## Reply to: Anonymous Referee #1

### General Comments to the Author

The author gives a very detailed critique of the operational MODIS cloud mask (MOD35) aggregation strategy using the CALIOP lidar cloud detection product as "ground truth". Collocated 1-km CALIOP and Aqua MODIS data is used to assign mean cloud amounts to the four output cloud mask categories (confident clear, probably clear, probably cloudy, confident cloudy) for various illumination and surface types.

The manuscript is very well written and organized with tables and figures that add detail and understanding to the text. The main object of the paper is to ascertain the suitability of the operational aggregation method from pixel level (Level 2) to temporal and spatial averages (Level 3). The current method simply compares numbers of cloudy pixels to the total number when calculating cloud amounts, where "cloudy" includes confident cloudy and probably cloudy categories and "clear" is confident clear and probably clear categories. This makes the implicit assumption that the two cloudy categories indicate 100% cloudiness while the two clear categories imply 0% cloudiness. The author concludes that this method leads to significant errors in regional level 3 cloud amounts and reassigns cloudiness values to each mask output category based on the collocated CALIOP cloud detection data.

My only objection is that the author makes the assumption that variability in the confidence of clear sky depends only on cloud fractions within 1-km pixels. However, there are other possibilities. Some of them are: 1) optically thin clouds that cover an entire pixel (thin cirrus) 2) surface brightness approaches that of clouds 3) orientation of clouds relative to the sun (scattering angle, 3-D effects) 4) variability in surface characteristics (brightness, topography, shadows, land/water boundaries).

The stated philosophy of the MODIS cloud mask is to be clear sky conservative (see cloud mask ATBD), i.e., if there is any hint of cloudiness in a given pixel, it should be considered "not clear". In this sense then, the practice of considering both confident and probably cloudy pixels to be "cloudy" during aggregation seems reasonable to me, given that no information about cloud morphology is available. However, this is not to say that the present study is not useful as an error analysis or otherwise beneficial to users.

I am recommending the manuscript be accepted for publication with minor revisions as outlined in the specific comments below.

*I am aware that the confidence in detecting clear sky is not only a function of cloud fraction within the IFOV. Thermal and reflectance contrast between a cloud and a background is controlled by many factors, including those listed by the Reviewer. The role of the cloud masking algorithm is to account for all of these factors as closely as possible, and maximize cloud detection success. In this study, only the resulting product of cloud detection (the Level 2 cloud mask) is evaluated. There is no attempt to investigate which factor – and to what degree – impacts the performance of the L2 algorithm. It is assumed that the algorithm, and the resulting L2 product, come "as they are", and are not 100% perfect.*

*The goal is to use the MODIS L2 product with independent data (namely CALIOP) to evaluate how the further processing of Level 2 observations impacts the uncertainty level of the L3 product (gridded monthly cloud amount). The study does not focus on generating the Level 2 product itself, but on calculating the Level 3 data, which is why cloud detection conditions are not explored in detail. We only consider day-time and night-time differences, latitudinal variability, and individual algorithm paths – assuming the latter reflect variation in the background's brightness temperature, reflectance, topography, etc. Differences in MODIS and CALIOP sensitivity to cirrus are discussed.*

**Specific Comments to the Author**

Many high, very optically thin clouds detected by CALIOP have no chance to be categorized as cloudy by observations from a passive radiometer such as MODIS. Please mention this as a partial reason for the 21.5% cloud amount associated with confident clear.

*Explanation added (in the Discussion section), as suggested.*

Abstract: What is the meaning of "uncertainties were related to the efficiency of the cloud masking algorithm"? Please clarify or delete.

*This has been deleted in order to keep the abstract concise (clarification at this point would make the Abstract too discursive).*

Please delete "Until the algorithm can be significantly modified". After 20 years, algorithm issues notwithstanding, large and potentially disruptive modifications to the cloud mask are unlikely and probably unwise.

*Deleted, as suggested.*

Line 133: "IFOV" should be "scan lines".

*In my opinion, 'IFOV' is the correct term. The CALIOP-MODIS matching procedure is IFOV-based, not scan line-based. MODIS is a whiskbroom-type scanner, meaning one rotation of the instrument's mirror results in a scan of 10 lines (considering 1 km detections only). Each line is then divided into 1354 instantaneous fields of view. Only a few, those located close to the MODIS ground track (~10 IFOVs per scan event, ~2030 per data granule) can be matched with CALIOP detections. The use of 'scan lines' would be misleading in this context.*

Line 183: "probably cloudy" should be "probably clear".

*Corrected.*

Line 140: The first two sentences beginning at line 140 are probably better placed at the beginning of Section 3.2. I would eliminate the last sentence of this section as it seems superfluous.

*Changed, as suggested.*

Lines 152-154: This would be a good place to insert a few words about the difficulty of cloud detection from passive instruments during polar night. Thermal contrast is almost nil in these situations and what does exist is often due to temperature inversions, many times multiple ones, that exist with or without clouds being present. Please explain that in polar night, CALIOP has an even bigger advantage in detecting clouds than in warmer climes.

*The discussion has been added – as suggested – however, not in the Results, but in the Discussion section.*

Line 258: What is "the most modest version" of the MODIS cloud mask? Please explain or eliminate the phrase.

*Clarified, as suggested (changed to Collection 061).*

Lines 265-268: Given the errors inherent in remote sensing of cloud properties in general, and in the difficulty of accurate cloud detection in particular, I am surprised that the author would ever expect 100% accuracy from any type of cloud amount calculation. All algorithms are inadequate in some way and to some extent. I strongly advise the author to eliminate the section beginning with “We found the approach to be inadequate” and ending with “environmental conditions”. The statistics given here are just a restatement of previously reported results. Of course there are limitations to the cloud masking procedure and undoubtedly “certain cloud regimes and/or environmental conditions” are more difficult than others. On the other hand, it is quite fair to report the results of the 100% clear/cloud assumptions in the MODIS cloud amount calculations, as is done immediately following.

*The MODIS cloud detection algorithm is only one of many methods, and no method is completely free of limitations. 100% and 0% are only points of reference – the theoretical cloud fraction that an ideal, perfect algorithm would give (if it existed). However, because the Reviewer found this paragraph to be a restatement of previously-reported results, I have followed the recommendation and deleted the suggested part.*

Line 275: I think you mean Table 3. Please add a sentence or two justifying the use of collocated near-nadir CALIOP data on entire swaths of MODIS data or a description of a corrective measure.

*Corrected and additional information added.*

Line 310: The sentence beginning with “Therefore, the standard” is unnecessary.

*Deleted, as suggested.*

Line 327: Variability is to be expected within algorithm paths as they are necessarily very general categories. The statement that the same thresholds are applied in widely varying locations is not completely true. The important 0.65 µm daytime land cloud test is a function of background NDVI and scattering angle. Please include this information.

*Information included, as suggested.*

Line 329: Again, the text “Until significant modifications are made to the MODIS cloud masking algorithm,” is unnecessary and a bit high-handed. I would simply begin with “CALIOP-based : : :”.

*Changed, as suggested.*

References Line 63: Fontana et al., 2013 is missing from the reference list.

*Reference added.*

Figures and Tables Table 3: Caption should indicate that the cloud fractions listed are CALIOP-based.

*Information added, as suggested.*

## Reply to: Anonymous Referee #2

**Major comments**:

1. Line 71: the author mentioned that "no research-based, objective alternatives to the 0/0/100/100 interpretation currently in use have been put forward". However Line 250 reviewed the validation work conducted by Wang et al. (2016) which also estimated the cloud fraction for four MODIS cloud masks. The reviewer suggested reviewing Wang's work in the introduction and also emphasizing what's new in this work. For example, Wang's work focused on daytime only, this work included daytime, night time and both day and night time. Moreover, this work examined how those four cloud fractions changed for different MODIS cloud mask algorithm paths and latitude regions.

*Wang et al. (2016) is an excellent study. It validates the MODIS cloud mask (daytime only), with a focus on multilayer clouds, and considering different cloud regimes (with 2D histograms). However, the latter study does not provide a CALIOP-based cloud fraction for each of the four MODIS cloud mask classes. It may be possible to obtain these fractions (global values only) by an analysis of the confusion matrix (Table 2) presented by the authors. However, no direct information is provided about these values, how such statistics could be derived, or why (while this is the main objective of our study). For this reason, I prefer not to change the Introduction. I do, however, fully acknowledge the work of Wang et al. (2016), and refer to their results in the Discussion.*

2. The cloud fractions were derived with two months data, i.e., January and July 2015. While the author demonstrated the fractions could have a large variability depending on environmental conditions. Could they also have a seasonal variation? How valid to apply the same numbers to different seasons for the whole MODIS mission?

*The seasonality of the CALIPSO-based cloud fraction for MODIS cloud mask classes can be expected wherever environmental conditions are dominated by a strong seasonal cycle, in particular regions where the cloud regime changes noticeably. On the other hand, seasonal environmental change is consistent with changes in the frequency of per-location MODIS algorithm paths. Therefore, when regional CALIOP-based fractions per algorithm path are used (instead of fixed global fractions) the seasonality effect is balanced (at least partially). An operational use of CALIPSO-based factions would require the development of a relevant 'climatology'. An investigation of such a climatology would be an interesting extension of this study.*

3. The author considered CALIPSO data as "ground truth" by including all cloud layers detected by CALIOP. As CALIOP data reported quality flags, it is possible to choose confident clouds only. For example, including clouds with cloud-aerosol discrimination score between 20 and 100 (low, middle and high confidence) or 70 and 100 (high confidence only) by specifying the range of parameter CAD_Score. Not sure how this filter might change the current findings in the paper.

*Our study found that 95.6% of analysed CALIOP observations had CAD confidence of at least 70%, and confidence was below 20% for only 1.5% of data. These statistics did not differ between day and night, or January and July. High, stable CAD values makes it possible to conclude that filtering for data with CAD >70% or >80% would have no impact on the results. On the other hand, CAD results varied slightly more in the tropics, and this issue is discussed in the paper.*

4. In the paper, the cloud fractions are further estimated for each cloud mask algorithm path and day/night conditions. It is noted that the CALIOP has different detection sensitivity during day and night, i.e., CALIOP is able to detect more thin cirrus clouds around the tropical region at night than during the day. This might help understand the day/night discrepancies in Figure1-3.

*Additional information about CALIOP daytime/ night-time sensitivity has been added to the Discussion.*

5. As briefly touched by the author in Line 238, the level 2 CALIOP cloud layer product reported detected cloud layers only. It is very possible there are aerosol layers detected and those aerosol layers would be reported in aerosol products but not in cloud products. In this scenario, the sky is not exactly "clear". To avoid confusions, some researchers use "cloud free" instead "clear".

*I agree that 'cloud-free' is much more accurate in the context of this research, and I have changed 'clear' to 'cloud-free' whenever possible. Nonetheless, I have retained 'clear' in the name of the MODIS cloud mask, since these names are widely (and officially) used in MODIS product documentation.*

**Minor comments:**

1. Abstract: keep consistency when describing four cloud fraction numbers and cloud mask categories. Line 7: "confident cloudy", "probably cloudy", "probably clear", "confident clear". Line 14: 21.5%, 27.7%, 66.6%, 94.7%.

*Corrected, as suggested.*

2. Line 16: "selected locations"? Please give a few locations as examples.

*Examples added, as suggested.*

3. Line 17: "error" → "uncertainty"?

*Changed – 'uncertainty' is the more relevant term.*

4. Line 18: What is "our method"?

*The method used in the study to calibrate MODIS cloud amount. This sentence has been rephrased.*

5. Line 19: "robust" is a strong word. Does the author would like to say something like "We recommend using the cloud fraction ratios found in this work to improve MODIS estimates."

*The sentence has been rephrased.*

6. Line 20: "other mission"? Other passive missions?

*Passive cloud imagers – the sentence has been rephrased.*

7. Line 24: "W m-2" should be "W $m^{-2}$".

*Corrected.*

8. Line 48: "The procedure implemented by NASA…" → The procedure implemented by MODIS science working group?

*I agree. The procedure was developed by the MODIS Science Team or – more precisely – the Atmosphere Discipline Group within the MODIS Science Team. As the MODIS Science Team is a collaboration coordinated by NASA, I used NASA, but I agree that MODIS Science Team is more accurate. NASA has been changed to MODIS Science Team throughout the manuscript.*

9. Line 51: "- see, for example, " → e.g. ?

*Changed.*

10. Line 54: "NASA's approach" → standard procedure? It is not an approach from an agency. Instead, it is from MODIS science working group.

*Changed. See reply to comment 8.*

11. Line 54: "… are both allowed and in use." → "… are adopted by other groups." ?

*Changed (shortened) to: "... are in use."*

12. Line 63: Moved "in Switzerland" after "observations". It would be nice to specify the number of ground-based observations, i.e., "… compared MODIS data with n ground ground-based observations…".

*Changed, as suggested.*

13. Line 70: "NASA standard approach" → standard procedure or standard approach?

*Changed to 'procedure'.*

14. Line 71: "… currently in use have been put forward" is confusing. Does the author mean "… currently widely used are still missing" or something like that?

*Rephrased.*

15. Line 72: "… based on quantitative, empirical lidar observations" is confusing. Does the author mean "… based on a quantitative analysis with lidar observations"?

*Rephrased.*

16. Line 75: The CALIPSO was launched in 2006 instead of 2016.

*Corrected.*

17. Line 77-78: Consider removing "This is because" and "which means that" to make a concise and formal statement.

*The sentence justifies why the study uses CALIOP as a reference. The phrase, "Furthermore, the use of short..." at the beginning of the following sentence is a logical continuation. Therefore, I prefer to leave the paragraph as it is.*

18. Line 83: Add "with CALIOP observations" after "… correspond to".

*Added, as suggested.*

19. Line 83: Again it is not an approach from an agency. The author probably meant "current standard approach" or "current standard procedure".

*Changed. See also reply to comment 8.*

20. Line 84: Does the author mean "Finally, we evaluate whether the MODIS Level 3 standard approach is reliable"?

*Clarified, as suggested.*

21. Line 101: Consider removing “This is made available”.

*Rephrased, as suggested.*

22. Line 103: Consider replacing “product; this was used to assign” with “with”.

*Rephrased, as suggested.*

23. Line 108: Below 8.2 km, CALIOP has a horizontal resolution 0.333 km not 0.33 km.

*‘0.33 km’ corrected to ‘0.333 km’*

24. Line 109: Between 20.2 km and 30.1 km, CALIOP has a horizontal resolution 5/3 km and vertical resolution 180 m. From 30.1 km to 40 km, the horizontal resolution is 5 km and the vertical resolution is 300 m. Please refer to Table 2 in Winker et al. [2006].

*Corrected and clarified, as suggested.*

25. Line 114: “CAL_LID_L2” → level 2 cloud layer products.

*Changed, as suggested.*

26. Line 115: (version 4.20) → (version 4.20, CAL_LID_L2_01kmCLay-Standard-V4-20)?

*Product codename added, as suggested.*

27. Line 119: “Number Layers Found” variable → “Number_Layers_Found” parameter

*Changed, as suggested.*

28. Line 130: “… January and July 2005 …” should be “… January and July 2015 …” Any special reasons to choose these two months?

*Yes, these two months represent atmospheric conditions for summer (July) and winter (January) in the northern hemisphere. The selection of these months makes it possible to investigate contrasting cloud regimes in mid-latitudes (more cumuliform in summer, more stratiform in winter) and season-dependent conditions for cloud detection (e.g. snow cover).*

29. Line 141: Add “MODIS” after “perfect” would help a reader understand.

*Added, as suggested.*

30. Line 147: Based on Table 1, should the number “86.7%” be “64.2%” at night?

*In fact, it should be ‘84.2% at night’ (as in Table 1) – corrected.*

31. Line 151: Should the number “77.4%” be “73.3%”?

*Corrected.*

32. Line 157: Is this region “ITCZ”? Does this high frequency misdetections due to high sensitivity of CALIOP? In other words, CALIOP detected very thin cirrus clouds which are invisible to MODIS.

*Yes, it is the intertropical convergence zone. I have expanded on cloud detection by MODIS and CALIOP at low latitudes in the Discussion.*

33. Line 159: "… MODIS tended to falsely detect cloud rather than fail to detect it". This sentence is confusing. Does this mean higher percentage occurrence or larger area spatial extent? Should "Only" be removed?

*The statement was deleted.*

34. Line 166: It is not exactly "every fifth MODIS" even though the percentage is about 20%.

*Changed to "one fifth of MODIS".*

35. Line 172-173: "no significant day/night difference" even though it is 12.3% for 'probably cloud'?

*Clarified.*

36. Figure 3g and 3h: What does black color over Southern Ocean mean?

*It means there were no confident clear detections by MODIS in these regions at that time.*

37. Line 183: Should  probably cloudy' be 'probably clear'?

*Corrected – 'probably clear' is the correct term.*

38. Line 186: What does "this" in '…, but this was …" mean?

*Rephrased and clarified.*

39. Table 3: Use same terms to describe snow-covered conditions in the context and table caption. For example, use "Snow-free" and "Snow-covered" or "No snow" and "Snow".

*Corrected, as suggested ('snow-covered' and 'snow-free' are now used consistently).*

40. Line 205- 215: The author chose three cloud masking algorithm paths for detailed discussion. It would help a reader understand why those three if providing some explanations. Explain "Results" in Line 205 and "A similar pattern" in Line 211. Which results? Which pattern?

*Four algorithm paths are described in the text. The first is "the combination of night, an oceanic background and snow-cover (or sea ice)". This scenario is notable because it "constituted the 'most cloudy' scenario". The second is "snow-free land at night", this was chosen because: "Results [for it] were most consistent with the standard Level 3" (already mentioned in the manuscript). The two other scenarios are "snow-free land during the day", and "ice-free oceans". The choice of the latter is justified in the paper: it is "the most frequent algorithm path". I agree that the justification of the choice of "snow-free land during the day" was missing. Therefore, following the Reviewer's suggestion, I have added an explanation (it is of particular interest for land/ vegetation MODIS remote sensing).*

*Lines 205 and 211 have been clarified, as suggested.*

41. Line 223: Add a dot between MODIS collection "6" and "1"?

*There are two conventions in use: a three-digit name with leading zero (005, 055, 006, 061, etc.), or to divide a collection number by 100 and use a coma (5.0, 5.5, 6.0, 6.1, etc.). I prefer to use the first, hence '61' has been changed to '061'.*

42. Line 225: It is confusing to discuss level 3 product here since no plots or work on level 3 clouds presented so far.

*Clarified. The implications for Level-3 data are presented in the Discussion, but not before. The first paragraph of the section only introduces issues that are discussed in the following paragraphs. I have made this point clearer in the new version of the manuscript.*

43. Line 235 and Line 240: The author claimed that temporal and spatial separations between Aqua and CALIPSO do not impact the results significantly. If not complicated, it is a good idea to show the plots when using different time and range shifts.

*I have prepared the plots, as suggested. I also agree that they might be interesting for some readers. However, I leave it to the Editor to decide whether they should be included in the main text, or as additional/ supporting online material (the latter would be my choice).*

44. Line 246: Explain acronym "AVHRR".

*Explained, as suggested.*

45. Line 316: What is the spatial grid used to plot Figure 8?

*All figures use the equirectangular projection with 2.5°×2.5° spatial resolution.*

46. Line 321: The author drew a conclusion "Whenever MODIS cloud amount is estimated at a spatial resolution of ~10 degrees of finer, …". There seems no evidence in the paper to support this conclusion. Something missing?

*Ten degrees longitude/ latitude was the approximate area of cloud amount uncertainties in China, along the coast of the Arabian Peninsula, north-west Africa, and some locations in North America. However, I agree that the figure could be misleading when considering, for example, polar regions where the area is much larger. Consequently, the reference to "10 degrees" has been deleted, and replaced by "regional/local".*

47. Line 324: Discussions on MODIS level 3 cloud product could be moved from "Summary and Conclusions" section to previous "Discussion" section

*I prefer not to move the discussion about Level 3 data from the Discussion to the Results. The key 'technical' objective of the study was to derive CALIOP-based cloud fraction from MODIS. The outcome of this work is reported in the Results section. A discussion of the implications of these results for calculating global cloud amounts is a different matter. In my opinion, the present structure of the manuscript clearly separates the results of the study's calculations from a discussion of their impact.*

# Calibration of global MODIS cloud amount using CALIOP cloud profiles

Andrzej Z. Kotarba[1]

[1]Space Research Centre, Polish Academy of Sciences, 00-716 Warsaw, Poland

*Correspondence to*: Andrzej Z. Kotarba (akotarba@cbk.waw.pl)

**Abstract.** The Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection procedure classifies instantaneous fields of view (IFOV) as either 'confident ~~cloudy'~~clear', 'probably ~~cloudy'~~clear', 'probably ~~clear'~~cloudy', or 'confident ~~clear'~~cloudy'. The cloud amount calculation requires quantitative cloud fractions to be assigned to these classes. The operational procedure used by the ~~NASA~~MODIS Science Team assumes that 'confident clear' and 'probably clear' IFOV are cloud-free (cloud fraction 0%), while the remaining categories are completely filled with clouds (cloud fraction 100%). This study demonstrates that this 'best guess' approach is unreliable, especially on a regional/ local scale. We use data from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument flown on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) mission, collocated with MODIS/ Aqua IFOV. Based on 33,793,648 paired observations acquired in January and July 2015, we conclude that actual cloud fractions to be associated with MODIS cloud mask categories are 21.5%, 27.7%, 66.6%, and 94.7%. Spatial variability is significant, even within a single MODIS algorithm path, and the operational approach introduces uncertainties of up to 30% of cloud amount, notably in ~~the~~ polar regions at night, and in selected locations over the northern hemisphere (e.g. China, the north-west coast of Africa, and eastern parts of the United States). Consequently, a~~A~~pplications of MODIS data on a regional/ local ~~at ~10 degrees resolution (or finer)~~scale should first assess the extent of the uncertainty. ~~error. Uncertainties were related to the efficiency of the cloud masking algorithm.~~ We suggest using CALIPSO-based cloud fractions ~~Until the algorithm can be significantly modified, our method is a robust way~~ to ~~calibrate~~ improve ~~(correct)~~ MODIS cloud amount estimates. This approach ~~It~~ can also be ~~also~~ used for MODIS/ Terra data, and other passive cloud imagers, ~~missions~~ where the footprint is collocated with CALIPSO.

## 1 Introduction

Cloud plays a key role in distributing solar energy in the Earth's atmosphere (Trenberth et al., 2009). Consequently, research into the present and future state of the climate system requires accurate information about cloud amount. Depending on its frequency and physical properties, cloud can both heat (greenhouse effect: +30 $Wm^{-2}$) and cool (albedo effect: –48 $Wm^{-2}$) the atmosphere. Their net effect on the planetary radiation budget is negative, meaning the Earth would be warmer if all cloud disappeared (Ramanathan and Kiehl, 2006).

The Global Climate Observing System identifies 13 Essential Climate Variables. This set of critical environmental parameters characterize the Earth's climate (Hollmann et al., 2013); they not only include cloud properties, but also highlight that our knowledge of cloud relies largely on satellite remote sensing. Satellite cloud climatology starts with a cloud mask. The aim is to decide whether cloud is present in a sensor's instantaneous field of view (IFOV), or whether it is cloud free. Input data includes at-sensor registered radiances, along with other auxiliary information that aims to maximize cloud detection.

Efficient cloud detection algorithms have to consider the technical limitations of sensors, available computing power, and environmental factors such as the background (e.g. water, land, snow) and solar illumination (day and night). The resulting cloud mask takes the form of a map that divides IFOV into at least two categories: 'cloud free', and 'cloud contaminated' (or 'cloud filled'). Many masking algorithms introduce additional categories in order to reflect the level of uncertainty in cloud detection (Derrien and Le Gléau, 2005; Dybbroe et al., 2005; Kopp et al., 2014).

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a cloud imaging instrument that is flown onboard NASA's ~~paired~~ polar orbiting satellites: Terra and Aqua. Circling the Earth in the morning orbit (10:30 local solar time; Terra), and afternoon orbit (13:30 local solar time, Aqua), these twin sensors provide a global picture of cloud four times each day, at 1 km/pixel resolution (Guenther et al., 2002; Platnick et al., 2003). With 36 spectral channels, continuous correction for orbital drift, and precisely-calibrated detectors, MODIS has set a new standard in cloud remote sensing, and is still considered to be a state-of-the-art cloud imager, despite being launched in 1999 (Terra), and 2002 (Aqua).

MODIS's cloud detection scheme results in four cloud mask categories: 'confident cloudy', 'probably cloudy', 'probably clear', and 'confident clear' (Frey et al., 2008). The fact that these classes are presented as qualitative, text-based labels rather than a numeric probability causes the technical problem of how to quantitatively interpret these labels. A numeric interpretation is mandatory when instantaneous observations (Level 2 products) are aggregated spatially and/ or temporally to provide climatological information such as mean monthly cloud amount (Level 3 products).

The procedure implemented by NASA's MODIS Science Team (hereinafter the 'standard' or 'operational' procedure) is to assume that IFOV declared 'confident cloudy' and 'probably cloudy' are, in fact, 100% cloud filled, while 'confident clear' and 'probably clear' are completely cloud free (cloud fraction of 0%) (Hubanks et al., 2008). The approach is widely used whenever there is a need to make a binary distinction between cloudy and ~~clear~~ cloud-free pixels – ~~see, for example,~~e.g., Gao et al. (2008), Remer et al. (2012), Wilson and Oreopoulos (2013), Wilson, Parmentier, and Jetz (2014), Kraatz, Khanbilvardi, and Romanov (2017), Gomis-Cebolla, Jimenez, and Sobrino (2020).

However, since the ~~NASA's~~ MODIS Science Team (ST) approach is only a 'best guess', alterative assumptions are ~~both allowed and in~~ also used. For instance, it can be assumed that only 'confident cloudy' pixels are 'cloudy', while all remaining classes are 100% cloud free. Similarly, only 'confident clear' detections can be considered as truly ~~clear~~cloud-free, while all other classes are assumed to be 100% cloud filled (Li et al., 2005). Krijger et al. (2007) argue that the latter approach leads to the false detection of small clouds, while cloud is frequently overlooked if the first method is applied. Another approach is simply to exclude 'probably clear' and 'probably cloudy' detections from the analysis. This strategy was adopted by Chan and Comiso (2013), whose work was based on only 'confident clear' and 'confident cloudy' categories of MODIS data.

Quantitative studies have shown that only considering the ‘confident cloudy’ class as cloudy may be more consistent with other cloud data such as Landsat observations (Melchiorre et al., 2020), or visual observations at meteorological stations (Kotarba, 2015). On the other hand~~, in Switzerland~~, Fontana et al. (2013) compared MODIS data with ground-based observations in Switzerland (4 stations, 12 years of data), and found that results varied from station to station.

The theoretical range of uncertainty related to various interpretations of the MODIS cloud mask was investigated by Kotarba (2015). The latter study found that the global cloud amount estimates may differ by up to 14%, depending on whether only ‘confident cloudy’ detections are considered to be ‘cloudy’, or whether the definition is extended to include intermediate classes. The discrepancy was found to increase by up to 40–60% regionally, suggesting that MODIS cloud estimates are very uncertain in these areas. Such a wide range of uncertainty makes it difficult to run reliable studies on the climate system.

Neither the ~~NASA~~ MODIS ST standard ~~approach~~procedure, nor any other ‘best guess’ variants have been validated on a global scale. Most importantly: no research-based, objective alternatives to the 0/0/100/100 interpretation ~~currently in use~~ have been ~~put forward~~suggested. This study addresses this problem. Specifically, it provides global cloud fractions based on quantitative analysis of ~~, empirical~~ CALIOP lidar observations.

CALIOP (the Cloud-Aerosol Lidar with Orthogonal Polarization) is a cloud profiling instrument flown onboard the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) spacecraft. Launched in 2016, CALIPSO flies in close formation with the Aqua satellite, therefore both instruments – MODIS and CALIOP – sample the same fragment of the atmosphere tenths of seconds apart (Stephens et al., 2018). In this study, CALIPSO data is considered as ground truth. This is because CALIOP is an active remote sensing instrument, which means that it can sample the atmosphere during the day and at night with comparable sensitivity. Imaging radiometers (such as MODIS) perform less effectively at night, when solar channels are missing. Furthermore, the use of short wavelengths makes CALIOP very sensitive to cloud of low optical thickness (e.g. sub-visual cirrus) that is often missed by imagers (Ackerman et al., 2008).

In the following sections we seek to answer the questions: 1) What quantitative cloud fractions (based on CALIOP observations) should be applied to ~~do qualitative~~ MODIS thematic cloud mask classes~~correspond to~~? and 2) What uncertainties in global cloud amount are introduced by the MODIS ST~~NASA~~ standard ~~approach~~procedure? Finally, ~~our conclusions help to answer the question of~~ we evaluate whether the ~~MODIS Level 3~~ standard ~~approach~~ procedure for calculating the MODIS Level 3 cloud amount is reliable.

## 2. Data & Methods

### 2.1 MODIS data

The MODIS cloud detection scheme is based on thresholds that are applied to brightness temperature (thermal bands), and reflectance (solar channels), derived from observations in 22 spectral bands ranging from 0.66 µm to 13.9 µm. Ackerman et al. (1998), Frey et al. (2008), and Baum et al. (2012) provide very detailed descriptions of the cloud masking procedure. The general concept is as follows.

The algorithm executes a series of tests, each of which results in a confidence level (ranging from 0 to 1) that a particular IFOV is cloud free. Tests to detect similar cloud types are grouped. The lowest confidence level for a test within a group is set as the confidence level for the whole group. Confidence levels for groups are then multiplied to determine the final confidence level (Q). Finally, the IFOV is assigned to one of four cloud mask classes: 'confident clear' (Q > 0.99), 'probably clear' (Q > 0.95), 'probably cloudy' (Q > 0.66), or 'confident cloudy' (Q ≤ 0.66). The exact number of spectral tests varies from a few to over a dozen, depending on the path through the algorithm. Paths reflect different environmental conditions, and are introduced to maximize success. Dedicated sets of spectral tests are executed for land, ocean, desert and coastal areas, for both day and night conditions. The presence of snow and/ or ice is taken into account, as is sunglint over oceans. Separate thresholds have been introduced for polar regions, which are defined as land and ocean within 30 degrees of each pole.

Cloud detection results are stored in the 48-bit 'Cloud Mask' product, codenamed MYD35 (Aqua) and MOD35 (Terra) following the MODIS nomenclature. In this study, we evaluated the latest version of MYD35 (Collection 061) data, ~~. This is made~~ available in the form of 5-minute granules, at 1 km per pixel spatial resolution (at nadir), with native satellite projection. Each MYD35 file is accompanied by a MYD03 'Geolocation file' product, that ~~; this was used to assign longitude~~stores longitude and latitude information ~~to~~ for individual cloud mask IFOV.

## 2.2 CALIOP data

CALIOP operates at 532 nm and 1064 nm. The instrument's pencil-like beam only scans locations along the satellite's ground track, as a trade-off for information on the vertical structure of cloud/ aerosols. Its spatial resolution is a function of the satellite's altitude. Resolution is finest – 0.333 km horizontal, 30 m vertical – in the troposphere, up to 8.2 km. Between 8.2 km and 20.2 km, vertical resolution falls to 60 m, and horizontal sampling to 1 km. ~~Above~~ Between 20.2 km ~~,~~ and 30.1 km, data are even coarser: ~~5~~ 1.667 km horizontal and ~~300~~ 180 m vertical resolution. Higher in the atmosphere (30.1 km to 40.0 km) horizontal resolution decreases to 5 km, while vertical resolution is 300 m (Hunt et al., 2009; Winker et al., 2006).

CALIOP detects cloud by applying thresholds to 532 nm attenuated scattering ratios. The aim is to separate the cloud signal from the clear air background (molecular scattering), aerosols, and instrument noise. The algorithm calculates cloud base height, cloud top height, and – as a consequence – the number of cloud layers within a profile. Up to 10 layers can be reported. The procedure is fully automatic (Vaughan et al., 2009). The output is stored in the Level 2 Cloud Layer Data ~~CAL_LID_L2~~ product, available at ~~⅓~~333 ~~k~~m, 1 km, and 5 km along-track sampling intervals. Here, we use the 1 km interval (version 4.20; CAL_LID_L2_01kmCLay-Standard-V4-20), as its resolution matches the spatial resolution of the MODIS cloud mask. Furthermore, 1 km is the highest available level of detail for CALIOP data within the troposphere.

In order to use the CALIOP product to evaluate MODIS data, 3-dimensional cloud layer data was reduced to column-integrated, binary cloud/ no cloud information. Specifically, we focused on the 'Number_Layers_Found' ~~variable~~ parameter provided in the CAL_LID_L2 product. 'No cloud' was recorded when the latter variable was set to 0 (i.e. zero layers found), and 'cloud' otherwise (i.e. at least one layer was reported). Geolocation was based on longitude and latitude arrays included in the product at 1 km spatial resolution.

In some cases, cloud and aerosol can appear similar to CALIOP. The cloud-aerosol discrimination (CAD) score, which is a numerical index stored in the CAL_LID_L2 product, provides information about the algorithm's uncertainty in separating cloud and aerosol. In the case of cloud, CAD values vary between 0% (it is unclear whether aerosol or cloud was observed) and 100% (cloud detected with the highest confidence). The index is calculated for each cloud layer found in the CALIOP atmospheric profile. Since our study focuses on column-integrated information of cloud presence, we selected the highest CAD value within a profile. Statistics for January and July 2015 showed that 95.6% of considered CALIOP observations were characterized by a CAD score of at least 70%, while it was below 20% for only 1.5% of data. Therefore, the selected CALIOP data can be considered as a reliable reference for MODIS. See Supplementary Online Materials, Fig. S1 for more detailed statistics about CAD scores.

**2.3 Matching CALIOP and MODIS data**

Matching CALIOP data is a well-established method for the calibration/ validation of atmospheric products from various space missions. It has already been widely used for MODIS/ Aqua (Baum et al., 2012; Holz et al., 2009; Sun-Mack et al., 2014; Wang et al., 2016; Xie et al., 2010), and sensors flown onboard Suomi-NPP, NOAA, and MetOp polar orbiting spacecraft, which occasionally synchronize their orbital configuration with CALIPSO (Hutchison et al. 2014; Heidinger et al. 2012; Karlsson and Johansson 2013; Karlsson and Dybbroe 2010). CALIPSO also passes within the field of view of other geostationary satellites, and CALIOP data is used to assess their atmospheric products (Sèze et al., 2015; Shang et al., 2018).

In this study, Aqua/ MODIS data for January and July 2005 were paired with corresponding CALIPSO/ CALIOP observations. The matching procedure selected a MODIS IFOV and compared it with the corresponding CALIOP profile (where the geometric centre was within the selected MODIS IFOV). Although very straightforward, the procedure was time-consuming since a single MODIS granule contains ~2030 IFOV, and a full day of Aqua observations produces 288 granules.

The final database consisted of 33,793,648 MODIS–CALIOP paired observations. Average spatial separation between the centres of MODIS and CALIOP IFOV was 418 m, and 19% had a separation of less than 250 m. Temporal differences between lidar and imager observations were determined using the spacecrafts' on-orbit separation, and ranged from 60 sec to 97 sec (81 sec on average). Our dataset excluded one MODIS cloud mask processing path: sunglint. This was because CALIPSO's orbit has been intentionally designed to avoid sunglint areas, in order to avoid the lidar being 'blinded' by solar reflection from the ocean.

Our empirical calculation of cloud fraction in each MODIS cloud mask class was based on the ratio of CALIOP cloudy detections to all detections within a class. A perfect MODIS cloud detection algorithm would categorise a 0% cloud fraction as 'confident clear', while a 100% cloud fraction would be categorised as 'confident cloudy'. ~~Our results showed that this was not true.~~

## 3 Results

### 3.1 Misdetection of cloud and clear sky by MODIS

The first key point that emerged from the matched MODIS–lidar observations was the accuracy of MODIS cloud detections. Overall accuracy in January and July 2015, compared to reference CALIOP data, was 89.4% during the day, and ~~86.7~~84.2% at night (Tab. 1). This statistic assumes that ‘probably clear’ detections are merged with ‘confident clear’, and ‘probably cloudy’ detections are combined with ‘confident cloudy’. If a less tolerant approach is applied, i.e. only ‘confident clear’ and ‘confident cloudy’ detections are considered (‘probably clear’/ ‘probably cloudy’ classes are interpreted as misdetections), overall accuracy fell to 81.9% during the day, and ~~77.4~~73.3% at night.

Clouds missed by MODIS, but detected by CALIOP were most frequent during the polar night, regardless of the hemisphere (Fig. 1c, d). Up to 40% of MODIS ‘confident clear’ and ‘probably clear’ detections were found to be incorrect around Antarctica in July, and the Arctic in January. ~~Both polar oceans and the continental shelf were affected.~~ Globally, daytime (Fig. 1a, b) misdetections were around half of those at night. They only exceeded 30% locally, and polar regions were less affected. A notable observation was July in the northern hemisphere, where only a few small regions of misdetection were observed. The analysis highlighted a belt of relatively higher frequency misdetections (15–25%) in the equatorial zone; here the magnitude of the effect was similar for both day and night.

~~It was noticeable that MODIS tended to falsely detect cloud rather than fail to detect it.~~ Only a few occasions were identified when over 10–15% of MODIS ‘confident cloudy’ and ‘probably cloudy’ observations were identified as ~~clear~~ cloud-free by CALIOP (Fig. 2). Further analysis showed that although false detection was rare in polar regions, it was significant in specific regions of the northern hemisphere. North-east China emerged as the most problematic area (Fig. 2a). Here, 50–70% of MODIS ‘confident cloudy’ and ‘probably cloudy’ detections were cloud-free ~~clear~~ according to CALIOP. However, this high rate of false detection was only observed in January, and only during the day.

### 3.2 Cloud fraction for cloud mask classes

Our empirical calculation of cloud fraction in each MODIS cloud mask class was based on the ratio of CALIOP cloudy detections to all detections within a class. A perfect MODIS cloud detection algorithm would categorise a 0% cloud fraction as ‘confident clear’, while a 100% cloud fraction would be categorised as ‘confident cloudy’.

On average, ~~every~~ one fifth of MODIS ‘confident clear’ detections were~~as~~ found to be cloudy by CALIOP. Consequently, the average cloud fraction for this class was 21.5%, instead of the theoretically expected 0% (Tab. 2). At night, the fraction was over twice the daytime value (29.5% compared to 12.7%). On the other hand, pixels flagged by the MODIS algorithm as ‘confident cloudy’ were, almost always, contaminated with some cloud, and were sometimes cloud-filled. Regardless of the time of day, the actual CALIOP-based cloud fraction for ‘confident cloudy’ detections was close to 100%, reaching 94.7%.

MODIS intermediate classes constituted 13.3% of all detections. CALIOP cloud fractions were 27.7% and 66.6% for ‘probably clear’, and ‘probably cloudy’ classes respectively. The statistics revealed a difference of up to 17% ~~no significant day/ night~~

between day and night conditions, and it was especially small for the ‘probably clear’ class (1.3%) and the ‘confident cloudy’ class (daytime had no impact at all).~~, where the difference was only 1.3% (compared to 12.3% for ‘probably cloudy’).~~ The parameters reported in Table 2 are global averages (means), and spatial diversity was observed. Differences were smallest for the ‘confident cloudy’ class – both during the day (Fig. 3a) and at night (Fig. 3b) – and the CALIOP-based cloud fraction exceeded 90% at almost every location. North-east China, the southern Arabian Peninsula and Eastern Antarctica were the only significant exceptions; here cloud fraction decreased to 50–70%.

The cloud fraction distribution was homogeneous for the ‘confident clear’ class, however, only during the day (Fig. 3g). At night (Fig. 3h) it increased substantially in polar regions, especially over the oceans of the southern hemisphere, along the coast of Antarctica. Unlike polar regions, no noticeable day/ night difference was observed for mid- and low-latitudes (< 10%); here, the cloud fraction was very low (< 20%), and very few MODIS ‘confident clear’ detections were identified as cloudy by CALIOP.

Among the MODIS intermediate classes, ‘probably ~~cloudy’~~ clear’ differed most from CALIOP-based data. First, a high cloud fraction (> 70%) was observed at night along the equator and in polar regions (both oceanic and continental; Fig. 3f). At mid-latitudes the cloud fraction for MODIS ‘probably clear’ observations was relatively low (< 10–20%). This pattern was inverted during the day (Fig. 3e). At this time a higher (50–75%) cloud fraction was noted for mid-latitudes, ~~but this was only~~ and over ~~partially~~ parts (typically land) of the ~~true in~~ polar regions.

### 3.3 Cloud fraction as a function of the algorithm path

The MODIS cloud detection algorithm distinguishes between day and night (Tab. 1, Tab. 2), and four types of background (land, desert, coast, ocean), each of which can be either snow-covered or snow-free. CALIOP-based cloud fractions for all algorithm paths are reported in Table 3. These values give a detailed understanding of MODIS cloud detection results. Data are given for each class of the MODIS cloud mask separately. In our study, we structured the paths through the algorithm in more detail. Snow-covered conditions were considered for land, desert, ocean and coast separately, while in the MODIS algorithm they are grouped as snow/ ice. This greater level of detail allowed us to observe how the presence of snow impacted the cloud mask over different backgrounds.

Per-class estimates of cloud fraction were very consistent for all algorithm paths for the ‘confident cloudy’ category (Tab. 3). Final values ranged between 97.7% (night, snow-free, land) and 86.4% (night, snow-covered, desert), and were close to the standard Level 3 assumption of 100%. This finding contrasted with cloud fractions found for the ‘confident clear’ category. While MODIS recorded cloud-free conditions, CALIOP data revealed that the actual cloud fraction ranged from 8.0% (night, snow-free, land) to 49.7% (night, snow-covered, ocean).

The combination of night, an oceanic background and snow-covered (or sea ice) constituted the ‘most cloudy’ scenario (Fig. 4). Here, a very high cloud fraction was found for not only the ‘confident cloudy’ category (96.8%, Fig. 4a), but also all remaining classes: 82.5% (‘probably cloudy’; Fig. 4b), 73.3% (‘probably clear’; Fig. 4c) and, surprisingly, up to 49.7% for ‘confident clear’ (Fig. 4d).

~~Results~~ CALIOP-based cloud fractions were most consistent with the standard Level 3 interpretation for snow-free land at night (Fig. 5). Here, the cloud fraction for 'confident clear' was low (8.0%; Fig. 5d), and very high for 'confident cloudy' (97.7%; Fig. 5a). At the same time, intermediate classes were well-separated: 68.5% for 'probably cloudy' (Fig. 5b), and 25.6% for 'probably clear' (Fig. 5c). Globally, no significant difference was found for cloud fraction values for the night/ ~~no~~ snow-free/ land algorithm path. A small exception was noted for the 'probably clear' type, where the cloud fraction was 10–30% higher in the tropics compared to the rest of the world.

A similar ~~pattern~~ spatial distribution of CALIOP-based cloud fractions was observed for snow-free land during the day – the scenario of particular interest for land/ vegetation remote sensing with MODIS. The two notable differences were related to 'probably clear' (Fig. 6c) and 'confident clear' categories (Fig. 6d). The latter occurred twice as often during the day (15.6%) than at night (8.0%). Similarly, cloud was more frequent in the 'probably clear' class. However, this was only found in the tropics and at high latitudes, which mirrored a zonal pattern that was only weakly seen at night.

As ice-free oceans represent the majority of Earth's surface, cloud detection over ocean is the most frequent algorithm path. Daytime conditions make detection easier (due to the availability of solar channels). Under such circumstances, CALIOP detected cloud in 10.5% of MODIS's 'confident clear' observations (Fig. 7d), and confirmed 95.2% of 'confident cloudy' detections (Fig. 7a). Cloud fractions for intermediate classes (daytime over ice-free ocean) were 54.5% and 28.4% for 'probably cloudy' (Fig. 7b) and 'probably clear' (Fig. 7c) categories, respectively. 'Probably clear' was the only class where there was a clear latitude-dependent cloud fraction: values increased by 30–60% along a path ~30–40 degrees north/ south.

## 4 Discussion

Our investigation of spatially and temporally collocated MODIS (cloud imager) and CALIOP (cloud profiling lidar) observations for January and July 2015 revealed that MODIS Collection 061 global cloud amount estimates are imperfect ~~in two ways~~. ~~First, d~~During the generation of the Level 2 product, the masking algorithm fails to accurately report cloud over polar regions, and over selected locations at lower latitudes. Consequently (as discussed in this section) ~~Second,~~ the Level 3 product generation underestimates cloud fractions for cloud mask classes in numerous regions. The reliability of these results depends on several factors, most notably the spatial and temporal accuracy of Aqua/ MODIS and CALIPSO/ CALIOP collocation.

Temporal differences between Aqua and CALIPSO observations varied from 60 seconds to 97 seconds. In this time, cloud can develop and move, introducing the risk that CALIOP observes a different state of the atmosphere compared to MODIS. Várnai and Marshak (2009) evaluated the problem by comparing MODIS reflectance with that collected by the Wide Field Camera. The latter is an imaging instrument flown onboard CALIPSO, along with CALIOP. They found that for low cloud, radiance differed only slightly over 72 seconds, and it was reasonable to ignore any discrepancies when focusing on aerosol properties (they gave no particular conclusions for cloud). In order to test how sensitive our results were to the time shift, we calculated the overall accuracy of the cloud detection algorithm as a function of the time between Aqua and CALIPSO passes. The results were very consistent: despite the shift, accuracy remained at 86.7±0.1%. This finding confirmed that the temporal separation between Aqua and CALIPSO had no significant impact on the results of our study.

Another potential source of uncertainty is the geometric mismatch between MODIS and CALIOP IFOV. They are not aligned perfectly: 66% of collocated IFOV were separated by less than 0.5 km, and 82% by less than 0.6 km. Similar statistics – 75% and 93%, respectively – were found by Wang et al. (2016) in their investigation of cloud based on MODIS and CALIOP observations. To investigate whether geometric conditions did have an impact on our results, we calculated the overall accuracy of the MODIS cloud mask as a function of the distance between MODIS and CALIOP IFOV. For ranges up to 1 km with a 100 m step, the change in accuracy was insignificant: –87.0±0.3% on average. (See Supplementary Online Materials' Fig. S2 for more detailed statistics about the spatial and temporal separation between MODIS and CALIOP).

It is possible that agreement between MODIS and CALIOP data is affected by cloud optical thickness (τ) or, more precisely, by the higher sensitivity of CALIOP in detecting optically-thin cloud. Ackerman et al. (2008) estimated the MODIS limit for τ to be approximately 0.4. A similar improvement in agreement with CALIOP as a consequence of increasing τ was observed by Karlsson and Håkansson (2017) for the Advanced Very High Resolution Radiometer (AVHRR) instrument. The latter study demonstrated that the imager's probability of detection changed in the range $0.0<\tau<1.0$. We calculated the same statistic, and found that the probability distribution for MODIS was identical to AVHRR – although MODIS values were higher. This finding strongly suggests that cloud thickness has the same impact on our results as that found in previous studies.

Collection 006 of MODIS data was investigated by Wang et al. (2016), who used lidar–radar (CALIPSO–CloudSat) profiles to focus on daytime multi-layered clouds. Our findings are consistent with those reported by Wang et al. (2016), despite the fact that the latter authors used a dataset of 267 million IFOV, while our study relied on around 33 million profiles. Their validation of Collection 006 reliability found overall agreement of 77.8% compared to our study, which found 81.9%. The difference may be due to the different sample sizes. Our result for cloud-free ~~clear~~ sky detection was slightly higher than in Wang et al. (2016): 25.5% compared to 20.9%. On the other hand, results for cloudy sky detection were very similar: 56.9% compared to 56.4% in our study.

~~As reported by Wang et al. (2016), and previously by Baum et al. (2012) and Ackerman et al. (2008), cloud detection in polar regions remains an unsolved issue for MODIS.~~ Our study revealed that even for ~~the~~ Collection 061, i.e. the most recent ~~modest~~ (July 2020) version of the MODIS cloud mask, up to 40% of cloud-free ~~clear~~ skies ~~recorded~~ detected during the polar night were actually cloudy. Daytime accuracy was lowest over China (in January), the USA/ Canada (in January) and over tropical ocean along the west (January) and east (July) coasts of Africa. In these cases, MODIS detected cloud that did not exist according to CALIOP. False detections may be due to snow cover (the USA/ Canada), high aerosol content over China (Zhang et al. 2019; Tan, Zhang, and Shi 2019), and ocean bordering desert regions in North Africa (Weinzierl et al., 2017; Zuluaga et al., 2012).

As reported by Wang et al. (2016), and previously by Baum et al. (2012) and Ackerman et al. (2008), cloud detection in polar regions remains an unresolved issue for MODIS, and similar passive imaging radiometers. Polar night is especially challenging. Successful discrimination between cloud and the underlying surface requires radiance measurements in ice absorption bands (e.g. 1.6 µm or 2.1 µm). But as these are only available in daytime, night-time detection has to rely on thermal infrared data. As thermal inversion in the polar tropopause decreases the thermal contrast between cloud and the background, the thermal signatures of cloud and the land/ ocean surface become indistinguishable, leading to cloud masking errors (Liu et al., 2004). CALIOP, however, does not require solar illumination to operate. As it uses light emitted by the instrument itself, its performance is far less affected by day-night conditions. CALIOP’s night-time data are of even higher quality, because solar illumination introduces an additional background signal and, thus, decreases the signal-to-noise ratio (Hunt et al., 2009). Furthermore, MODIS tends to miss up to ~20% of cloud along the Intertropical Convergence Zone (ITCZ), regardless of the time of year (January or June), and the time of day. This can be partially explained by the fact that MODIS is less sensitive to optically thin cloud than CALIPSO, and the ITCZ is the region where cirrus is most frequently observed (Sassen et al., 2009). The higher sensitivity of CALIOP to optically thin cirrus, and the higher sensitivity of the lidar during night-time, also explains why CALIOP-based cloud fractions for MODIS ‘confident clear’ and ‘probably clear’ classes are higher along the ITCZ  at night (Fig. 3f, h) than during the day (Fig. 3e, g).

The main goal of our study was to investigate the validity of the standard (operational) approach to the quantitative ~~interpolation~~ interpretation of MODIS cloud mask classes. ~~We found the approach to be inadequate. Our study found that it is unreasonable to assume that ‘confident cloudy’ and ‘probably cloudy’ IFOV are always cloudy. Similarly, ‘confident clear’ and ‘probably clear’ IFOV cannot be assumed to be clear. Our findings show that actual cloud fractions for these classes varied significantly, and never reached the expected 100% or 0%. The most accurate assumption was that ‘confident cloudy’ was actually cloudy. ‘Confident cloudy’ detections were confirmed as cloudy by CALIOP in at least 90% of cases (97.7% for night/ snow-free/ land, and 97.6% for day/ snow-covered/ land). For all remaining categories, cloud fractions reflected the limitations of the cloud masking procedure for certain cloud regimes and/ or environmental conditions.~~

The most important consequence of calculating empirical cloud fractions for MODIS cloud mask categories is the ability to recalculate global cloud amount with new weights. Therefore, instead of using global fractions (reported in Table 2), we derived a set of dedicated fractions for each algorithm path, and each 2.5-degree grid box (i.e. a local equivalent to the data given in Table ~~2~~3). This considers MODIS IFOV within the full swath (excluding sunglint), and not only those collocated with CALIOP. Full-swath data were used because the MODIS L3 cloud amount product applies the same cloud mask interpretation to all IFOVs, regardless of their off-nadir angle. On the other hand, the use of nadir-only MODIS observations would result in CALIOP-like spatial coverage of the data, creating significant gaps due to CALIOP’s pencil-like viewing geometry. Figure 8 illustrates the results of the calculation and reports differences in cloud amount between the MODIS ST ~~NASA~~ operational product, and the product generated using the fractions presented in this study.

The outcome of the simulation shows that the use of current operational cloud fractions introduces significant errors. In some locations, MODIS underestimates cloud amount by 20–40%, most notably in polar regions at night. An overestimation of similar magnitude is observed mostly over the northern hemisphere: the USA/ Canada and China in January (both day and night), and the tropical coasts of Africa during the day (both in January and July). Consequently, MODIS Level 3 estimates of cloud amount should be used with great caution in those regions. This is especially important for the Arctic, which is undergoing a rapid change in climatic conditions (Serreze and Barry, 2011), and where cloud has been found to be an essential element in feedback (Kay et al. 2008; Vavrus 2004; Shupe and Intrieri 2004; Tan and Storelvmo 2019).

The availability of collocated MODIS and CALIOP observations also allowed us to examine which of the three 'best guess' interpretations of cloud mask categories is most accurate: the one when only 'confident cloud' IFOV are 'cloudy', the one that only considers 'confident clear' as 'clear', or the operational approach? We therefore calculated merged global cloud amount for January and July 2015. Our results show that, on the global scale, the standard approach is closest to CALIOP reference data, although only during the day (Tab. 4). At night, it is more accurate to assume that only 'confident clear' is actually cloud-freeclear. The global result is biased by the polar night. In these conditions, all three 'best guess' interpretations noticeably underestimate cloud amount. At low- and mid-latitudes the standard (operational) approach differs from CALIOP data by ±2%. However, it should be noted that these statistics relate to large areas. As our study shows, regional differences are orders of magnitude larger.

Our study assumed that CALIOP's 'cloudy' IFOV was always completely cloud filled. This assumption is common when interpreting cloud masks based on data from the majority of imaging radiometers flown onboard meteorological and land-imaging satellites. However, studies by Zhao and Di Girolamo (2006), and Kotarba (2010) suggest that this postulate may not be true. Both of the latter studies took advantage of a rare collocation between a meteorological imager (MODIS) and the high-resolution land imager (ASTER) flown onboard the Terra satellite. Nearly 3,000 ASTER IFOV were located within each MODIS pixel. Kotarba (2010) showed that for sunglint-free, oceanic scenes in the tropics, actual cloud coverage for the 'confident cloudy' MODIS category was 79.2% (mean) or 99.8% (median), instead of the assumed 100%. Comparable statistics for CALIOP are not available, as the CALIPSO spacecraft does not carry a high-resolution imager. Given the lack of alternatives, we must accept the hypothesis that 'cloudy' means 100% cloud filled.

## 5 Summary and Conclusion

This study investigated 33,793,648 collocated MODIS (cloud imager) and CALIOP (cloud profiling lidar) observations, acquired in January and July 2015. Our evaluation of the dataset allowed us to answer three, essential questions, related to global estimates of cloud amount resulting from the MODIS/ Aqua mission. These questions are:

1. *What are the actual cloud fractions corresponding to MODIS cloud mask classes?* We found that these fractions are 21.5%, 27.7%, 66.6%, and 94.7%, rather than the MODIS Science TeamNASA-assumed values of 0%, 0%, 100% ,and 100% for 'confident clear', 'probably clear', 'probably cloudy', and 'confident cloudy' categories, respectively (Tab. 2). ~~Therefore, the standard (operational) approach used to generate MODIS Level 3 cloud amounts is inaccurate.~~ Importantly, we found that the percentage of cloud cover to be assigned to MODIS cloud mask classes varied spatially (Fig. 3), and recommend that global fractions should be avoided, in favour of local alternatives.

2. *How significant are uncertainties in global cloud amount estimates calculated using the MODIS ST ~~NASA~~ operational approach?* We found that uncertainties were up to −30% of cloud amount in the polar regions at night, and up to +30% of cloud amount in selected locations over the northern hemisphere, more frequently during the day (Fig. 8).
3. *Is the MODIS Level 3 standard approach reliable?* Our results showed that when a global cloud amount value is required (day and night, for all latitudes), the standard approach can be considered reliable (Tab. 4). We found that, in this case, it was more accurate than other 'best guess' approaches – namely only 'confident clear' is 'clear' (other classes are 'cloudy'), and 'confident cloudy' is 'cloudy' (other classes are 'clear'). However, on a regional scale the standard approach fails (Fig. 8). Whenever MODIS cloud amount is estimated regionally or locally ~~at a spatial resolution of ~10 degrees or finer,~~ it is necessary to assess whether a particular location might be affected by an error of up to ±30%.

Errors and uncertainties related to the generation of the MODIS Level 3 cloud amount product originate in the Level 2 product: the cloud mask (Fig. 1–2 vs. Fig. 8). The cloud detection algorithm is more-or-less accurate depending on environmental conditions, which are approximated as algorithm paths (Tab. 3). However, conditions within paths are not constant (Fig. 4–7): for instance, the same radiance/ reflectance thresholds are applied to Europe, the USA and China, while environmental conditions in these locations are not the same (e.g. different aerosol loads, different aerosol optical properties). The MODIS Science Team have attempted to discriminate between these conditions. For instance, since Collection 006 the 0.86 μm reflectance test over land considers thresholds that are a function of the Normalized Difference Vegetation Index (NDVI) and scattering angle. Although cloud misclassification is less frequent than in previous Collections, it still occurs, and impacts the degree of uncertainty of L3 cloud amount estimates, as shown in this study.

CALIOP-based estimates of cloud fraction are a robust way to adjust (and correct) MODIS estimates. The method described in this paper can be used globally, with the exception of sunglint regions (which are not sampled by CALIOP). In these areas 'best guess' findings can, potentially, be applied. The polar regions benefit most from the new method. Cloud fractions derived for MODIS/ Aqua may be also adopted for MODIS/ Terra, since the two sensors are expected to produce comparable and homogenous records. Moreover, the occasional collocation of the CALIPSO satellite with AVHRR and VIIRS instruments makes it possible to calculate similar cloud fractions for these missions, and produce more reliable cloud climatologies.

## Data availability

MODIS data are available from the Level 1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) at NASA's Goddard Space Flight Center (https://earthdata.nasa.gov/eosdis/daacs/laads). CALIPSO products are available from the Atmospheric Science Data Center (ASDC) at NASA's Langley Research Center (https://eosweb.larc.nasa.gov/).

## Author contribution

AZK designed the research, carried it out, and prepared the manuscript.

## Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgements

The organization and pre-processing of the CALIPSO database was part of a project funded by the National Science Institute of Poland under contract no. UMO-2017/25/B/ST10/01787. Processing of the MODIS-CALIPSO database and the resulting analyses was funded by the Space Research Centre of the Polish Academy of Sciences, as a part of the research theme ‘Satellite monitoring of geophysical processes in the atmosphere and Earth’s surface, related to climate and light pollution’.

## References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C. and Gumley, L. E.: Discriminating clear sky from clouds with MODIS, J. Geophys. Res. Atmos., doi:10.1029/1998JD200032, 1998.

Ackerman, S. A., Holz, R. E., Frey, R., Eloranta, E. W., Maddux, B. C. and McGill, M.: Cloud detection with MODIS. Part II: Validation, J. Atmos. Ocean. Technol., doi:10.1175/2007JTECHA1053.1, 2008.

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K. and Yang, P.: MODIS cloud-top property refinements for collection 6, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-11-0203.1, 2012.

Chan, M. A. and Comiso, J. C.: Arctic cloud characteristics as derived from MODIS, CALIPSO, and cloudsat, J. Clim., doi:10.1175/JCLI-D-12-00204.1, 2013.

Derrien, M. and Le Gléau, H.: MSG/SEVIRI cloud mask and type from SAFNWC, Int. J. Remote Sens., 26(21), 4707–4732, doi:10.1080/01431160500166128, 2005.

Dybbroe, A., Karlsson, K. G. and Thoss, A.: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: Algorithm description, J. Appl. Meteorol., doi:10.1175/JAM-2188.1, 2005.

Fontana, F., Lugrin, D., Seiz, G., Meier, M., and Foppa, N.: Intercomparison of satellite- and ground-based cloud fraction over Switzerland (2000–2012), Atmospheric Research, doi: 10.1016/j.atmosres.2013.01.013, 2013.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R. and Wang, X.: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5, J. Atmos. Ocean. Technol., doi:10.1175/2008JTECHA1052.1, 2008.

Gao, M., Qin, Z., Zhang, H., Lu, L., Zhou, X. and Yang, X.: Remote sensing of agro-droughts in Guangdong Province of China using MODIS satellite data, Sensors, doi:10.3390/s8084687, 2008.

Gomis-Cebolla, J., Jimenez, J. C. and Sobrino, J. A.: MODIS probabilistic cloud masking over the Amazonian evergreen tropical forests: a comparison of machine learning-based methods, Int. J. Remote Sens., doi:10.1080/01431161.2019.1637963, 2020.

Guenther, B., Xiong, X., Salomonson, V. V., Barnes, W. L. and Young, J.: On-orbit performance of the Earth Observing System Moderate Resolution Imaging Spectroradiometer; First year of data, Remote Sens. Environ., doi:10.1016/S0034-4257(02)00097-4, 2002.

Heidinger, A. K., Evan, A. T., Foster, M. J. and Walther, A.: A naive Bayesian cloud-detection scheme derived from Calipso and applied within PATMOS-x, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-11-02.1, 2012.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M. and Wagner, W.: The ESA climate change initiative: Satellite data records for essential climate variables, Bull. Am. Meteorol. Soc., 94(10), 1541–1552, doi:10.1175/BAMS-D-11-00254.1, 2013.

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., Vaughan, M. A. and Baum, B.: Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP, J. Geophys. Res. Atmos., doi:10.1029/2008JD009837, 2009.

Hubanks, P. A., King, M. D., Platnick, S. and Pincus, R.: MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document No. ATBD-MOD-30 for Level-3 Global Gridded Atmosphere Products (08_D3, 08_E3, 08_M3). [online] Available from: https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/L3_ATBD_C6_C61_2019_02_20.pdf, 2008.

Hunt, W. H., Vaughan, M. A., Powell, K. A. and Weimer, C.: CALIPSO lidar description and performance assessment, J. Atmos. Ocean. Technol., doi:10.1175/2009JTECHA1223.1, 2009.

Hutchison, K. D., Heidinger, A. K., Kopp, T. J., Iisager, B. D. and Frey, R. A.: Comparisons between VIIRS cloud mask performance results from manually generated cloud masks of VIIRS imagery and CALIOP-VIIRS match-ups, Int. J. Remote Sens., doi:10.1080/01431161.2014.932465, 2014.

Karlsson, K.-G. and Håkansson, N.: Detailed characterisation of AVHRR global cloud detection performance of the CM SAF CLARA-A2 climate data record based on CALIPSO-CALIOP cloud information, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-307, 2017.

Karlsson, K. G. and Dybbroe, A.: Evaluation of arctic cloud products from the EUMETSAT climate monitoring satellite application facility based on CALIPSO-CALIOP observations, Atmos. Chem. Phys., doi:10.5194/acp-10-1789-2010, 2010.

Karlsson, K. G. and Johansson, E.: On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data: Example investigating the CM SAF CLARA-A1 dataset, Atmos. Meas. Tech., doi:10.5194/amt-6-1271-2013, 2013.

Kay, J. E., L'Ecuyer, T., Gettelman, A., Stephens, G. and O'Dell, C.: The contribution of cloud and radiation anomalies to the 2007 Arctic sea ice extent minimum, Geophys. Res. Lett., doi:10.1029/2008GL033451, 2008.

Kopp, T. J., Thomas, W., Heidinger, A. K., Botambekov, D., Frey, R. A., Hutchison, K. D., Iisager, B. D., Brueske, K. and Reed, B.: The viirs cloud mask: Progress in the first year of s-npp toward a common cloud detection scheme, J. Geophys. Res., doi:10.1002/2013JD020458, 2014.

Kotarba, A. Z.: Estimation of fractional cloud cover for Moderate resolution imaging Spectroradiometer/Terra cloud mask classes with high-resolution over ocean ASTER observations, J. Geophys. Res. Atmos., 115(22), doi:10.1029/2009JD013520, 2010.

Kotarba, A. Z.: Impact of Moderate Resolution Imaging Spectroradiometer (MODIS) cloud mask interpretation on cloud amount estimation, J. Geophys. Res., 120(17), 8971–8986, doi:10.1002/2015JD023277, 2015.

Kraatz, S., Khanbilvardi, R. and Romanov, P.: A comparison of MODIS/VIIRS cloud masks over ice-bearing river: On achieving consistent cloud masking and improved river ice mapping, Remote Sens., doi:10.3390/rs9030229, 2017.

Krijger, J. M., Van Weele, M., Aben, I. and Frey, R.: Technical Note: The effect of sensor resolution on the number of cloud-free observations from space, Atmos. Chem. Phys., doi:10.5194/acp-7-2881-2007, 2007.

Li, J., Liu, C. Y., Huang, H. L., Schmit, T. J., Wu, X., Menzel, W. P. and Gurka, J. J.: Optimal cloud-clearing for AIRS radiances using MODIS, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2005.847795, 2005.

Liu, Y., Key, J.R., Frey, R.A., Ackerman, S.A., and Menzel, W.P.: Nighttime polar cloud detection with MODIS, Remote Sens., doi:10.1016/j.rse.2004.06.004, 2004

Melchiorre, A., Boschetti, L. and Roy, D. P.: Global Evaluation of the Suitability of MODIS-Terra Detected Cloud Cover as a Proxy for Landsat 7 Cloud Conditions, Remote Sens., doi:10.3390/rs12020202, 2020.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C. and Frey, R. A.: The MODIS cloud products: Algorithms and examples from terra, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2002.808301, 2003.

Ramanathan, V. and Kiehl, J. T.: The Radiative Forcing due to Clouds and Water Vapour, in Frontiers of Climate Modelling, pp. 119–151., 2006.

Remer, L. A., Mattoo, S., Levy, R. C., Heidinger, A., Pierce, R. B. and Chin, M.: Retrieving aerosol in a cloudy environment: Aerosol product availability as a function of spatial resolution, Atmos. Meas. Tech., doi:10.5194/amt-5-1823-2012, 2012.

Sassen, K., Wang, Z. and Liu, D.: Cirrus clouds and deep convection in the tropics: Insights from CALIPSO and CloudSat, J. Geophys. Res. Atmos., 114, D00H06, doi:10.1029/2009JD011916.

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Glob. Planet. Change, doi:10.1016/j.gloplacha.2011.03.004, 2011.

Sèze, G., Pelon, J., Derrien, M., Le Gléau, H. and Six, B.: Evaluation against CALIPSO lidar observations of the multi-geostationary cloud cover and type dataset assembled in the framework of the Megha-Tropiques mission, Q. J. R. Meteorol. Soc., doi:10.1002/qj.2392, 2015.

Shang, H., Letu, H., Nakajima, T. Y., Wang, Z., Ma, R., Wang, T., Lei, Y., Ji, D., Li, S. and Shi, J.: Diurnal cycle and seasonal variation of cloud cover over the Tibetan Plateau as determined from Himawari-8 new-generation geostationary satellite data, Sci. Rep., doi:10.1038/s41598-018-19431-w, 2018.

Shupe, M. D. and Intrieri, J. M.: Cloud radiative forcing of the Arctic surface: The influence of cloud properties, surface albedo, and solar zenith angle, J. Clim., doi:10.1175/1520-0442(2004)017<0616:CRFOTA>2.0.CO;2, 2004.

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., L'Ecuyer, T. and Lebsock, M.: Cloudsat and calipso within the a-train: Ten years of actively observing the earth system, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-D-16-0324.1, 2018.

Sun-Mack, S., Minnis, P., Chen, Y., Kato, S., Yi, Y., Gibson, S. C., Heck, P. W. and Winker, A. D. M.: Regional apparent boundary layer lapse rates determined from CALIPSO and MODIS data for cloud-height determination, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-13-081.1, 2014.

Tan, I. and Storelvmo, T.: Evidence of Strong Contributions From Mixed-Phase Clouds to Arctic Climate Change, Geophys. Res. Lett., doi:10.1029/2018GL081871, 2019.

Tan, S., Zhang, X. and Shi, G.: MODIS cloud detection evaluation using CALIOP over polluted eastern China, Atmosphere (Basel)., doi:10.3390/atmos10060333, 2019.

Trenberth, K. E., Fasullo, J. T. and Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 90(3), 311–323, doi:10.1175/2008BAMS2634.1, 2009.

Várnai, T. and Marshak, A.: MODIS observations of enhanced clear sky reflectance near clouds, Geophys. Res. Lett., doi:10.1029/2008GL037089, 2009.

Vaughan, M. A., Powell, K. A., Kuehn, R. E., Young, S. A., Winker, D. M., Hostetler, C. A., Hunt, W. H., Liu, Z., Mcgill, M. J. and Getzewich, B. J.: Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements, J. Atmos. Ocean. Technol., doi:10.1175/2009JTECHA1228.1, 2009.

Vavrus, S.: The impact of cloud feedbacks on Arctic climate under Greenhouse forcing, J. Clim., doi:10.1175/1520-0442(2004)017<0603:TIOCFO>2.0.CO;2, 2004.

Wang, T., Fetzer, E. J., Wong, S., Kahn, B. H. and Yue, Q.: Validation of MODIS cloud mask and multilayer flag using CloudSat-CALIPSO cloud profiles and a cross-reference of their cloud classifications, J. Geophys. Res. Atmos., 121(19), 11620–11635, doi:10.1002/2016JD025239, 2016.

Weinzierl, B., Ansmann, A., Prospero, J. M., Althausen, D., Benker, N., Chouza, F., Dollner, M., Farrell, D., Fomba, W. K., Freudenthaler, V., Gasteiger, J., Groß, S., Haarig, M., Heinold, B., Kandler, K., Kristensen, T. B., Mayol-Bracero, O. L., Müller, T., Reitebuch, O., Sauer, D., Schäfler, A., Schepanski, K., Spanu, A., Tegen, I., Toledano, C. and Walser, A.: The Saharan aerosol long-range transport and aerosol-cloud-interaction experiment: Overview and selected highlights, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-D-15-00142.1, 2017.

Wilson, A. M., Parmentier, B. and Jetz, W.: Systematic land cover bias in Collection 5 MODIS cloud mask and derived products - A global overview, Remote Sens. Environ., doi:10.1016/j.rse.2013.10.025, 2014.

Wilson, M. J. and Oreopoulos, L.: Enhancing a simple MODIS cloud mask algorithm for the ~~landsat~~ Landsat ~~data~~ Data ~~continuity~~ Continuity ~~mission~~Mission, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2012.2203823, 2013.

Winker, D., Vaughan, M. and Hunt, B.: The CALIPSO mission and initial results from CALIOP, in Lidar Remote Sensing for Environmental Monitoring VII., 2006.

Xie, Y., Qu, J. J. and Xiong, X.: Improving the CALIPSO VFM product with Aqua MODIS measurements, Remote Sens. Lett., doi:10.1080/01431161003720387, 2010.

Zhang, X., Tan, S. C., Shi, G. Y. and Wang, H.: Improvement of MODIS cloud mask over severe polluted eastern China, Sci. Total Environ., doi:10.1016/j.scitotenv.2018.10.369, 2019.

Zhao, G. and Di Girolamo, L.: Cloud fraction errors for trade wind cumuli from EOS-Terra instruments, Geophys. Res. Lett., doi:10.1029/2006GL027088, 2006.

Zuluaga, M. D., Webster, P. J. and Hoyos, C. D.: Variability of aerosols in the tropical Atlantic Ocean relative to African Easterly Waves and their relationship with atmospheric and oceanic environments, J. Geophys. Res. Atmos., doi:10.1029/2011JD017181, 2012.

**Tables**

**Table 1**. Agreement in cloud detection between MODIS and CALIOP (% of all cases). Overall accuracy (given in brackets) refers to ‘confident clear’ and ‘confident cloudy’ detections. In other cases, ‘confident clear’ and probably clear’ were merged, as were ‘probably cloudy’ and ‘confident cloudy’.

| | | **MODIS** | | | | **Overall accuracy** |
|---|---|---|---|---|---|---|
| | | confident clear | probably clear | probably cloudy | confident cloudy | |
| | | *Day+Night* | | | | |
| CALIOP | clear | 22.7 | 5.4 | 1.9 | 3.1 | 86.7% (77.3%) |
| | cloudy | 6.2 | 2.1 | 3.9 | 54.6 | |
| | | *Day only* | | | | |
| CALIOP | clear | 25.5 | 5.1 | 1.7 | 3.2 | 89.4% (81.9%) |
| | cloudy | 3.7 | 2.0 | 2.4 | 56.4 | |
| | | *Night only* | | | | |
| CALIOP | clear | 20.2 | 5.7 | 2.2 | 3.0 | 84.2% (73.3%) |
| | cloudy | 8.5 | 2.1 | 5.2 | 53.1 | |

**Table 2**. Global cloud fractions for MODIS cloud mask classes derived from CALIOP observations ('This study'), and used in the operational MODIS Science Team Level 3 product ('Operational'). Numbers in brackets refer to class frequency ($n$ = 33, 793, 648).

| **Source of cloud fractions for cloud mask classes** | | **Cloud fractions (%) for MODIS cloud mask class** (class frequency, % of *n*) | | | |
|---|---|---|---|---|---|
| | | **confident clear** (28.9%) | **probably clear** (7.5%) | **probably cloudy** (5.8%) | **confident cloudy** (57.8%) |
| Operational | Day+Night | 0.0 | 0.0 | 100.0 | 100.0 |
| This study | Day+Night | 21.5 | 27.7 | 66.6 | 94.7 |
| | Day only | 12.7 | 28. 4 | 58.4 | 94.7 |
| | Night only | 29.5 | 27.1 | 70.7 | 94.7 |

**Table 3**. CALIOP-based ~~Cloud~~ cloud fractions for MODIS cloud mask classes, calculated individually for each MODIS algorithm path. Note that more paths are reported here than in the MODIS project. Snow-covered ocean, land, desert and coast constitute a single path in the operational algorithm, while here they are reported individually to highlight how snow impacts the results. The sunglint path is missing as CALIOP does not sample over sunglint areas. Numbers in brackets refer to how frequently (% of *n*) a given algorithm path was executed, *n* = 33, 793, 648.

| **Cloud masking algorithm path** | | | | **CALIOP-based c~~C~~loud fractions [%] for MODIS cloud mask class** | | | |
|---|---|---|---|---|---|---|---|
| | | | | **confident clear** | **probably clear** | **probably cloudy** | **confident cloudy** |
| Day (47.2) | Snow-covered (5.5) | Land | (0.2) | 13.8 | 67.0 | 56.0 | 97.6 |
| | | Desert | (3.9) | 12.6 | 32.6 | 71.8 | 96.6 |
| | | Coast | (0.2) | 15.3 | 55.5 | 61.8 | 93.8 |
| | | Ocean | (1.1) | 20.5 | 76.3 | 69.7 | 88.6 |
| | ~~No~~ Snow-free (41.7) | Land | (6.7) | 15.6 | 32.3 | 63.9 | 93.4 |
| | | Desert | (3.4) | 9.1 | 19.1 | 45.5 | 90.0 |
| | | Coast | (1.6) | 19.0 | 33.8 | 59.8 | 93.0 |
| | | Ocean | (30.1) | 10.5 | 28.4 | 54.5 | 95.2 |
| Night (52.8) | Snow-covered (15.8) | Land | (2.6) | 31.4 | 65.0 | 80.9 | 93.9 |
| | | Desert | (4.7) | 34.3 | 65.3 | 75.9 | 86.4 |
| | | Coast | (0.9) | 29.8 | 60.8 | 75.0 | 93.7 |
| | | Ocean | (7.6) | 49.7 | 73.7 | 82.5 | 96.8 |
| | ~~No~~ Snow-free (37.0) | Land | (5.4) | 8.0 | 25.6 | 68.5 | 97.7 |
| | | Desert | (2.6) | 8.2 | 23.5 | 55.8 | 95.4 |
| | | Coast | (0.9) | 10.9 | 23.0 | 60.9 | 96.4 |
| | | Ocean | (28.1) | 22.9 | 22.4 | 61.8 | 94.6 |

**Table 4**. Global cloud amount (%) calculated with different 'best guess' interpretations of the MODIS cloud mask product. Only MODIS IFOV collocated with CALIOP are considered.

| | **CALIOP** | **MODIS cloud mask interpretation scenario** | | |
|---|---|---|---|---|
| | | Only 'confident cloudy' is 'cloudy' | 'Confident clear' and 'probably clear' are clear, while the rest is cloudy | Only 'confident clear' is 'clear' |
| | *Global* | | | |
| Day+Night | 66.7 | 57.7 | 63.5 | 71.0 |
| Day | 64.3 | 59.3 | 63.4 | 70.6 |
| Night | 68.9 | 56.1 | 63.5 | 71.3 |
| | *Polar regions (latitudes above 60ºN/S)* | | | |
| Day+Night | 66.9 | 50.5 | 57.6 | 61.0 |
| Day | 64.8 | 59.0 | 62.6 | 66.4 |
| Night | 68.5 | 44.1 | 53.9 | 57.0 |
| | *Equatorial region (latitudes between 30ºN and 30ºS)* | | | |
| Day+Night | 59.8 | 52.8 | 58.0 | 67.4 |
| Day | 56.2 | 49.9 | 54.8 | 65.4 |
| Night | 63.5 | 55.7 | 61.2 | 69.4 |
| | *Mid-latitudes (between polar and equatorial)* | | | |
| Day+Night | 73.3 | 68.9 | 74.2 | 83.6 |
| Day | 72.0 | 69.1 | 72.7 | 79.1 |
| Night | 74.6 | 68.8 | 75.7 | 87.9 |

**Figures**

Figure 1. Observations declared ‘confident clear’ or ‘probably clear’ by the MODIS cloud masking algorithm, but identified as ‘cloudy’ by CALIOP.

Figure 2. Observations declared ‘confident cloudy’ or ‘probably cloudy’ by the MODIS cloud masking algorithm, but identified as ‘clear’ by CALIOP.

Figure 3. CALIOP-based cloud fraction for MODIS cloud mask classes.

Figure 4. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘nighttime, snow(ice)-covered ocean’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 5. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘nighttime snow-free land’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 6. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘daytime, snow-free land’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 7. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘daytime, snow-free ocean’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 8. Difference between the MODIS Science Team (MODIS ST) Level 3 cloud amount product, and cloud amount calculated with the cloud fractions found in this study. Positive values indicate that the MODIS operational product overestimates cloud amount (with respect to CALIOP), while negative values indicate a MODIS underestimate. All MODIS observations refer to the full swath, not only those collocated with CALIOP.

## Supplementary Online Material

Figure S1. The average cloud-aerosol discrimination (CAD) score for CALIOP cloud data used in the study. Maps show spatial variation in the CAD score during the day (b), at night (c), and regardless of the time of the day (a). These plots demonstrate the high stability of CAD scores at various latitudes during the day (d) and at night (e).

Figure S2. Overall accuracy of MODIS cloud detection as a function of the temporal (a, c) and spatial (b, d) separation of MODIS and CALIOP IFOVs. Top plots show the frequency of observations for individual time (a) and distance (b) ranges, while bottom plots report accuracy for these ranges. MODIS detections are validated using CALIOP cloud profiles as a reference. Accuracy is defined as the ratio of MODIS true detections (true positive and true negative) to all MODIS observations (see Table 1 in the main text for details).

# Calibration of global MODIS cloud amount using CALIOP cloud profiles

Andrzej Z. Kotarba[1]

[1]Space Research Centre, Polish Academy of Sciences, 00-716 Warsaw, Poland

*Correspondence to*: Andrzej Z. Kotarba (akotarba@cbk.waw.pl)

**Abstract.** The Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection procedure classifies instantaneous fields of view (IFOV) as either ‘confident clear’, ‘probably clear’, ‘probably cloudy’, or ‘confident cloudy’. The cloud amount calculation requires quantitative cloud fractions to be assigned to these classes. The operational procedure used by the MODIS Science Team assumes that ‘confident clear’ and ‘probably clear’ IFOV are cloud-free (cloud fraction 0%), while the remaining categories are completely filled with clouds (cloud fraction 100%). This study demonstrates that this ‘best guess’ approach is unreliable, especially on a regional/ local scale. We use data from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument flown on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) mission, collocated with MODIS/ Aqua IFOV. Based on 33,793,648 paired observations acquired in January and July 2015, we conclude that actual cloud fractions to be associated with MODIS cloud mask categories are 21.5%, 27.7%, 66.6%, and 94.7%. Spatial variability is significant, even within a single MODIS algorithm path, and the operational approach introduces uncertainties of up to 30% of cloud amount, notably in polar regions at night, and in selected locations over the northern hemisphere (e.g. China, the north-west coast of Africa, and eastern parts of the United States). Consequently, applications of MODIS data on a regional/ local scale should first assess the extent of the uncertainty. We suggest using CALIPSO-based cloud fractions to improve MODIS cloud amount estimates. This approach can also be used for MODIS/ Terra data, and other passive cloud imagers, where the footprint is collocated with CALIPSO.

## 1 Introduction

Cloud plays a key role in distributing solar energy in the Earth’s atmosphere (Trenberth et al., 2009). Consequently, research into the present and future state of the climate system requires accurate information about cloud amount. Depending on its frequency and physical properties, cloud can both heat (greenhouse effect: +30 Wm$^{-2}$) and cool (albedo effect: –48 Wm$^{-2}$) the atmosphere. Their net effect on the planetary radiation budget is negative, meaning the Earth would be warmer if all cloud disappeared (Ramanathan and Kiehl, 2006).

The Global Climate Observing System identifies 13 Essential Climate Variables. This set of critical environmental parameters characterize the Earth’s climate (Hollmann et al., 2013); they not only include cloud properties, but also highlight that our knowledge of cloud relies largely on satellite remote sensing. Satellite cloud climatology starts with a cloud mask. The aim is to decide whether cloud is present in a sensor's instantaneous field of view (IFOV), or whether it is cloud free. Input data includes at-sensor registered radiances, along with other auxiliary information that aims to maximize cloud detection.

Efficient cloud detection algorithms have to consider the technical limitations of sensors, available computing power, and environmental factors such as the background (e.g. water, land, snow) and solar illumination (day and night). The resulting cloud mask takes the form of a map that divides IFOV into at least two categories: 'cloud free', and 'cloud contaminated' (or 'cloud filled'). Many masking algorithms introduce additional categories in order to reflect the level of uncertainty in cloud detection (Derrien and Le Gléau, 2005; Dybbroe et al., 2005; Kopp et al., 2014).

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a cloud imaging instrument that is flown onboard NASA's polar orbiting satellites: Terra and Aqua. Circling the Earth in the morning orbit (10:30 local solar time; Terra), and afternoon orbit (13:30 local solar time, Aqua), these twin sensors provide a global picture of cloud four times each day, at 1 km/pixel resolution (Guenther et al., 2002; Platnick et al., 2003). With 36 spectral channels, continuous correction for orbital drift, and precisely-calibrated detectors, MODIS has set a new standard in cloud remote sensing, and is still considered to be a state-of-the-art cloud imager, despite being launched in 1999 (Terra), and 2002 (Aqua).

MODIS's cloud detection scheme results in four cloud mask categories: 'confident cloudy', 'probably cloudy', 'probably clear', and 'confident clear' (Frey et al., 2008). The fact that these classes are presented as qualitative, text-based labels rather than a numeric probability causes the technical problem of how to quantitatively interpret these labels. A numeric interpretation is mandatory when instantaneous observations (Level 2 products) are aggregated spatially and/ or temporally to provide climatological information such as mean monthly cloud amount (Level 3 products).

The procedure implemented by NASA's MODIS Science Team (hereinafter the 'standard' or 'operational' procedure) is to assume that IFOV declared 'confident cloudy' and 'probably cloudy' are, in fact, 100% cloud filled, while 'confident clear' and 'probably clear' are completely cloud free (cloud fraction of 0%) (Hubanks et al., 2008). The approach is widely used whenever there is a need to make a binary distinction between cloudy and cloud-free pixels – e.g., Gao et al. (2008), Remer et al. (2012), Wilson and Oreopoulos (2013), Wilson, Parmentier, and Jetz (2014), Kraatz, Khanbilvardi, and Romanov (2017), Gomis-Cebolla, Jimenez, and Sobrino (2020).

However, since the MODIS Science Team (ST) approach is only a 'best guess', alterative assumptions are also used. For instance, it can be assumed that only 'confident cloudy' pixels are 'cloudy', while all remaining classes are 100% cloud free. Similarly, only 'confident clear' detections can be considered as truly cloud-free, while all other classes are assumed to be 100% cloud filled (Li et al., 2005). Krijger et al. (2007) argue that the latter approach leads to the false detection of small clouds, while cloud is frequently overlooked if the first method is applied. Another approach is simply to exclude 'probably clear' and 'probably cloudy' detections from the analysis. This strategy was adopted by Chan and Comiso (2013), whose work was based on only 'confident clear' and 'confident cloudy' categories of MODIS data.

Quantitative studies have shown that only considering the 'confident cloudy' class as cloudy may be more consistent with other cloud data such as Landsat observations (Melchiorre et al., 2020), or visual observations at meteorological stations

(Kotarba, 2015). On the other hand,, Fontana et al. (2013) compared MODIS data with ground-based observations in Switzerland (4 stations, 12 years of data), and found that results varied from station to station.
The theoretical range of uncertainty related to various interpretations of the MODIS cloud mask was investigated by Kotarba (2015). The latter study found that the global cloud amount estimates may differ by up to 14%, depending on whether only ‘confident cloudy’ detections are considered to be ‘cloudy’, or whether the definition is extended to include intermediate classes. The discrepancy was found to increase by up to 40–60% regionally, suggesting that MODIS cloud estimates are very uncertain in these areas. Such a wide range of uncertainty makes it difficult to run reliable studies on the climate system.
Neither the MODIS ST standard procedure, nor any other ‘best guess’ variants have been validated on a global scale. Most importantly: no research-based, objective alternatives to the 0/0/100/100 interpretation have been suggested. This study addresses this problem. Specifically, it provides global cloud fractions based on quantitative analysis of CALIOP lidar observations.
CALIOP (the Cloud-Aerosol Lidar with Orthogonal Polarization) is a cloud profiling instrument flown onboard the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) spacecraft. Launched in 2016, CALIPSO flies in close formation with the Aqua satellite, therefore both instruments – MODIS and CALIOP – sample the same fragment of the atmosphere tenths of seconds apart (Stephens et al., 2018). In this study, CALIPSO data is considered as ground truth. This is because CALIOP is an active remote sensing instrument, which means that it can sample the atmosphere during the day and at night with comparable sensitivity. Imaging radiometers (such as MODIS) perform less effectively at night, when solar channels are missing. Furthermore, the use of short wavelengths makes CALIOP very sensitive to cloud of low optical thickness (e.g. sub-visual cirrus) that is often missed by imagers (Ackerman et al., 2008).
In the following sections we seek to answer the questions: 1) What quantitative cloud fractions (based on CALIOP observations) should be applied to MODIS thematic cloud mask classes? and 2) What uncertainties in global cloud amount are introduced by the MODIS ST standard procedure? Finally, we evaluate whether the standard procedure for calculating the MODIS Level 3 cloud amount is reliable.

## 2. Data & Methods

### 2.1 MODIS data

The MODIS cloud detection scheme is based on thresholds that are applied to brightness temperature (thermal bands), and reflectance (solar channels), derived from observations in 22 spectral bands ranging from 0.66 µm to 13.9 µm. Ackerman et al. (1998), Frey et al. (2008), and Baum et al. (2012) provide very detailed descriptions of the cloud masking procedure. The general concept is as follows.
The algorithm executes a series of tests, each of which results in a confidence level (ranging from 0 to 1) that a particular IFOV is cloud free. Tests to detect similar cloud types are grouped. The lowest confidence level for a test within a group is set as the confidence level for the whole group. Confidence levels for groups are then multiplied to determine the final confidence level

(Q). Finally, the IFOV is assigned to one of four cloud mask classes: ‘confident clear’ (Q > 0.99), ‘probably clear’ (Q > 0.95), ‘probably cloudy’ (Q > 0.66), or ‘confident cloudy’ (Q ≤ 0.66). The exact number of spectral tests varies from a few to over a dozen, depending on the path through the algorithm. Paths reflect different environmental conditions, and are introduced to maximize success. Dedicated sets of spectral tests are executed for land, ocean, desert and coastal areas, for both day and night conditions. The presence of snow and/ or ice is taken into account, as is sunglint over oceans. Separate thresholds have been introduced for polar regions, which are defined as land and ocean within 30 degrees of each pole.

Cloud detection results are stored in the 48-bit ‘Cloud Mask’ product, codenamed MYD35 (Aqua) and MOD35 (Terra) following the MODIS nomenclature. In this study, we evaluated the latest version of MYD35 (Collection 061) data, available in the form of 5-minute granules, at 1 km per pixel spatial resolution (at nadir), with native satellite projection. Each MYD35 file is accompanied by a MYD03 ‘Geolocation file’ product, that stores longitude and latitude information for individual cloud mask IFOV.

## 2.2 CALIOP data

CALIOP operates at 532 nm and 1064 nm. The instrument’s pencil-like beam only scans locations along the satellite’s ground track, as a trade-off for information on the vertical structure of cloud/ aerosols. Its spatial resolution is a function of the satellite’s altitude. Resolution is finest – 0.333 km horizontal, 30 m vertical – in the troposphere, up to 8.2 km. Between 8.2 km and 20.2 km, vertical resolution falls to 60 m, and horizontal sampling to 1 km. Between 20.2 km and 30.1 km, data are even coarser: 1.667 km horizontal and 180 m vertical resolution. Higher in the atmosphere (30.1 km to 40.0 km) horizontal resolution decreases to 5 km, while vertical resolution is 300 m (Hunt et al., 2009; Winker et al., 2006).

CALIOP detects cloud by applying thresholds to 532 nm attenuated scattering ratios. The aim is to separate the cloud signal from the clear air background (molecular scattering), aerosols, and instrument noise. The algorithm calculates cloud base height, cloud top height, and – as a consequence – the number of cloud layers within a profile. Up to 10 layers can be reported. The procedure is fully automatic (Vaughan et al., 2009). The output is stored in the Level 2 Cloud Layer Data product, available at 333 m, 1 km, and 5 km along-track sampling intervals. Here, we use the 1 km interval (version 4.20; CAL_LID_L2_01kmCLay-Standard-V4-20), as its resolution matches the spatial resolution of the MODIS cloud mask. Furthermore, 1 km is the highest available level of detail for CALIOP data within the troposphere.

In order to use the CALIOP product to evaluate MODIS data, 3-dimensional cloud layer data was reduced to column-integrated, binary cloud/ no cloud information. Specifically, we focused on the ‘Number_Layers_Found’ parameter provided in the CAL_LID_L2 product. ‘No cloud’ was recorded when the latter variable was set to 0 (i.e. zero layers found), and ‘cloud’ otherwise (i.e. at least one layer was reported). Geolocation was based on longitude and latitude arrays included in the product at 1 km spatial resolution.

In some cases, cloud and aerosol can appear similar to CALIOP. The cloud-aerosol discrimination (CAD) score, which is a numerical index stored in the CAL_LID_L2 product, provides information about the algorithm’s uncertainty in separating cloud and aerosol. In the case of cloud, CAD values vary between 0% (it is unclear whether aerosol or cloud was observed) and 100% (cloud detected with the highest confidence). The index is calculated for each cloud layer found in the CALIOP atmospheric profile. Since our study focuses on column-integrated information of cloud presence, we selected the highest CAD value within a profile. Statistics for January and July 2015 showed that 95.6% of considered CALIOP observations were characterized by a CAD score of at least 70%, while it was below 20% for only 1.5% of data. Therefore, the selected CALIOP data can be considered as a reliable reference for MODIS. See Supplementary Online Materials, Fig. S1 for more detailed statistics about CAD scores.

### 2.3 Matching CALIOP and MODIS data

Matching CALIOP data is a well-established method for the calibration/ validation of atmospheric products from various space missions. It has already been widely used for MODIS/ Aqua (Baum et al., 2012; Holz et al., 2009; Sun-Mack et al., 2014; Wang et al., 2016; Xie et al., 2010), and sensors flown onboard Suomi-NPP, NOAA, and MetOp polar orbiting spacecraft, which occasionally synchronize their orbital configuration with CALIPSO (Hutchison et al. 2014; Heidinger et al. 2012; Karlsson and Johansson 2013; Karlsson and Dybbroe 2010). CALIPSO also passes within the field of view of other geostationary satellites, and CALIOP data is used to assess their atmospheric products (Sèze et al., 2015; Shang et al., 2018).

In this study, Aqua/ MODIS data for January and July 2005 were paired with corresponding CALIPSO/ CALIOP observations. The matching procedure selected a MODIS IFOV and compared it with the corresponding CALIOP profile (where the geometric centre was within the selected MODIS IFOV). Although very straightforward, the procedure was time-consuming since a single MODIS granule contains ~2030 IFOV, and a full day of Aqua observations produces 288 granules.

The final database consisted of 33,793,648 MODIS–CALIOP paired observations. Average spatial separation between the centres of MODIS and CALIOP IFOV was 418 m, and 19% had a separation of less than 250 m. Temporal differences between lidar and imager observations were determined using the spacecrafts' on-orbit separation, and ranged from 60 sec to 97 sec (81 sec on average). Our dataset excluded one MODIS cloud mask processing path: sunglint. This was because CALIPSO's orbit has been intentionally designed to avoid sunglint areas, in order to avoid the lidar being 'blinded' by solar reflection from the ocean.

Our empirical calculation of cloud fraction in each MODIS cloud mask class was based on the ratio of CALIOP cloudy detections to all detections within a class. A perfect MODIS cloud detection algorithm would categorise a 0% cloud fraction as 'confident clear', while a 100% cloud fraction would be categorised as 'confident cloudy'.

## 3 Results

### 3.1 Misdetection of cloud and clear sky by MODIS

The first key point that emerged from the matched MODIS–lidar observations was the accuracy of MODIS cloud detections. Overall accuracy in January and July 2015, compared to reference CALIOP data, was 89.4% during the day, and 84.2% at night (Tab. 1). This statistic assumes that ‘probably clear’ detections are merged with ‘confident clear’, and ‘probably cloudy’ detections are combined with ‘confident cloudy’. If a less tolerant approach is applied, i.e. only ‘confident clear’ and ‘confident cloudy’ detections are considered (‘probably clear’/ ‘probably cloudy’ classes are interpreted as misdetections), overall accuracy fell to 81.9% during the day, and 73.3% at night.

Clouds missed by MODIS, but detected by CALIOP were most frequent during the polar night, regardless of the hemisphere (Fig. 1c, d). Up to 40% of MODIS ‘confident clear’ and ‘probably clear’ detections were found to be incorrect around Antarctica in July, and the Arctic in January. Globally, daytime (Fig. 1a, b) misdetections were around half of those at night. They only exceeded 30% locally, and polar regions were less affected. A notable observation was July in the northern hemisphere, where only a few small regions of misdetection were observed. The analysis highlighted a belt of relatively higher frequency misdetections (15–25%) in the equatorial zone; here the magnitude of the effect was similar for both day and night.

Only a few occasions were identified when over 10–15% of MODIS ‘confident cloudy’ and ‘probably cloudy’ observations were identified as cloud-free by CALIOP (Fig. 2). Further analysis showed that although false detection was rare in polar regions, it was significant in specific regions of the northern hemisphere. North-east China emerged as the most problematic area (Fig. 2a). Here, 50–70% of MODIS ‘confident cloudy’ and ‘probably cloudy’ detections were cloud-free according to CALIOP. However, this high rate of false detection was only observed in January, and only during the day.

### 3.2 Cloud fraction for cloud mask classes

Our empirical calculation of cloud fraction in each MODIS cloud mask class was based on the ratio of CALIOP cloudy detections to all detections within a class. A perfect MODIS cloud detection algorithm would categorise a 0% cloud fraction as ‘confident clear’, while a 100% cloud fraction would be categorised as ‘confident cloudy’.

On average, one fifth of MODIS ‘confident clear’ detections were found to be cloudy by CALIOP. Consequently, the average cloud fraction for this class was 21.5%, instead of the theoretically expected 0% (Tab. 2). At night, the fraction was over twice the daytime value (29.5% compared to 12.7%). On the other hand, pixels flagged by the MODIS algorithm as ‘confident cloudy’ were, almost always, contaminated with some cloud, and were sometimes cloud-filled. Regardless of the time of day, the actual CALIOP-based cloud fraction for ‘confident cloudy’ detections was close to 100%, reaching 94.7%.

MODIS intermediate classes constituted 13.3% of all detections. CALIOP cloud fractions were 27.7% and 66.6% for ‘probably clear’, and ‘probably cloudy’ classes respectively. The statistics revealed a difference of up to 17% between day and night conditions, and it was especially small for the ‘probably clear’ class (1.3%) and the ‘confident cloudy’ class (daytime had no impact at all).

The parameters reported in Table 2 are global averages (means), and spatial diversity was observed. Differences were smallest for the ‘confident cloudy’ class – both during the day (Fig. 3a) and at night (Fig. 3b) – and the CALIOP-based cloud fraction exceeded 90% at almost every location. North-east China, the southern Arabian Peninsula and Eastern Antarctica were the only significant exceptions; here cloud fraction decreased to 50–70%.

The cloud fraction distribution was homogeneous for the ‘confident clear’ class, however, only during the day (Fig. 3g). At night (Fig. 3h) it increased substantially in polar regions, especially over the oceans of the southern hemisphere, along the coast of Antarctica. Unlike polar regions, no noticeable day/ night difference was observed for mid- and low-latitudes (< 10%); here, the cloud fraction was very low (< 20%), and very few MODIS ‘confident clear’ detections were identified as cloudy by CALIOP.

Among the MODIS intermediate classes, ‘probably clear’ differed most from CALIOP-based data. First, a high cloud fraction (> 70%) was observed at night along the equator and in polar regions (both oceanic and continental; Fig. 3f). At mid-latitudes the cloud fraction for MODIS ‘probably clear’ observations was relatively low (< 10–20%). This pattern was inverted during the day (Fig. 3e). At this time a higher (50–75%) cloud fraction was noted for mid-latitudes, and over parts (typically land) of the polar regions.

### 3.3 Cloud fraction as a function of the algorithm path

The MODIS cloud detection algorithm distinguishes between day and night (Tab. 1, Tab. 2), and four types of background (land, desert, coast, ocean), each of which can be either snow-covered or snow-free. CALIOP-based cloud fractions for all algorithm paths are reported in Table 3. These values give a detailed understanding of MODIS cloud detection results. Data are given for each class of the MODIS cloud mask separately. In our study, we structured the paths through the algorithm in more detail. Snow-covered conditions were considered for land, desert, ocean and coast separately, while in the MODIS algorithm they are grouped as snow/ ice. This greater level of detail allowed us to observe how the presence of snow impacted the cloud mask over different backgrounds.

Per-class estimates of cloud fraction were very consistent for all algorithm paths for the ‘confident cloudy’ category (Tab. 3). Final values ranged between 97.7% (night, snow-free, land) and 86.4% (night, snow-covered, desert), and were close to the standard Level 3 assumption of 100%. This finding contrasted with cloud fractions found for the ‘confident clear’ category. While MODIS recorded cloud-free conditions, CALIOP data revealed that the actual cloud fraction ranged from 8.0% (night, snow-free, land) to 49.7% (night, snow-covered, ocean).

The combination of night, an oceanic background and snow-covered (or sea ice) constituted the ‘most cloudy’ scenario (Fig. 4). Here, a very high cloud fraction was found for not only the ‘confident cloudy’ category (96.8%, Fig. 4a), but also all remaining classes: 82.5% (‘probably cloudy’; Fig. 4b), 73.3% (‘probably clear’; Fig. 4c) and, surprisingly, up to 49.7% for ‘confident clear’ (Fig. 4d).

CALIOP-based cloud fractions were most consistent with the standard Level 3 interpretation for snow-free land at night (Fig. 5). Here, the cloud fraction for ‘confident clear’ was low (8.0%; Fig. 5d), and very high for ‘confident cloudy’ (97.7%; Fig. 5a). At the same time, intermediate classes were well-separated: 68.5% for ‘probably cloudy’ (Fig. 5b), and 25.6% for ‘probably clear’ (Fig. 5c). Globally, no significant difference was found for cloud fraction values for the night/ snow-free/ land algorithm path. A small exception was noted for the ‘probably clear’ type, where the cloud fraction was 10–30% higher in the tropics compared to the rest of the world.

A similar spatial distribution of CALIOP-based cloud fractions was observed for snow-free land during the day – the scenario of particular interest for land/ vegetation remote sensing with MODIS. The two notable differences were related to 'probably clear' (Fig. 6c) and 'confident clear' categories (Fig. 6d). The latter occurred twice as often during the day (15.6%) than at night (8.0%). Similarly, cloud was more frequent in the 'probably clear' class. However, this was only found in the tropics and at high latitudes, which mirrored a zonal pattern that was only weakly seen at night.

As ice-free oceans represent the majority of Earth's surface, cloud detection over ocean is the most frequent algorithm path. Daytime conditions make detection easier (due to the availability of solar channels). Under such circumstances, CALIOP detected cloud in 10.5% of MODIS's 'confident clear' observations (Fig. 7d), and confirmed 95.2% of 'confident cloudy' detections (Fig. 7a). Cloud fractions for intermediate classes (daytime over ice-free ocean) were 54.5% and 28.4% for 'probably cloudy' (Fig. 7b) and 'probably clear' (Fig. 7c) categories, respectively. 'Probably clear' was the only class where there was a clear latitude-dependent cloud fraction: values increased by 30–60% along a path ~30–40 degrees north/ south.

## 4 Discussion

Our investigation of spatially and temporally collocated MODIS (cloud imager) and CALIOP (cloud profiling lidar) observations for January and July 2015 revealed that MODIS Collection 061 global cloud amount estimates are imperfect. During the generation of the Level 2 product, the masking algorithm fails to accurately report cloud over polar regions, and over selected locations at lower latitudes. Consequently (as discussed in this section) the Level 3 product generation underestimates cloud fractions for cloud mask classes in numerous regions. The reliability of these results depends on several factors, most notably the spatial and temporal accuracy of Aqua/ MODIS and CALIPSO/ CALIOP collocation.

Temporal differences between Aqua and CALIPSO observations varied from 60 seconds to 97 seconds. In this time, cloud can develop and move, introducing the risk that CALIOP observes a different state of the atmosphere compared to MODIS. Várnai and Marshak (2009) evaluated the problem by comparing MODIS reflectance with that collected by the Wide Field Camera. The latter is an imaging instrument flown onboard CALIPSO, along with CALIOP. They found that for low cloud, radiance differed only slightly over 72 seconds, and it was reasonable to ignore any discrepancies when focusing on aerosol properties (they gave no particular conclusions for cloud). In order to test how sensitive our results were to the time shift, we calculated the overall accuracy of the cloud detection algorithm as a function of the time between Aqua and CALIPSO passes. The results were very consistent: despite the shift, accuracy remained at 86.7±0.1%. This finding confirmed that the temporal separation between Aqua and CALIPSO had no significant impact on the results of our study.

Another potential source of uncertainty is the geometric mismatch between MODIS and CALIOP IFOV. They are not aligned perfectly: 66% of collocated IFOV were separated by less than 0.5 km, and 82% by less than 0.6 km. Similar statistics – 75% and 93%, respectively – were found by Wang et al. (2016) in their investigation of cloud based on MODIS and CALIOP observations. To investigate whether geometric conditions did have an impact on our results, we calculated the overall accuracy of the MODIS cloud mask as a function of the distance between MODIS and CALIOP IFOV. For ranges up to 1 km with a

100 m step, the change in accuracy was insignificant: 87.0±0.3% on average. (See Supplementary Online Materials' Fig. S2 for more detailed statistics about the spatial and temporal separation between MODIS and CALIOP).

It is possible that agreement between MODIS and CALIOP data is affected by cloud optical thickness ($\tau$) or, more precisely, by the higher sensitivity of CALIOP in detecting optically-thin cloud. Ackerman et al. (2008) estimated the MODIS limit for $\tau$ to be approximately 0.4. A similar improvement in agreement with CALIOP as a consequence of increasing $\tau$ was observed by Karlsson and Håkansson (2017) for the Advanced Very High Resolution Radiometer (AVHRR) instrument. The latter study demonstrated that the imager's probability of detection changed in the range $0.0<\tau<1.0$. We calculated the same statistic, and found that the probability distribution for MODIS was identical to AVHRR – although MODIS values were higher. This finding strongly suggests that cloud thickness has the same impact on our results as that found in previous studies. Collection 006 of MODIS data was investigated by Wang et al. (2016), who used lidar–radar (CALIPSO–CloudSat) profiles to focus on daytime multi-layered clouds. Our findings are consistent with those reported by Wang et al. (2016), despite the fact that the latter authors used a dataset of 267 million IFOV, while our study relied on around 33 million profiles. Their validation of Collection 006 reliability found overall agreement of 77.8% compared to our study, which found 81.9%. The difference may be due to the different sample sizes. Our result for cloud-free sky detection was slightly higher than in Wang et al. (2016): 25.5% compared to 20.9%. On the other hand, results for cloudy sky detection were very similar: 56.9% compared to 56.4% in our study.

Our study revealed that even for Collection 061, i.e. the most recent (July 2020) version of the MODIS cloud mask, up to 40% of cloud-free skies detected during the polar night were actually cloudy. Daytime accuracy was lowest over China (in January), the USA/ Canada (in January) and over tropical ocean along the west (January) and east (July) coasts of Africa. In these cases, MODIS detected cloud that did not exist according to CALIOP. False detections may be due to snow cover (the USA/ Canada), high aerosol content over China (Zhang et al. 2019; Tan, Zhang, and Shi 2019), and ocean bordering desert regions in North Africa (Weinzierl et al., 2017; Zuluaga et al., 2012).

As reported by Wang et al. (2016), and previously by Baum et al. (2012) and Ackerman et al. (2008), cloud detection in polar regions remains an unresolved issue for MODIS, and similar passive imaging radiometers. Polar night is especially challenging. Successful discrimination between cloud and the underlying surface requires radiance measurements in ice absorption bands (e.g. 1.6 μm or 2.1 μm). But as these are only available in daytime, night-time detection has to rely on thermal infrared data. As thermal inversion in the polar tropopause decreases the thermal contrast between cloud and the background, the thermal signatures of cloud and the land/ ocean surface become indistinguishable, leading to cloud masking errors (Liu et al., 2004). CALIOP, however, does not require solar illumination to operate. As it uses light emitted by the instrument itself, its performance is far less affected by day-night conditions. CALIOP's night-time data are of even higher quality, because solar illumination introduces an additional background signal and, thus, decreases the signal-to-noise ratio (Hunt et al., 2009). Furthermore, MODIS tends to miss up to ~20% of cloud along the Intertropical Convergence Zone (ITCZ), regardless of the time of year (January or June), and the time of day. This can be partially explained by the fact that MODIS is less sensitive to optically thin cloud than CALIPSO, and the ITCZ is the region where cirrus is most frequently observed (Sassen et al., 2009).

The higher sensitivity of CALIOP to optically thin cirrus, and the higher sensitivity of the lidar during night-time, also explains why CALIOP-based cloud fractions for MODIS 'confident clear' and 'probably clear' classes are higher along the ITCZ at night (Fig. 3f, h) than during the day (Fig. 3e, g).

The main goal of our study was to investigate the validity of the standard (operational) approach to the quantitative interpretation of MODIS cloud mask classes. The most important consequence of calculating empirical cloud fractions for MODIS cloud mask categories is the ability to recalculate global cloud amount with new weights. Therefore, instead of using global fractions (reported in Table 2), we derived a set of dedicated fractions for each algorithm path, and each 2.5-degree grid box (i.e. a local equivalent to the data given in Table 3). This considers MODIS IFOV within the full swath (excluding sunglint), and not only those collocated with CALIOP. Full-swath data were used because the MODIS L3 cloud amount product applies the same cloud mask interpretation to all IFOVs, regardless of their off-nadir angle. On the other hand, the use of nadir-only MODIS observations would result in CALIOP-like spatial coverage of the data, creating significant gaps due to CALIOP's pencil-like viewing geometry. Figure 8 illustrates the results of the calculation and reports differences in cloud amount between the MODIS ST operational product, and the product generated using the fractions presented in this study.

The outcome of the simulation shows that the use of current operational cloud fractions introduces significant errors. In some locations, MODIS underestimates cloud amount by 20–40%, most notably in polar regions at night. An overestimation of similar magnitude is observed mostly over the northern hemisphere: the USA/ Canada and China in January (both day and night), and the tropical coasts of Africa during the day (both in January and July). Consequently, MODIS Level 3 estimates of cloud amount should be used with great caution in those regions. This is especially important for the Arctic, which is undergoing a rapid change in climatic conditions (Serreze and Barry, 2011), and where cloud has been found to be an essential element in feedback (Kay et al. 2008; Vavrus 2004; Shupe and Intrieri 2004; Tan and Storelvmo 2019).

The availability of collocated MODIS and CALIOP observations also allowed us to examine which of the three 'best guess' interpretations of cloud mask categories is most accurate: the one when only 'confident cloud' IFOV are 'cloudy', the one that only considers 'confident clear' as 'clear', or the operational approach? We therefore calculated merged global cloud amount for January and July 2015. Our results show that, on the global scale, the standard approach is closest to CALIOP reference data, although only during the day (Tab. 4). At night, it is more accurate to assume that only 'confident clear' is actually cloud-free. The global result is biased by the polar night. In these conditions, all three 'best guess' interpretations noticeably underestimate cloud amount. At low- and mid-latitudes the standard (operational) approach differs from CALIOP data by ±2%. However, it should be noted that these statistics relate to large areas. As our study shows, regional differences are orders of magnitude larger.

Our study assumed that CALIOP's 'cloudy' IFOV was always completely cloud filled. This assumption is common when interpreting cloud masks based on data from the majority of imaging radiometers flown onboard meteorological and land-imaging satellites. However, studies by Zhao and Di Girolamo (2006), and Kotarba (2010) suggest that this postulate may not be true. Both of the latter studies took advantage of a rare collocation between a meteorological imager (MODIS) and the high-resolution land imager (ASTER) flown onboard the Terra satellite. Nearly 3,000 ASTER IFOV were located within each

MODIS pixel. Kotarba (2010) showed that for sunglint-free, oceanic scenes in the tropics, actual cloud coverage for the 'confident cloudy' MODIS category was 79.2% (mean) or 99.8% (median), instead of the assumed 100%. Comparable statistics for CALIOP are not available, as the CALIPSO spacecraft does not carry a high-resolution imager. Given the lack of alternatives, we must accept the hypothesis that 'cloudy' means 100% cloud filled.

## 5 Summary and Conclusion

This study investigated 33,793,648 collocated MODIS (cloud imager) and CALIOP (cloud profiling lidar) observations, acquired in January and July 2015. Our evaluation of the dataset allowed us to answer three, essential questions, related to global estimates of cloud amount resulting from the MODIS/ Aqua mission. These questions are:

1. *What are the actual cloud fractions corresponding to MODIS cloud mask classes?* We found that these fractions are 21.5%, 27.7%, 66.6%, and 94.7%, rather than the MODIS Science Team-assumed values of 0%, 0%, 100% ,and 100% for 'confident clear', 'probably clear', 'probably cloudy', and 'confident cloudy' categories, respectively (Tab. 2). Importantly, we found that the percentage of cloud cover to be assigned to MODIS cloud mask classes varied spatially (Fig. 3), and recommend that global fractions should be avoided, in favour of local alternatives.
2. *How significant are uncertainties in global cloud amount estimates calculated using the MODIS ST operational approach?* We found that uncertainties were up to −30% of cloud amount in the polar regions at night, and up to +30% of cloud amount in selected locations over the northern hemisphere, more frequently during the day (Fig. 8).
3. *Is the MODIS Level 3 standard approach reliable?* Our results showed that when a global cloud amount value is required (day and night, for all latitudes), the standard approach can be considered reliable (Tab. 4). We found that, in this case, it was more accurate than other 'best guess' approaches – namely only 'confident clear' is 'clear' (other classes are 'cloudy'), and 'confident cloudy' is 'cloudy' (other classes are 'clear'). However, on a regional scale the standard approach fails (Fig. 8). Whenever MODIS cloud amount is estimated regionally or locally it is necessary to assess whether a particular location might be affected by an error of up to ±30%.

Errors and uncertainties related to the generation of the MODIS Level 3 cloud amount product originate in the Level 2 product: the cloud mask (Fig. 1–2 vs. Fig. 8). The cloud detection algorithm is more-or-less accurate depending on environmental conditions, which are approximated as algorithm paths (Tab. 3). However, conditions within paths are not constant (Fig. 4–7): for instance, the same radiance/ reflectance thresholds are applied to Europe, the USA and China, while environmental conditions in these locations are not the same (e.g. different aerosol loads, different aerosol optical properties). The MODIS Science Team have attempted to discriminate between these conditions. For instance, since Collection 006 the 0.86 μm reflectance test over land considers thresholds that are a function of the Normalized Difference Vegetation Index (NDVI) and scattering angle. Although cloud misclassification is less frequent than in previous Collections, it still occurs, and impacts the degree of uncertainty of L3 cloud amount estimates, as shown in this study.

CALIOP-based estimates of cloud fraction are a robust way to adjust (and correct) MODIS estimates. The method described in this paper can be used globally, with the exception of sunglint regions (which are not sampled by CALIOP). In these areas 'best guess' findings can, potentially, be applied. The polar regions benefit most from the new method. Cloud fractions derived for MODIS/ Aqua may be also adopted for MODIS/ Terra, since the two sensors are expected to produce comparable and homogenous records. Moreover, the occasional collocation of the CALIPSO satellite with AVHRR and VIIRS instruments makes it possible to calculate similar cloud fractions for these missions, and produce more reliable cloud climatologies.

## Data availability

MODIS data are available from the Level 1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) at NASA's Goddard Space Flight Center (https://earthdata.nasa.gov/eosdis/daacs/laads). CALIPSO products are available from the Atmospheric Science Data Center (ASDC) at NASA's Langley Research Center (https://eosweb.larc.nasa.gov/).

## Author contribution

AZK designed the research, carried it out, and prepared the manuscript.

## Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgements

The organization and pre-processing of the CALIPSO database was part of a project funded by the National Science Institute of Poland under contract no. UMO-2017/25/B/ST10/01787. Processing of the MODIS-CALIPSO database and the resulting analyses was funded by the Space Research Centre of the Polish Academy of Sciences, as a part of the research theme 'Satellite monitoring of geophysical processes in the atmosphere and Earth's surface, related to climate and light pollution'.

## References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C. and Gumley, L. E.: Discriminating clear sky from clouds with MODIS, J. Geophys. Res. Atmos., doi:10.1029/1998JD200032, 1998.

Ackerman, S. A., Holz, R. E., Frey, R., Eloranta, E. W., Maddux, B. C. and McGill, M.: Cloud detection with MODIS. Part II: Validation, J. Atmos. Ocean. Technol., doi:10.1175/2007JTECHA1053.1, 2008.

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K. and Yang, P.: MODIS cloud-top property refinements for collection 6, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-11-0203.1, 2012.

Chan, M. A. and Comiso, J. C.: Arctic cloud characteristics as derived from MODIS, CALIPSO, and cloudsat, J. Clim., doi:10.1175/JCLI-D-12-00204.1, 2013.

Derrien, M. and Le Gléau, H.: MSG/SEVIRI cloud mask and type from SAFNWC, Int. J. Remote Sens., 26(21), 4707–4732, doi:10.1080/01431160500166128, 2005.

Dybbroe, A., Karlsson, K. G. and Thoss, A.: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: Algorithm description, J. Appl. Meteorol., doi:10.1175/JAM-2188.1, 2005.

Fontana, F., Lugrin, D., Seiz, G., Meier, M., and Foppa, N.: Intercomparison of satellite- and ground-based cloud fraction over Switzerland (2000–2012), Atmospheric Research, doi: 10.1016/j.atmosres.2013.01.013, 2013.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R. and Wang, X.: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5, J. Atmos. Ocean. Technol., doi:10.1175/2008JTECHA1052.1, 2008.

Gao, M., Qin, Z., Zhang, H., Lu, L., Zhou, X. and Yang, X.: Remote sensing of agro-droughts in Guangdong Province of China using MODIS satellite data, Sensors, doi:10.3390/s8084687, 2008.

Gomis-Cebolla, J., Jimenez, J. C. and Sobrino, J. A.: MODIS probabilistic cloud masking over the Amazonian evergreen tropical forests: a comparison of machine learning-based methods, Int. J. Remote Sens., doi:10.1080/01431161.2019.1637963, 2020.

Guenther, B., Xiong, X., Salomonson, V. V., Barnes, W. L. and Young, J.: On-orbit performance of the Earth Observing System Moderate Resolution Imaging Spectroradiometer; First year of data, Remote Sens. Environ., doi:10.1016/S0034-4257(02)00097-4, 2002.

Heidinger, A. K., Evan, A. T., Foster, M. J. and Walther, A.: A naive Bayesian cloud-detection scheme derived from Calipso and applied within PATMOS-x, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-11-02.1, 2012.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M. and Wagner, W.: The ESA climate change initiative: Satellite data records for essential climate variables, Bull. Am. Meteorol. Soc., 94(10), 1541–1552, doi:10.1175/BAMS-D-11-00254.1, 2013.

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., Vaughan, M. A. and Baum, B.: Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP, J. Geophys. Res. Atmos., doi:10.1029/2008JD009837, 2009.

Hubanks, P. A., King, M. D., Platnick, S. and Pincus, R.: MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document No. ATBD-MOD-30 for Level-3 Global Gridded Atmosphere Products (08_D3, 08_E3, 08_M3). [online]

Available from: https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/L3_ATBD_C6_C61_2019_02_20.pdf, 2008.

Hunt, W. H., Vaughan, M. A., Powell, K. A. and Weimer, C.: CALIPSO lidar description and performance assessment, J. Atmos. Ocean. Technol., doi:10.1175/2009JTECHA1223.1, 2009.

Hutchison, K. D., Heidinger, A. K., Kopp, T. J., Iisager, B. D. and Frey, R. A.: Comparisons between VIIRS cloud mask performance results from manually generated cloud masks of VIIRS imagery and CALIOP-VIIRS match-ups, Int. J. Remote Sens., doi:10.1080/01431161.2014.932465, 2014.

Karlsson, K.-G. and Håkansson, N.: Detailed characterisation of AVHRR global cloud detection performance of the CM SAF CLARA-A2 climate data record based on CALIPSO-CALIOP cloud information, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-307, 2017.

Karlsson, K. G. and Dybbroe, A.: Evaluation of arctic cloud products from the EUMETSAT climate monitoring satellite application facility based on CALIPSO-CALIOP observations, Atmos. Chem. Phys., doi:10.5194/acp-10-1789-2010, 2010.

Karlsson, K. G. and Johansson, E.: On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data: Example investigating the CM SAF CLARA-A1 dataset, Atmos. Meas. Tech., doi:10.5194/amt-6-1271-2013, 2013.

Kay, J. E., L'Ecuyer, T., Gettelman, A., Stephens, G. and O'Dell, C.: The contribution of cloud and radiation anomalies to the 2007 Arctic sea ice extent minimum, Geophys. Res. Lett., doi:10.1029/2008GL033451, 2008.

Kopp, T. J., Thomas, W., Heidinger, A. K., Botambekov, D., Frey, R. A., Hutchison, K. D., Iisager, B. D., Brueske, K. and Reed, B.: The viirs cloud mask: Progress in the first year of s-npp toward a common cloud detection scheme, J. Geophys. Res., doi:10.1002/2013JD020458, 2014.

Kotarba, A. Z.: Estimation of fractional cloud cover for Moderate resolution imaging Spectroradiometer/Terra cloud mask classes with high-resolution over ocean ASTER observations, J. Geophys. Res. Atmos., 115(22), doi:10.1029/2009JD013520, 2010.

Kotarba, A. Z.: Impact of Moderate Resolution Imaging Spectroradiometer (MODIS) cloud mask interpretation on cloud amount estimation, J. Geophys. Res., 120(17), 8971–8986, doi:10.1002/2015JD023277, 2015.

Kraatz, S., Khanbilvardi, R. and Romanov, P.: A comparison of MODIS/VIIRS cloud masks over ice-bearing river: On achieving consistent cloud masking and improved river ice mapping, Remote Sens., doi:10.3390/rs9030229, 2017.

Krijger, J. M., Van Weele, M., Aben, I. and Frey, R.: Technical Note: The effect of sensor resolution on the number of cloud-free observations from space, Atmos. Chem. Phys., doi:10.5194/acp-7-2881-2007, 2007.

Li, J., Liu, C. Y., Huang, H. L., Schmit, T. J., Wu, X., Menzel, W. P. and Gurka, J. J.: Optimal cloud-clearing for AIRS radiances using MODIS, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2005.847795, 2005.

Liu, Y., Key, J.R., Frey, R.A., Ackerman, S.A., and Menzel, W.P.: Nighttime polar cloud detection with MODIS, Remote Sens., doi:10.1016/j.rse.2004.06.004, 2004

Melchiorre, A., Boschetti, L. and Roy, D. P.: Global Evaluation of the Suitability of MODIS-Terra Detected Cloud Cover as a Proxy for Landsat 7 Cloud Conditions, Remote Sens., doi:10.3390/rs12020202, 2020.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C. and Frey, R. A.: The MODIS cloud products: Algorithms and examples from terra, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2002.808301, 2003.

Ramanathan, V. and Kiehl, J. T.: The Radiative Forcing due to Clouds and Water Vapour, in Frontiers of Climate Modelling, pp. 119–151., 2006.

Remer, L. A., Mattoo, S., Levy, R. C., Heidinger, A., Pierce, R. B. and Chin, M.: Retrieving aerosol in a cloudy environment: Aerosol product availability as a function of spatial resolution, Atmos. Meas. Tech., doi:10.5194/amt-5-1823-2012, 2012.

Sassen, K., Wang, Z. and Liu, D.: Cirrus clouds and deep convection in the tropics: Insights from CALIPSO and CloudSat, J. Geophys. Res. Atmos., 114, D00H06, doi:10.1029/2009JD011916.

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Glob. Planet. Change, doi:10.1016/j.gloplacha.2011.03.004, 2011.

Sèze, G., Pelon, J., Derrien, M., Le Gléau, H. and Six, B.: Evaluation against CALIPSO lidar observations of the multi-geostationary cloud cover and type dataset assembled in the framework of the Megha-Tropiques mission, Q. J. R. Meteorol. Soc., doi:10.1002/qj.2392, 2015.

Shang, H., Letu, H., Nakajima, T. Y., Wang, Z., Ma, R., Wang, T., Lei, Y., Ji, D., Li, S. and Shi, J.: Diurnal cycle and seasonal variation of cloud cover over the Tibetan Plateau as determined from Himawari-8 new-generation geostationary satellite data, Sci. Rep., doi:10.1038/s41598-018-19431-w, 2018.

Shupe, M. D. and Intrieri, J. M.: Cloud radiative forcing of the Arctic surface: The influence of cloud properties, surface albedo, and solar zenith angle, J. Clim., doi:10.1175/1520-0442(2004)017<0616:CRFOTA>2.0.CO;2, 2004.

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., L'Ecuyer, T. and Lebsock, M.: Cloudsat and calipso within the a-train: Ten years of actively observing the earth system, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-D-16-0324.1, 2018.

Sun-Mack, S., Minnis, P., Chen, Y., Kato, S., Yi, Y., Gibson, S. C., Heck, P. W. and Winker, A. D. M.: Regional apparent boundary layer lapse rates determined from CALIPSO and MODIS data for cloud-height determination, J. Appl. Meteorol. Climatol., doi:10.1175/JAMC-D-13-081.1, 2014.

Tan, I. and Storelvmo, T.: Evidence of Strong Contributions From Mixed-Phase Clouds to Arctic Climate Change, Geophys. Res. Lett., doi:10.1029/2018GL081871, 2019.

Tan, S., Zhang, X. and Shi, G.: MODIS cloud detection evaluation using CALIOP over polluted eastern China, Atmosphere (Basel)., doi:10.3390/atmos10060333, 2019.

Trenberth, K. E., Fasullo, J. T. and Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 90(3), 311–323, doi:10.1175/2008BAMS2634.1, 2009.

Várnai, T. and Marshak, A.: MODIS observations of enhanced clear sky reflectance near clouds, Geophys. Res. Lett., doi:10.1029/2008GL037089, 2009.

Vaughan, M. A., Powell, K. A., Kuehn, R. E., Young, S. A., Winker, D. M., Hostetler, C. A., Hunt, W. H., Liu, Z., Mcgill, M. J. and Getzewich, B. J.: Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements, J. Atmos. Ocean. Technol., doi:10.1175/2009JTECHA1228.1, 2009.

Vavrus, S.: The impact of cloud feedbacks on Arctic climate under Greenhouse forcing, J. Clim., doi:10.1175/1520-0442(2004)017<0603:TIOCFO>2.0.CO;2, 2004.

Wang, T., Fetzer, E. J., Wong, S., Kahn, B. H. and Yue, Q.: Validation of MODIS cloud mask and multilayer flag using CloudSat-CALIPSO cloud profiles and a cross-reference of their cloud classifications, J. Geophys. Res. Atmos., 121(19), 11620–11635, doi:10.1002/2016JD025239, 2016.

Weinzierl, B., Ansmann, A., Prospero, J. M., Althausen, D., Benker, N., Chouza, F., Dollner, M., Farrell, D., Fomba, W. K., Freudenthaler, V., Gasteiger, J., Groß, S., Haarig, M., Heinold, B., Kandler, K., Kristensen, T. B., Mayol-Bracero, O. L., Müller, T., Reitebuch, O., Sauer, D., Schäfler, A., Schepanski, K., Spanu, A., Tegen, I., Toledano, C. and Walser, A.: The Saharan aerosol long-range transport and aerosol-cloud-interaction experiment: Overview and selected highlights, Bull. Am. Meteorol. Soc., doi:10.1175/BAMS-D-15-00142.1, 2017.

Wilson, A. M., Parmentier, B. and Jetz, W.: Systematic land cover bias in Collection 5 MODIS cloud mask and derived products - A global overview, Remote Sens. Environ., doi:10.1016/j.rse.2013.10.025, 2014.

Wilson, M. J. and Oreopoulos, L.: Enhancing a simple MODIS cloud mask algorithm for the Landsat Data Continuity Mission, IEEE Trans. Geosci. Remote Sens., doi:10.1109/TGRS.2012.2203823, 2013.

Winker, D., Vaughan, M. and Hunt, B.: The CALIPSO mission and initial results from CALIOP, in Lidar Remote Sensing for Environmental Monitoring VII., 2006.

Xie, Y., Qu, J. J. and Xiong, X.: Improving the CALIPSO VFM product with Aqua MODIS measurements, Remote Sens. Lett., doi:10.1080/01431161003720387, 2010.

Zhang, X., Tan, S. C., Shi, G. Y. and Wang, H.: Improvement of MODIS cloud mask over severe polluted eastern China, Sci. Total Environ., doi:10.1016/j.scitotenv.2018.10.369, 2019.

Zhao, G. and Di Girolamo, L.: Cloud fraction errors for trade wind cumuli from EOS-Terra instruments, Geophys. Res. Lett., doi:10.1029/2006GL027088, 2006.

Zuluaga, M. D., Webster, P. J. and Hoyos, C. D.: Variability of aerosols in the tropical Atlantic Ocean relative to African Easterly Waves and their relationship with atmospheric and oceanic environments, J. Geophys. Res. Atmos., doi:10.1029/2011JD017181, 2012.

**Tables**

**Table 1**. Agreement in cloud detection between MODIS and CALIOP (% of all cases). Overall accuracy (given in brackets) refers to ‘confident clear’ and ‘confident cloudy’ detections. In other cases, ‘confident clear’ and probably clear’ were merged, as were ‘probably cloudy’ and ‘confident cloudy’.

| | | **MODIS** | | | | **Overall accuracy** |
|---|---|---|---|---|---|---|
| | | confident clear | probably clear | probably cloudy | confident cloudy | |
| | | *Day+Night* | | | | |
| CALIOP | clear | 22.7 | 5.4 | 1.9 | 3.1 | 86.7% (77.3%) |
| | cloudy | 6.2 | 2.1 | 3.9 | 54.6 | |
| | | *Day only* | | | | |
| CALIOP | clear | 25.5 | 5.1 | 1.7 | 3.2 | 89.4% (81.9%) |
| | cloudy | 3.7 | 2.0 | 2.4 | 56.4 | |
| | | *Night only* | | | | |
| CALIOP | clear | 20.2 | 5.7 | 2.2 | 3.0 | 84.2% (73.3%) |
| | cloudy | 8.5 | 2.1 | 5.2 | 53.1 | |

**Table 2**. Global cloud fractions for MODIS cloud mask classes derived from CALIOP observations ('This study'), and used in the operational MODIS Science Team Level 3 product ('Operational'). Numbers in brackets refer to class frequency (*n* = 33, 793, 648).

| **Source of cloud fractions for cloud mask classes** | | **Cloud fractions (%) for MODIS cloud mask class** (class frequency, % of *n*) | | | |
|---|---|---|---|---|---|
| | | **confident clear** (28.9%) | **probably clear** (7.5%) | **probably cloudy** (5.8%) | **confident cloudy** (57.8%) |
| Operational | Day+Night | 0.0 | 0.0 | 100.0 | 100.0 |
| This study | Day+Night | 21.5 | 27.7 | 66.6 | 94.7 |
| | Day only | 12.7 | 28. 4 | 58.4 | 94.7 |
| | Night only | 29.5 | 27.1 | 70.7 | 94.7 |

**Table 3**. CALIOP-based cloud fractions for MODIS cloud mask classes, calculated individually for each MODIS algorithm path. Note that more paths are reported here than in the MODIS project. Snow-covered ocean, land, desert and coast constitute a single path in the operational algorithm, while here they are reported individually to highlight how snow impacts the results. The sunglint path is missing as CALIOP does not sample over sunglint areas. Numbers in brackets refer to how frequently (% of *n*) a given algorithm path was executed, *n* = 33, 793, 648.

| **Cloud masking algorithm path** | | | | **CALIOP-based cloud fractions [%] for MODIS cloud mask class** | | | |
|---|---|---|---|---|---|---|---|
| | | | | **confident clear** | **probably clear** | **probably cloudy** | **confident cloudy** |
| Day (47.2) | Snow-covered (5.5) | Land | (0.2) | 13.8 | 67.0 | 56.0 | 97.6 |
| | | Desert | (3.9) | 12.6 | 32.6 | 71.8 | 96.6 |
| | | Coast | (0.2) | 15.3 | 55.5 | 61.8 | 93.8 |
| | | Ocean | (1.1) | 20.5 | 76.3 | 69.7 | 88.6 |
| | Snow-free (41.7) | Land | (6.7) | 15.6 | 32.3 | 63.9 | 93.4 |
| | | Desert | (3.4) | 9.1 | 19.1 | 45.5 | 90.0 |
| | | Coast | (1.6) | 19.0 | 33.8 | 59.8 | 93.0 |
| | | Ocean | (30.1) | 10.5 | 28.4 | 54.5 | 95.2 |
| Night (52.8) | Snow-covered (15.8) | Land | (2.6) | 31.4 | 65.0 | 80.9 | 93.9 |
| | | Desert | (4.7) | 34.3 | 65.3 | 75.9 | 86.4 |
| | | Coast | (0.9) | 29.8 | 60.8 | 75.0 | 93.7 |
| | | Ocean | (7.6) | 49.7 | 73.7 | 82.5 | 96.8 |
| | Snow-free (37.0) | Land | (5.4) | 8.0 | 25.6 | 68.5 | 97.7 |
| | | Desert | (2.6) | 8.2 | 23.5 | 55.8 | 95.4 |
| | | Coast | (0.9) | 10.9 | 23.0 | 60.9 | 96.4 |
| | | Ocean | (28.1) | 22.9 | 22.4 | 61.8 | 94.6 |

**Table 4**. Global cloud amount (%) calculated with different ‘best guess’ interpretations of the MODIS cloud mask product. Only MODIS IFOV collocated with CALIOP are considered.

| | **CALIOP** | **MODIS cloud mask interpretation scenario** | | |
|---|---|---|---|---|
| | | Only ‘confident cloudy’ is ‘cloudy’ | ‘Confident clear’ and ‘probably clear’ are clear, while the rest is cloudy | Only ‘confident clear’ is ‘clear’ |
| | *Global* | | | |
| Day+Night | 66.7 | 57.7 | 63.5 | 71.0 |
| Day | 64.3 | 59.3 | 63.4 | 70.6 |
| Night | 68.9 | 56.1 | 63.5 | 71.3 |
| | *Polar regions (latitudes above 60ºN/S)* | | | |
| Day+Night | 66.9 | 50.5 | 57.6 | 61.0 |
| Day | 64.8 | 59.0 | 62.6 | 66.4 |
| Night | 68.5 | 44.1 | 53.9 | 57.0 |
| | *Equatorial region (latitudes between 30ºN and 30ºS)* | | | |
| Day+Night | 59.8 | 52.8 | 58.0 | 67.4 |
| Day | 56.2 | 49.9 | 54.8 | 65.4 |
| Night | 63.5 | 55.7 | 61.2 | 69.4 |
| | *Mid-latitudes (between polar and equatorial)* | | | |
| Day+Night | 73.3 | 68.9 | 74.2 | 83.6 |
| Day | 72.0 | 69.1 | 72.7 | 79.1 |
| Night | 74.6 | 68.8 | 75.7 | 87.9 |

**Figures**

Figure 1. Observations declared ‘confident clear’ or ‘probably clear’ by the MODIS cloud masking algorithm, but identified as ‘cloudy’ by CALIOP.

Figure 2. Observations declared ‘confident cloudy’ or ‘probably cloudy’ by the MODIS cloud masking algorithm, but identified as ‘clear’ by CALIOP.

Figure 3. CALIOP-based cloud fraction for MODIS cloud mask classes.

Figure 4. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘nighttime, snow(ice)-covered ocean’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 5. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘nighttime snow-free land’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 6. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘daytime, snow-free land’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 7. CALIOP-based cloud fraction for MODIS cloud mask classes for the ‘daytime, snow-free ocean’ algorithm path, and corresponding histograms (red vertical line indicates the mean value).

Figure 8. Difference between the MODIS Science Team (MODIS ST) Level 3 cloud amount product, and cloud amount calculated with the cloud fractions found in this study. Positive values indicate that the MODIS operational product overestimates cloud amount (with respect to CALIOP), while negative values indicate a MODIS underestimate. All MODIS observations refer to the full swath, not only those collocated with CALIOP.

**Supplementary Online Material**

Figure S1. The average cloud-aerosol discrimination (CAD) score for CALIOP cloud data used in the study. Maps show spatial variation in the CAD score during the day (b), at night (c), and regardless of the time of the day (a). These plots demonstrate the high stability of CAD scores at various latitudes during the day (d) and at night (e).

Figure S2. Overall accuracy of MODIS cloud detection as a function of the temporal (a, c) and spatial (b, d) separation of MODIS and CALIOP IFOVs. Top plots show the frequency of observations for individual time (a) and distance (b) ranges, while bottom plots report accuracy for these ranges. MODIS detections are validated using CALIOP cloud profiles as a reference. Accuracy is defined as the ratio of MODIS true detections (true positive and true negative) to all MODIS observations (see Table 1 in the main text for details).