Reply to review #1:

We thank the reviewer for the helpful for comments. Based on the comments of both reviewers, the paper has been substantially restructured and revised. In particular, the discussion of technical details has been reduced (with a lot of material being shifted into the electronic supplement), while the discussion of new scientific insights has been extended. Please find below the point-by-point response.

General comments:

After reading the abstract and introduction, it appears that MIPclouds will add cloud retrieval capabilities to the MIPAS mission which so far has not provided cloud and aerosol information in their level 2 products. This would be a very important contribution, but later we learn that Hurley et al. (2011), and earlier work by the first author of this paper have already published about this.

The introduction includes now a paragraph where the reference to former publications and the objectives of this specific piece of work (intention) are explained in more detail. The manuscript has been checked and changed where the reference to former and parallel publication was unclear. We are confident that we have explained better in the new version how far the paper is complementary and that it is not only focusing to the ADP quantity.

New paragraph in the introduction:
"*To date, the focus of most scientific analysis of MIPAS cloud observations is either on the analysis of PSC (Spang et al., 2005a, Höpfner et al., 2006a/b, Höpfner et al., 2009) or technical aspects of the cloud parameter retrieval (Spang et al.. 2005, Greenhough et al., 2005, Hurley et al., 2009, 2011, Sembhi et al., 2012). For the first time this paper presents a combination of the best suited detection methods and new developed cloud parameter retrievals in one processing system with corresponding data products. A stringent cross-validation of the applied detection methods with cloud products of other satellite instruments is performed to optimise, quantify, and validate the detection sensitivity of the new, combined approach. First attempts for a more adequate comparison of cloud products of different instrument types and viewing geometries are presented and applied to the data. Finally, this paper introduces to the scientific community a new and – after the completion of the processing of the full measurement period – a 10 year data record of cloud products for future comparisons with model data (e.g. GCM or ECMWF re-analyses) or with combined cloud datasets of operational nadir instruments.* "

The introduction also promises the introduction of a robust cloud detection threshold. If understood correctly, this is one of the utilities of the newly introduced ADP. The detection threshold is introduced in terms of ADP. However, while I suspect that this is more or less the reasoning of the authors, I don't see this clearly stated anywhere.
The ADP may also be used in cross-platform sensitivity comparisons, although the "translation" to nadir-viewing sensitivities is not well (or not at all) described.

We agree with the interpretation of the reviewer. This point is discussed in Section 4.3.2 which has been revised to bring this point across more clearly. Further, it is stated now in the summary and in the abstract, with a the clarification that ADP is *"used for a more objective definition of detection thresholds of the applied detection methods."* .
A concept of how to compare nadir and limb is now presented in a new subsection at the end of the GEWEX comparison:

5.3.3 A concept on how to compare limb and nadir cloud occurrences

"Usually, in a COF analysis nadir instruments underestimate the true COF due to a lack of sensitivity. Limb based COF analyses will overestimate the amount of clouds seen in the nadir

direction due to the limb path smearing effect (Sect. 4.3.1). The ATSR `MIPAS-like' product is a first step to make limb and nadir measurements better comparable.
Similar to the blind test in Sect. 4 where the `true state' ADP is computed from ECMWF data and compared with the retrieved CTH and ADP, it would be better to raytrace the MIPAS limb paths through the horizontally high resolved cloud parameter fields of the nadir instrument and to estimate, for example, an effective limb CTH from the nadir CTHs along the LOS. A similar approach was successfully applied to global nadir BTR composites to create realistic high resolved fields of CTHs (Adams et al., 2009). The study quantified the benefit of limb cloud imaging for infrared limb sounding of tropospheric trace gases.
Additional nadir information on R_{eff}, IWC or vertical IWP would offer the potential to estimate a limb equivalent ADP from the nadir data. Consequentially, the most realistic approach to make cloud occurrence frequencies from nadir and limb measurements comparable would be based on retrieved ADP values for both instrument types, even if the nadir information is only a rough estimate (e.g. as a first approximation any ADP > 0 is equivalent to a cloud detection in the limb). Detailed analyses and refinements will be necessary to test and quantify the benefit of the new concept. However, the results of the blind tests in Sect. 4 already show that the concept should be applied for comparisons of limb measurements and 3D model data with incorporated cloud physics."

While I indicated "minor revisions" above, I think that quite a bit of work (between "minor" and "major") will be necessary to make the manuscript flow better. When revising the manuscript, I would also strongly revise to shorten it, i.e., removing unnecessary sections that don't contribute to the main core of the paper, while extending/adding sections that do help the core statements.

In response to this comment, the paper is revised and shortened now. Technical details (e.g. altitude correction and cloud type classification) are moved to the supplement material, excessive details or less important comparison have been deleted from the manuscript or are just cited.

Major comments:

* The 3D aspect, stressed various times in this paper, should be more clearly demonstrated by showing some of the 3D/2D cloud fields that were used as input in, e.g., the blind tests. Were the input fields as shown in Figure 7 used by SHDOM which calculated the radiance spectra, which then formed the basis for the blind test?

The discussion of 2D and 3D was not consistent in the previous version of the manuscript. The aspect of 3D is only used in the analyses when transects of MIPAS orbits are applied through the ECMWF 3D model output. The radiative transfer calculations are 'only' in 2D. This is now corrected. The 2D fields of Fig. 7 are input for the SHDOM, this is now stated more clearly and is also mentioned in the figure caption.
Generally, the sections on the blind test retrieval and the corresponding setup of the 'true state' cloud parameter for the comparison with the BTR results are revised in the sense of better clarification for what the extracted parameters are used. For example, the term 'true state' ADP is now introduced to highlight that these quantity is based on the best of knowledge true state of the atmosphere. A 'perfect' retrieval should reproduce the corresponding ADP values.

* Describe the outcome of the blind test in terms of 3D variability. In the abstract and introduction, the promise is given that ADP will address these problems, and that is good, but has to be shown with examples (see comment above).

As already mentioned above, the 3D variability is considered by the 2D transects through the 3D model fields of IWC and LWC. This is a much more realistic approach than the usually assumed 2D homogeneity in radiative transfer calculations, with homogeneous cloud fields along the whole limb path in the tangent height layer.

The ensemble of the ECWMF based ‘true state’ ADP includes all these in-homogeneities of the cloud field integrated along the line of sight and is the best suited parameter for a comparison with the limb measurements. The measurement technique and the applied detection methods have an unambiguous sensitivity with respect to ADP but not to IWC. We revised a paragraph in Sect. 4.3 to point out the 2D-effect more clearly:

“The effect of using more realistic 2D cloud distributions instead of idealised 2D-homogeneity becomes obvious when comparing the original IWC fields (Figure 7) and the limb-integrated ‘true state’ ADP fields (Figure 8). Even small extended cirrus layers as in Scenario 2 (at ~12 km and between -5° and 2° latitude) affect a significantly larger area of potential limb measurements in the horizontal and vertical domain than suggested by the IWC distribution.”

* Figure 2: I think this Figure shows a very interesting result that should be further discussed. While the bi-modality of thick/thin clouds is mentioned, the implications are not. To my knowledge, this bi-modality has never been discovered in climatologies based on nadir-viewing instruments. Discussing this in more detail could be one added science aspect of this paper. In the caption, "number density" should be renamed to "probability density", the same on page 33022,l18.

The observed bi-modality is not related to thick/thin clouds but to thick/no cloud conditions. Optically thin clouds are filling the range between the two maxima. This fact is described in section 3.1.1. We modified the corresponding text and changed the figure caption to avoid misinterpretation.

* Figure 4: The meaning of the confidence classes "disputable", "likely" is not defined. Explain. Also, add "The" before "Bottom figure"

A reference to Table 3 with the definition of the confidence classes is now given in the figure caption.

* Figure 9, and discussion around it. Explain better what was actually done here. Were ECMWF fields used as input to radiative transfer calculations in the context of a blind comparison? Is there a "true" ADP that the reader can compare to? Otherwise, I see no point in figures 9 or 10.

We changed and added various subsection in section 4 (Validation on simulated data) to point out more pronounced which parameters are described in the so-called ‘true state’ of the atmosphere, where these are retrieved from, and where these parameter are used to validate the blind test results. The term ‘true state’ is now always used in the manuscript for the ECMWF based input parameter (e.g. ADP) for the radiative transport calculations for the blind test data and the following blind test retrieval. The figure caption is changed accordingly.

* The same is true for Figure 8; while the "truth" (i.e., IWC) is given in Figure 7, the outcome in terms of ADP cannot be compared to any truth unless ADP is directly derived from the input fields, circumventing the algorithm, and *then* retrieved by the algorithm, unless I am missing something here.

See item above. The figure caption is also changed accordingly.

* In the interest of making the manuscript easier to digest, I recommend removing Figure 20 and 21

We followed the reviewers suggestion and moved the figure now to the supplement material. For reducing the amount of technical information in the paper, we also moved the technical sections of the appendix and the altitude correction (section 2.1) into the electronic supplement.

* What is a "naive" Bayes scheme? (p33030,l10)

The term ‘naive’ is now explained.

“The latter method is a simple probabilistic classifier based on applying Bayes’ theorem with strong (‘naive') independence assumption (Hanson et al., 1991).”

The Appendix C (classification methods) is moved to the supplement to shorten the manuscript.

* p33038,l5-7: Unclear

The SUM_CLOUD information is now reworded and moved to Sect. 3.2.1, where it is better suited and directly linked to the CTH, CTP and CTT parameters:

“For a simple use in scientific applications combined (summary) information of all methods for each macroscopic parameter (CTH, CTP and CTT) is introduced in the processing (prefix: SUM_CLOUD). Currently, the parameters are dominated by the optimal estimation retrieval result. If there is a successful retrieval then this is used for the SUM_CLOUD information, if this is not the case then weighted CTH information of all other detection methods is applied in a manner similar to the cloud confidence parameter in Section 3.1.4. The SUM_CLOUD information is part of an ongoing optimisation procedure for the definition of the best possible combination of different detection methods based on the validation results (see section 5).”

* p33038,l21, and following lines. Use more accurate language than saying "FOV case". What is meant is that unlike in the other calculation, the FOV is taken into account in these blind test. Unclear what the next lines convey.

We change the term ‘FOV case’ to MIPAS-FOV blind test retrieval. Details on physical background were already presented in the last paragraph of section 4.2.

* The discussion of the threshold on p33039,l12 is only qualitative and should be made quantitative.

We already did a more quantified analysis of the detection threshold by the PDF analysis in the following sentences and paragraphs of this subsection. The PDF gives the information how frequently the detection methods is able to detect a cloud with a corresponding true-state ADP value in the blind test retrieval. We revised the paragraph and the figure caption for better explanation of the PDF analysis.

* p33039,l23 and following. Explain what the percentages refer to: Likelihood, fraction of retrieved spectra?

We included: ‘*up to 99.5% of all true state cloudy profiles (ADP>0) for the SUM_CLOUD parameter*’. The term ‘true state’ is introduced above and refers to the defined input parameter of the radiative transfer calculation, which represents the ‘real’ atmospheric state from model data (ECMWF).

* Structural comment: For example, at the end of section 5.1.1 one would expect the presentation of data and plots, the same is true after 5.1.2 and 5.1.3. Instead, this is delayed until later when a confusing amount of plots is presented. This could be restructured by showing the data earlier, here in these sections where they belong.

We followed the suggestion by the referee changed the structure of section 5 by moving the instrument specific results closer to the instrument description. We reduced the number of information

* p33046,l6-8: Why can subvisible and opaque clouds not be separated by MIPAS? After all, they should be at opposite ends of the ADP range!

Correct, the ADP values correspond to different optical thicknesses, but the MIPAS measurements have also an upper detection limit for ADP (respectively a lower threshold for CI) where the contrast in sensitivity is lost. This threshold e.g. for CI ~ 1.2 is not identical with the SAGE II extinction threshold for opaque clouds in the visible wavelength region. Therefore the differentiation of opaque and thin clouds is so far not installed in the MIPAS detection. A comment is now given in the manuscript.

* Table 4 - These results are discussed as "good" agreement. I think table 4 needs to be better explained in general. To the untrained eye, it appears that not only the table reveals substantial differences between the MIPAS techniques, but also substantial differences when applying the technique to different data (sets). This is an important table that needs to be interpreted much better.

Table 4 is now explained in a different way. For better clearance of the main subject of the table we deleted the number for the less sensitive detection method CI_OPER from the table. Differences between the MIPAS techniques are not a surprising fact because the sensitivity is different. This is already presented in detail in other sections and figures.

We agree with the referee, the table also suggest not always "good agreement" between the instruments. Therefore explanations for a good agreement and discrepancies are now presented in Sect 5.1.5:

"Table 4 summarises the coincidence statistics. All combinations of coincidences are accounted for: cloudy to non-cloudy for SAGE II and MIPAS, respectively, both cloudy, and both non-cloudy. Percentages are given with respect to the total number of coincidence events (bottom row).

For high northern latitudes the agreement is excellent, only 12% and 13% of the coincident profiles are flagged as cloudy when the other instrument detects no cloud. At low latitudes there is less agreement. Yet the number of coincidences is also significantly smaller than at high and mid-latitudes and consequently the statistical significance is reduced.

Coincidences at high southern latitudes are mainly observed during polar winter and are dominated by high occurrence rates of PSC observations in the MIPAS measurements. The detection of PSCs is not covered by the SAGE II detection SVC flag. Consequently, the discrepancies for the cloud/no-cloud events are not meaningful. But there is a significant correspondence for non-cloudy conditions between the instruments. No clouds are detected by SAGE II when MIPAS predicts non-cloudy conditions for the coincidence.

In the tropics, MIPAS detects in 39% of the coincidences a cloud where SAGE II seems cloud-free, whilst for 8% of the profiles SAGE II detected a cloud for cloud-free conditions with MIPAS. The imbalance is basically caused by MIPAS detections with CTHs above 15 km. This is indicating a higher detection sensitivity for MIPAS than SAGE II for cirrus clouds around the tropical tropopause (see also Sect. 5.1.4).

At mid-latitudes the percentages of in-conclusive results are reduced, but there are still significantly higher detection rates for MIPAS. In this case the events are not caused by detections around the tropopause (~12 km) but due to detections at lower altitudes (6 and 9 km). In this region the analysis is biased by missing opaque clouds in the SAGE II statistic. The amount of both-cloudy coincidences is higher than in the tropics and indicates generally better agreement for mid-latitudes.

Although both SAGE II and MIPAS use the limb technique, the satellite orbit for the occultation measurements results in a quite different viewing direction than for the emission instrument when probing the same air mass. This fact, together with the relatively large coincidence criteria (400km/4h), may produce misleading results (reduced number of both-cloudy coincidences) especially in the mid-troposphere where the horizontal cloud scales are smaller. These restrictions seem to affect especially the mid- and tropical latitude analysis. An unbalance in the amount of inconclusive (cloudy/non-cloudy) coincidences between the instruments is always an indication for differences in the detection sensitivity or a hypersensitivity of a specific detection method. This becomes obvious for MIPAS in the mid-latitudes and tropopause region of the tropics."

* Summary: Quantify "excellent" correspondence with other instruments. Table 4 makes the reader think otherwise.

We changed 'excellent' to 'good' correspondence and highlighted briefly that differences were observed as well. Some details are given later on in the summary.

Minor comments/technical corrections:

We changed all minor comments and correction accordingly.