Author Response in boldface
Changes in the paper in italic

Referee 3

MAX-DOAS measurements are highly influenced by clouds, therefore a simple method to identify cloud contaminated spectra based only on the measurement itself is very useful. This paper presents a novel simple method applicable to many measurements. Once the method is calibrated for a certain instrument, it is easy to add the cloud information automatically to the spectra. Thus it is suited very well for a publication in AMT. The authors should address some minor changes/corrections as detailed below:

p. 5885 l. 11-12: We focus on 90deg elevation observations since simulations show these are the most sensitive to the sky conditions. Please give a reference/proof.
Reference to section of the paper added: *(see Sect. 3).*

p. 5889 l. 6-8: For high SZA clear sky can appear in different colours than blue
Sentence added: *Since, during the daytime, the sky colour changes from blue during clear skies to white/gray when clouds or aerosols are present, we can use the CI to qualify the sky condition. This becomes increasingly difficult for high SZA values, as the sky colour varies, even for clear skies.*

p. 5889 l. 10: Please give a reason for averaging the intensity in this big window.
This was done to to reduce the effect of spectral noise on the derived intensity values. Sentence added in the paper: *to reduce the effect of spectral noise on the derived intensity values..*

p. 5889 l. 13-22: - Please mention that here are results from Xhianghe are shown. Do the other stations look similar?
Mention added. Other stations show similar trends, with the exception that the extreme aerosol events only occur in Xianghe.

- The in the text mentioned letters a-d in Fig. 1 do not appear in Fig. 1
Changed in text.

p. 5889 concerning DAK: - What albedo is used?
Added to text: *a surface albedo of 0.05 was used.*

- What is the height of the cloud? (Bottom height at 1km, but what is the top height?)
1km thickness, added to text: *For the aerosols a homogeneous layer up to 1 km with a single scattering albedo of 0.9 and asymmetry parameter of 0.7 was used, for the clouds these values are respectively 1.0 and 0.85. The cloud base height was set at 1 km, with a total thickness of 1 km, a surface albedo of 0.05 was used, and atmospheric Rayleigh scattering and ozone absorption were included.*

p. 5890 l. 1.3: I would prefer a short example.
It is unclear what the referee refers to.

p. 5890 l. 6: ... from spectra with low elevations angles shows a very narrow spread regarding different aerosol and cloud settings
Adapted in text.

p. 5890 l. 11: eliminate maybe better exclude
Changed in text.

p. 5890 l. 19-23: Did any instrumental issues occur? If so, how could they be fixed?
For Brussels an instrumental failure caused a shift in the observed CI values. We determined the CI-value distribution histograms before and after the failure and shifted the data

after the defect in such a way that the histograms coincided. We added this and a figure as illustration to the text: ***An example of this can be seen in Fig. 1: an instrumental failure at Brussels on the 20th of May resulted in a strong downward shift of the CI values. We corrected for this by shifting the CI values after the failure in such a way that the peak values of the histograms of CI values before and after the incident coincide. It is clear that sufficient data needs to be present to make an accurate correction.***

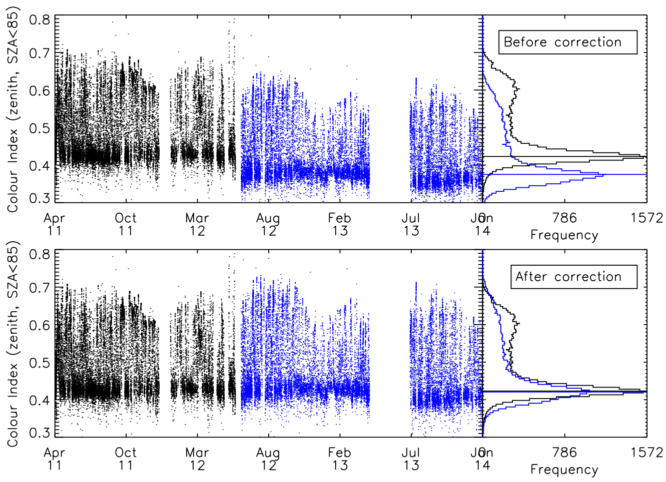

Figure 1: Illustration of the CI correction for Brussels due to an instrumental failure on May 20th 2012. The top panel shows the CI values and histograms before and after the incident in black and blue respectively. The bottom panel shows the corrected CI values. This was done by shifting the blue points in such a way that both histograms have the same peak value.

p. 5892 l. 1: arbitrarily Something led to this choice, please explain a little further.
Added to text: ***We then take the scaled simulation made with AOD= 0.15 and COD= 0.0 (green-diamond line) as the limit to separate the 'good' and 'mediocre' region, as the simulation predicts that data above this curve was taken under cloud-free conditions with an extremely low aerosol load. This is further corroborated by comparison with co-located AOD measurements, as explained above.***

p. 5892 l. 8: meteorological -¿better visibility
Changed in text.

p. 5892 l. 14: Please define scattered clouds
Added to text: ***To determine the presence of broken (semi-continuous cloud cover) or scattered clouds (predominantly clear sky) in the line-of-sight of measurement, the temporal variability of the CI is studied.***

p. 5892 l. 21: How were the different Cs determined?
We initially chose to use a relative limit for the O4 DSCD, as these have been in a way scaled by removing the 90deg elevation DSCD. This makes the sample much more homogeneous over time and removes the large diurnal trend. For the CI values this is not the case, which means the CI values can, especially for good days at Xianghe, reach both very low values (morning/evening) and quite high (noon). Using a relative-change cut-off limit, this gives a predominance to flag broken clouds during morning/evening. Also, for overcast measurements, the CI values can be extremely low, again resulting in too many points flagged if a minor change in CI occurs.
However, since this problem typically occurs for very low CI values, these data will in any case already be flagged as 'bad' and removed from the sample. This means the influence on our overall results is only minimal. We have adapted our study to use also for the broken

clouds a relative change limit. But, one needs to keep this in mind if he wants to use solely the broken-cloud flag for some statistics. This new limit was determined by investigation the typical variability observed for a selection of clear days, and those with evidence for scattered clouds. We found that for the latter jumps typically spanned a 0.1 CI value. The text has been changed accordingly: ***This value was derived by investigating those days with rapid temporal variability in the CI. For these days it was found that the observed jumps in CI predominantly fall above these cut-off value***

p. 5894 l. 2: How was the threshold of 0.2 determined?
The cutoff values where determined by investigating the typical change in O4 DSCD value for days with evidence for multiple scattering (i.e. quick jumps in O4 DSCD). We found that quick temporal variability gives an typical shift above 0.2, whereas smooth variability stays below this value.

p. 5894 l. 1-3: I would appreciate a similar example as for the temporal variation of the CI (fig. 6).
Figure added.

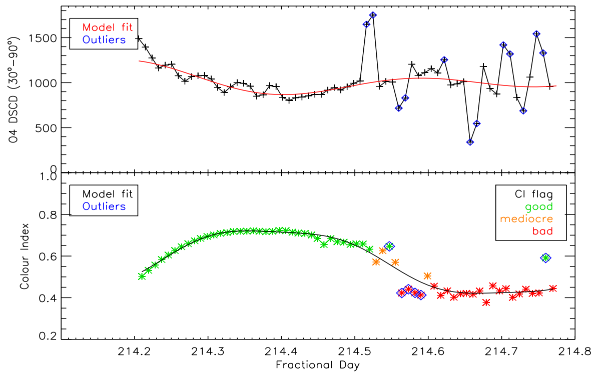

Figure 2: Top panel: Results of the O_4 DSCD modelling (green line) to the measured O_4 values (black crosses) and outlier detection (blue diamonds) for the multiple-scattering flagging, for an example day in Brussels. For comparison we also show the results of the CI flagging (bottom panel). The CI values (asterisks) are coloured according to their CI-flag. Outliers from the broken-cloud flagging are marked with a blue diamond.

p. 5894 l. 8: A new section should be started here, from here on the introduced flags are compared. p. 5894 l. 15-19. This section is misleading, for Brussels and Jungfraujoch is another effect discussed than for Xianghe. Please separate this more clearly.
We have adjusted the text to accommodate this.

p. 5895 l. 22: Please introduce MS flag (best would be to do this in section 4.3, and BC flag in 4.2)
Adapted in text.

p. 5896 l. 14-19: It is not clear to me, if for Xianghe the bad condition is now used or not. A small sketch for the different conditions and how they are used for which dataset could help here.
When using the colour index flag, the bad data is not used. However, when the multiple scattering flag is used, this will include data that is flagged as bad by the CI flag. The legends in the different figures always clearly state which flag criteria are used.

p. 5900 l. 13: screws colloquial English
Typo, needed to be skews.

Fig. 1: - Is the fractional day in UTC?

Yes, also added to caption. - It could be useful for the further interpretation if the CI for the 90deg measurements would be connected with a line (e.g. the scattered cloud effect would be better visible)
adjusted - It is not clear, if the boxes Extreme aerosol load in fig. 1b and the legend in fig. 1d are covering data. This may be checked.
adjusted Fig. 2: - Printed the colours in the left legend box are not very good visible on the dark grey background.
adjusted Fig. 3: - This figure is not shown properly. The left side is missing completely.
Can you be more specific, I cannot find any issue with the figure. Fig. 5: - This figure is not shown properly. The dots on the left side are missing on the 2nd and 3rd figure.
For the three different sites, the data start at different SZA values, due to the different latitudes. Fig. 7: - Please draw the black line thicker.
adjusted

Author remark: We have adapted the paper to use non-cloud-screened AERONET and Brewer data for the correlation study with our AOD retrievals. This more clearly shows the effect of our cloud-screening, but does not change the general conclusions of our study.