Response to Referee #2

We want to thank the referee for appreciating our work and for the thoughtful comments and suggestions. Most of them have been taken into account to improve the manuscript. We apologize for the difficulties associated with the length of the manuscript and excessively long sentences. We have re-worked the manuscript and addressed each comment. In the text below, the reviewerôs comments are marked in italics blue and our answers are given in normal font.
As the referee has correctly pointed out, the method itself is not new (the first author developed it in the 1990ôs for TOVS), and there exist several publications, which are referenced in the article. Indeed it was a difficult task to select what should be presented and what left out, which is reflected in differing opinions of the 2 referees.
Since both referees suggested to shorten the manuscript, we have done our best to do it without losing the message we wanted to deliver to the community. Here is the list of actions performed
1) shortening section 2 ôData and methodsô and moving a shortened version of section 3.1 ôCollocated AIRS-CALIPSO-CloudSat dataô to this section
2) simplifying Table 2, taking out 5 figures / 22 figure panels (3 figures moved to supplement)
3) taking out the ENSO discussion in section 5 (together with Fig. 16) and
4) revising the remaining applications in section 5

We do not agree with the suggestion of a complete removal of section 5 ôApplicationsô, as the presented method is not new and one of the goals of this article was to present scientific applications (as indicated in the title).
Since the results similar to those presented in new Fig. 12 have recently been published for other data sets, it would be difficult to use the presented material in a separate publication. We compare our results to one of them and point out an interesting extension. We plan to work on a more complex analysis to pursue this subject further, but we think itôs important to present these results in the current publication.

5 particular issues that need further explanation:

- the role of CALIPSO-CALIOP data for tuning the method

The cloud property retrieval was originally developed for TOVS data (Stubenrauch et al. 1996, 1999, 2006); at that time the cloud detection, which indeed was applied before the cloud retrieval, was essentially based on interchannel regression tests using a combination of IR sounder and microwave (MSU) brightness temperatures.
When we adapted the cloud retrieval to AIRS, channel 7 of AMSU did not work, so we could not adapt the cloud detection. However the retrieval itself provides cloud pressure and emissivity for each measurement (only about 5% of the data do not give a solution, these are declared immedeately as clear sky). We then considered it more interesting to develop a cloud detection which could be applied after the retrieval. The idea was to test the reliability of the results to decide if a footprint is cloudy. By comparing clear sky and cloudy scenes determined within time synchroneous samples from CALIPSO L2 5km cloud data, provided by NASA, we found that the relative spectral spread of cloud emissivities determined at atmospheric window wavelengths is small if the footprint contains a cloud for which the cloud height and emissivity are well determined (both are used in the computation of the spectral emissivities), while most clear sky scenes lead to very large values. These distributions have been published in Stubenrauch et al. 2010, and for the retrievals with new ancillary data in Fig. S1. These distributions show a nice distinction between clear and cloudy, but the thresholds themselves have been determined by examining many different aspects, like maps and comparison with other datatsets, distributions separately over tropics, midlatitudes and polar regions. One important aspect was also to test that AIRS, using two different ancillary data sets, together with IASI gave coherent answers, day and night.

So, the CALIPSO-CloudSat data have been essential to guide us in the cloud detection, but they were not used to tune it.

- the exact description of the used CALIPSO dataset for tuning and for evaluation of cloud properties

Again we want to stress that we did not use CALIPSO for tuning.

We have moved the section of the collocated AIRS-CALIPSO-CloudSat data forward, so that the description is placed before the description of the cloud detection. It was well written that we used version 3 of the NASA CALIPSO L2 cloud data averaged over 5 km (Winker et al. 2009) ; and we explained the procedures how we used the data (for example excluding subvisible cirrus). By the way, we published comparisons with lidar already in 2005, when we compared TOVS Path B cloud properties with LITE (Stubenrauch et al. 2005) where we also investigated subvisible cirrus. In this paper we just wanted to show that the CIRS cloud data are of slightly better quality than the AIRS-LMD cloud climatology, and the effect of ancillary data, which in our opinion has not been stressed with other cloud climatologies.

- the consequence of using some unphysical assumptions in the retrieval

We accept cloud emissivities up to a value of 1.5, due to noise. This is explained in the reference Stubenrauch *et al.* 1999, which is cited :

As in Eq 2 the denominator includes two terms (Icld and Iclr) which get very close to each other in the case of low-level clouds, the cloud emissivity can get larger than 1 when taking into account uncertainties. In Stubenrauch et al. (1999), it was shown that the original method, which excluded values larger than 1, underestimated the amount of low-level clouds considerably.

The limit larger than 1 has been chosen to compensate for radiation noise and ancillary data uncertainties and this leads to a better identification of low-level clouds.

- the balance between finding spectral coherence in the solutions and still maintain physically reasonable emissivity differences

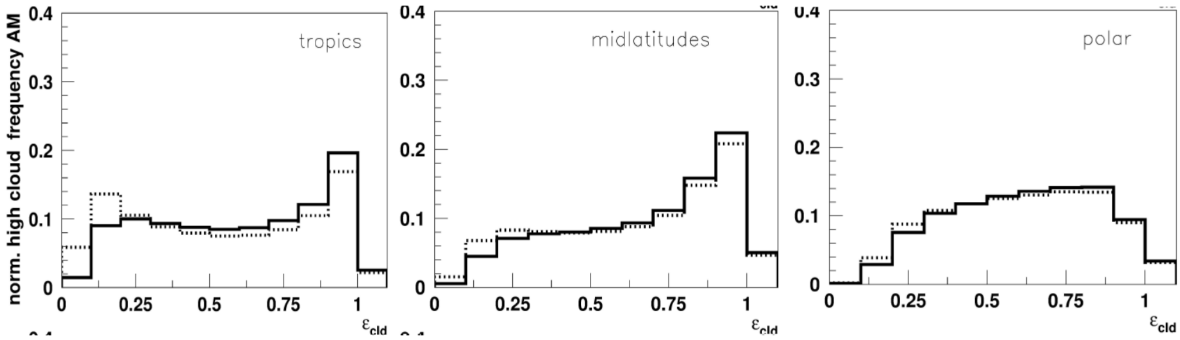
The multi-spectral cloud detection is indeed based on wavelengths in an interval which is sensitive to thermodynamical phase and ice crystal sizes. As can be seen in Fig. 3 of Guignard et al. (2012), the relative cloud emissivity difference between 9 μm and 12 μm can go up to 0.3 for small IWP and ice crystal size. However, instead of using a spectral difference, we use a standard deviation between 6 wavelengths, divided by retrieved cloud emissivity. This should be always smaller than 0.15, even in the case of small IWP and ice crystal sizes which produce the largest slope (we have studied that in detail when developing the method in 2010). In this empirical method, the error one makes, if the used cloud pressure does not correspond to the real pressure, is larger, and Fig. S1 (of the supplement) illustrates nicely, that this relative standard deviation is larger than 0.3 for clear sky scenes, while for cloudy scenes distributions the distributions are really narrow, using CALIPSO-GEOPROF to separate cloudy and clear sky scenes.

- justification of the statement of achieving successful cloud detection down to IR cloud optical thicknesses of 0.1

optical thickness can be deduced from cloud emissivity as COD = $- \ln(1-\varepsilon_{cld})$

As we present clouds with $\varepsilon_{cld} > 0.1$, this corresponds to clouds with IR COD > 0.1 (or with VIS COD > 0.2 as VIS COD = $-2\ln(1-\varepsilon_{cld})$.

To reduce misidentification of clear sky as high-level clouds, only clouds with ε_{cld} Ó0.10 are considered. Indeed, this came out of a study with CALIPSO-CloudSat :



The above figures present normalized ε_{cld} distributions of high-level clouds, after multi-spectral cloud detection, but leaving clouds with $0.05 < \varepsilon_{cld} < 0.10$ as clouds, separately for cloudy scenes defined by GEOPROF and CALIPSO (full line) and for all scenes (dotted line). The first bin includes scenes with $0.05 < \varepsilon_{cld} < 0.10$; in the tropics this bin has more clear sky than high-level clouds. Therefore we have moved the threshold to 0.1. As the contribution of the first bin is small compared to the integral, this seemed a reasonable choice.

Specific comments

1. Page 1, Abstract, line 19, ***ñ****to evaluate****ò****:*
The term ***ñ****to evaluate****ò*** *should be changed to* ***ñ****to design and evaluate****ò****. You used A-train data to find your* ***ó****a posteriori****ô*** *cloud masking thresholds, right? Then you should be clear in your description that A-train data is not completely independent from your data/method. This is important for the reader to know.*

We do not quite agree with this comment; the cloud retrieval was originally developed for TOVS data (Stubenrauch et al. 1996, 1999, 2006); at that time the cloud detection, which indeed was applied before the cloud retrieval, was essentially based on interchannel regression tests using a combination of IR sounder and microwave (MSU) brightness temperatures.

When we adapted the cloud retrieval to AIRS, channel 7 of AMSU did not work, so we could not adapt the cloud detection. However the retrieval itself provides cloud pressure and emissivity for each measurement (only about 5% of the data do not give a solution, these are declared immedeately as clear sky). We then considered it more interesting to develop a cloud detection which could be applied after the retrieval. The idea was to test the reliability of the results to decide if a footprint is cloudy. By comparing clear sky and cloudy scenes determined within time synchroneous samples from CALIPSO L2 5km cloud data, provided by NASA, we found that the relative spectral spread of cloud emissivities determined at atmospheric window wavelengths is small if the footprint contains a cloud for which the cloud height and emissivity are well determined (as both are used in the computation), while most clear sky scenes lead to very large values. These distributions have been published in Stubenrauch et al. 2010, and for the retrievals with new ancillary data in Fig. S1. These distributions show a nice distinction between clear and cloudy, but the thresholds themselves have been determined by examining many different aspects, like maps and comparison with other datatsets, distributions separately over tropics, midlatitudes and polar regions. One important aspect was also to test that AIRS, using two different ancillary data sets, together with IASI gave coherent answers, day and night.

So, the CALIPSO-CloudSat data have been essential to guide us in the cloud detection, but they were not used to tune it.

2. Page 1, Abstract, line 23, ***ñ****coincides****ò****:*
To use the term ***ñ****coincides****ò*** *here is a too strong conclusion from your results. Figure 6 (lower right panel) clearly shows a rather broad distribution of results where frequencies at the two extremes (0 and 1) are still about 20-25 % of the frequency for the value 0.5 (representing the middle of the defined layer).Therefore you can possibly only state that the cloud height can be* ***ñ****approximated****ò*** *by the middle of the defined layer. Also* ***ñ****middle****ò*** *could possibly be replaced by* ***ñ****the mean layer height****ò*** *to make the description scientifically stricter.*
3. Page 1, Abstract, line 27, ***ñ****apparent vertical cloud extent****ò****:*

The explanation here is confusing, indicating that upper level clouds generally have higher cloud emissivities than lower level clouds. This cannot be true. I guess the authors mean something else. Please clarify!
Rewritten as :
CIRS cloud height can be approximated by the mean layer height (for optically thin clouds) or the mean between cloud top and the height at which the cloud reaches opacity. For high-level clouds, especially in the tropics, this height lies on average 1 km to 3 km below cloud top.

*4. Page 2, Abstract, lines 5-8, **ñ**response to climate change**ò** + Page 3, Section 1, lines 23- 25 and the entire section 5: The last sentence in the abstract, the sentence about Section 5 in Section 1 and the entire section 5 could possibly be removed for shortening the paper (see also comment 25!).*
We have considerable shortened section 5, but have left two main studies, which have been described in a more concise manner. The latter study is also compared to recent results using other data.
Changed last part of abstract to :
The 5% annual mean excess in high-level cloud amount in the Northern compared to the Southern hemisphere has a pronounced seasonal cycle with a maximum of 25% in boreal summer, in accordance with the moving of the ITCZ peak latitude, with annual mean of 4°N, to a maximum of 12°N. This suggests that this excess is mainly determined by the position of the ITCZ. Considering interannual variability, tropical cirrus are more frequent relative to all clouds when the global (or tropical) mean surface gets warmer. Changes in relative amount of tropical high opaque and thin cirrus with respect to mean surface temperature show different geographical patterns, suggesting that their response to climate change might differ.

*5. Page 2, Section 1, line 11, **ñ**70 % cloud cover**ò**:*
Although this is a widely used and accepted figure for global cloudiness, I would like to point out that a value of global cloud cover cannot be stated without first defining what you mean by a cloud. The figure 70 % is kind of representing clouds which have a significant impact on radiation budgets and it could possibly be relevant if you define that clouds should have at least a cloud optical thickness of approximately 0.2. But if including also the thinnest clouds (often called sub-visible clouds and so far only observed by high sensitive instruments like CALIPSO-CALIOP) the figure may increase to values well above 80 %. I think it would be appropriate to at least make a short statement on what clouds are considered when stating that global cloudiness is about 70 %.
Indeed, in the GEWEX Cloud Assessment we found out that global cloud amount is about 0.68±0.03 when considering clouds with VIS optical depth of larger than 0.2, and additional 0.06 arise from subvisible clouds detected by CALIPSO (Stubenrauch et al. 2013), which brings it to 0.74. This is written in Section 4.
It seems for us appropriate to leave the about 70%, as this sentence is the first in the introduction and is just meant to bring up the importance of clouds because of their large coverage. 7 lines further the reader finds more detail on the threshold (IR optical depth > 0.1).

*6. Page 3, Section 1, line 3: **ñ**optical depth less than 3**ò***
My impression is that the capability is better than that, i.e., the capability of having reasonable cloud optical depth estimations from CALIOP data covers the interval 0-5. Please check that the value of 3 is really justified.
The optical depth at which clouds are opaque is difficult to determine. In an earlier publication (Lamquin *et al.* 2008), we wrote that the upper limit lies between 3 and 5. One should not forget that the uncertainty is easily 20% due to uncertainty in multiple scattering contributions (Lamquin *et al.* 2008).
We have rewritten this in accordance :
Whereas the lidar can detect sub-visible cirrus, its beam can only penetrate the cloud down to optical depth of about 3 to 5 (in visible range). For optically thicker clouds, the radar provides the cloud base.

*7. Page 7, Section 2.4, line 4, **ñ**emissivities larger than 1**ò**:*
*I must say that it is quite disturbing to **ñ**be forced**ò** to use unphysical values in the retrieval. I understand that uncertainties can lead to this but I am not sure that this is then the best way of handling these uncertainties. Why not restrict emissivities to 1 in the optimization/minimization process when knowing*

that this is physically correct? I canôt see why your present method gives better uncertainty descriptions of the retrieved cloud pressures than when using a restricted emissivity value. Donôt inconsistencies give rise to new inconsistencies? Please explain and motivate.

The reason is explained in the reference Stubenrauch *et al.* 1999 which is cited:

As in Eq 2 the denominator includes two terms (Icld and Iclr) which get very close to each other in the case of low-level clouds, the cloud emissivity can easily get unphysical when taking into account uncertainties. In Stubenrauch et al. (1999), it was shown that the original method, which excluded values larger than 1, underestimated the amount of low-level clouds considerably.

The limit larger than 1 has been chosen to compensate for radiation noise and ancillary data uncertainties and this leads then to a better identification of low-level clouds.

8. Page 7, Section 2.4, lines 22-28, ña posteriori cloud detectionò:

The ña posteriori cloud detectionò has already been briefly introduced (page 4, lines 7- 11). Why repeating this information here? Delete these lines or move part of this to the relevant section 2.5.

deleted

9. Page 9, Section 2.4.1, lines 18-20, ñocean cloud amounts larger during nightò:

To find larger ocean cloud amounts at night than during day is found in many regions (e.g. over marine stratocumulus areas). What made you think this was a problem specifically for ERA-Interim? Please explain.

The problem is not that the cloud amount is larger during night than during day, but that results are different when using two different sets of ancillary data ; we had to find out which dataset had a problem, and after some time we found that the amplitude of the ERA-Interim SST diurnal cycle is not in agreement with observations. It is reassuring that after applying a correction, this had a positive effect on the cloud amounts, as now the diurnal variation of cloud amount is more similar.

Rewritten to : Without this correction, the cloud amount (CA) at night / early afternoon was 78% / 71%, compared to 71% / 71% when using AIRS ancillary data. The correction led to 76% / 73%, closer to the results using AIRS ancillary data.

10. Page 10, Section 2.4.2:

The CO_2 correction appears to be a very relevant change (also visualized nicely in Figure 13. This appears to be one of the most important improvements of the methodology. Should become mandatory in all sounding-based retrievals for climate datasets, in my opinion.

Thank you for the compliment ☺ In our case this was necessary, as the spectral transmissivities came from look-up tables computed for a fixed CO_2 concentration.

Actually, Menzel *et al.* (2016) also use a varying CO_2 concentration adjustment, for a 35-year HIRS cloud climatology.

11. Page 11, Section 2.5, general comment on the ña posteriori cloud detectionò:

The methodology appears a bit awkward compared to many other cloud retrieval methods in that cloud properties are first derived and then a determination whether a FOV is cloudy is carried out as a second step. Most common otherwise is that a cloud screening is done first and then followed by a cloud property retrieval. So, could you confirm that after having performed the cloud property retrieval, all FOVs are still assumed to be cloudy? Does it mean that you will always find a solution to Equation 2? You have already mentioned some problems in finding a distinct minimum for lowlevel clouds (page 7, lines 2-3) but what happens in obviously cloud-free situations?

Actually, we see this method as an advantage, because the method tests if the retrieved values are coherent, whereas most cloud detection methods use many different threshold tests, mostly based on brightness temperatures. We would have liked to adapt the cloud detection which was based on the comparison of temperatures (after correction for water vapour effects) obtained from HIRS to those of the microwave sounding unit MSU (developed for TOVS) to AIRS. Unfortunately, the AMSU channel which sounded closest to the surface did not work from the beginning. Therefore we have developed this method. Indeed, the χ^2 method provides in most cases (95%) a solution. The cloud detection is based on the coherence of spectral emissivities which are calculated using the retrieved cloud pressure. If the

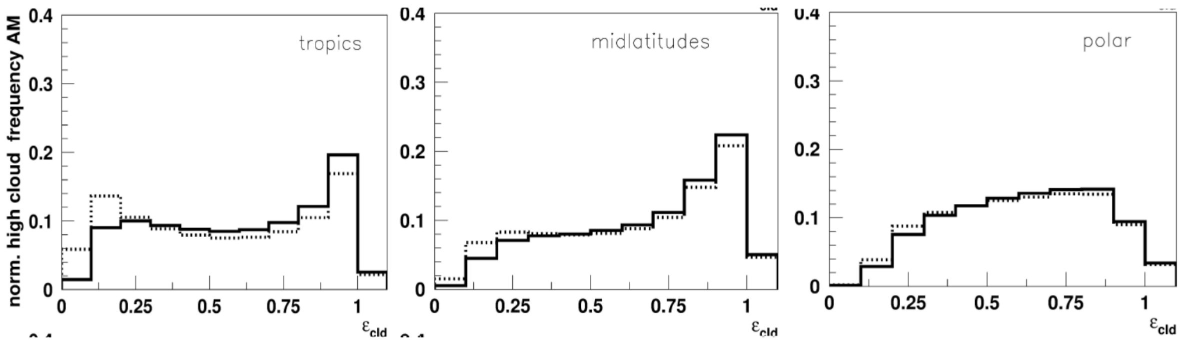
retrieved cloud pressure does not correspond to reality (as for clear sky or partly cloudy situations), the spectral variability gets large, as illustrated in Fig. S1.
We have now moved section 2.5 to section 2.4.3 and have rewritten part of the text.

12. Page 11, Section 2.5, line 16 + lines 20-21, ñmeaning of spectral coherenceò:
I am a bit concerned about the concept indicating that, for a cloud to be identified, the differences between emissivities in the six infrared channels should be small. In this wavelength region we know that the refractive indices of water and ice, respectively, varies considerably. For example, this is one of the fundamental properties that allows separating water clouds from ice clouds in passive imagery (e.g. as introduced by Pavolonis et al., 2005, J. Appl. Meteorol.). This fact would also certainly introduce considerable differences in cloud emissivities depending on if it is a water or ice cloud in addition to variations in optical thickness or partial coverage within each FOV. So, isnât there a risk that the demand on spectral coherence is in conflict with reality? Or are you able to find a balanced and optimized method based on reference observations from CALIPSO-CALIOP data and still retain reasonable resulting emissivity differences? I guess that the access to CALIPSO-CALIOP data here is essential since it would be difficult otherwise (e.g. through detailed cloud model simulations) to find an optimal way here. Please comment.
The multi-spectral cloud detection is indeed based on wavelengths in an interval which is sensitiv to thermodynamical phase and ice crystal sizes. As can be seen in Fig. 3 of Guignard et al. (2012), the relative cloud emissivity difference between 9 μm and 12 μm can go up to 0.3 for small IWP and ice crystal size. However, instead of using a spectral difference, we use a standard deviation between 6 wavelengths, divided by retrieved cloud emissivity. This should be always smaller than 0.15, even in the case of small IWP and ice crystal sizes which produce the largest slope (we have studied that in detail when developing the method in 2010). In this empirical method, the error one makes, if the used cloud pressure does not correspond to the real pressure, is larger, and Fig. S1 (of the supplement) illustrates nicely, that this relative standard deviation is larger than 0.3 for clear sky scenes, while for cloudy scenes distributions the distributions are really narrow, using CALIPSO-GEOPROF to separate cloudy and clear sky scenes.

13. Page 11, Section 2.5, line 25, ñstandard deviationò:
How do you calculate the standard deviation here? Do you use all values in the ñAIRS golf ballò (i.e., 9 values) for the calculation for each wavelength? The current description is not clear enough on this.
It is a standard deviation over all 6 emissivities per AIRS footprint.

14. Page 11, Section 2.5, line 27, ñCALIPSO samplesò:
Unfortunately, here you introduce the use of CALIPSO data without having described what data you actually used (this description comes later in Section 3.1). More clearly, it is not obvious to the reader that you will get three CALIPSO samples in the AIRS golf ball. For this, you need to know that you use 5 km CALIPSO data. Because of the importance of A-train data for your method and study, I am of the opinion that you should have introduced them already in Section 2 on ñData and Methodsò. Can you consider changing this?
Section 3.1 now moved to section 2.4

15. Page 12, Section 2.5, lines 18-19, ñminimum optical depthò:
In the introduction section you mention that with IR vertical sounding data ñreliable detection of cirrus with IR optical depths as low as 0.1ò is possible indicating that this is much better than what can be achieved from other sensors (except from active sensors). I wonder what this restriction in order ñto reduce noiseò means in this context? Have you estimated further the minimum cloud optical depths being detected after introducing this restriction? CALIPSO-CALIOP offers the possibility to do such in-depth studies.
We made this sentence more explicit : To reduce misidentification of clear sky as high-level clouds, only clouds with ε_{cld} Ó0.10 are considered.
Indeed, this came out of a study with CALIPSO-CloudSat :



The above figures present normalized ε_{cld} distributions of high-level clouds, after multi-spectral cloud detection, but leaving clouds with $0.05 < \varepsilon_{cld} < 0.10$ as clouds, separately for cloudy scenes defined by GEOPROF and CALIPSO (full line) and for all scenes (dotted line). The first bin includes scenes with $0.05 < \varepsilon_{cld} < 0.10$; in the tropics this bin has more clear sky than high-level clouds. Therefore we have moved the threshold to 0.1. As the contribution of the first bin is small compared to the integral, this seemed a reasonable choice.

16. Page 13, Section 3.1, lines 16-19, ñCALIPSO and CloudSat dataò:
This requirement should mean (?) that you require that both CloudSat and CALIPSO say it is cloudy. But what about the fact that CALIPSO sees much more of the very thin cirrus clouds being available? Does it mean that these cirrus cases are not included in your evaluation study despite the fact that you several times have emphasized the capability of your method to detect very thin cirrus? Or is it different for studies of cloud amount (as indicated by description in lines 7-15) and cloud top height? Please comment!
We use CloudSat-lidar GEOPROF data, which detect a cloud layer when either CALIPSO or CloudSat detect a cloud layer (footprint 2.5 km x 1.5 km), and to add a different sampling (and because we needed a few other variables like COD) we use the CALIPSO 5km cloud data. In the latter we exclude subvisible cirrus (admitting only clouds detected with horizontal averaging < =5 km) for the evaluation, as we know that IR sounders are not sensitive to those. This corresponds to clouds with COD > 0.05 to 0.1, according to Winker *et al.* (2008).
Then, we require that both samplings detect a cloud, just to be sure that the sampling is coherent. These data are then used for all studies in this paper. We have tried to explain it better in the new section 2.4 :
é .The CALIPSO cloud data also indicate at which horizontal averaging along the track the cloud was detected (1 km, 5 km or 20 km), which is a measure of the COD. As in Stubenrauch et al. (2010), for a direct comparison with AIRS cloud data, we use clouds detected at **horizontal averaging over 5 km or less**. **This corresponds to clouds with visible COD larger than about 0.05 to 0.1** (Winker *et al.*, 2008). The scene type of an AIRS footprint is estimated as cloudy when the CALIPSO sample as well as the GEOPROF sample include at least one cloud layer. Clear sky is defined by cloud-free CALIPSO and GEOPROF samples within the AIRS footprint.

17. Page 13, Section 3.1, line 23, ñunderestimated CODò:
Just for your information: The latest version of the CALIPSO-CALIOP dataset (version 4.1) gives indeed higher CODs. This change can possibly be connected to what you write here (currently I do not know the details behind this change).
Thanks for this information!

18. Page 14, Section 3.2, lines 2-3, ñagreementò:
I have to ask you to specify better what you mean by ñagreementò. There are so many skill scores around so youôd better be strict in describing exactly the measure you use. I guess you refer to what is normally called ñHit Rateò which is the number of correct cloudy AND clear cases divided by the total number of cases.
Indeed, it is the hit rate which we have calculated. We have changed this in the text :
The hit rates between the ôa posterioriô cloud detection and the CALIPSO-CloudSat cloud detection are 85% (84%) over ocean, 82% (79%) over land and 70% (73%) over ice / snow.

19. Page 14, Section 3.3, generally on results in Figure 4 (Page 40):
First, please revise the wording of the caption of this figure. The first sentence here is too complicated and the description should possibly be made more clear (the same is actually true for Figure 5). Also make clear (in all figures) what you mean by ***ñ****1:30 LT****ò*** *(AM or PM??). The question raised in the previous comment 16 remains: Are thin cirrus detected by CALIPSO but not by CloudSat part of this study or not?*
If not, what can be said about the quality of these retrieved cloud heights (as compared to CALIPSO data alone)?
1 :30 is 1 :30AM, as defined in section 2.1 (1 :30 and 13 :30) ; however, as this leads to confusion with American readers, we will change this in the whole paper to 1 :30AM and 1 :30PM etcé
As explained before, for this comparison CALIPSO cloud data with COD > 0.05 to 0.1 are used.
The other referee suggested to take out the right panels of Figure 4 (which look very similar to the results published in Subenrauch et al. 2010). We have worked on all figure captions ;
Compared to the publication of Kahn et al. 2008 about the NASA AIRS Science team results of cloud height from Version 5, we show that in both cases, high-level clouds as well as mid- and low-level clouds the height is determined without bias, if one consideres the cloud height given by AIRS as the height of maximum lidar backscatter (Stubenrauch *et al.*, 2010), by the mean layer height (for optically thin clouds) or the mean between cloud top and the height at which the cloud reaches opacity, as shown in Figure S2 (considering mid-p_{cld}), or by $z_{COD0.5}$ (Figure 3).

21. Page 16, Section 3.3, lines 5-24, Figure 7:
Very interesting and impressive results shown here! Results for medium and high clouds are probably quite superior to those being presented from passive imagery in other CDRs. Only for low-level clouds we still see quite some discrepancies which is understandable for several reasons. This indicates that the best representation of the true vertical distribution of cloudiness in a climate sense could be a combination of sounding and passive imagery data. Do you agree? Maybe you should mention this. Interesting is that problems for low clouds for sounding applications is not showing up very clearly later in Figure 9, except possibly during night for the land-ocean difference. Maybe you should explain why?
Indeed, a combination of IR sounder and passive imagery would increase the quality during day. During night, sounding provides better results, though the large footprints are a handicap for the identification of low-level cloud fields (as shown in the analysis of new Fig. 5). The concept of the CIRS retrieval was guided by the goal to create a cloud climatology with small biases, also for low-level clouds. Indeed, the noise is much larger for low-level clouds than for high-level clouds, but the biases are small compared to other IR sounder cloud climatologies. The comparison with CALIPSO-CloudSat comes to its limit in the analysis of new Fig 5, as the size of the footprints is very different.

20. Page 15, Section 3.3, line 9, ***ñ****coincides****ò****:*
See previous comment 2.
22. Page 16, Section 3.3, line 32, ***ñ****coincides****ò****:*
See previous comment 2.
26. Page 26, Section 6, line 1, ***ñ****coincides****ò****:*
See previous comment 2.
Replaced by ócan be approximatedô

23. Page 18, Section 4, lines 15-16, ***ñ****sensitivity of lidars****ò****:*
You write that ***ñ****active lidar is the most sensitive****ò****. Quite true but you haven****ôt*** *explained whether CALIPSO results in Figure 9 are already* ***ñ****filtered****ò*** *(so that the thinnest clouds as given by the original CALIOP CLAY product are removed) or not. Has there been any filtering of* ***ó****sub-visible clouds****ô*** *(I assume there has)? This is a relevant question to ask also for the statement in the Conclusions section on page 25, line 25. We need to know exactly what is the used CALIPSO dataset used as reference!*
In section 4, the CALIPSO L3 data of the GEWEX Cloud Assessment data base are used ; two teams have provided their data, with the main difference by vertical (CALIPSO-GOCCP) or horizontal averaging (CALIPSO-ST), as mentioned in the text. The details of the GEWEX Cloud Assessment data base are found in (Stubenrauch et al. 2013) and especially in the WCRP report (Stubenrauch et al. 2012),

where each team gave details how they created the L3 data. As I remember, CALIPSO-ST includes subvisible cirrus, which explains the larger CA, compared to all other datasets.
In section 3, L2 products have been used, as descibed in the new section 2.4.

*24. Page 21, Section 4, line 4, Figure 14, **ñ**Seasonal cycle of cloud temperatures**ò**:*
How come there is a rather large consensus between different methods when studying cloud temperatures for the polar areas (leftmost and rightmost columns) when the spread is very large when it comes to cloud amount (top row of the same columns)? I suspect it is an indication of that cloud temperatures and surface temperatures are very similar here. This implies (in my opinion) that the separation of cloudy and cloud-free areas is indeed not very accurate. So, where is really the truth as regards polar cloudiness? Apart from this reflection, I consider Figure 14 as a very nice compilation of global cloudiness and its variation.
This actually shows that cloud amount, depending on thesholds, might be different by 10%, while the averages of retrieved cloud properties, which only can be given when a cloud is detected, are more similar. (Missing 10% does not mean that the average properties of the clouds are completely different).
In addition the polar regions are to be considered with care, as written in the discussions : the CALIPSO data does not conform with the other data sets in the GEWEX Cloud Assessment data base, because they exclude measurements from 1:30PM during polar night (polar winter) and from 1:30AM during polar day (polar summer).
As a similar figure was already published in Stubenrauch *et al.* (2013) (though not CT), we moved this Fig. to the supplement, in order to shorten the paper, and as suggested by referee#1.

*25. Pages 21-24, Section 5, **ñ**beyond scope??**ò***
In my opinion, Section 5 feels like out of scope of this study. Although introducing highly interesting topics (especially section 5.2), this work would benefit from being presented as a separate (or companion) publication. This manuscript is very, very long and it will put the readers (as it truly has for reviewers!) to a real test when digesting it. I would say that especially section 5.2 on the ENSO effects and its coupling to cloud/radiation feedbacks also requires a different category of expertize for reviewing it with more focus on modelling and studies of climate change and climate feedback effects. Consequently, I have not provided specific comments on this section and I suggest that it is removed for the shortening of this paper.
We do not agree with the suggestion of a complete removal of section 5 ÔApplicationsÕ, as the presented method is not new and one of the goals of this article was to present scientific applications (as indicated in the title).
However, we have considerably shortened the section by removing the introduction on ENSO and the discussion about Fig. 16 as well as Fig. 16 itself.
Since the results similar to those presented in new Fig. 12 have recently been published using other data sets, it would be difficult to use the presented material in a separate publication. We plan to work on a more complex analysis to pursue this subject further, but we think itÕs important to present these results in the current publication.

27. Page 24-27, Section 6, general comment:
A very comprehensive and good summary of the content of the paper. However, it could be shortened (page 26, lines 14-32) as a consequence of comment 25 above.
Thank you ! We have revised the part considering section 5.

Technical corrections
1. Page 1, Abstract, line 11-14:
The current introductory sentences assumes that the reader already knows about the LMD cloud retrieval scheme. I suggest a slight reformulation to make it less unclear, e.g. like the following
***ñ**The Laboratoire de Météorologie Dynamique (LMD) cloud retrieval scheme CIRS (Clouds from IR Sounders) has been adapted to cope with any Infrared (IR) sounding instrument. This has been accomplished by applying improved radiative transfer calculations as well as by introducing an original*

*method accounting for atmospheric spectral transmissivity changes associated with varying CO2 concentrations***ò**.

This is not fully correct, as the cloud retrieval developed in the 1990ôs did not have the name óCIRSô; this name corresponds to the adapted version.

We have rewritten the beginning as:

Global cloud climatologies have been built from 13 years of Atmospheric IR Sounder (AIRS) and 8 years of IR Atmospheric Interferometer (IASI) observations, using an updated Clouds from IR Sounders (CIRS) retrieval. The CIRS software can handle any Infrared (IR) sounder data. Compared to the original retrieval, it uses improved radiative transfer modelling, accounts for atmospheric spectral transmissivity changes associated with CO_2 concentration and incorporates the latest ancillary data (atmospheric profiles, surface temperature and emissivities).

2. Page 2, Abstract, line 3, ***ñ****5 % asymmetry****ò****:*

Please clarify better what you mean with asymmetry. Does it mean that there is generally 5 % more high clouds in the Northern Hemisphere? I assume this is what you mean (supported also by Figure 10) but you should make it crystal clear for the reader in the Abstract!

Rewritten as :

The 5% annual mean excess in upper tropospheric cloud amount in the Northern compared to the Southern hemisphere has a pronounced seasonal cycle with a maximum of 25% in boreal summer, in accordance with the moving of the ITCZ peak latitude to a maximum of 10°N.

3. Page 2, Section 1, line 17, ***ñ****properties****ò****:*

Do you really mean ***ñ****properties****ò****? I would rather say* ***ñ****cloud detection****ò****.*

Yes : we meant here that in addition to identification (which means detection), also their properties (height and emissivity) are well determined (even better than those for low-level clouds)

4. Page 2, Section 1, line 32, ***ñ****determine****ò****:*

Like the previous comment, I am not sure about the correct wording here. The word ***ñ****determine****ò*** *is very strong and almost indicates that the CALIPSO and CloudSat satellites together are creating/defining the clouds. Rather, you should express that they* ***ñ****are capable of observing the cloud vertical structure***ò**.

Changed according to suggestion

5. Page 3, Section 1, line 5, ***ñ****the cloud retrieval method****ò****:*

Be a bit more specific, e.g. write ***ñ****the evolution of the original cloud retrieval method***ò**.

changed

6. Page 3, Section 1, line 9, ***ñ****radiative transfer****ò****:*

I think you should write ***ñ****radiative transfer calculations****ò*** *or* ***ñ****radiative transfer modelling****ò****. To only write* ***ñ****radiative transfer****ò*** *is too general and (I guess) just a shortening of more correct terms.*

changed

7. Page 3, Section 1, line 11, ***ñ****initial****ò****: See 5 above (consider using same notation).*

Changed to original

8. Page 3, Section 1, line 11, ***ñ****radiative transfer****ò****: See 6 above (consider using same notation).*

changed

9. Page 4, Section 2.1, line 11, ***ñ****The NASA Science team****é*** *.****ò****:*

I would recommend to start a new paragraph here to increase the readability.

done

10. Page 4, Section 2.1, line 15, ***ñ****Susskind et al, 2003****ò****:*

I see inconsequent reference formulations on several places in the manuscript. When you make a direct reference to other publications directly in the text (like here) you should (according to my experience) preferably write: ***ñ****The methodology is essentially unchanged from that described in Susskind et al. (2003).****ò*** *You have done this correctly in other places (e.g., Page 5, line 27). I think you should be*

consistent here. Use the formulation above when specifically discussing a publication and use reference in parenthesis when not making a direct statement of the referred publication (a ***ñ****softer****ò*** *reference). Check also the following references for the same reason:*
- Page 4, line 27
- Page 6, line 5
Thanks, all changed

11. Page 4, Section 2.1, line 20, ***ñ****shortwave window channels****ò****:*
Please write ***ñ****shortwave infrared window channels****ò*** *since* ***ñ****shortwave****ò*** *most often is reserved to define visible channels.*
changed

12. Page 4, Section 2.1, line 22, ***ñ****partial cloud cover****ò****:*
A better formulation is probably ***ñ****under partially cloudy conditions****ò****.*
changed

13. Page 4, Section 2.1, line 24, ***ñ****snow or ice****ò****:*
Maybe a better formulation is ***ñé*** *snow or ice covered surfaces also provided by NASA L2 data****ò****.*
changed

14. Page 4, Section 2.1, line 26, ***ñ****ideology****ò****:*
I would suggest using the term ***ñ****concept****ò*** *rather than* ***ñ****ideology****ò****.*
changed

15. Page 4, Section 2.1, line 27, ***ñ****and allow****ò****:*
I suggest replacing this with ***ñ****which allows****ò****.*
Rewritten to : The CIRS cloud retrieval allows cloud levels up to 30 hPa above the tropopause.

16. Page 5, Section 2.2, line 1, ***ñ****12 km****ò****:*
Is the 12 km valid for each individual footprint or the 2x2 array?
For each individual footprint, clarified in text

17. Page 5, Section 2.2, line 9, ***ñ****the cloud retrieval****ò****:*
You should write ***ñ****the CIRS cloud retrieval****ò****.*
changed

18. Page 5, Section 2.2, lines 9-10, ***ñ****retrieved atmospheric profiles****ò****:*
Be more specific. You should write ***ñ****IASI-retrieved atmospheric profiles****ò****.*
changed

19. Page 5, Section 2.2, line 15, ***ñ****Therefore****ò****:*
You should not start a new paragraph here if you refer directly to what was written in the previous sentences. Make it also very clear that you never (well, not in time for your development) got access to EUMETSAT Version 6 data otherwise this statement appears rather strange.
We could have gotten access after the development and evaluation of the cloud climatologies were nearly at the end. Since it would have taken another year to build the ancillary data from this data set and evaluate again the IASI cloud climatology (also in combination with AIRS), we opted for ERA-Interim ancillary data to build the combined AIRS-IASI cloud climatologies.
As the sentence about V6 EUMETSAT retrievals seems to cut the flow, we took it out.

20. Page 5, Section 2.2, line 21, ***ñ****same source****ò****:*
I guess you rather mean a ***ñ****less instrument-dependent source****ò****?*
We think it is more óretrieval quality-dependent sourceô, but this would be difficult to write, as the different Science Teams are doing the best with the fundings they have available. (In the case of NOAA for example, the team had to move working on CrIS).

21. Page 6, Section 2.3, line 1, ñproxyò:
I donôt like the word ñproxyò in this context. It indicates that it is a kind of simulation or approximation of the real vertical velocity. The vertical pressure velocity ɤ is just another formulation of the vertical velocity which arises when you use pressure as your vertical coordinate instead of the standard geometrical height in meters. So, to my knowledge, itôs the ñreal thingò and not a ñproxyò.
But I guess you refer to the fact that the direct calculation of ɤ is difficult without making approximations. The most common here is the geostrophic assumption leading to the so-called ñɤ-equationò. In this sense, I guess you may be correct in interpreting it as an approximation. But still, present day NWP models are capable of calculating ɤ so I just wonder what value you are using here? On the other hand, the approximated value at the 500 hPa level is probably quite accurate anyway (conditions here are largely quasi-geostrophic on the large scale) so perhaps this discussion is less important. Anyway, give it a thought.
We needed the vertical velocity for the interpretation in the ENSO analysis. Since Fig. 16 and its interpretation is taken out according to the referees suggestion, this sentence is also taken out.

22. Page 7, Section 2.4, line 12, ñariseò:
Maybe reformulate to ñthese cases occur in about 7 to 15 % of all casesò?
Changed to : these cases occur in about 7 to 15 % of all cloudy cases

23. Page 8, Section 2.4.1, line 14, ñless than ..?..ò:
Strange formulation. Youôd better write ñ0.99 for wavelengths less than 10 ɛm and 0.98 for wavelengths larger than 10 ɛmò.
Changed to : the surface emissivity is set to 0.99 for λ_i < 10 µm and 0.98 for λ_i ≥10 µm

24. Page 13, Section 3.1, line 6, ñspatial resolution CALIPSOò:
Shouldnôt it be ñ5 km x 0.3 kmò? I thought the basic FOV of CALIOP was 300 meter.
I have understood that the diameter of the spots is 90m, and the sampling along track is 333 m.
For example : https://calipso.cnes.fr/en/CALIPSO/lidar.htm or Winker *et al.* (2009), p. 2312

25. Page 15, Section 3.3, Figure 5 (Page 41):
I suggest that you try to include some additional explanatory features or legends in the figure (e.g., legend with the three coloured dots explained). To look for all explanations in the caption is not very reader-friendly. Try to speed up the correct interpretation of figures with the use of more graphical legends or marks. This remark is probably valid for many other figures in the manuscript.
We have taken into account the refereeôs suggestion and revised all figures accordingly.

26. Page 15, Section 3.3, line 27, ñConsideringé ò:
I suggest starting a new paragraph here in order to avoid too long chunks of text (unnecessary tiring for the reader).
This whole paragraph has been rewritten (as Fig. 6 has been taken out, and Fig. 5 has been rebuilt with medians and interquartiles to show the width of the distributions within the same figure). We hope that it is now much easier to read.

27. Page 15, Section 3.3, line 28; Figure 6 (Page 42):
In the caption you describe one of the curves as ñbroken lineò. I am not sure whether this is the most common way of describing such a curve. More often the term ñdashed lineò is used. Consider changing to ñdashedò. This suggestion is valid for many other figures in the manuscript.
Thanks ; changed everywhere ; though dashed lines seems also to exist, at least according to google ;)

28. Page 16, Section 3.3, lines 28-29, ñheight of CODò:
Semantically, it sounds strange (or even incorrect) to express COD as representing a height. Of course, I understand what you mean but it can actually be misinterpreted. Since you have already defined zCOD0.5 why not use this terminology here, e.g. ñthe retrieved cloud height exceeds zCOD0.5 for optically thin clouds while it is lower than zCOD0.5 for optically thick cloudsò.

This is obvious from the figure, but we want to stress the following :
In that case, z_{cld} of thin cirrus should be approximated to a height at which COD reaches a value < 0.5 and z_{cld} of opaque high clouds to a height at which COD reaches a value > 0.5.

29. Page 20, Section 4, line 17, "three CIRS datasets?"
It is not obvious what three datasets you mean (not explained in text)! Please clarify.
three CIRS climatologies (AIRS, using AIRS-NASA and ERA-Interim ancillary data, as well as IASI, using ERA-Interim ancillary data)

References

Guignard, A., C. J. Stubenrauch, A. J. Baran, and R. Armante, Bulk microphysical properties of semi-transparent cirrus from AIRS: a six year global climatology and statistical analysis in synergy with geometrical profiling data from CloudSat-CALIPSO, *Atmos, Chem. Phys.*, **12**, 503-525 (2012)

Lamquin N., C. J. Stubenrauch, and J. Pelon, Upper tropospheric humidity and cirrus thickness: a statistical analysis using one year of collocated AIRS-CALIPSO data, *J. Geophys. Res.*, 113, D00A08, doi:10.1029/2008JD010012 (2008).

Menzel, W.P., R.A. Frey, E.E. Borbas, B.A. Baum, G. Cureton, and N. Bearson, Reprocessing of HIRS Satellite Measurements from 1980 to 2015: Development toward a Consistent Decadal Cloud Record. *J. Appl. Meteor. Climatol.*, 55, 2397–2410, https://doi.org/10.1175/JAMC-D-16-0129.1 (2016)

Stubenrauch, C. J., N. A. Scott, and A. Chédin: Cloud Field Identification for Earth Radiation Budget Studies: I) Cloud Field Classification using HIRS/MSU Sounder Measurements. *J. Appl. Meteor.*, 35, 416-427 (1996).

Stubenrauch, C. J., A. Chédin, R. Armante, and N. A. Scott: Clouds as Seen by Satellite Sounders (3I) and Imagers (ISCCP): II) A New Approach for Cloud Parameter Determination in the 3I Algorithms. *J. Climate*, 12, 2214-2223 (1999).

Stubenrauch C. J., F. Eddounia, and L. Sauvage: Cloud heights from TOVS Path-B: Evaluation using LITE observations and distributions of highest cloud layers. *J. Geophys. Res.*,110, D19203,doi:10.1029/2004JD005447 (2005).

Stubenrauch C. J., A. Chédin, G. Rädel, N. A. Scott, and S. Serrar: Cloud properties and their seasonal and diurnal variability from TOVS Path-B. J. Climate, 19,5531-5553 (2006).

Stubenrauch. C. J.. S. Cros. A. Guignard, and N. Lamquin, A six-year global cloud climatology from the Atmospheric InfraRed Sounder aboard the Aqua Satellite: statistical analysis in synergy with CALIPSO and CloudSat. *Atmos. Chem. Phys.*, 10, 7197–7214 (2010).

Winker, D., Getzewitch, B., and Vaughan, M.: Evaluation and Applications of Cloud Climatologies from CALIOP, Proc. Int. Laser Radar Conference (ILRC), 2008.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., and Powell, K. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, J. Atmos. Oceanic. Technol., 26, 2310-2323 (2009).