences in the rest of the globe. In some places, the paper gives the impression that previous studies where flawed, when in reality many of them did not use data polewards of 60 deg latitude to avoid large uncertainties.

Answer: Granted that, in this introductory paragraph in the conclusion section, we did not highlight the regionally variable impact of choosing a POD-approach (method 3) compared to using a static global optical depth threshold -approach (method 1). We have expanded this paragraph to share the overall regional impacts, rather than just a global assessment as we do now.
We agree that the wording of this paragraph also gave the impression that there are many incorrect studies out there that have assessed simulated clouds in regions where it is inappropriate. This impression is not intentional, and we have rewritten this paragraph to make sure that we are not implying this. We want to send the message that our approach can avoid sizable uncertainties.
We now mention that as long as model evaluations are carried out between +/- 60 degrees, and they usually are, the negative impact of using method one is

1

not very large. However, we also stress that by using a simulator that employs method 3, users need not limit their evaluation to +/- 60, especially not during the polar summer.

See the first paragraph of the conclusion in the marked-up version of the paper to see our reformulation

Reviewer 2: Section 4.2. The observational pattern of trends is regionally inhomogeneous, and therefore Figure 9 is not very informative. Does EC-Earth show smaller trends due to compensation of regional patterns? It would be interesting to show the regional patterns from EC-Earth, perhaps replacing Figure 9 by a figure like Figure 8 but for EC-Earth.

Answer: The EC-Earth pattern of trends is also regionally inhomogeneous. There is some cancellation between the regions but the main reasons for the smaller EC-Earth trend in Figure 9 is due to EC-Earth trends being smaller than observed, especially for the interior Arctic. We have now included the cloud trends for the climate model as we have for CLARA-A2. The results are, however, difficult to interpret. We describe in the text that we only have access to one realization of the model, and therefore no access to the model spread, which would be essential to assess cloud trends correctly here. We decided to keep Figure 9 in order to still raise the point that the choice of simulator does not seem to impact the cloudiness trend in the model.

Most of Sect 4.2 has changed to reflect this

Specific comments

Reviewer 1: Line 7, "compared to the simulators in CFMIP". It should probably read "comparable to the simulators in CFMIP". It took me a few more lines until I understood what the usual approach was. Please clarify.

Answer: We have now reworded this sentence to:

The first method, comparable to the simulators in COSP, relies on a single τ_c-threshold applied globally to delineate cloudy and cloud-free conditions.

Reviewer 1: Line 15, "Method three ...": Isn't this sentence just rewording the statement of the sentence before?

Answer: We agree. We changed: "such as over the Arctic region during the polar night. Method three has the added advantage that it indirectly takes into account that cloud retrievals in some areas are more likely than others to miss some clouds. This situation is common in cold regions where even thick clouds may be inseparable from cold, snow-covered surfaces and also in areas" to

such as in cold regions at night, where thick clouds may be inseparable from cold, snow-covered surfaces, as well as in areas

Reviewer 1: Line 23, "the simulated cloud mask of CLARA-A2": Please add "based on EC-Earth" for clarity.

Answer: OK

Reviewer 1: Line 30: In the abstract I'm missing the information how the location-illumination dependent POD is found/ how the method is calibrated. Please add this information.
Answer: We have now added this sentence earlier on:
The gridded POD values are from the CLARA-A2 validation study by Karlsson and Håkansson (2018)

Reviewer 2: L36-41. This sentence is hard to read, please rewrite.
Answer: We have split this super long sentence into several instead:
Currently, there are only a few CDRs derived from imaging sensors that span more than 30 years. The ISCCP CDR (Young et al., 2018) was the first such dataset and mainly based geostationary satellite data, complemented with data from polar orbiting satellites at high latitudes. The three other CDRs are based on data from the polar-orbiting meteorological satellites from the National Oceanic and Atmospheric Administration (NOAA) and Meteorological Operational Satellite (METOP) series. They are the Pathfinder Atmospheres-Extended (PATMOS-x) (Heidinger et al., 2014), the Cloud_cci (Stengel et al., 2017), and the CLARA-A2 CDR.

Reviewer 1: Lines 94/95: Should read "trends are inverstigated", "Summary and conclusion are given".
Answer: Fixed

Reviewer 1: Line 104, "five pixels from the first scan line and none from the next two scan lines are used to create the GAC measurement.": Please explain why, with another sentence.
Answer: OK. We have added this sentence to clarify the situation:
Saving the data on a GAC pixel resolution was a compromise to drastically reduce the data, a necessity due to limited bandwidth.

Reviewer 1: Line 105: Here you cite Figure 1 OF Karlsson and Hakansson 2018 and not Figure 1 IN THIS manuscript, right? Maybe "(Fig. 1 in Karlsson and Hakansson, 2018) " might be clearer.
Answer: Yes, I understand how this was confusing. Fixed

Reviewer 1: Line 127 and 141: The use of the acronym "SNO" seems unnecessary. You just mention it twice and, at least for me, it's not a very common acronym and thus not easy to read.
Answer: I have remove them

Reviewer 1: Line 140: The same again. Better write "(Fig. 9 in ...".
Answer: Fixed

Reviewer 2: L141-142. What's the difference between gridbox size and area?
Answer: The Fibonacci grid is points spread approximately evenly over the

globe, with the pixels matched to the closest point. The form is not quite round, nor is it a lat/long grid. To avoid confusion, we removed the word 'size' and call it:

a nearly equal-area grid.

Reviewer 2: L148-150. This statement is slightly optimistic. Only subtropical deserts show PODs below 0.4 like most of the Arctic region. Most of the continental regions show larger PODs than the Arctic, and comparable or larger than the Antarctic region.

Answer: We agree that the statement was too broad here. PODs in the polar regions are improving considerably during the polar summer (day), but they are still not reaching values representative of most continental land surfaces. However, a strong point for the situation in the polar summer is that if plotting a somewhat higher COT interval than shown here (e.g., 0.5-0.6), the differences decrease significantly between polar regions and most continental surfaces. This decrease is because of the higher skill in detecting liquid water clouds in the polar summer. The reason why this is not reflected in the current figure is that the very thin clouds in the COT interval 0.20-0.25 mostly consist of thin ice clouds, which are still difficult to detect over ice and snow surfaces in the polar summer. We changed and expanded our statement to:

Another significant result in Fig. 1 is the high POD in the Arctic and Antarctic during the summer months. CLARA-A2 has nearly comparable skill in detecting clouds in these regions during the sunlit months as it has over non-polar land regions. Additionally, in the polar summer, for a somewhat higher COT interval than shown here (e.g., 0.5-0.6), the POD in polar regions increases more than most continental surfaces. This is due to a high skill in detecting liquid water clouds in the polar summer. The POD shown in Fig. 1 is somewhat lower here since clouds in the τ_c interval 0.20-0.25 mostly consist of thin ice clouds which are still difficult to detect over ice and snow surfaces. Overall though, this...

Reviewer 1: Line 150: Can you please comment on the lowest tau detected by CALIOP and its impact on a comparison with the model clouds.

Answer: OK. I added this sentence:

By comparison, the reference dataset, CALIOP can detect clouds with $\tau_c > 0.01$ (Winker et. al., 2009) and is generally stable across any surface.

Reviewer 1: Line 157: Why "IR" instead of a wavelength? Are they different? Then please give a wavelength range. Line 165 and again in line 327, "198307–201506" Please change the date format to something more readable: E.g. "July, 1983 – June 2015"

Answer: I have included 11 micron and updated the date format

~~IR~~

infrared (11 µm)

~~198307–201506~~

July 1983 to June 2015

4

Answer: Yes, this is clearly missing. This subsection is now rewritten to describe all the simulated variables (Section 3). Note, we choose to only shortly explain the simulation for the effective radius as this is described in detail in Pincus et. al., (2012). We reference also as such.

Answer: It may not be that the average cloud phase is irrelevant, but we have decided not to include this quantity. We have removed this confusing sentence. In future versions of the CLARA simulator, we may decide to include it

Answer: OK

Answer: OK, we removed the numbered list

Answer: Thanks, we now point this out as well as pointing out the added advantage of this new simulator approach. That is, we added that the CLARA simulator can simulate cloud fraction and cloud top products all times of the year. We also point out that the CLARA simulator does not produce COT, water path, or 2D CTP-COT histogram products during night time conditions

Answer: Yes, the underlying results that are base for these figures and table where created for the Karlsson et.al., (2018) paper. I can add this information in the captions of the figures and see that it is clear in the text.

Answer: To us, the color scale is OK. However, we changed the top color from violet to dark brown as a compromise, and hopefully, it will be less confusing

Answer: I see your point, that this paragraph seems out of place. I think it may fit better at the beginning of this subsection before we go into detail about the τ_c intervals, illumination etc.

Answer: We have to admit that we probably do not understand this question clearly. The CLARA-A2 simulator is a tool that should be used to facilitate model-to-satellite inter-comparisons and in this particular case, inter-comparisons with the results from the CLARA-A2 climate data record. So we are discussing clouds in the real world and not the cloud situation in a particular future or another scenario. If modeled clouds (channeled through the simulator) deviate from CLARA-A2 observations, it should be an indication of a model problem. This is the main goal for the simulator development.

However, the reviewer is possibly asking how to interpret cases where models systematically place clouds incorrectly in space and then being subject to (potentially) other PODs than what they should have been in the CLARA-A2 simulator. The consequences here should not be large except for the extreme cases when a model place clouds over ice- and snow-covered areas in the polar night (with very low PODs) instead of over adjacent ice-free ocean areas (with very high PODs). Knowing about the unique problems over snow- and ice-covered regions (especially for the polar night) it will be hard to cover this situation adequately knowing about the specific cloud detection issues occurring over snow and ice during night conditions for AVHRR observations.

So, yes, under these particular circumstances, this might be a problem, and perhaps other observational datasets (e.g., from active sensors) would be more suitable to use here. However, for more normal situations, we do not believe this to be a big problem. Geographical mismatches between modeled and observed clouds should be possible to detect as long as the POD variability in the area of interest is not extreme.

We added a brief discussion on this at the end of section 3.2. We plan on leaving the purely lat/long approach in future releases and preferably base the PODs on something like climate zones or surface conditions.

Answer: Yes, the top subplot should be named (a). Also, subplot (d), soon to be (e), should be labeled "EC Earth (#1) - EC Earth (#3)" for clarity. The figures have now been updated.

Answer: I changed the order of the sentences so that I can reference ISCCP-H again as well as the ISCCP simulator

Answer: By three datasets, I am referring to CLARA-A2, ISCCP-H, and EC Earth. For clarity we wrote:
The two satellite datasets and the climate model are limited to ...

Answer: Fixed

Answer: This is a good point. We swapped out the simulated datasets showing absolute cloudiness to showing absolute difference compared to the observations. After reexamining the data, we updated the stated biases since they seemed exaggerated. However, the message stayed the same

Answer: It appears to me that especially lines 333-335 more or else repeat what is said in the information from lines 331-333. I removed the second duplicate and moved the sentence about ISCCP-H underestimating cloudiness to the earlier paragraph, and now I think it reads much better. Otherwise, to me, I think the latter information from lines 336-341 is necessary to explain why the IS-CCP simulator produces more clouds in the Arctic summer that the CLARA simulator as seen in Fig. 7

Answer: Well, yes, this is what we are saying. We are using the wrong notation here and have fixed the unit and description

7

Answer: We have fixed the units, but instead of changing the unit to 1/decade, we decided to keep (%/decade) and describe clearly in the text that the trends are trends in an absolute sense.

Answer: OK

A simulator for the CLARA-A2 cloud climate data record and its application to assess EC-Earth polar cloudiness

Salomon Eliasson[1], Karl-Göran Karlsson[1], and Ulrika Willén[1]

[1]Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 601 76 Norrköping, Sweden

Correspondence: Salomon Eliassson (salomon.eliasson@smhi.se)

Abstract.

This paper describes a new satellite simulator for the Satellite Application Facility on Climate Monitoring (CM SAF) cLoud, Albedo and RAdiation dataset (CLARA), Advanced Very High Resolution Radiometer (AVHRR)-based, version 2 (CLARA-A2) Climate Data Record (CDR). This simulator takes into account the variable skill in cloud detection in the CLARA-A2

5   CDR by using a different approach to other similar satellite simulators to emulate the ability to detect clouds.

In particular, the paper describes three methods to filter out clouds from climate models undetectable by observations. The first method is comparable to the current simulators in Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP), since it relies on a single visible cloud optical depth at 550nm (τ_c) threshold applied globally to delineate cloudy and cloud-free conditions. Method two and three apply long/lat -gridded values separated by day and nighttime

10   conditions. Method two uses gridded varying τ_c as opposed to method one that uses just a τ_c threshold, and method three uses a cloud Probability of Detection (POD) depending on the model τ_c. The gridded POD values are from the CLARA-A2 validation study by Karlsson and Håkansson (2018)

Method two and three replicate the relative ease or difficulty for cloud retrievals depending on the region and illumination. They increase ~~by increasing~~ the cloud sensitivity where the cloud retrievals are relatively straightforward, such as over mid-

15   latitude oceans, and ~~by decreasing~~decrease the sensitivity where cloud retrievals are notoriously tricky. This is the situation for cold regions, especially at night, where ~~, such as over the Arctic region during the polar night. Method three has the added advantage that it indirectly takes into account that cloud retrievals in some areas are more likely than others to miss some clouds. This situation is common in cold regions where even~~in cold regions at night, where thick clouds may be inseparable from cold, snow-covered surfaces ~~and also~~, as well as in areas with an abundance of broken and small scale cumulus clouds

20   such as the atmospheric subsidence regions over the ocean.

The simulator, together with the International Satellite Cloud Climatology Project (ISCCP) simulator of COSP, is used to assess Arctic clouds in the EC-Earth climate model compared to the CLARA-A2 and ISCCP-H CDRs. Compared to CLARA-A2, EC-Earth is shown to generally underestimate cloudiness in the Arctic. However, compared to ISCCP and its simulator, the opposite conclusion is reached. ~~Previous studies have found that the CLARA-A2 CDR performs well in the Arctic during~~

25   ~~the summer months.~~ Based on EC-Earth, this paper shows that the simulated cloud mask of CLARA-A2 using method three is more representative of the CDR than method one which is used for the ISCCP simulator ~~using a global τ_c threshold to simulate clouds~~.

The simulator substantially improves the simulation of the CLARA-A2 detected clouds, especially in the polar regions, by accounting for the variable cloud detection skill over the year. The approach to cloud simulation based on the POD of clouds depending on their τ_c, location, and illumination is the preferred one as it reduces cloudiness over a range of cloud optical depths. Climate model comparisons with satellite-derived information can be significantly improved by this approach, mainly by reducing the risk of misinterpreting problems with satellite retrievals as cloudiness features. Since previous studies found that the CLARA-A2 CDR performs well in the Arctic during the summer months, and the method three is more representative than method one, the conclusion is that EC-Earth likely underestimates clouds in the Arctic summer.

1 Introduction

Clouds constitute one of the most significant sources of uncertainties for projecting the future climate (IPCC, 2014). Therefore, countless studies have been made testing and improving the skill of climate models in this regard over the years (e.g., Waliser et al., 2009). As more and more information on cloud climatologies from satellite sensors are available in CDRs, climate models have been able to improve their representation of clouds continuously, and hence their description of the climate system itself.

~~To date~~Currently, there are only a few CDRs derived from imaging sensors that span more than 30 years. The ISCCP CDR (Young et al., 2018) was the first such dataset and mainly based geostationary satellite data, complemented with data from polar orbiting satellites at high latitudes. The three other CDRs are based on data from the polar-orbiting meteorological satellites from the National Oceanic and Atmospheric Administration (NOAA) and Meteorological Operational Satellite (METOP) series. They are the Pathfinder Atmospheres- Extended (PATMOS-x) (Heidinger et al., 2014), the Cloud_cci (Stengel et al., 2017), and the CLARA-A2 (Karlsson et al., 2017) CDRs. The long length of these CDRs make them ideal for assessing the cloud climatologies of climate models.

However, to directly compare model clouds to cloud observations from satellites is akin to comparing "apples to oranges" as is explained in Waliser et al. (2009); Eliasson et al. (2011), and many others. Two of the primary considerations to make when comparing climate models to satellite observations is their very different horizontal and vertical scales, and the observations' finite sensitivity to clouds. Therefore, nowadays, in order to utilize the CDRs from satellite data, the CDRs usually need to be simulated from the model atmosphere with these attributes/limitations in mind.

In general, satellite simulators create cloud products or brightness temperatures that would have been made from satellite measurements if the model atmosphere was the real atmosphere. The simulators' objective is to emulate the inherent limitations, sensitivity, and geometry of the real retrievals. One of the main tasks for these simulators, among others, is to filter out model clouds that would not be detected by the instrument behind the cloud CDR. These simulated satellite products can then be compared directly to the observations.

Satellite simulators are primarily used to validate Earth System Models (ESMs)~~earth system models~~ such as climate models. Although satellite simulators bridge the gap between models and observations by significantly reducing the comparison uncertainties, they do not eliminate them, and this should be taken into account when comparing satellite product simulations to the

observations (Pincus et al., 2012). This paper introduces the CLARA-A2 satellite simulator v1.0, for use in model validations compared to the CLARA-A2 CDR.

The COSP (Bodas-Salcedo et al., 2011; Swales et al., 2018) was developed to gather and provide a suite of satellite simulators. These simulators provide column-integrated cloud retrievals, just as the datasets they represent, and therefore they need the cloud averages on the coarse grid of climate models to be translated into many smaller subcolumns for each model long/lat-grid box (Jakob and Klein, 1999; Pincus et al., 2006). The number of subcolumns per grid depends on the host models' resolution, and typically number around $100\times$ the model resolution in degrees. Therefore, if a model has a resolution of $0.7°$, the simulator will generate 70 subcolumns per horizontal grid. ~~As described in Jakob and Klein (1999), the subcolumns in a grid produce a horizontal cloud distribution, and each subcolumn has a cloud vertical structure determined according to the cloud overlap assumptions of the host model.~~ The cloud retrieval simulations are further carried out on each of these subcolumns.

The ISCCP (Jakob and Klein, 1999), the MODerate resolution Imaging Spectroradiometer (MODIS) (Pincus et al., 2012), and the Multi-angle Imaging SpectroRadiometer (MISR) simulators are the visible/infrared (VIS/IR) satellite dataset simulators in COSP. The CLARA-A2 cloud products are also retrieved using an instrument that measures in this frequency range, and hence the CLARA-A2 simulator has many similarities with these. Other VIS/IR satellite simulators not included in COSP to date are the Spinning Enhanced Visible Infrared Imager (SEVIRI) (Bugliaro et al., 2011) and the Cloud_cci (Eliasson et al., 2019) simulators.

All satellite datasets based on VIS/IR data have regionally varying skill in detecting clouds, and all retrievals suffer when clouds are too tenuous to detect, or obscured. The removal of would-be undetectable clouds from the model is an essential feature of satellite simulators and to date is being carried out by comparing the τ_c of a subcolumn to some threshold value. To date, the simulators in COSP and the Cloud_cci simulator rely on a global static τ_c value to reclassify subcolumns, with an optical depth less than this threshold, as cloud free. It is well established that all cloud masks based on the AVHRR channels have a variable skill, mainly depending on the underlying surface and the illumination conditions (e.g., Karlsson and Håkansson, 2018). Karlsson and Håkansson (2018) studied the performance of the CLARA-A2 cloud mask against Cloud-Aerosol Lidar with Orthogonal Polarisation (CALIOP) measurements in detail and produced global statistics for different τ_c thresholds, the probability of cloud detection, and the rate of falsely detected clouds (false alarm rate), on a global and regional basis. For instance, they showed that the general likelihood of detecting clouds is much higher over warm ocean surfaces than over perpetually ice-covered regions and likewise that in some regions, e.g., deserts and other dry surfaces, retrievals there are relatively susceptible to producing false clouds.

It is clear that the use of a fixed τ_c threshold, applied globally to modeled cloud fields in order to simulate satellite-based cloud detection limitations, is a substantial simplification of the actual observation conditions. Therefore a completely new approach is introduced in this paper describing a simulator for the CLARA-A2 CDR applying spatially and temporally varying cloud detection thresholds. Employing this novel approach to simulating observed cloud cover, should put the confidence in cloud cover comparisons between the climate models and the CLARA-A2 CDR on a stronger footing. The CLARA-A2 simulator also incorporates a method of model temporal sampling in order to reduce errors potentially introduced by not taking

the different and changing equatorial overpass times of the satellites used in the CLARA-A2 CDR, into account. This approach

95   is also used in the Cloud_cci simulator and is motivated and described in Eliasson et al. (2019).

The article structure is as follows: Sect. 2.1, Sect. 2.2, and Sect. 2.3 describe the CLARA-A2 CDR, ISCCP-H series (ISCCP-H) CDR, and the EC-Earth climate model (Hazeleger et al., 2010) respectively. Section 3 describes the CLARA-A2 simulator and the simulated variables and Sect. 3.1 offers a description and demonstration of the simulated cloud masks. The CLARA-A2 simulator approach is demonstrated and tested over the Arctic region where trends in polar summer cloudiness are investigated

100   using simulations from the EC-Earth climate model in Sect. 4. A summary and conclusion are given in Sect. 5.

2   Data

2.1   The CLARA-A2 climate data record

The CLARA-A2 CDR (Karlsson et al., 2017) is based on long term measurements from the AVHRR instrument operated onboard polar orbiting NOAA satellites as well as onboard the MetOp polar orbiters operated by European Organisation for

105   the Exploitation of Meteorological Satellites (EUMETSAT) since 2006. AVHRR measures in five spectral channels (two visible and three infrared channels) with an original horizontal field of view (FOV) resolution at the nadir of $1.1\,\mathrm{km}$. However, the data used in CLARA-A2 is a reduced resolution ($5\,\mathrm{km}$) resampled version of these measurements, called global area coverage (GAC), where three consecutive scanlines made up of 3x5 original FOVs make one GAC pixel. Saving the data on a GAC pixel resolution was a compromise to drastically reduce the data amount, a necessity due to limited bandwidth and onboard storage

110   capacity.

Specifically, the average radiance from four out of five pixels from the first scan line and none from the next two scan lines are used to create the GAC measurement. Thus, only about $27\,\%$ of the nominal GAC FOV is actually used (see Fig. 1 in Karlsson and Håkansson (2018)). Only GAC data is available globally (i.e., being archived) over the full period since the introduction of the AVHRR sensor in space.

115   The visible radiances were inter-calibrated and homogenized, using MODIS data as a reference before applying the multiple parameter retrievals. The inter-calibration uses the method introduced by Heidinger et al. (2010), which is now updated using MODIS Collection 6 as well extended by six years. The calibration of infrared AVHRR channels is based on the standard NOAA calibration methodology utilizing an onboard blackbody reference (Rao et al., 1993). CLARA-A2 is an improved and extended follow-up of the first version, CLARA, AVHRR-based, version 1 (CLARA-A1) of the record (Karlsson et al., 2013)

120   and is extended to cover 34 years (1982–2015).

CLARA-A2 features a range of cloud products: cloud mask (cloud amount), cloud top temperature/pressure/height, cloud thermodynamic phase, and for liquid and ice clouds separately, cloud optical thickness, particle effective radius, and cloud water path. Cloud products are available as monthly and daily averages in a $0.25°$ latitude-longitude grid and also as daily resampled global products (Level 2b) on a $0.05°$ grid for individual satellites. The CDR also includes multi-parameter distributions (i.e.,

125   joint frequency histograms of cloud optical thickness, cloud top pressure, and cloud phase) for daytime conditions. Besides

cloud products, CLARA-A2 also includes surface radiation budget and surface albedo products. Examples of CLARA-A2 products can be found in Karlsson et al. (2017).

In this study, we focus exclusively on the AVHRR GAC cloud mask because of its central importance for the quality of all other CLARA-A2 products. Validation results for other CLARA-A2 products can be found in Karlsson et al. (2017) and CMSAF1 (2017). The method for generating the CLARA-A2 cloud mask originates from Dybbroe et al. (2005), but significant improvements and adaptations since then were made to enable reliable processing of the historic AVHRR GAC record (CMSAF2, 2017).

2.1.1 The skill of the CLARA-A2 CDR

As mentioned earlier, Karlsson and Håkansson (2018) performed an extensive validation of the CLARA-A2 cloud mask against ~~SNO~~simultaneous nadir observations of CALIOP retrievals, and following is a recap of their main results. The goal was to find out at which optical depth thin clouds were thick enough to have a 50% probability of being detected. They investigated the global performance of the CLARA-A2 cloud mask on a global equal-area grid with a $300\,\mathrm{km}$ resolution, covering different surface types, and separately for daytime and nighttime conditions. This detection level can be considered the baseline for any cloud mask; i.e., the smallest τ_c threshold where the cloud mask detects more clouds than it misses. They found that the global mean minimum cloud optical thickness was $\tau_c = 0.225$. However, importantly, their results showed that the global mean is far from being representative of all local conditions. For instance, a τ_c threshold value of 0.07 is a better approximation over ice-free oceanic regions at mid-latitudes, whereas a τ_c threshold value as high as 4.5 is suitable for some ice-capped regions such as over Greenland and Antarctica. By comparison, the reference dataset, CALIOP can detect clouds with $\tau_c > 0.01$ (Winker et al., 2009) and is generally stable across any surface.

However, the capability of the cloud mask in CLARA-A2 is better described by the POD of clouds rather than a τ_c threshold. Karlsson and Håkansson (2018) showed that even if all thin clouds with a τ_c less than 0.225 are removed from the comparison, i.e., by reclassifying such CALIOP reference clouds as cloud-free, the POD varies considerably per region. Additionally, they showed that for most regions in the world, the probability of detecting clouds with a τ_c near the average of 0.225 is higher than 50% (see Fig. 9 in Karlsson and Håkansson (2018)).

Through their validation studies, POD was calculated for τ_c-intervals (or bins) based on these ~~SNO~~simultaneous nadir observation validations on an equal-area Fibonacci grid with about a $300\,\mathrm{km}$ radius. A Fibonacci grid is a type of grid where each grid box is nearly equal ~~in size and~~ area (see Karlsson and Håkansson (2018) and references therein for more information). Figure 1 shows the different POD for clouds that have an optical depth that falls in the optical depth interval centered around 0.225 ($0.2 < \tau_c < 0.25$) for daytime, nighttime and all conditions. The figure shows that the POD of clouds in this optical depth range is dependent on whether clouds are sunlit[1] or not, especially in the polar regions. The global average POD in this interval, but also all POD-intervals (not shown), is somewhat skewed towards lower values due to the poor performance in the Polar regions during night time. Another significant result in Fig. 1 is the ~~exceptionally good POD results~~high POD in the Arctic and Antarctic during the summer months. CLARA-A2 has ~~equal~~nearly comparable skill in detecting clouds in these regions during

---

[1]Sunlit refers to when the solar zenith angle is less than ~~80°~~84°.
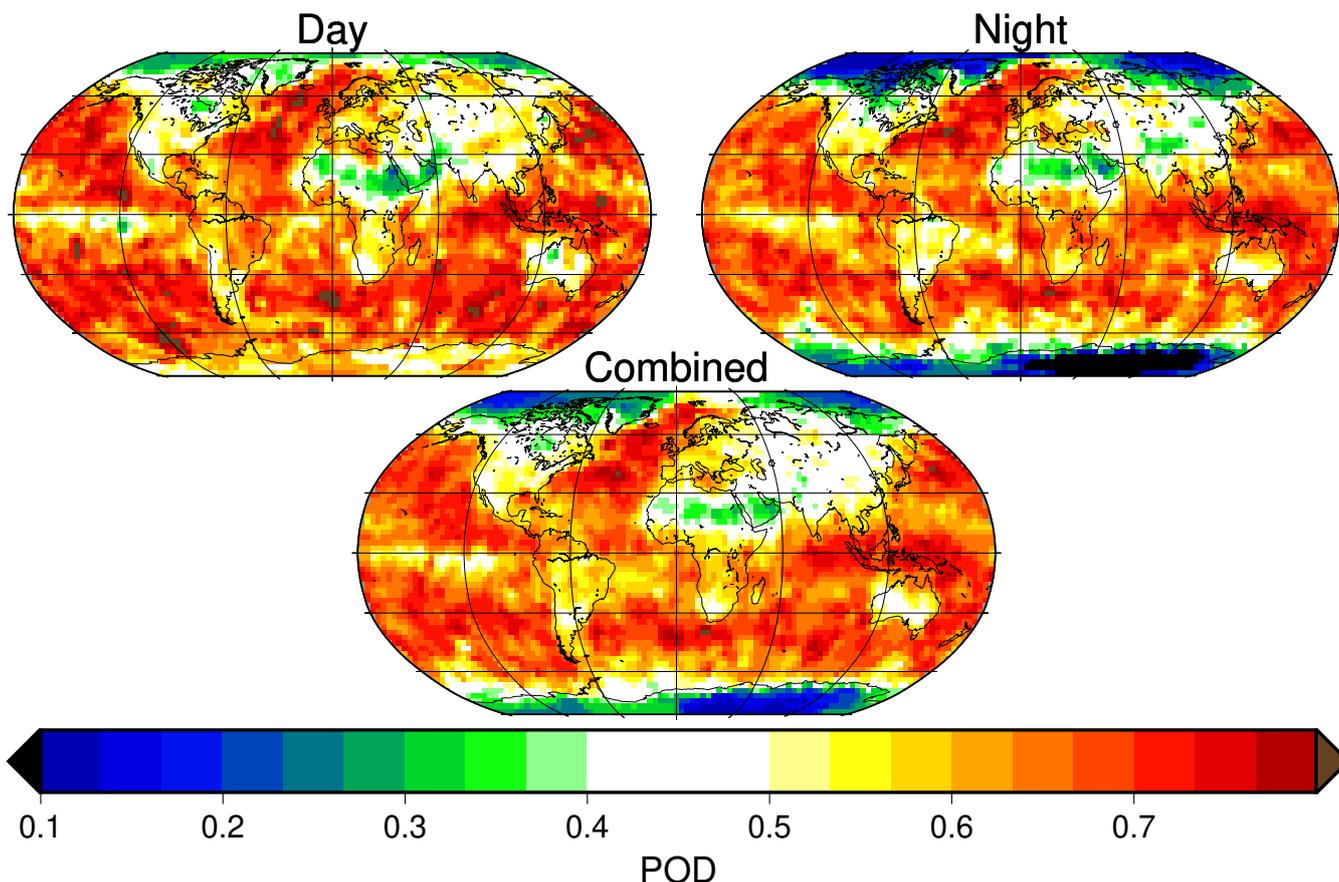
Figure 1. Probability of detection of clouds having an optical depth between $0.2 \leq \tau_c < 0.25$. The τ_c in the center of this interval, 0.225, is the global average of the smallest τ_c threshold where the CLARA-A2 cloud mask detects more clouds than it misses according to CALIOP.(see text)

the sunlit months as it has over non-polar land regions. Additionally, in the polar summer, for a somewhat higher COT interval than shown here (e.g., 0.5-0.6), the POD in polar regions increases more than most continental surfaces. This is due to a high skill in detecting liquid water clouds in the polar summer. The POD shown in Fig. 1 is somewhat lower here since clouds in the τ_c interval 0.20-0.25 mostly consist of thin ice clouds which are still difficult to detect over ice and snow surfaces. Overall though, this result further establishes the CLARA-A2 CDR as very suitable for cloud studies in the polar summer.

2.2 ISCCP-H

The ISCCP-H CDR (Young et al., 2018) is a recently released high resolution version of the ISCCP CDR (Rossow and Schiffer, 1999) that starts in July 1983 and ends in June 2015 due to data availability at the time of this study. The ISCCP CDR comprises of geostationary and polar-orbiting satellites, where data from the geostationary satellites have precedence at low

Table 1. The cloud variables produced by the simulator. The middle column specifies the separate categories available for each variable, and the third column indicates under which illumination conditions the variables are available.

Cloud variable	Categories	day/night
Cloud fraction	total, ice, liquid, low, mid and high	day and night
Cloud top	height, temperature, pressure	day and night
τ_c	liquid, ice	day only
cloud particle effective radius (r_e)	liquid, ice	day only
cloud water path (CWP)	liquid, ice	day only
cloud top pressure (CTP)-τ_c 2D histograms	liquid, ice	day only

and mid-latitudes (absolute latitude $< 55°$). The main improvement of ISCCP-H CDR is that it is on a higher resolution spatial grid compared to its predecessor and covers a longer period. Otherwise the ISCCP-H CDR is quite similar to previous ISCCP
170   versions. The CDR uses bi-spectral radiances, with one channel in the visible ($0.6\,\mu m$) and one in the ~~IR~~infrared ($11\,\mu m$). This CDR is described at more length in Karlsson and Devasthale (2018) and Tzallas et al. (2019).

2.3   The EC-Earth model

The EC-Earth climate model (Hazeleger et al., 2010, 2012) is an ESM with its atmospheric component based on the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The version used for this
175   study is 3.3, based on IFS cycle 36r4 is on a horizontal resolution of T255 with 91 vertical layers. The variant used in this study is the EC-Earth-Veg3 Atmospheric Model Inter-comparison Project (AMIP) simulation with prescribed monthly sea surface temperatures and sea ice conditions to enable comparisons with atmospheric observations. The temporal range used to demonstrate the simulator covers 1982 to 2015 when compared only to the CLARA-A2 CDR and covers ~~198307–201506~~July 1983 to June 2015 when ISCCP-H is involved in the comparison. EC-Earth simulated ISCCP clouds are produced at run time
180   through the COSP. In terms of cloudiness, EC Earth has no lower or upper limit to cloud optical thickness aside from numerical precision. Therefore any satellite simulator, should~~will~~ always produce less cloudiness than the direct model output.

3   Description of the CLARA-A2 simulator

Tab. 1 lists the variables simulated by the CLARA simulator, and this section provides an overview of them. As briefly described in the introduction and detailed in Bodas-Salcedo et al. (2011) and Jakob and Klein (1999), the CLARA-A2 simulator relies on
185   subcolumns within the climate model grid, as all COSP simulators do, to simulate the observational datasets' cloud variables. The subcolumns created in each model grid together produce the horizontal and vertical cloud structure that preserves the

internal cloud overlap assumption of the host model. Each subcolumn has the same number of layers as the model, and each layer in a subcolumn is either completely cloudy or clear.

The next stage in the simulation is to map the average model layer in-cloud [2] optical depth, water content, and effective radius, both liquid and ice phase, to the cloudy layers of each subcolumn. Every subcolumn is determined to be either cloud-free or cloudy, and the simulator performs cloud retrievals on each 'cloudy' subcolumn, and these represent the column-integrated retrievals of CLARA-A2. Finally, the simulated cloud parameters are averaged to the climate model grid so that they are ready to be directly compared to observations. Table 1 provides an overview of the simulated variables included in this simulator. The CLARA-A2 satellite simulator can currently only be run in an "offline"-mode, meaning that it relies on access to pre-processed model output files. Following is a short description of the simulated cloud retrieval simulation:

Cloud microphysics

The cloud microphysical retrievals τ_c, r_e, Water Path (WP), and cloud phase are simulated using the same method described in Eliasson et al. (2019), which very closely resembles the method described in Pincus et al. (2012). The dominant cloud water phase of the top optical depth of the cloud determines the simulated cloud water phase. The simulation of the effective radius r_e is calculated by comparing the top of the atmosphere reflectance, calculated by the adding–doubling technique, to lookup tables of reflectance versus cloud effective radius. The lookup tables for the effective radius simulation rely on the same microphysical model as the CLARA-A2 CDR (see details in Karlsson et al. (2017)). The simulated optical depth and cloud water path is the sum in the column. For consistency with observations, if a cloud parameter requires sunlight for ~~its'~~ its retrieval, it will only be simulated if the calculated solar zenith angle is less then ~~80°~~84°. These include the cloud microphysical retrievals τ_c, r_e, WP, and the CTP-τ_c histograms.

Cloud top

The simulated CTP, cloud top height (CTH), and cloud top temperature (CTT) are calculated by two methods depending on if the clouds are optically thick or not. If a subcolumn has a simulated cloud optical depth, $\tau_c \geq 5$, it is considered opaque, and finding the cloud top is achieved by matching a calculated brightness temperature to the model temperature profile, i.e., precisely the same approach to simulate cloud top as used by the ISCCP simulator (Jakob and Klein, 1999).

However, acknowledging that optically thinner, i.e., semi-transparent clouds, are more difficult to accurately determine the cloud top (Håkansson et al., 2018), the CLARA simulator has a different method for simulating thin clouds. First, the simulator finds the CTH that is one optical depth down from the physical top of the model cloud. This is the same manner as the MODIS simulator (Pincus et al., 2012) and Cloud_cci simulator Eliasson et al. (2019) finds the retrieved cloud top. The CLARA simulator then offsets this height using the median error in CTH for semi-transparent clouds in the CLARA CDR (Table 13 in NWCSAF, 2018). This approach emulates the real world performance of the CLARA cloud top retrievals more closely for semi-transparent clouds than treating all clouds as opaque. The offsets used are $257\,\mathrm{m}$, $-145\,\mathrm{m}$, and $-3336\,\mathrm{m}$ for low (CTP$\geq$680 hPa), middle (440 hPa$\leq$CTP$>$680 hPa), and high clouds (CTP$<$440 hPa) respectively.

---

[2]The climate model provides the combined cloud free and cloudy component average for cloud variables. The simulator needs the in-cloud amounts, i.e., average/cloud fraction

3.1 Simulating CLARA-A2 cloud masks

As mentioned in Sect. 1, the main feature of the CLARA-A2 simulator is a more sophisticated simulation of the observational datasets' cloud mask. It is possible to choose between one of the three methods of cloud mask simulation described below:

1) ~~To use a global static τ_c, and treat all subcolumns with a τ_c less than 0.225 as cloud free. This method is used for the equivalent simulators in COSP, except that they use a slightly higher τ_c threshold of 0.3.~~

2) ~~To use gridded optical depth thresholds separately for day and night conditions.~~

3) ~~Make use of the gridded POD for cloud retrievals separately for day and night conditions.~~

~~Following is a short description of these three approaches.~~

3.1.1 A globally static optical depth threshold

Method one is to simulate the cloud mask by using one global minimum cloud optical depth value. This is the classical approach used by the ISCCP, MODIS, MISR, and the Cloud_cci simulators. For the ISCCP, MODIS, and MISR simulators, this global limit is set to $\tau_c = 0.3$ (Pincus et al., 2012), and for the Cloud_cci simulator (Eliasson et al., 2019), to 0.2. As mentioned earlier, the global average τ_c threshold for the CLARA-A2 CDR is 0.225, and thus the threshold value used in method one of the CLARA-A2 simulator.

By the approach used in this method, $100\,\%$ of the cloudy subcolumns with an optical thickness less than the global average τ_c limit are treated as being cloud-free and $100\,\%$ of the subcolumns above this threshold are treated as cloudy. Since the threshold is a global average, this method does not consider the illumination conditions or the geographical location of the retrieval. The advantage of this approach is its robustness and simplicity. However, as mentioned in Sect. 2.1, this approach can lead to very misrepresentative cloud mask simulations in some geographical regions.

The cloud retrieval simulations in COSP are only carried out during sunlit conditions. However, the next two approaches described below also simulate the cloud amount and the cloud top retrievals also during night time conditions. This doesn't apply to r_e, τ_c, WP, or the CTP–τ_c 2D histograms.

3.1.2 Gridded optical depth thresholds

The second method uses varying gridded optical depth thresholds. This method also relies on the robust and straightforward approach of reclassifying subcolumns with a small optical depth as cloud-free, while keeping those above this threshold cloudy. However, this method is designed to also take into account that the τ_c-threshold, or cloud detection limit, varies geographically and depends on the solar illumination. This method relies on the gridded data that are used in Fig. 12 in Karlsson and Håkansson (2018) that shows the smallest τ_c threshold where the CLARA-A2 cloud mask detects more clouds than it misses (see Sect. 2.1).

Figure 2 shows the detection limits used in the simulator according to this method. As shown by the figure, the τ_c threshold varies quite strongly regionally and also depends on if the CLARA-A2 cloud mask can make use of solar channels or not. The
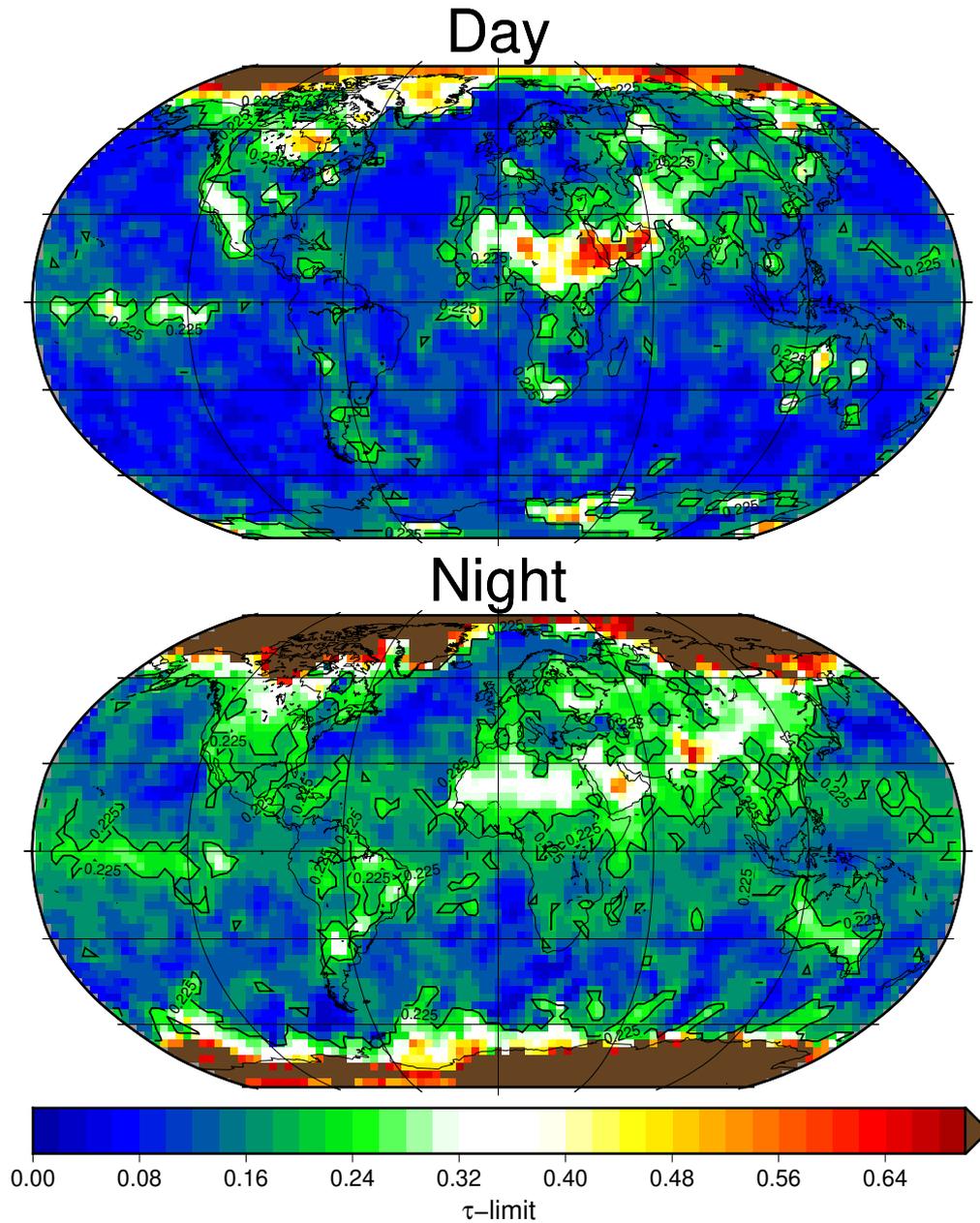
Figure 2. The gridded cloud detection limit, i.e., the smallest τ_c threshold where the CLARA-A2 cloud mask detects more clouds than it misses according to CALIOP for sunlit (top) and nighttime (bottom) conditions. For reference, the global average τ_c -threshold = 0.225 is shown as contour lines. These results are based on the results from the Karlsson and Håkansson (2018) study.

global average τ_c-threshold, included for reference in the figure, clearly shows that during sunlit conditions, the cloud mask is much more sensitive to thin clouds than a global average value of $\tau_c = 0.225$ suggests.

During sunlit conditions, the regions with the least cloud sensitivity are over the Arctic and the desert regions of the Sahara and Arabia, as well as a large patch in the central Pacific. During nighttime conditions, especially over the oceans, the cloud mask is generally less sensitive and is particularly degraded in the ice-covered regions. However, there is an improvement in cloud sensitivity in some regions during nighttime conditions. For instance, in the desert regions of Northern Africa and the Arabian Peninsula, and the worst performing areas in the central Pacific, the cloud mask is somewhat surprisingly better than when these regions are sunlit. A more in-depth validation study on CLARA-A2 is provided in Karlsson et al. (2017) and Karlsson and Håkansson (2018).

Their results demonstrate that using two sets of gridded detection-limits gives a more realistic cloud mask, one for sunlit and one for nighttime conditions. Method two is more realistic than the global static minimum optical depth approach of method one (Sect. 3.1.1). However, the authors of this paper advocate the further improved simulated cloud mask based on the use of PODs described in the next section that also emulates some of the expected variability in cloud detection over a range of cloud optical depths.

3.1.3 Probability of cloud detection

The third method is an approach to simulate the CLARA-A2 cloud mask using the POD, provided on a roughly $300\,\text{km}$ grid, as a function of the cloud's optical thickness. These POD, discussed in Sect. 2.1, are treated as the likelihood that the cloud mask would detect the model cloud given its optical thickness, geographical location, and whether or not it is sunlit.

inserted here: The simulator uses computer-generated random numbers for comparison to the gridded POD value found in a lookup table, where one set of optical depth dependent- PODs is for sunlit, and one is for nighttime conditions. The simulator assigns a random number between 0–1 to each subcolumn at the initiation. After the simulated τ_c is computed, the column integrated τ_c, latitude, and longitude are used to find the POD value from the lookup table for comparison. A subcolumn is cloudy, only if its assigned random number is less than the POD. Therefore, if the probability of detection of a cloud with a specific optical depth is 0.05, even though it is very transparent, there is still a $5\,\%$ chance the subcolumn will be considered cloudy. Conversely, regardless of how optically thick a cloud is in a subcolumn, there is a non-zero chance this subcolumn will not be flagged as cloudy, and hence not included in any further cloud simulations.

The look up table of gridded POD used by the simulator contains separate values for each of the ~~is reported separately for the set of~~ τ_c- intervals listed in Tab. 2. The main purpose of Tab. 2 is to list all of the POD intervals used to simulate the cloud mask, but it also provides a summary of average POD separated into Global, Ocean, Land outside the polar regions, and the Polar regions during sunlit conditions (nighttime in braces). As is completely intuitive, the POD increases for optically thicker clouds for all regions, and in general, the cloud mask is more sensitive to clouds over ice-free oceans. Additionally, nowhere, and not even for the thickest clouds, does the POD reach 1. The reasons for this seeming paradox are discussed at length in Karlsson and Håkansson (2018), and here is a summary:

11

1. Thick clouds are likely undetectable if they have the same temperature as the underlying surface during nighttime conditions when solar reflectivity measurements are not available.

2. Collocation errors between CALIOP and AVHRR can cause a mismatch between the datasets. Some collocation error is unavoidable due to the maximum time difference of 3 minutes, and that sometimes the geo-location data for AVHRR itself may not be sufficiently accurate.

3. Even if the data is ideally collocated, the FOVs of the measurements most likely differ somewhat due to how the GAC footprint is made (see Fig. 1 in Karlsson and Håkansson (2018) and Sect. 2.1 here).

In fairness, only the first point directly has to do with the skill of the CLARA-A2 cloud mask and thus should be simulated. The next two bullets have to do with imperfections in the validation process, and therefore should not be simulated. Unfortunately, at this moment, the POD is reduced by all three points, and in the future, it could make sense to estimate and take into account the impact of all three of these considerations in the simulator.

On the other hand, results from Tab. 2 indicate that the impact of points two and three may not be that strong after all. Over global oceans during the daytime, where highest POD values are found, the detection rate for the most optically thick clouds is $98\,\%$ indicating, on average, that the combined error from points two and three is probably less than $2\,\%$. However, in some oceanic regions where relatively thick inhomogeneous clouds are prevalent, such as the stratocumulus-dominated regions off the west coast of South America and southern Africa, POD values are slightly below 0.9, hence the impact of points two and three may not be negligible in these regions.

To illustrate the global distribution of POD, Fig. 3 contrasts two τ_c-intervals used by the simulator. Clouds that fall in the interval centered at $\tau_c = 0.125$, which are translucent clouds at only half the global average τ_c -limit (see Sect. 3.1.1), generally have a low POD. The POD is particularly low in this interval over land and during nighttime conditions. However, take notice that especially over ocean areas and especially during sunlit hours there is at least a $50\,\%$ POD despite the clouds being so thin.

For clouds centered at $\tau_c = 0.55$, which is about twice the global average detection limit, the PODs are predictably quite high in general. However, again, this is not true globally. Even though the clouds are relatively thick, in areas such as Northern Africa, the Arabian peninsula, and the Polar regions, the POD is only around $50\,\%$. Another striking feature is that for these semi-transparent clouds, the POD over nearly all regions, except the poles, are higher for cloud retrievals made during nighttime conditions. This result is demonstrated further in Fig. 4. Outside the polar regions, clouds in the τ_c-intervals from 0.2 to 0.6 have a higher POD during nighttime conditions overall (especially in the Tropics), whereas for clouds thinner or thicker than this interval, the daytime cloud masks have better success.

That this slightly improved detectability at night for clouds in the τ_c range 0.5-1.0 is a robust feature is supported by intercomparisons made between CLARA-A2 and other AVHRR-based datasets (e.g., Karlsson et al., 2017; Karlsson and Devasthale, 2018). They found (although not explicitly reported in the papers) the same behavior for results from PATMOS-x and Cloud-cci compared to CALIOP observations. Whether to interpret this as an indeed improved nighttime detectability for AVHRR-based methods or something caused by the CALIOP observation reference (e.g., enhanced daytime problems due to
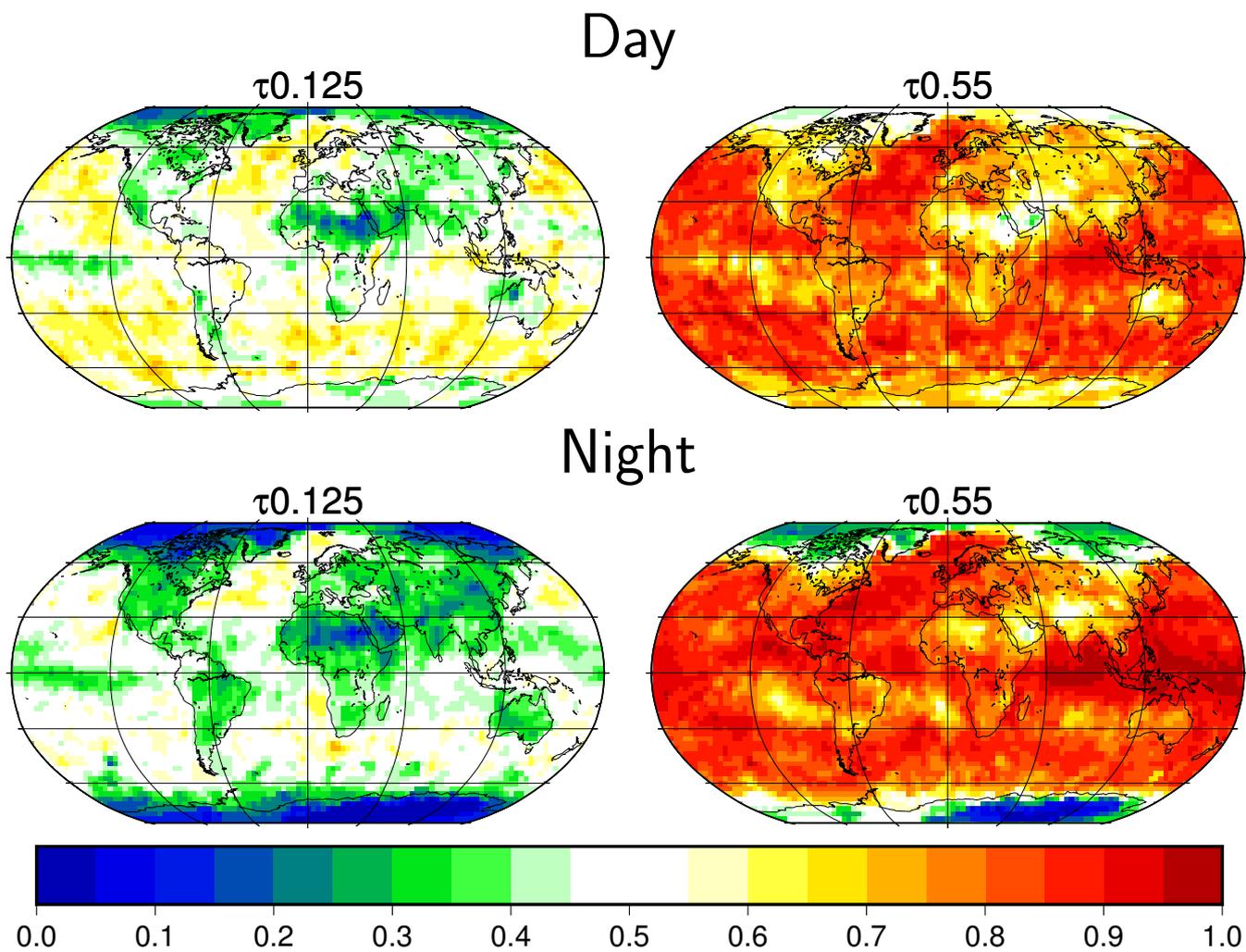
Figure 3. The probability of detection at two τ_c-intervals centered at 0.125 and 0.55 for day and night conditions. These results are based on the results from the Karlsson and Håkansson (2018) study.

Table 2. The probability of cloud detection for the CLARA-A2 cloud mask separated by intervals of CALIOP cloud optical thickness. This table shows the regional averages based on the POD values used in the simulator of large geographical regions. Note that the simulator makes use of gridded POD values on a 300 km equal-area grid (see Fig. 3) and not the POD regional averages provided here for reference. The Polar region here refers to latitude > 75° N/S. The values apply to daytime (nighttime) conditions. These results are based on the results from the Karlsson and Håkansson (2018) study.

τ_c-centers	τ_c-range	Global	Ocean	Land	Polar
0.025	$0.00 < \tau_c \le 0.05$	0.31 (0.23)	0.34 (0.32)	0.30 (0.14)	0.22 (0.08)
0.075	$0.05 < \tau_c \le 0.10$	0.44 (0.29)	0.49 (0.38)	0.40 (0.22)	0.33 (0.11)
0.125	$0.10 < \tau_c \le 0.15$	0.49 (0.36)	0.56 (0.47)	0.43 (0.30)	0.38 (0.13)
0.175	$0.15 < \tau_c \le 0.20$	0.55 (0.43)	0.62 (0.55)	0.48 (0.38)	0.43 (0.17)
0.225	$0.20 < \tau_c \le 0.25$	0.59 (0.50)	0.67 (0.63)	0.51 (0.46)	0.46 (0.20)
0.275	$0.25 < \tau_c \le 0.30$	0.62 (0.56)	0.70 (0.70)	0.54 (0.52)	0.49 (0.23)
0.325	$0.30 < \tau_c \le 0.35$	0.64 (0.60)	0.73 (0.75)	0.57 (0.57)	0.51 (0.25)
0.375	$0.35 < \tau_c \le 0.40$	0.67 (0.64)	0.75 (0.78)	0.59 (0.61)	0.53 (0.28)
0.425	$0.40 < \tau_c \le 0.45$	0.69 (0.66)	0.78 (0.81)	0.62 (0.64)	0.55 (0.30)
0.475	$0.45 < \tau_c \le 0.50$	0.72 (0.68)	0.80 (0.82)	0.65 (0.66)	0.58 (0.32)
0.550	$0.50 < \tau_c \le 0.60$	0.74 (0.70)	0.83 (0.84)	0.68 (0.68)	0.60 (0.34)
0.650	$0.60 < \tau_c \le 0.70$	0.77 (0.72)	0.85 (0.85)	0.71 (0.70)	0.62 (0.37)
0.750	$0.70 < \tau_c \le 0.80$	0.79 (0.73)	0.87 (0.85)	0.74 (0.72)	0.65 (0.39)
0.850	$0.80 < \tau_c \le 0.90$	0.82 (0.74)	0.89 (0.86)	0.77 (0.74)	0.67 (0.42)
0.950	$0.90 < \tau_c \le 1.00$	0.84 (0.76)	0.90 (0.86)	0.80 (0.76)	0.71 (0.47)
1.500	$1.00 < \tau_c \le 2.00$	0.87 (0.78)	0.92 (0.87)	0.83 (0.79)	0.76 (0.53)
2.500	$2.00 < \tau_c \le 3.00$	0.90 (0.81)	0.94 (0.89)	0.87 (0.82)	0.82 (0.59)
3.500	$3.00 < \tau_c \le 4.00$	0.94 (0.84)	0.97 (0.91)	0.93 (0.86)	0.88 (0.66)
4.500	$4.00 < \tau_c \le 5.00$	0.97 (0.88)	0.98 (0.93)	0.96 (0.90)	0.92 (0.70)

315 lower signal-to-noise ratios) is currently unclear. However, this feature is not critical to the CLARA-A2 simulator but merits a more in-depth investigation in the future.

~~The simulator uses computer-generated random numbers for comparison to the gridded POD value found in a lookup table, where one set of optical depth dependent-PODs is for sunlit, and one is for nighttime conditions. The simulator assigns a random number between 0–1 to each subcolumn at the initiation. After the simulated τ_c is computed, the column integrated τ_c,~~
320 ~~latitude, and longitude are used to find the POD value from the lookup table for comparison. A subcolumn is cloudy, only if its assigned random number is less than the POD. Therefore, if the probability of detection of a cloud with a specific optical depth is 0.05, even though it is very transparent, there is still a 5% chance the subcolumn will be considered cloudy. Conversely, as mentioned earlier, regardless of how optically thick a cloud is in a subcolumn, there is a non-zero chance this subcolumn will not be flagged as cloudy, and hence not included in any further cloud simulations.~~ moved up

POD day-night
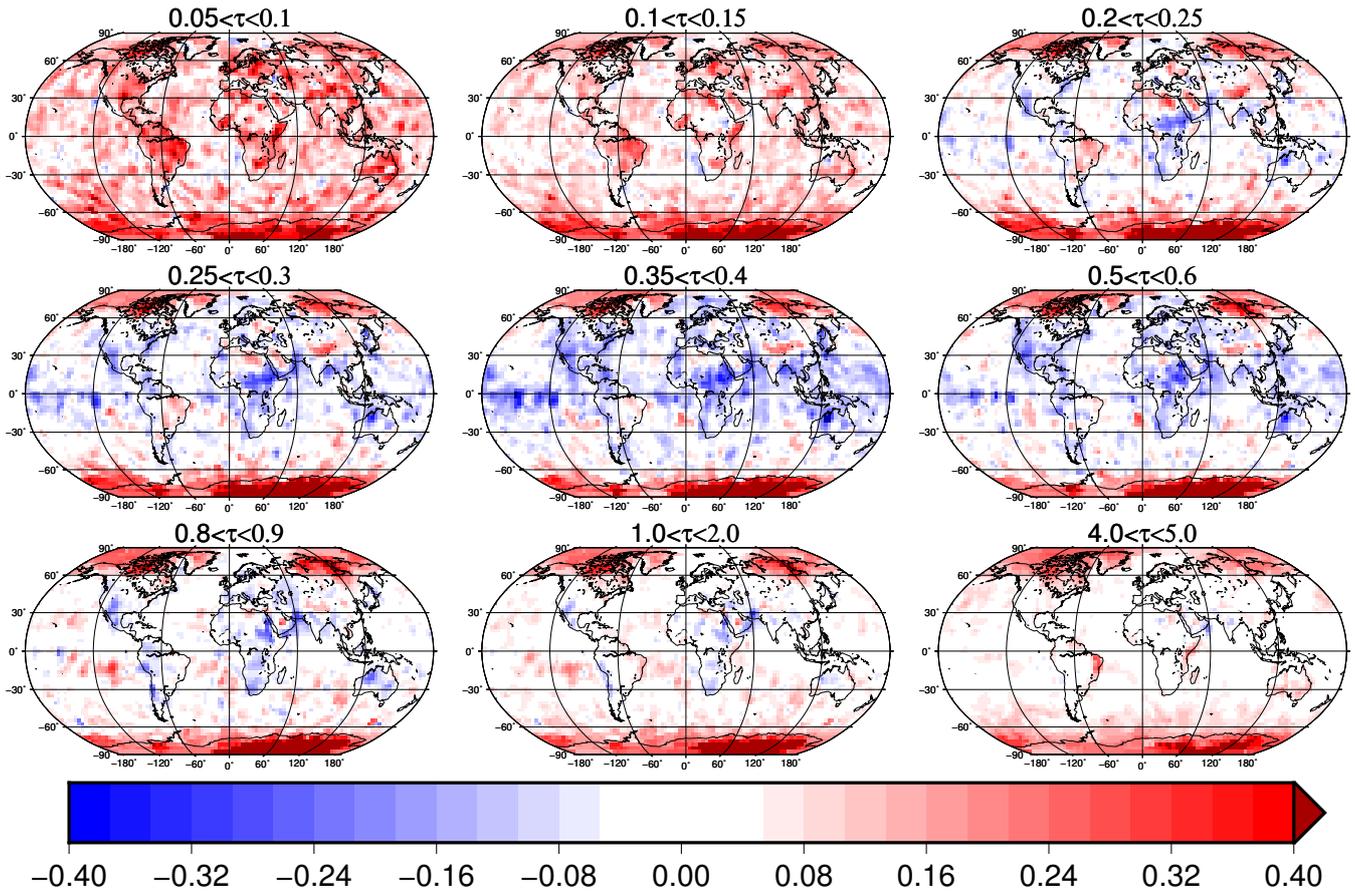
Figure 4. The difference in the POD of the cloud mask during sunlit and nighttime conditions for selected cloud optical depth intervals. These results are based on the results from the Karlsson and Håkansson (2018) study.

3.2   The choice of simulated cloud mask

In this section we refer to figures 5 and 6 to illustrate how the choice of cloud mask simulation method affects the comparison of cloud cover of EC-Earth to CLARA-A2. The results are separated into seasons here since it is essential to understand the seasonal impact of choosing one method over another. The ~~top figure in~~ Fig. 5 a and Fig. 6 a show the cloud cover according to CLARA-A2 for 1982–2015 during Southern Hemisphere summers and the Northern Hemisphere summers respectively. EC-Earth minus CLARA-A2 based on the first method (Sect. 3.1.1) is subplot ~~(a)~~(b), based on the second method (Sect. 3.1.2) is subplot ~~(b)~~(c), and based on the third method (Sect. 3.1.3) is subplot ~~(c)~~(c). Subplot ~~(d)~~(e) shows the difference between the simulated cloud mask based on method one, a global static τ_c threshold, and method three, based on POD thresholds (first method minus the third method).

15
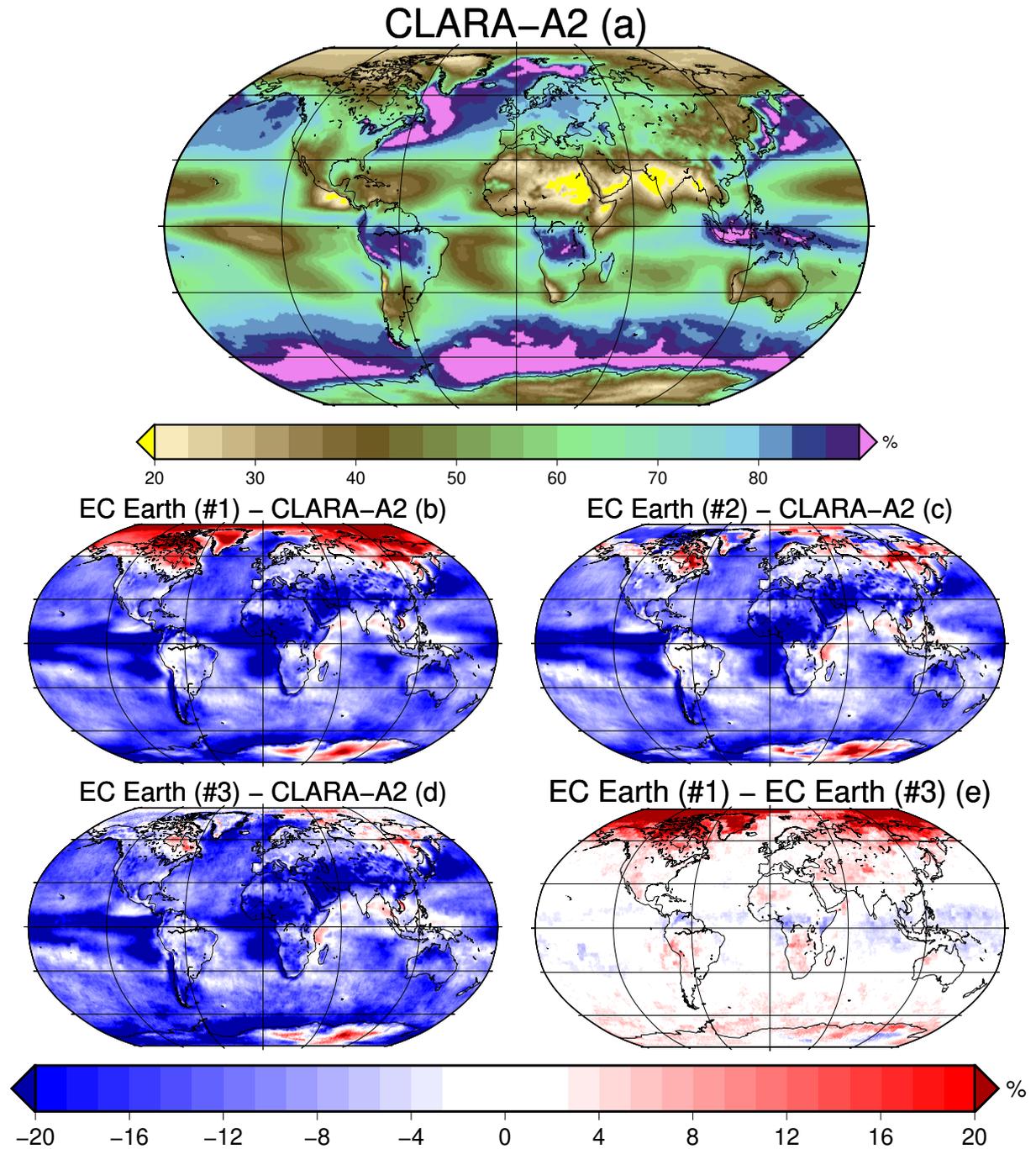
Figure 5. Total cloud cover during the DJF- seasons of 1982–2015. This figure shows a comparison of EC-Earth to CLARA-A2 using three different methods of cloud mask simulation. The reference figure at the top (a) is the cloud fraction from CLARA-A2. Subfigure (a)(b) shows the simulated observations using method one, based on a global static τ_c-limit, minus CLARA-A2. Subfigure (b)(c) shows the same comparison using method two, based on gridded τ_c-limits, and (c)(d) shows the same using method three, based on POD. Sub-figure (d)(e)) shows the difference between the simulated CF based on method one minus method three. See Sect. 3.2 for a wider description of the figure.
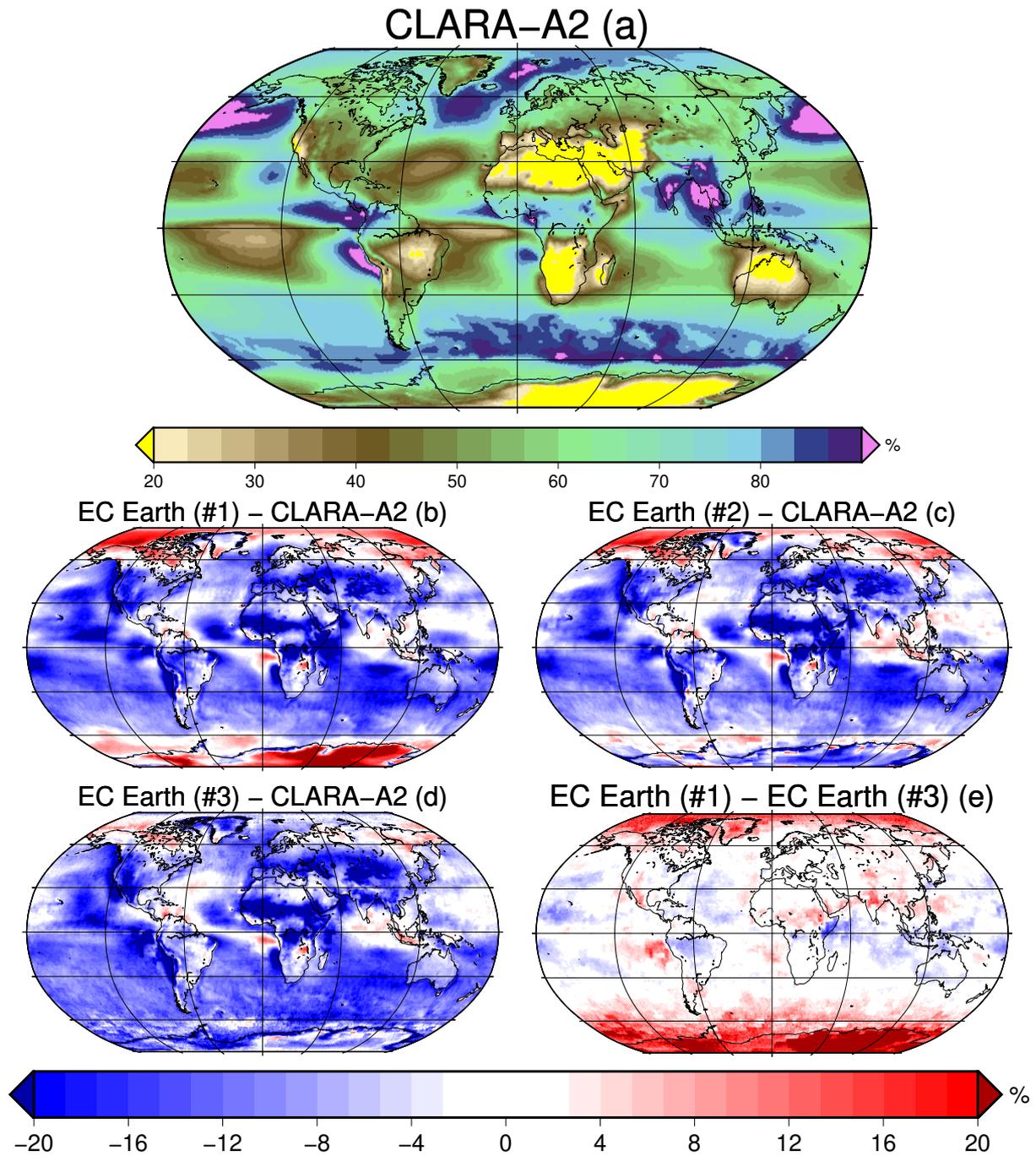
Figure 6. Total cloud cover during the JJA -seasons of 1982–2015. See the description in Fig. 5 for a description of the layout in this figure

Globally, the overall impression is that EC-Earth underestimates cloud fraction. In most regions of the world, within a few
percent, this is the conclusion one would reach regardless of which of the three methods was used to simulate the CLARA-A2
cloud mask. However, as described in Sect. 2.1, the CLARA-A2 CDR is systematically and substantially less skillful under
certain conditions than on average.

As discussed in Sect. 2.1.1, CLARA-A2 is skillful at detecting clouds in the polar regions during sunlit conditions, but not
so during the polar winter. This is why the apparent overestimation of clouds in these regions by EC-Earth ~~(Fig. 5,6a)~~(Fig. 5 b
and Fig. 6 b) is likely strongly exaggerated. Without prior knowledge of the retrieval difficulties in cold dark locations, i.e.,
when only passive infrared channels are available, if method one is used to simulate clouds, one could erroneously conclude
that EC-Earth places too many clouds in polar regions. This problem is especially salient during winter months, but it also has
a considerable impact on cumulative averages over these regions. Therefore cloud mask simulations based on method one are
notably unsuitable in the Polar regions and, to lesser extent, desert areas.

However, and what is the main point of this innovation, if one uses the second or third method to simulate clouds, the
apparent bias in cloudiness in these regions is mostly removed in the problematic regions. The second and third methods do a
much better job at reproducing the limitations of cloud datasets than the first method, and the size of the difference between
method three and one is substantial and seasonally dependent in the problematic regions ~~(Fig. 5,6d)~~(Fig. 5 e and Fig. 6 e).

Notice also from ~~(Fig. 5,6 b)~~(Fig. 5 c and Fig. 6 c) and ~~(Fig. 5,6 c)~~(Fig. 5 d and Fig. 6 d) that the second and third methods
produce similar results, and hence both do well in this regard. However, there are some subtle differences. One is that during
the northern hemisphere summer months a model validation based on the second method leads to the conclusion that EC-Earth
overestimates clouds in the Arctic, yet if the comparison were made based on the third method, one would conclude only a
slight overestimation here.

The third method gives the most accurate description of the cloud detection limitations since it describes the likelihood of
detecting/missing clouds over the full range of cloud optical thicknesses for day and night conditions. Also, method three can
emulate the non-zero probability that even thick clouds might be undetectable under certain conditions. This approach better
describes the skill of the cloud retrievals of a satellite dataset than using gridded static values of $\tau_{\min}$ in method two, and
especially instead of using a single global $\tau_{\min}$ value used by method one. Overall, therefore, the recommendation is to choose
method three to simulate the cloud mask.

However, the advantage of tying statistics to geographical regions, may also be a weakness in some situations. If a models'
cloud distribution is systematically misplaced, the model clouds may be subject to (potentially) other PODs than what they
should have been in the CLARA-A2 simulator. The consequences here should not be large, except for the extreme cases when a
model places clouds over ice- and snow-covered areas in the polar night (with very low PODs) instead of over adjacent ice-free
ocean areas (with very high PODs). Additionally, the underlying statistics used in method two and three are, as mentioned in
Sect. 2.1.1, derived from collocations that cover the time period between 2006–2015. Therefore in some regions, such as in the
marginal ice regions, the conditions for cloud detection may have changed appreciably from those during the validation period,
for instance, due to a changing climate, rendering the statics less representative than in more climatically stable regions.

## 4 Application of the simulator to Arctic case studies

4.1 Average cloudiness during summer months

370 Karlsson et al. (2017) asserted, and the POD maps shown in Fig. 4 suggest, that the CLARA-A2 CDR is reasonably ~~particularly~~ skillful at detecting clouds in the Arctic during sunlit conditions. Therefore, to demonstrate the utility of the CLARA-A2 simulator, we assessed the cloud cover in these conditions over the full length of the datasets. We added the ISCCP-H CDR (Young et al., 2018) to the comparison since it is an equivalent CDR with a well-established satellite simulator used in many previous model studies (e.g., Webb et al., 2001; Norris et al., 2016; Terai et al., 2016; Tan et al., 2017). However, Karlsson and

375 Devasthale (2018) found the cloud cover of ISCCP-H too low in the polar summer and early autumn.

The cloudiness from ISCCP-H should be compared to the simulated cloudiness using the ISCCP simulator (Jakob and Klein, 1999), and the cloudiness, according to CLARA-A2, is compared to the CLARA-A2 simulator. Figure 7 shows the average cloudiness in Arctic summer months according to CLARA-A2 (Fig. 7a) and ISCCP-H (Fig. 7b). EC-Earth's representation of overall cloudiness during Arctic summer is tested using the simulated CLARA-A2 and simulated ISCCP-H, shown in Fig. 7c

380 and Fig. 7d respectively. ~~The cloudiness from ISCCP-H should be compared to the simulated cloudiness using the ISCCP simulator, and the cloudiness, according to CLARA-A2, is compared to the CLARA-A2 simulator.~~ As mentioned in Sect. 3.1.1, the simulated cloud mask for ISCCP-H uses a global τ_c threshold ($\tau_c = 0.3$) for the simulated cloud mask (method one, different threshold), and the CLARA-A2 simulator uses the POD-based approach for the simulated cloud mask (method three). The two satellite datasets and the climate model are limited to July 1983 to June 2015 to match the availability of the ISCCP-H period

385 to date.

Fig. 7 demonstrates that using simulators that do not take the variable skill of the cloud mask into account, such as the ISCCP simulator, could easily lead to false conclusions about EC-Earth cloud cover in the Arctic summer. Compared to the ISCCP-H observations, the simulated ISCCP-H observations indicate that EC-Earth has a strong positive cloud bias in the Arctic of ~~about 40-50%~~more than 30 %. However, CLARA-A2, shown to have a high skill in the polar summer (see Fig. 5b in Karlsson et al.

390 (2017)), indicates that EC-Earth under-predicts the cloudiness in large parts of this region by ~~20-30%~~more than 10 %. ~~Similarly, Karlsson et.al, (2019) found the cloud cover of ISCCP-H too low in the polar summer and early autumn. The CLARA-A2 simulator shows that rather than EC-Earth massively overestimating cloudiness, especially over the central Arctic regions, EC-Earth has a similar amount of clouds in the Arctic, and rather tends to under-represent clouds in the sunlit Arctic conditions.~~ These large differences between the simulated ISCCP-H and CLARA-A2 are mainly due to the ISCCP simulator being too

395 sensitive to thin clouds here. As shown in Fig. 2, during daytime conditions in the Arctic, a more appropriate daytime τ_c-limit would be around 0.5 or more, which is higher than the global average of 0.3 assumed by the ISCCP simulator. Therefore in the Arctic summer, the ISCCP simulator retrieves clouds in between these cloud optical thicknesses that the CLARA-A2 simulator, and most likely the observations, do not. As a consequence, anyone assessing cloudiness in the Arctic will reach the opposite conclusion using the CLARA-A2 CDR and simulator compared to the ISCCP-H counterpart.

400 Overall, based on CLARA-A2 as the reference, EC-Earth has a smaller average cloud fraction over most of the region between 50N–90N during the summer months. The difference is more substantial over ocean areas than over land, with the
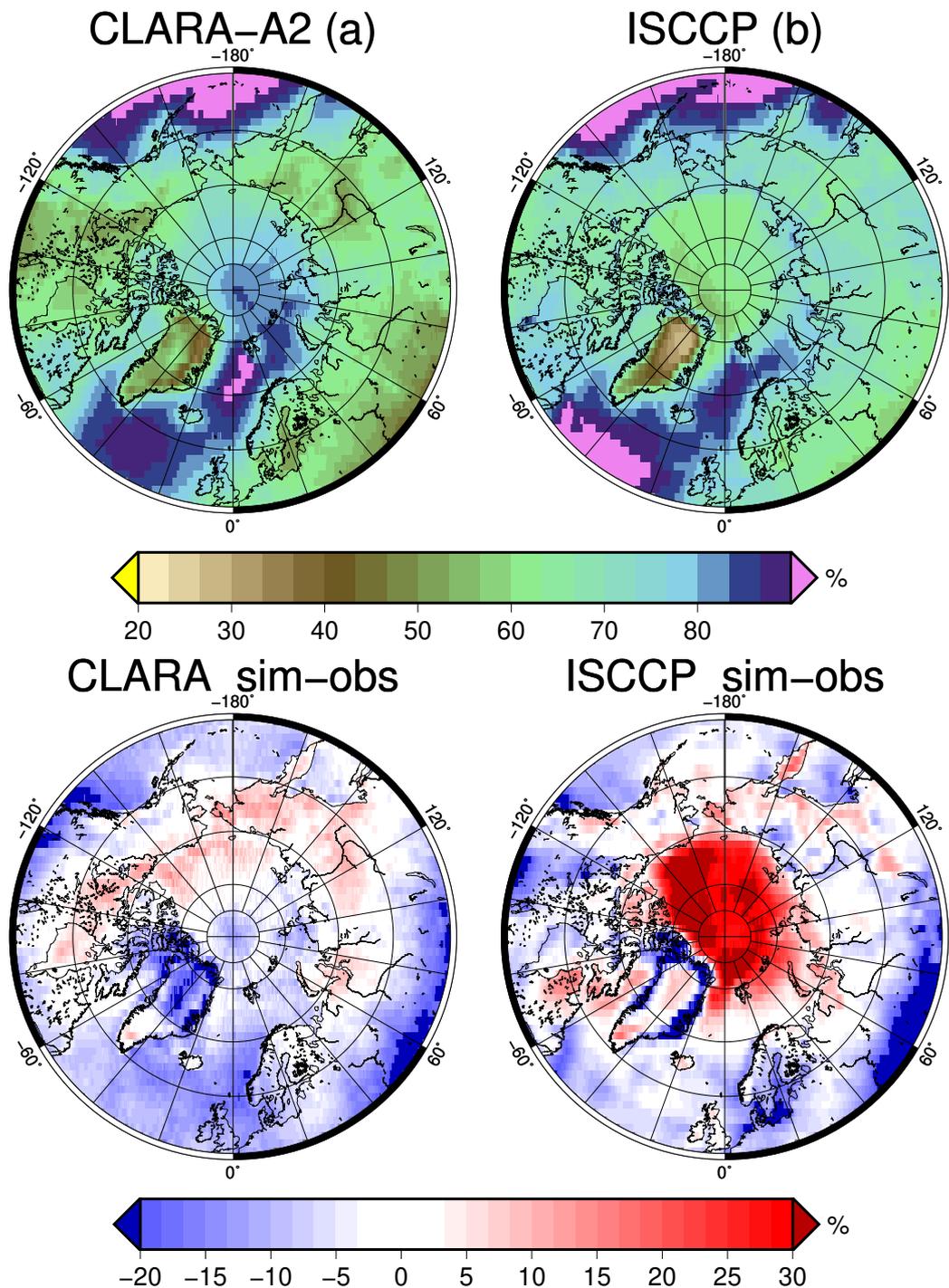
Figure 7. The Total Cloud Fraction (TCF) in the Arctic summer. The top row contains the observations from two equivalent CDRs, CLARA-A2 (a) and ISCCP-H (b). The bottom row contains the ~~difference between the~~ simulated ~~CDRs based on the model atmosphere of EC-Earth, simulated CLARA-A2 (c) and simulated ISCCP (d)~~ CDR minus the CDR for CLARA-A2 (c) and ISCCP-H (d). The period is July 1983 to June 2015.

largest under-representation of cloudiness at these latitudes is over the North Atlantic and following the Gulf Stream north of Norway. However, globally, the most considerable negative cloud biases between the model and observations are in the Tropics and subtropics (see Fig. 6).
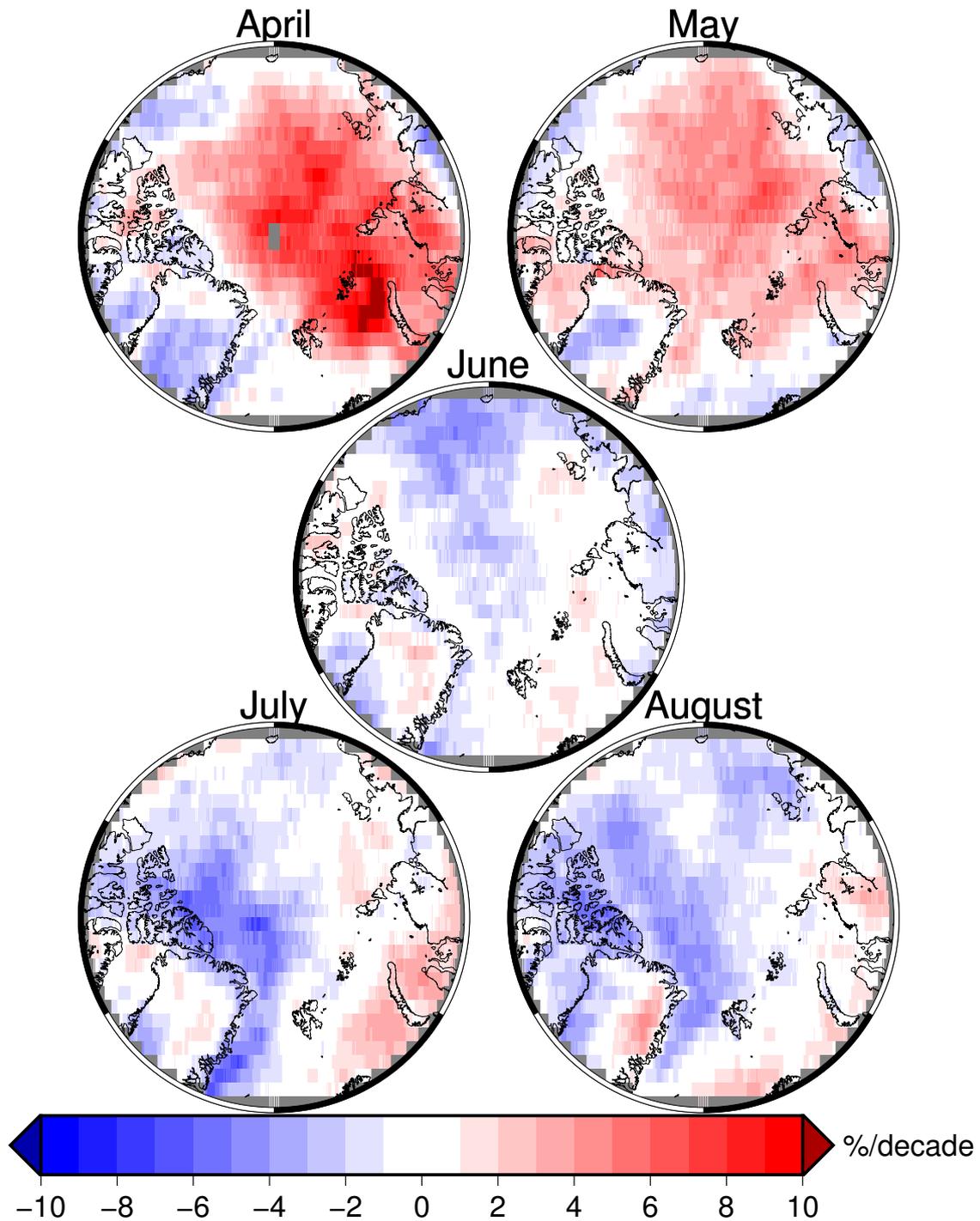
405 ## 4.2 Trends in cloudiness

~~This section assesses trends in cloudiness during the Arctic sunlit months, according to CLARA-A2, and EC-Earth.~~ The CLARA-A2 CDR is particularly suitable for cloud trend analysis in the Arctic summer due to its length and~~since it is long enough to make statements about cloud trends and is reported to have~~ high cloud detection skills there (Karlsson and Dev-asthale, 2018). Here is an assessment of the cloud trends from the months that have enough sunlight, i.e., where the solar zenith
410 angle is less than ~~80°~~84° in the Arctic above 70°N for CLARA-A2 and EC-Earth. These trends are ~~decadal and~~ based on the linear regression of cloudiness from all data in 1982–2015 and expressed here as absolute change in cloudiness [$\frac{\%}{\text{decade}}$].

Fig. 8 shows the distribution of cloudiness trends, according to CLARA-A2. From this figure, some clear patterns emerge; in the spring months, there is an increase in cloudiness by more than 5 % in large parts of the Arctic and upwards of 10 % north of Novaja Zemlya, and in the summer to Autumn months the Arctic is dominated by a decrease in cloudiness. The increase in
415 cloudiness reaffirms observations previously reported in Kapsch et al. (2013) and Kapsch et al. (2019). Kapsch et al. (2013) asserted that the increase in cloudiness is likely due to an increased intrusion of water vapor into these regions during the spring months. The largest decrease in cloudiness seen in July and August is in the Beaufort, and especially the Lincoln Seas, north of the Canadian archipelago and Greenland. However, it is outside the scope of this study, whose main purpose is to describe the CLARA-A2 simulator, to further assess the possible reasons for the changing cloudiness seen in these observations.

420 Fig. 9 shows the average change in cloudiness of EC Earth over the same time period as in Fig. 8, using method three to simulate the cloud mask. The cloud trends in the model clearly differ from the observations. In particular, the size of the trends are in general much smaller than the observations indicate, but also the pattern of cloud trend is not in agreement (except in May). However, there are some important limiting factors to consider for this model evaluation.~~average cloudiness trends for the same conditions, aside from excluding land areas, as in Fig 8., for CLARA-A2, the three methods of simulated CLARA-A2~~
425 ~~cloud mask from the EC-Earth atmosphere, and the total cloudiness directly from EC-Earth without any simulator. The reason for the simplified analysis in Fig9 is to avoid over-emphasizing differences in the model cloudiness trends.~~

EC-Earth is represented here by only one model run, and although it employs prescribed sea surface temperatures and sea ice extent, the model atmosphere is free to meander. ~~The implication is that a perfectly valid atmospheric state based on one model run is hard to fairly compare to observations.~~ In order to assess if the model cloud trends agree with the observations,
430 ideally, several ensemble model runs are required to find a general trend ~~and~~ to assess whether or not the natural variability produced by the model is accurate (Koenigk et al., 2019).

Fig. 10 illustrates how the choice of cloud mask simulation ~~effects~~affects model cloud trend. Fig. 10 shows the average cloudiness trends for the same conditions, aside from excluding land areas, as in Fig. 8, for CLARA-A2, the three methods of simulated CLARA-A2 cloud mask from the EC-Earth atmosphere, and the total cloudiness directly from EC-Earth without

Figure 8. The average trend in cloudiness over the entire record [$\frac{\%}{\text{decade}}$] decadal cloudiness trend in the Arctic from the illuminated months of April to August according to CLARA-A2. Negative trends correspond to an average decrease in cloudiness over time. The trends are from all months in the period 1982–2015.
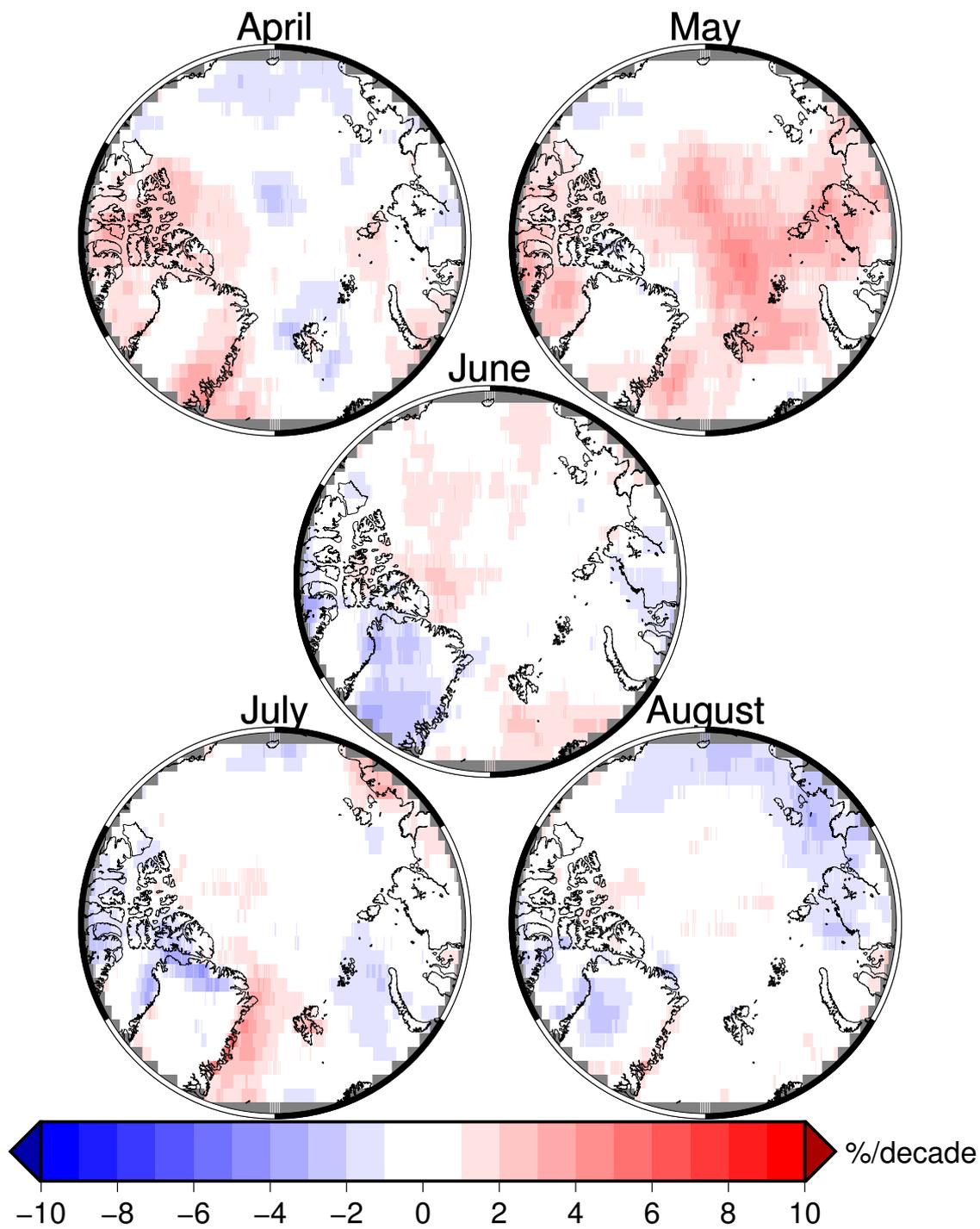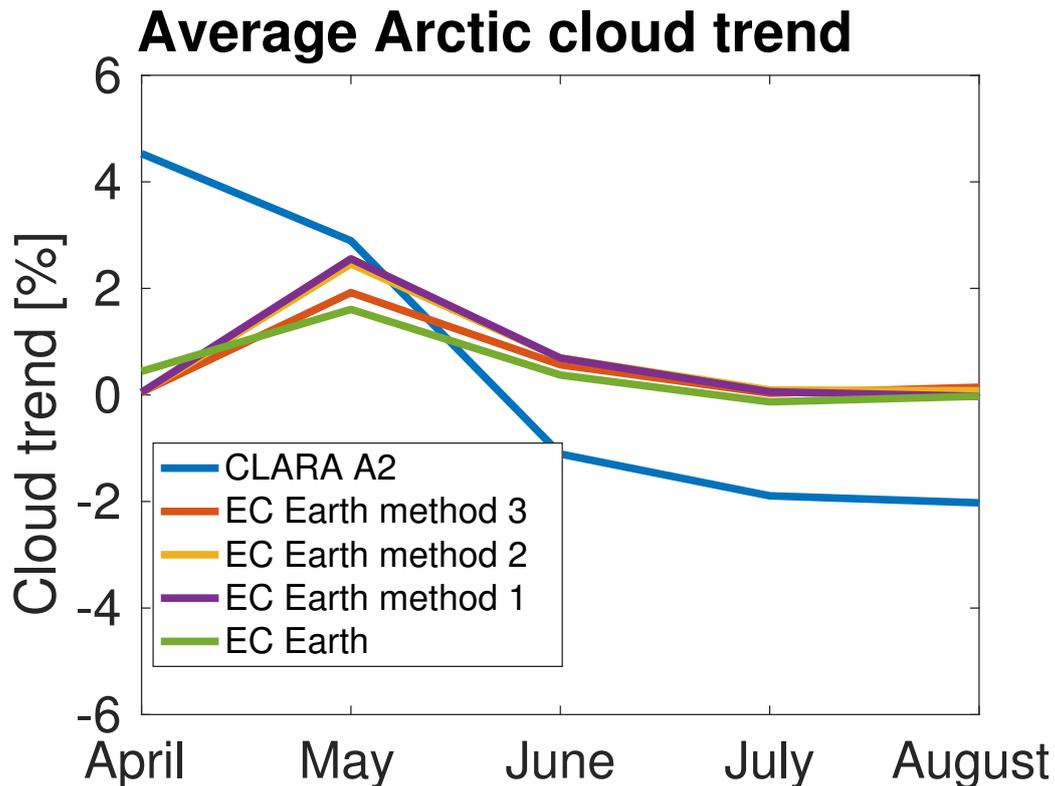
Figure 9. As for Fig. 8, but for the EC Earth climate model.

Figure 10. The average decadal cloudiness trend in the Arctic from the illuminated months of April to August only over the ocean (ice-free or ice-covered). The figure shows the reference dataset, CLARA-A2, the CLARA-A2 simulator, one line for each method, and the Total Cloud Cover (TCC) from the EC-Earth model without using any simulator. The trends are from all months in the period 1982–2015.

435   any simulator. ~~The reason for the simplified analysis in Fig9 is to avoid over-emphasizing differences in the model cloudiness trends~~

~~However,~~ Fig. 10 illustrates that regardless of which method is used to simulate cloudiness, or even using no simulator at all, ~~does not alter the average cloud trends in the Arctic~~ the simulators do not appear to alter the cloud trends in the Arctic summer. These results may indicate that the clouds in the model are not changing the average range and distribution of optical

440   thicknesses over time, even if the actual cloud amounts may change.

In summary ~~However~~, no definitive conclusions on model cloud trends in the Arctic can be drawn here for the reasons listed above, and a more thorough examination of whether or not EC-Earth reproduces realistic cloud trends is also outside the scope of this study. Although the choice of method does not appear at first glance to impact the model cloudiness trends, it still makes sense, in this case, to use method three to simulate clouds, since it more closely reflects the skill of the CLARA-A2 dataset.

5 Conclusions

This article describes a satellite simulator designed to enable comparisons between climate models and the CLARA-A2 CDR. Typically, satellite simulators simulate the satellite retrieved cloud fraction using one global cloud optical depth threshold~~, called method one in this paper,~~ to remove thin model clouds that are presumed undetectable by the instruments used to generate the CDR. There are more factors to consider that influence the ability to retrieve thin clouds. These include

- The optical thickness of the cloud

- How illuminated the clouds are

- The underlying surface properties and

- The temperature difference between the cloud and the surface

In this paper, we show that using one optical depth ~~value~~threshold for all conditions to emulate cloud sensitivity (method one) is inappropriate since the cloud detection skill of satellite retrievals may vary considerably. This is the method used in ~~some of~~ the COSP simulators, which many previous studies have relied on. Therefore, to avoid the largest uncertainties, many previous studies are limited to between $\pm 60\,°$ latitude. ~~, may have negatively impacted some model cloudiness analyses.~~ There is a need for a more realistic simulated cloud mask that better reflects the actual cloud detection ability of the CDR. We therefore propose two other methods that are both based on validations of the CLARA-A2 CDR using collocated cloud retrievals from CALIOP by Karlsson and Håkansson (2018).

Method two uses two maps of cloud detection thresholds on a $300\,\mathrm{km}$ grid, one for day and one for night conditions. These thresholds refer to the smallest cloud optical depth where there is a $50\,\%$ success rate in detecting clouds. The main improvement by this method is that in areas where the cloud retrievals are relatively straightforward, such as over mid-latitude oceans, the cloud sensitivity is generally increased, i.e., a lower cloud optical threshold. Conversely, in areas and conditions where cloud retrievals are notoriously difficult, a much higher optical depth threshold is suitable.

Method three, the recommended approach to simulating the cloud mask, is based on the POD of clouds depending on their τ_c. Instead of using a τ_c threshold to determine whether or not a model cloud would have been detected, with this approach, any model cloud could potentially be detected or missed. Maps of POD valid for separate optical depth ranges (see Tab. 2) are used together with a random number generated at run time for every model subcolumn~~subgrid column~~ to determine cloudiness. These are also provided on a $300\,\mathrm{km}$ grid and separated by day and night. The main improvement of this method is that it indirectly takes into account that retrievals in some regions are more likely than others to miss thick clouds. This situation is common in cold regions where thick clouds may be inseparable from cold snow-covered surfaces and also in regions with an abundance of broken and small scale cumulus clouds such as the atmospheric subsidence regions over the ocean.

Compared to method one, methods two and three allow for analyses to be carried out at high latitudes and during nighttime conditions. Although the largest improvements are at high latitudes, these new methods also account for the modestly improved cloud detection of CLARA-A2 over the global oceans compared to, especially, desert areas. Therefore, with these methods, model studies may also be improved for regions outside the polar regions.

This paper illustrates that these new approaches to cloud mask simulation bring the model and observations much closer to each other compared to using a fixed optical depth threshold globally to filter out clouds. They allow for a more realistic model to satellite comparison, and thus reduces the likelihood that incorrect conclusions from model assessments are reached simply due to cloud simulations not correctly representing the cloud retrievals of the CDR. Although methods two and three both significantly improve cloud mask simulations, method three, using the POD approach, is better since it realistically mimics the performance of the cloud mask of the CLARA-A2 CDR over the full range of cloud optical thicknesses.

The overall cloudiness in the Arctic during summer months from 1984–2014 is used to demonstrate the usefulness of the simulator and the new approach to cloud mask simulation. The ISCCP-H CDR here complemented the comparison as a second independent satellite dataset. Therefore, EC-Earth was assessed using both the ISCCP and CLARA-A2 simulators and compared to the CDRs they should simulate. This comparison shows that EC-Earth seems to produce too few clouds in and around the Arctic compared to CLARA-A2. However, despite the ISCCP-H CDR having more clouds than CLARA-A2 in the Arctic summer months, compared to ISCCP-H and using the ISCCP simulator, the assessment on EC-Earth cloudiness would lead to quite the opposite conclusion in some regions in the Arctic. The simulated ISCCP cloudiness is substantially higher than the ISCCP observations. This overrepresentation of clouds is mostly due to the ISCCP simulator using a global optical depth threshold that, in the Arctic is too generous. This example demonstrates the advantage of using the CLARA-A2 approach to cloud mask simulation compared to the traditional approach used by the ISCCP simulator and others. Although only demonstrated in the Arctic summer in this paper, the POD approach, method 3, is also the most appropriate globally.

In terms of trends in overall cloudiness in the Arctic for all months with sunlit conditions from 1982–2015, the observations from CLARA-A2 show a sharp increase in cloudiness over the years, especially in the ocean areas north of western Russia, in the spring months of April and May. In the summer and early autumn months, there is a large area of decreasing cloudiness in the seas just north of Canada and Greenland. Although only based on one model run, and therefore clear statements about cloud trends in the model cannot be made, one can deduce that the average cloudiness trends from the model are very similar using any simulator method, or no simulator at all.

In summary, the authors advocate an approach to cloud mask simulation based on the probability of detection of clouds depending on their optical depth, location, and illumination. This study suggests that evaluations of climate model simulations of cloudiness parameters would benefit substantially from using more advanced satellite simulators, which in a better way than today, accounts for weaknesses and strengths of satellite retrievals.

Competing interests. No competing interests

Data availability. The CLARA-A2 CDR (Karlsson et al., 2017) can be downloaded from https://wui.cmsaf.eu (last access: 30 July 2019). Data from the EC-Earth global climate model (Hazeleger et al., 2010) can be obtained from http://www.nextdataproject.it/?q=en/content/ec-earth-cmip5-data-extraction (last access: 30 July 2019). The ISCCP-H (Young et al., 2018) products and other ISCCP products are available from https://www.ncdc.noaa.gov/cdr/atmospheric/cloud-properties-isccp (last access: 30 July 2019)

515 *Code and data availability.* The simulator code itself is so far only available by contacting the authors.

References

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., Zhang, Y., Marchand, R., Haynes, J. M., Pincus, R., and John, V. O.: COSP: satellite simulation software for model assessment, Bull. Amer. Met. Soc., 92, 1023–1043, https://doi.org/10.1175/2011BAMS2856.1, 2011.

520   Bugliaro, L., Zinner, T., Keil, C., Mayer, B., Hollmann, R., Reuter, M., and Thomas, W.: Validation of cloud property retrievals with simulated satellite radiances: a case study for SEVIRI, Atmos. Chem. Phys., 11, 5603–5624, https://doi.org/10.5194/acp-11-5603-2011, 2011.

CMSAF1: Validation Report - CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2) – Cloud Products, SAF/CM/DWD/VAL/GAC/CLD version 2.3, Tech. rep., available at http://dx.doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002, 2017.

525   CMSAF2: Algorithm Theoretical Basis Document - CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2) – Cloud Fraction, SAF/CM/DWD/ATBD/CMA_AVHRR version 2.0, Tech. rep., available at http://dx.doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002, 2017.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative modelling — Part I: Algorithm description, J. Appl. Meteorol., 44, 39–54, 2005.

530   Eliasson, S., Buehler, S. A., Milz, M., Eriksson, P., and John, V. O.: Assessing observed and modelled spatial distributions of ice water path using satellite data, Atmos. Chem. Phys., 11, 375–391, https://doi.org/10.5194/acp-11-375-2011, 2011.

Eliasson, S., Karlsson, K. G., van Meijgaard, E., Meirink, J. F., Stengel, M., and Willén, U.: The Cloud_cci simulator v1.0 for the Cloud_cci climate data record and its application to a global and a regional climate model, Geoscientific Model Development, 12, 829–847, https://doi.org/10.5194/gmd-12-829-2019, https://www.geosci-model-dev.net/12/829/2019/, 2019.

535   Håkansson, N., Adok, C., Thoss, A., Scheirer, R., and Hörnquist, S.: Neural network cloud top pressure and height for MODIS, Atmospheric Measurement Techniques, 11, 3177–3196, https://doi.org/10.5194/amt-11-3177-2018, https://www.atmos-meas-tech.net/11/3177/2018/, 2018.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A. M. L., Christensen, J. H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., van Noije, T., Palmer, T., Parodi, J. A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P., and Willén, U.: EC-Earth, Bulletin of the American Meteorological Society, 91, 1357–1364, https://doi.org/10.1175/2010BAMS2877.1, 2010.

Hazeleger, W., Wang, X., Severijns, C., Stefanescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., van den Hurk, B., van der Linden, T. v. E.-C., and van der Wiel, K.: EC-Earth V2: description and validation of a new seamless Earth system prediction model, 545   Climate Dynamics, 39, 2611–2629, https://doi.org/10.1007/s00382-011-1228-5, 2012.

Heidinger, A. K., Straka, W. C., Molling, C. C., Sullivan, J. T., and Wu, X. Q.: Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record, Int. J. Remote Sensing, 31, 6493–6517, https://doi.org/10.1080/01431161.2010.496472, 2010.

Heidinger, A. K., Foster, M. J., Walther, A., and Zhao, X.: The Pathfinder Atmospheres–Extended AVHRR Climate Dataset, Bull. Amer. Met. Soc., 95, 909–922, https://doi.org/10.1175/BAMS-D-12-00246.1, https://doi.org/10.1175/BAMS-D-12-00246.1, 2014.

550   IPCC: Clouds and Aerosols, p. 571–658, Cambridge University Press, https://doi.org/10.1017/CBO9781107415324.016, 2014.

Jakob, C. and Klein, S. A.: 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, https://doi.org/10.1002/qj.49712555510, 1999.

Kapsch, M.-L., Graversen, R. G., and Tjernström, M.: Springtime atmospheric energy transport and the control of Arctic summer sea-ice extent, Nature Clim. Change, 3, 744–748, https://doi.org/10.1038/nclimate1884, 2013.

555 Kapsch, M.-L., Skific, N., Graversen, R. G., Tjernström, M., and Francis, J. A.: Summers with low Arctic sea ice linked to persistence of spring atmospheric circulation patterns, Climate Dynamics, 52, 2497–2512, https://doi.org/10.1007/s00382-018-4279-z, 2019.

Karlsson, K.-G. and Devasthale, A.: Inter-Comparison and Evaluation of the Four Longest Satellite-Derived Cloud Climate Data Records: CLARA-A2, ESA Cloud CCI V3, ISCCP-HGM, and PATMOS-x, Remote Sensing, 10, https://doi.org/10.3390/rs10101567, http://www.mdpi.com/2072-4292/10/10/1567, 2018.

560 Karlsson, K.-G. and Håkansson, N.: Characterization of AVHRR global cloud detection sensitivity based on CALIPSO-CALIOP cloud optical thickness information: demonstration of results based on the CM SAF CLARA-A2 climate data record, Atmos. Meas. Tech., 11, 633–649, https://doi.org/10.5194/amt-11-633-2018, https://www.atmos-meas-tech.net/11/633/2018/, 2018.

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J.-F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., and Wolters, E.: CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmos. Chem. Phys., 13, 5351–5367,
565 https://doi.org/10.5194/acp-13-5351-2013, 2013.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Meirink, J.-F., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G. J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmos. Chem. Phys., 17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017, https://www.atmos-chem-phys.net/17/5809/2017/, 2017.

570 Koenigk, T., Gao, Y., Gastineau, G., Keenlyside, N., Nakamura, T., Ogawa, F., Orsolini, Y., Semenov, V., Suo, L., Tian, T., Wang, T., Wettstein, J. J., and Yang, S.: Impact of Arctic sea ice variations on winter temperature anomalies in northern hemispheric land areas, Climate Dynamics, 52, 3111–3137, https://doi.org/10.1007/s00382-018-4305-1, https://doi.org/10.1007/s00382-018-4305-1, 2019.

Norris, J. R., Allen, R. J., Evan, A. T., Zelinka, M. D., O'Dell, C. W., and Klein, S. A.: Evidence for climate change in the satellite cloud record, Nature, 536, 72–75, https://doi.org/10.1038/nature18273, 2016.

575 NWCSAF: Scientific and Validation Report for the Cloud Product Processors of the NWC/PPS, Tech. rep., http://www.nwcsaf.org/web/guest/scientificdocumentation#NWC/PPS%20v2014, 2018.

Pincus, R., Hemler, R., and Klein, S.: Using Stochastically Generated Subcolumns to Represent Cloud Structure in a Large-Scale Model, Mon. Weather Rev., 134, 3644–3656, 2006.

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P.: Reconciling Simulated and Observed Views of Clouds:
580 MODIS, ISCCP, and the Limits of Instrument Simulators, J. Climate, 25, 4699–4720, https://doi.org/10.1175/JCLI-D-11-00267.1, 2012.

Rao, C. R. N., Sullivan, J. T., Walton, C. C., Brown, J. W., and Evans, R. H.: Non-linearity corrections for the thermal infrared channels of the Advanced Very High Resolution Radiometer: Assessment and corrections, Tech. rep., NOAA NESDIS 69, 1993.

Rossow, W. B. and Schiffer, R. A.: Advances in Understanding Clouds From ISCCP, Bull. Amer. Met. Soc., 80, 2261–2288, 1999.

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

Swales, D. J., Pincus, R., and Bodas-Salcedo, A.: The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2, Geosci. Model Dev., 11, 77–81, https://doi.org/10.5194/gmd-11-77-2018, https://www.geosci-model-dev.net/11/77/2018/, 2018.

590 Tan, J., Oreopoulos, L., Jakob, C., and Jin, D.: Evaluating rainfall errors in global climate models through cloud regimes, Climate Dynamics, https://doi.org/10.1007/s00382-017-3806-7, https://doi.org/10.1007/s00382-017-3806-7, 2017.

Terai, C. R., Klein, S. A., and Zelinka, M. D.: Constraining the low-cloud optical depth feedback at middle and high latitudes using satellite observations, JGR, 121, 9696–9716, https://doi.org/10.1002/2016JD025233, http://dx.doi.org/10.1002/2016JD025233, 2016.

Tzallas, V., Hatzianastassiou, N., Benas, N., Meirink, J. F., Matsoukas, C., Stackhouse, P., and Vardavas, I.: Evaluation of CLARA-
595 A2 and ISCCP-H Cloud Cover Climate Data Records over Europe with ECA&D Ground-Based Measurements, Remote Sensing, 11, https://doi.org/10.3390/rs11020212, http://www.mdpi.com/2072-4292/11/2/212, 2019.

Waliser, D. E., Li, J.-L. F., Woods, C. P., Austin, R. T., Bacmeister, J., Chern, J., Genio, A. D., Jiang, J. H., Kuang, Z., Meng, H., Minnis, P., Platnick, S., Rossow, W. B., Stephens, G. L., Sun-Mack, S., Tao, W.-K., Tompkins, A. M., Vane, D. G., Walker, C., and Wu, D.: Cloud ice: A climate model challenge with signs and expectations of progress, J. Geophys. Res., 114, D00A21,
600 https://doi.org/10.1029/2008JD010015, 2009.

Webb, M., Senior, C., Bony, S., and Morcrette, J.-J.: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, Climate Dynamics, 17, 902–922, https://doi.org/10.1007/s003820100157, 2001.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, J. Atmos. Oceanic Technol., 26, 2310–2323, https://doi.org/10.1175/2009JTECHA1281.1,
605 2009.

Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., and Rossow, W. B.: The International Satellite Cloud Climatology Project H-Series climate data record product, Earth Syst. Sci. Data, 10, 583–593, https://doi.org/10.5194/essd-10-583-2018, https://www.earth-syst-sci-data.net/10/583/2018/, 2018.

Appendix: Glossary

610 τ_c visible cloud optical depth at 550nm. 1–3, 5–14, 16, 19, 25

r_e cloud particle effective radius. 7–9

AMIP Atmospheric Model Inter-comparison Project. 7

AVHRR Advanced Very High Resolution Radiometer. 1, 3, 4, 11, 12

CALIOP Cloud-Aerosol Lidar with Orthogonal Polarisation. 3, 5, 6, 10–12, 14, 25

615 **CDR** Climate Data Record. 1–9, 15, 18–20, 24–26

CFMIP Cloud Feedback Model Intercomparison Project. 1

CLARA CM SAF cLoud, Albedo and RAdiation dataset. 1, 4, 8

CLARA-A1 CLARA, AVHRR-based, version 1. 4

CLARA-A2 CLARA, AVHRR-based, version 2. 1–12, 14–16, 18–22, 24–26