Reply to Ref. #1

First of all we want to thank this reviewer for the positive assessment of our manuscript and the contructive and helpful suggestions.
Before we respond to the reviewer comments in detail, we want to give a brief overview on the most important changes of the revised version:

A) We made the main focus of our manuscript more clear.
The main focus of this manuscript is not the further improvement of our original cloud classificaton scheme (Wagner et al., 2014). Actually, the original scheme was almost unchanged. The main aim of our study is the comparison to independent data sets.
Such a comparison to independent data sets was so far missing and is of high importance for the validation of the MAX-DOAS cloud classification scheme. We made this more clear in the revised version (in the abstract, introduction and conclusions).

B) We added the information that no direct and quantitative comparison to the independent data sets is possible.
From the MAX-DOAS observations and the satellite observations not the same quantities are derived: the MAX-DOAS classification results are complex quantities (e.g. categories of broken clouds or continuous clouds), which cannot directly be compared to the independent data sets. Thus only quanlitative conclusions can be drawn, e.g. that the probability for the detection of continuous clouds from MAX-DOAS are found increases if satellite observations indicate high cloud fractions.
Also the temporal and spatial resolution is different: the satellite observations typically cover an extended area, but are valid only for the time of the satellite overpass. Also the AERONET observations are not made at exactly the same location but about 18 km south-west of the Wuxi MAX-DOAS site. And the visibility meter covers only the layers close to the surface.
In spite of these fundamental difficulties, the comparison to the independent data sets is still very important, because from the general dependencies derived from the comparisons it can be concluded whether the classification scheme yields reasonable results or not. Here it should also be noted that to our knowledge, no data set exists, which could be used for a direct and quantitative validation of our classification scheme.
We added this information to the introduction (and also to the abstract and conclusions).

C) We added sensitivity studies with respect to the selected threshold values.
From the comparison to the independent data sets no quantitative conclusions on the accuracy of the cloud classification results can be drawn (see point B above). As a consequence, it is difficult to qantify the accuracy of the cloud classification results. Thus we chose a different way to assess the uncertainties of our classification method: we varied the different threshold values (by ±10%) and studied the corresponding changes of the classification results. Fortunatly, most classification results were only very weakly affected by these variations. This finding indicates that for these quantities the exact choice of the threshold values is not critical. In contrast, for the CI the exact value of the threshold value was found to have a significant impact on the classification results. In particular, it affects the assignment to clear sky conditions with either low or high aerosol. Fortunately, from the compariosn with smultaneously measured AERONET AOD we find that the chosen threshold value was a very reasonable choice. In particular we found that the transition from the categories clear sky with low aerosol to clear sky with high aerosols corresponds to an AOD of about 0.5.

Also the assignment to either broken clouds or cloud holes is strongly affected by the chosen threshold for the CI. But this ambiguity is of minor importance, because both classifications results basically belong to the same cloud category.
We added this informatuon to the new section 3.1.1.

D) We added sensitivity studies of the effect of temporal and spatial averaging
For the comparison of our cloud classification results we averaged the MAX-DOAS results over a period of one hour around the satellite overpass. Also the satellite observations represent averages over extended areas. We investigated the effect of the chosen ranges for the temporal and spatial averaging on the comparison results. Interestingly, the results hardly depend on the selected temporal and spatial averages. In particular the results for different time averages are almost identical. In contrast, a small effect of the spatial averaging on the classification results is found. If the satellite data are averaged over a larger spatial range a more clear assignment of the extreme values is found: The fraction of clear sky scenarios (with either low or high aerosol load) and of cloudy scenarios increases for MODIS observations of small (0 to 10%) and high (90 to 100%) cloud fractions, respectively. These findings indicate that if the requirement of either completely clear or cloudy sky are applied for a larger area the probability that the MAX-DOAS results around the time of the satellite overpass were really clear or cloudy, respectively, increases. The overall conclusion from these sensitivity studies is that the selected standard averaging criteria (temporal averaging: ±30 min around the overpass time; spatial averaging: 0.1 ° x 0.1 °) are well suited for the comparison of the MAX-DOAS results with satellite data. We added this information to the new section 3.2.4.

E) Most of the description of the O_4 analysis (section 2.2.2) was shifted to the supplement.
The old section 2.2.2 contained a lot of very detailed information on the O_4 analysis. To shorten the main part of the manuscript, we shifted most of these technical details to the supplement (including Fig. 5). The new section 2.2.2 contains only a brief summary of the O_4 analysis, which is necessary to understand the remaining part of our study.

General comments
The paper is a continuation of the previous work of Wagner et al. (2014) on using ground-based MAX-DOAS observations for characterizing cloudiness and aerosol presence in the atmosphere. The observations are performed in Wuxi, China, with very different atmospheric conditions than in Cabauw for which Wagner et al. developed their method. The results are compared to other ground-based measurements and to satellite measurements of aerosols and clouds.
The paper is useful as a refinement of the classification method of Wagner et al., and as a comparison of ground-based and satellite measurements of atmospheric conditions regarding clouds and aerosols. Especially sub-pixel cloudiness and the radiance similarity of clouds and aerosols at subpixel scale are important issues for satellite measurements. The interpretation of these effects is aided by using collocated MAXDOAS measurements from the ground. The referencing is very complete. The results shown in Figs. 12-17 are interesting, and deserve a good discussion.

Author reply:

Many thanks for the positive assessment! We added more information about the interpretation of the comparison results in the revised version of the paper (see point B above). We also added information on the effect of spatial and temporal averaging (see point D above).

Main comments

- The paper lacks clarity of text. The paper is too long and too technical. It is important to focus on method and results. The text can be more condensed. In fact, a major rewriting of the paper is necessary.

Author reply:
Already in the original version of our manuscript we had put a lot of 'technical details' into a supplement. In the revised version we also shifted the major part of section 2.2.2 (O_4 analysis) to the supplement. The new section 2.2.2 provides just a brief summary of the most relevant findings (see also general point E above).

- The use of the term radiance in the paper is confusing and often unspecific, e.g. missing information on viewing direction; calibrated or uncalibrated radiances; spectral dependence. The most problematic section of the paper is Sect. 2.2.1. Here in the first sentence radiances are given in units of counts per second: so these are not radiances but measured signals! From measured signals radiances can be derived (in radiance units, e.g. photons or Watts) by calibration. Please stick to common nomenclature.

Author reply:
We agree and changed the first sentence in section 2.2.1 to:
'The detector read-out of the MAX-DOAS measurements directly provides a spectrally resolved signal (in units of counts s^{-1}), which is proportional to the observed radiances (e.g. expressed in units of photons per area, time and solid angle). Here it should be noted that in the following we will use for the sake of simplicity the term radiance also for the measured signal (in units of counts s^{-1}).'

In this section first the calibration of the instrument is beinig performed using an RTM in which a simple Henyey-Greenstein model is used. This is of course too simplistic and can cause large deviations, as shown in the paper. It would be better to use the Mie phase function that belongs to the Aeronet microphysical aerosol retrievals (particle size, shape, refractive index) as a function of wavelength. So this remains to be done. The error due to the empirical calibration technique that is chosen should be assessed as good as possble. (p. 4661). It is essential for this cloud and aerosol classification method that a good calibration procedure is developed.

Author reply:
In principle we agree that a Mie phase function is probably better suited to describe the observed SZA dependence of the measured radiance during the selected clear days (based on the measurements at the nearby AERONET station at Taihu). However, we did not follow such an approach because of two reasons: first, the AERONET station is not fully representative for the MAX-DOAS measurements because of the distance of about 18 km to the Wuxi station. Second, the true aerosols might contain also non-spherical particles, and might also change with time. Thus even a Mie-phase function will probably not well describe the observed SZA dependence.

Moreover, for large SZA (about >50°) the HG phase functions well describes the observed radiances.
But there is a more fundamental aspect, which lessens the accuracy requirements for the radiance calibration: The chosen procedure for the determination of the radiance reference values has only a limited influence on the cloud classification results, because the cloud classification results also depend on the choice of the threshold values. And both steps are directly interconnected: In the case that e.g. the radiance reference value was determined too high, the corresponding threshold values will be also increased and thus the effects on the cloud classification results mainly cancel out.
We added this information to section 2.2.1.

- In the paper a good error analysis is missing. In general, a satellite-ground observation comparison suffers from many potential errors/uncertainties/representativeness differences. These errors should be discussed in the paper.

Author reply:
We agree and added a new sub-section:

3.1.1 Uncertainties of the Classification method
The MAX-DOAS cloud classification scheme is based on a set of thresholds for different quantities. When the values of the measured quantities fluctuate around these thresholds, the resulting sky conditions could be mixed with each other. This effect is referred to as “edge effect”, which is a well-known problem of threshold-based classification schemes. In order to estimate the effect of the specific choices of the threshold values, we compared the cloud classification results for slight variations of the threshold values. In Fig.12 the classification results for the original set of thresholds as well as for the modified thresholds are shown (changed by ±10%.). For most of the quantities, only very small changes of the classification results for the modified threshold values are found. This indicates that for these quantities the exact choice of the threshold values is not critical. In contrast, for the CI the exact value of the threshold value has a significant impact on the classification results. In particular, it affects the assignment to either clear sky with low or high aerosol load. From the results shown in Fig. 12, we can, however, directly conclude that the original threshold value for the CI is probably a very reasonable choice. If e.g. a smaller threshold value was used, almost no measurements are assigned to the condition of clear sky with high aerosol load (see Fig. 12 top), which is in contradiction to the frequent occurrence of rather high AOD at the Wuxi site. The comparison with AERONET AOD observations also indicates that for the used threshold value the transition between clear sky with either low or high aerosols corresponds to an AOD of about 0.5, which is very reasonable. Also the assignment to either broken clouds or cloud holes is strongly affected by the chosen threshold for the CI. But this ambiguity is of minor importance, because both classifications results basically belong to the same cloud category.

- Several figures are unreadable due to small texts. The figures are not nice. Please spend more time on making professional, well readable figures. Please give units for quantities along the axes.

Author reply:
We checked all figures and in several cases we increased the axis label size. We also added units to all axes. Please also note that due the specific format of the ‘discussion stage’ of the journal,

several figures were quite small in that version. In contrast, in the final version of our paper we will make sure that optimum use of the available page dimensions is made.

- The English text should be improved regarding clarity, grammar, and spelling. Many sentences are too long and unclear. An example: Abstract, line 21-24. Please shorten and clarify such sentences.

Author reply:
We carefully checked the Enlish writing of the whole manuscript and made several changes to improve the readability.
The sentence in the abstract was changed to: ‘for some cloud-free conditions, especially with high aerosol load, the coincident satellite observations indicated optically thin and low clouds. This finding indicates that the satellite cloud products contain valuable information on aerosols.’

Specific comments:

Title and rest of paper: 2 1
2 > 2.5

Corrected

Abstract, p. 4654: - l. 19: indicate > confirm;

Corrected

cloud classification > MAX-DOAS cloud classification –

Changed

l. 21: clear sky: do you mean cloud free? Because a high aerosol loading does not belong to the category “clear sky”.

Author reply:
We changed ‘clear sky conditions’ into ‘cloud-free conditions’.

p. 4656, l. 10: Please define CI in a separate formula, because it is an important quantity in this paper. For which directions are the radiances used in CI ? Please explain the behaviour of CI in words: if CI is large, the sky is blue, if CI is small the sky white due to multiple scattering by aerosols and clouds (etc.)

Author reply:
We added a formula for the difinition of the CI.
Concerning the basic dependencies we think that they were already well described in the introduction of the original manuscript. However, we added the following text:
‘In this definition high (low) CI values indicate blue (white) skies.’

We also added the following text:
‘Here it should be noted that this general dependence is strictly valid only for viewing directions close to the zenith. Thus for the discrimination of clear and cloudy sky, CI measurements made in zenith direction are used in our cassification scheme (for exceptions at small SZA, see section 2.3.3).’

p. 4657: sect. 2.1: please give the relevant geographical situation of Wuxi.

Author reply:
We added a new figure (Fig. 1) to the manuscript. It shows a picture of the instrument as well as a map with the location of Wuxi.

p. 4658:that - l. 9: clouds may change in a 12-minute cycle: could that be a problem in the interpretation?

Author reply:
Yes, this could be a problem, if the cloud cover changes on shorter time scales. In the worst case small clouds or cloud holes might be missed. We added the following information to section 2.3.1:
‘Here it should be noted that the typical time for a full elevation sequence is about 12 min. If the cloud cover above the instrument changes on shorter time scales, most of these variations will be missed. Thus in particular small clouds and also small cloud holes might be missed by our classification scheme.’

- l. 20: is this radiance absolutely calibrated?

Author reply:
Usually MAX-DOAS instruments are not radiometrically calibrated. This fact and the radiance calibration method used in our study are described in detail in section 2.2.1.

p. 4659, l. 14: radiance > zenith radiance (?)

Corrected

p. 4660, l. 20-23: please clarify this sentence

Author reply:
We changed this sentece into:
‘Because of the large discrepancy between measured and simulated radiances for SZA < 40 °, the simulated clear sky reference radiances cannot be used for the whole SZA range.’

p. 4661, l. 1âˇAˇ T5: “there are reasons to believe ...” is a very unscientific statement. Please remove this sentence, and give instead an error estimate of the empirical
method.

Author reply:
We changed this sentence into:

‘This choice may appear somewhat arbitrary, but it is justified by the fact that for SZA between 12 ° and 81 ° the average of the maximum and minimum measured radiances for all measurements (clear and cloudy conditions) (see the magenta curve in Fig. 5) is similar to the clear sky reference values, indicated by the green curve.’

It is not possible to derive an error estimate of the determined clear sky radiance reference values, because we can not say how representative the chosen clear days are for the whole measurement campaign. But we added the following information at the end of section 2.2.1:
‘Again it should be noted that the determination of the radiance reference values for SZA < 40 ° as described above is to some extent arbitrary. Thus future studies should aim to improve the calibration procedures for the measured radiance. Nevertheless, the chosen procedure for the determination of the radiance reference values has only limited influence on the cloud classification results, because the cloud classification results also depend on the choice of the threshold values. And both steps are directly interconnected: In the case that the radiance reference value was determined too high, the corresponding threshold values will be also increased and thus the effects on the cloud classification results mainly cancel out.’

p. 4662, l.17: which “both the measured spectra” are meant?

Author reply:
We changed ‘spectra’ to ‘spectrum’ to make this sentence more clear.

p. 4672, l. 20 – p. 4673, l. 5: please put these percentages in a table.

Author reply:
In our opinion, the chosen diagram type gives a nice and direct overview obout the relative frequencies of the different sky conditions. Thus we prefer to keep this figure.

p. 4677: l. 4-5: Since GOME-2 and OMI observe the cloud mid-level pressure (by means of the FRESCO O2 A-band and O2-O2 retrieval methods) and MODIS retrieves the cloud top pressure (by means of the IR method), this difference is expected.

Author reply:
We added the following sentence:
‘A systematic difference in the derived cloud pressures for UV/vis methods (OMI and GOME-2) and IR methods (MODIS) is expected because UV/vis methods are sensitive for the cloud center rather than for the cloud top.’

Table 1: what do the super and subscripts of the symbols mean? In principle acronyms should not be used as symbols. Please use commonly accepted symbols: nsigma for spread, L for radiance.

Author reply:
We prefer to keep the symbols as they are, because:
a) not for all quentities generally accepted symbols exist
b) in many publications also acronyms are used

Figure 1: Are these all data, unfiltered?

Author reply:
Yes the figure contains all (unfiltered) data.

Figure 4: which viewing direction? Calibrated radiance? Which unit?

Author reply:
We exchanged ‘radiance’ with ‘zenith radiance’.

Figure 5: caption: what is what? Give subplot numbers. What is FRS? Measured spectra are in red, I assume?

Author reply:
-We replaced FRS with Fraunhofer reference spectrum.
-We made clear what the red and black lines represent.
-We added the sentence: ‘The residual (top left) indicates the difference between the measured spectrum and all fitted spectra.’
Note that this figure was shifted from the main part of the manuscript into the supplement (new Fig. S1)

Figure 6: which direction?

Author reply:
We changed the figure caption to:
‘(a) The black, red and blue curves indicate the time series of the measured zenith CI_z, the corresponding reference CI_z and the normalised CI_z, respectively, on 8 July 2012. The dashed line presents the threshold of normalised CI of 0.84. The visual images in the morning (b) and in the afternoon (c) from MODIS indicate cloudy skies over the Wuxi site on the day.’

Figure 7: Give the quantity in the legend of each subplot. What is the explanation of the peak in CI around noon?

Author reply:
We added the information about the quantity in each sub-plot.
To explain why the CI is enhanced during neen, we added the sentence: ‘The normalised CI (top left) is the same as in Fig. 6.’

Figure 8: clear or cloudy ?

Author reply:
We added the following sentence:
‘As indicated in Fig. 9, this day was completely cloudy over Wuxi.’

Figure 13: please reverse the x-axis to agree with the x-axis order of Figure 12.

Author reply:
We followed this suggestion

Figure 18: what is the message of this figure? Please remove it, and give the numbers in a table.

Author reply:
We followed this suggestion