Author Response in boldface
Changes in the paper in italic

Refree 2

The study of A simple and versatile cloud-screening method for MAX-DOAS retrievals by C. Gielen et al. presents an important cloud screening method to be applied to large data set of passive remote DOAS instruments, in particular for zenith sky view. The method applied here follows up the study by Wagner et al (2014) about the effect of clouds on several MAX-DOAS observations such as radiances, colour index (CI), ring effect (Raman Scattering Probability, RSP) and O4 absorption (based on air mass factors, AMF). In this work the approach is simplified by using only CI, O4 dSCD, the zenith sky view, and is based in a general simulation of the CI. The method applied here has great potential, especially for long term zenith sky measurements, however the application of this method to off axis MAX-DOAS measurements might need more analysis regarding the primary objective of using several elevation angles with MAX- DOAS (as stated in the title of this work). I recommend publication after addressing the main points below:

One of the main observations is in the application of this method on MAX-DOAS measurements which use multiple elevation angles. This work focuses only in the effect of the CI, and O4 dSCDs using the 90 elevation angle, however the line of sight of 90deg can be completely different than the line of sight of lower elevation angles. The justification of using only 90deg for this work is that simulations with radiative transfer model show a higher contrast in the CI using 90deg than lower elevation angles (Fig 2).
It is indeed true that the line-of-sight meteorological conditions between different angles can be different, especially in the case of broken or scattered clouds. However, for MAX-DOAS profile retrievals, a full scan, consisting of all elevation angles will be used. The method could indeed be extended to include also the other elevation angles, if one takes into account that the lowest elevation angles CI values will make it very difficult to constrain the sky conditions. The 90degree elevation also has the benefit of not depending on the instrument viewing angle. Which means that the simulations do not need to be computed for different instrument pointing angles, which could drastically decrease the computational effort if one has data sets with different viewing angles. We had extended the text to include these remarks and also investigated the effect of including the 30degree elevation angle in our cloud-screening procedure:

Fig. 5, where simulations of the CI are presented for the three different wavelength ratios used for the different measurement sites, shows that the CI derived from spectra with low elevation angles shows a much more narrow spread regarding different aerosol settings, making it difficult to distinguish between the different parameters. These simulations furthermore show that the same problem of overlapping simulations occurs for observations taken at SZA > 85***degree. For this reason we exclude these data from our study. The simulations at 90degree elevation show a more narrow spread for lower SZA values (*** $\lesssim 40$***degree), compared to the 15 and 30degree elevation angle. However, at larger SZA (*** $\gtrsim 55$***degree) the situation is reversed. The same result is found for simulations made under different cloud optical depth settings.***

As we only have little observations made at low SZA (< 40***degree), we therefore choose the 90degree as the best elevation for our further study. In principal, the method can be extended in a similar way to include CI (and*** O_4 ***DSCD) information from multiple elevation angles, with the realisation that the higher elevation angles will give the best constraints. This is discussed briefly in Sect. ??. We restrict ourselves to only one elevation angle for the sake of simplicity and to show that the method already works well with this restriction. The zenith elevation further has the advantage of not depending on the viewing azimuth of the instrument, which simplifies the computational effort for the CI simulations if data sets from instrument with different pointing direction are used.***

We also tested the effect of extending the zenith-based cloud flagging to extend CI flagging with information from the 30degree elevation angle. Statistics on this procedure can be found in Table 1. On average we find that using both elevation angles results in a removal of $\sim 10\%$ ***more data points. When we then retain only those measurements which are flagged as cloud-free by both the 30 degree and zenith CI flags, we find only a minor improvement in the observed correlation with co-located measurements. The improvement is highest for Brussels, as one would expect, since the 30degree viewing angle will mainly provide additional information in the case of broken clouds. For Brussels we find an increase in the observed correlation of 0.06 to R = 0.68, for Jungfraujoch and Xianghe, the increase is 0.04 and 0.015 to R = 0.2 and R = 0.9 respectively. We do not see a change in the observed correlation slopes***

In general MAX-DOAS measurements spend more time in the off axis scan measurements, or even in only one low elevation angle to obtain boundary layer trace gas mixing ratios. If the method proposed here is applied to MAX-DOAS I would think that at the end of the screening method it will remain less/more points if using only the 90EA in order to filter all elevation angles.
We have extended the analysis by also including cloud screening based on the 30degree elevation angle and have included in the text. We find that when combining information from both elevations about 10% more data is flagged as cloudy, but the impact on the correlation with co-located measurements is only minor (see added text above).

This in fact is pointed out in section 6 where the cloud screening is applied to aerosol model retrievals and broken clouds are removed.
The authors should address the application of this method to MAX-DOAS inversion by studying the effect of broken clouds and aerosol in-homogeneity along the different EAs, otherwise they should state that this method can be applied to zenith sky DOAS.
In section 6 we compare AOD retrievals with AOD measurements, with and withoud removing data made under cloudy conditions (so not only broken clouds). For this retrieval a full MAX-DOAS scan is used, which also includes the 90deg elevation used for the cloud screening. The problem that could arise from only using this elevation angle is that the 90degree angle did not have clouds in the line-of-sight, whereas the lower elevation angles did. This means that we could overestimate the amount of non-cloudy data. ***As MAX-DOAS measurements have the benefit of multiple viewing angles, once could extend the zenith-viewing method proposed here to other elevation angles. From our simulations it is clear that the highest elevation angles are best suited for this. Different viewing angles will mainly be sensitive to broken or scattered clouds, as the flagging in the case of clear or overcast days will give the same results for each elevation angle. For sites which experience a lot of broken clouds including CI flags from for example the 30degree elevation angle will therefore be more likely to correctly identify the presence of clouds. We briefly discuss the effect on this on our retrievals in Sect. 6.***

Have you thought using the ratio or and indicator of two different EAs (for example 2 and 90deg).
At this point we opt to focus on one elevation angle, to show that the simplest procedure already gives very good results in terms of cloud screening. But, as mentioned before, one could opt to include as many angles as possible to more correctly mark the presence of broken or scattered clouds. We then suggest that the highest elevation angles be used, as they will provide the best constraints.
One has to be careful when using the ratio of the CI as an indicator, as it might introduce unwanted trends in for example the temporal variability, making it harder to detect broken clouds.

Also, Wagner et al (2013) studies the effect of the Ring effect (Raman scattering) but here it is not mentioned at all. I suggest an analysis in the advantages of using only CI and not the Raman Scattering.
This paper aims at giving a very simple cloud-screening method, which uses as little different parameters as possible. We therefore did not include other parameters such as the Ring effect, intensities, multiple elevation angles, etc. as can be found in Wagner 2013 (A first look at the comparison between the CI and the Ring effect for some example days can

also be found in that paper).

An extended cloud-screening method based on all available parameters, or a detailed comparison on the effects of the influence of these parameters on the cloud screening will be very interesting, but goes beyond the intent and scope of this paper.

The authors describe the impact on aerosol retrievals by comparing the AOD retrieved with MAX-DOAS and AERONET. This comparison is important to check compatibil- ity between the method applied here and the cloud screened filtering applied in the AERONET data L1.5. However the analysis in this work is focused in the AOD comparison. This comparison might not highlight the importance of the cloud screening method due to two main reasons (1) the air mass probed by the two instruments can be quite different, MAX-DOAS has a unique directionality, on the other hand the AERONET sun photometer tracks the sun; and (2) AERONET L1.5 is a cloud filtered product. In section 6.2 p5898 there is a basic description about the AOD products from AERONET, Brewer which reads we do find that data for which no measurements are available are generally flagged by our method as cloudy, showing the coherence between our cloud screening and that of the other instruments. It would be important if more detail is given about this comparison. I would suggest a thorough and quantitative comparison of the times where AERONET L1.5 or (Brewer) filtered data with their own cloud screening and can potentially coincide with the method applied here.
We agree with the referee that the comparison with level15 AERONET data was not ideal. We therefore adapted our study to compare with the unscreened level 10 AERONET data and unscreened Brewer data. Of course, the direct-sun constraint does remain also an issue with these data.
We also gave a more quantitative comparison between the non-presence of co-located measurements and our cloud flagging. These results can be found in Section 6.

To study the effect of our cloud-screening method, AOD values retrieved by MAX-DOAS are compared to co-located AOD measurements. For Xianghe and Brussels we use AERONET Level 1.0 (unscreened) (and 1.5 (cloud-screened)) data, and for the Brussels site we extend the comparison with co-located Brewer spectrophotometer measurements at 320 nm (Brewer instruments #16 and #178). A detailed description of the co-located instruments and measurements can be found in Cheymol2003,DeBock2010, and Holben2001.
In the supplementary material we also show the correlation between our AOD retrievals and co-located measurements, but now using cloud-screened AERONET level 15 and Brewer data. For Jungfraujoch no such cloud-screened data are available. We find that the AERONET cloud-screening procedure Smirnov2000, based on the stability of a measured AOD triplet over a 30 second interval and temporal AOD hourly and diurnal variability, removes more data compared to our cloud screening, leaving around 28% for Xianghe and 10% for Brussels. This results in better correlation and slope values for both Xianghe and Brussels, compared to the correlation with the non-screened level 10 data, with improvement on average of the order of $0.05 - 0.1$ ***for both*** R ***and*** S***. As the AERONET cloud screening is based only on temporal variability of the AOD, stable uniform clouds and aerosol plumes can be misidentified. This could account for differences between our cloud-screening method and the AERONET screening, as for example seen in the first plot of Fig. 1. For this day with a strong rise in aerosol load, the second half of the day is flagged as mainly cloudy by AERONET, whereas we do not.***

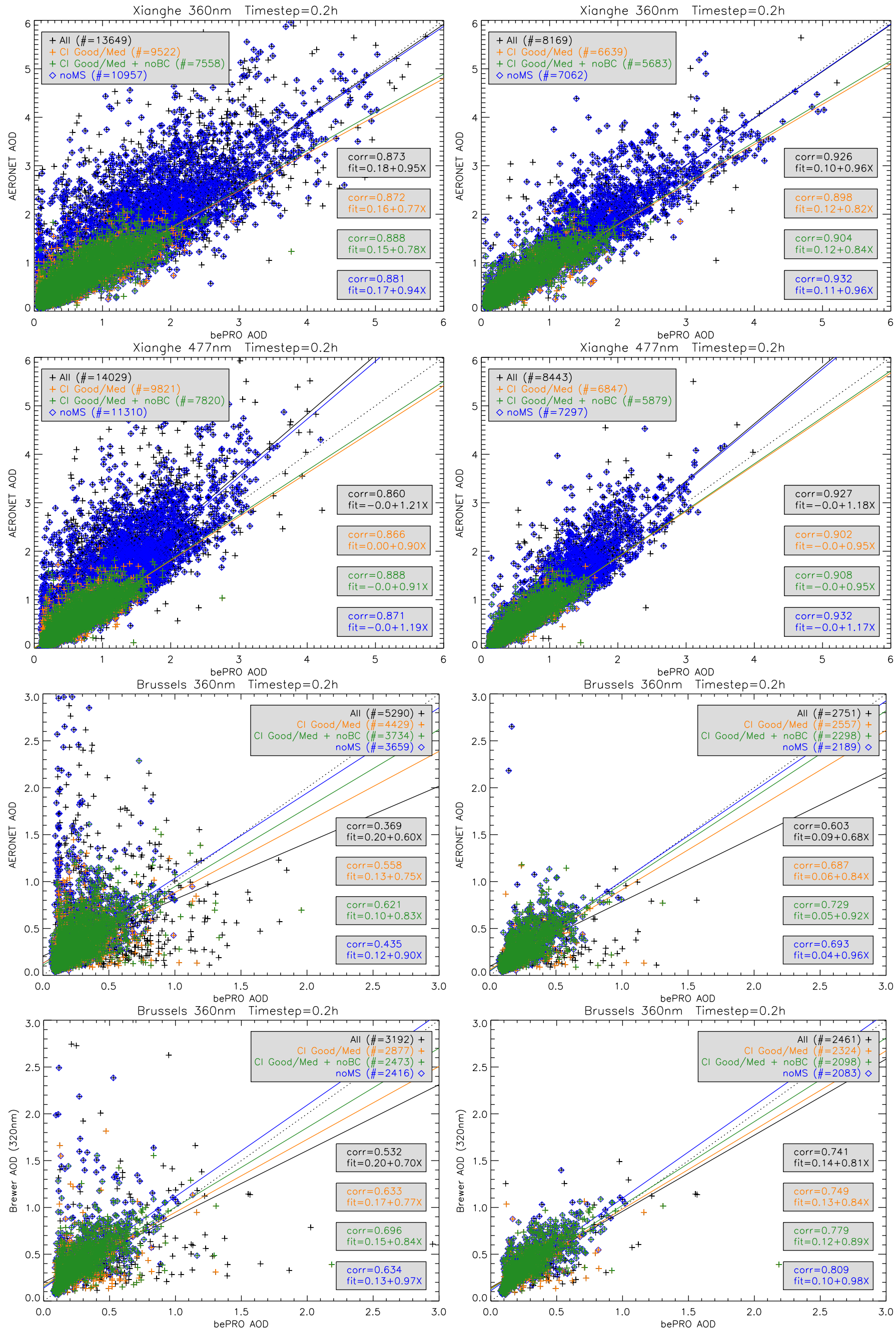

Figure 3: Correlation plots of our bePRO MAX-DOAS AOD retrievals and measured AOD values for the Xianghe and Jungfraujoch data set at 360 and 477 nm and for the Brussels at 360 nm, in time steps of 0.2 hour for Xianghe and Brussels and 1 hour for Jungfraujoch. The figures on the left use non-screened AERONET/Brewer data, whereas the figures on the right use cloud-screened measurements. The full non-cloud-screening data is given by black crosses. Cloud-screened data (based on the CI) with a 'good/mediocre' sky flag are marked in orange, data with 'good/mediocre' sky flag and no broken-cloud flag are marked in green crosses. Data with no multiple-scattering flag (based on the O_4 DSCDs) are marked with blue diamonds. For each sample set we also give the linear regression lines and correlation information.

The method applied here is based in the comparison of the CI using a based RTM simulation and measurements under different cloud conditions. In section 3 the simulation of the CI is introduced, however I did not find in this section or in figure 2 the wavelength range where the simulations were made.

The wavelengths used are visible in the title of the figure. We have additionally added them to the caption. Also we have added a figure showing the effect of different wavelengths on the CI, see Fig. 4 and Fig. 5 below.

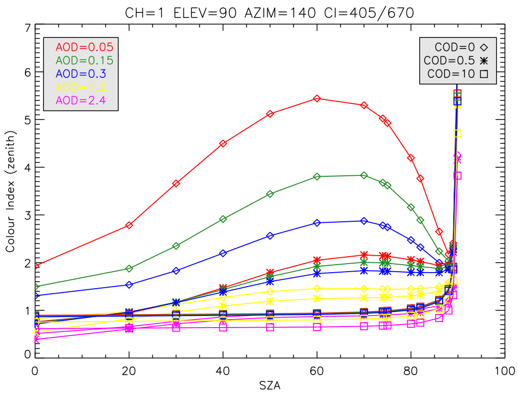

Figure 4: Top: Simulations of the colour index (I_{405}/I_{670}) under varying aerosol optical depth (AOD) and cloud optical depth (COD). The simulations were performed with the DAK plan-parallel radiative transfer model Stammes2001, using a cloud-layer height of 1 km.

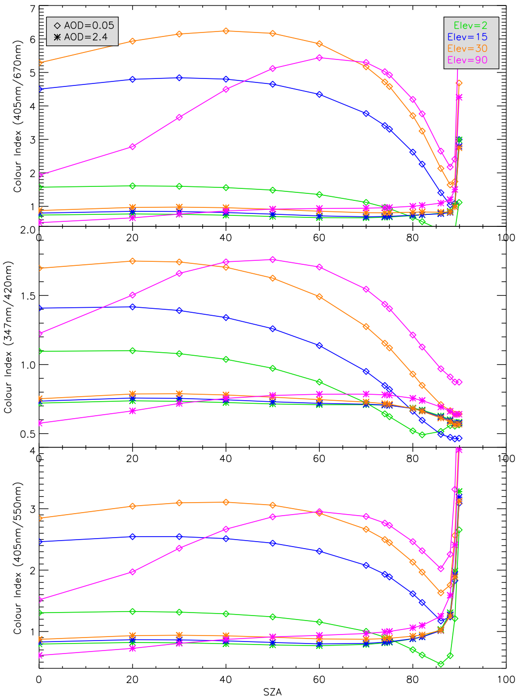

Figure 5: Simulations of the colour index under varying AOD and using the wavelengths corresponding to the three different measurement sites, respectively Xianghe, Brussels and Jungfraujoch from top to bottom. The simulations were performed with the DAK plan-parallel radiative transfer model Stammes2001, using a cloud-layer height of 1 km and a cloud optical depth of 0.

My point of this observation is regarding a comparison with measurements since most of the time filters are used to avoid stray light. Could you please expand the description of the three instruments, especially whether filters were used to eliminate stray light to the spectrometer, and what kind of effect do you expect in the CI if filters were used? If not filters were used it would be valuable for the application of this method to elaborate in the effect of filters which are very common in the passive remote sensing

observations. **Filters are used in the instruments to reduce the effect of stray light on the measurements. The use of filters or not could have an effect of the CI although this will mainly depend on the wavelength range of the spectrum, and the wavelengths that are affected by stray light. If no stray-light filters are present, and part of the wavelength range is affected by stray light, this will give an overestimation to one intensity value in the CI calculation making it difficult to compare with simulations (unless one could calculate a rough estimate of the stray-light contribution to the intensity and correct for it).**
Since this method is based on a more statistical approach of the observations, combined with a very general CI simulation model. The statistical part of our method (i.e. the cut-off of the CI histogram, the temporal variability of the CI) will still applicable in the case of stray light contribution, as one would expect the stray-light effect to be more or less consistent over the data set.
One needs to keep in mind that our simulations are not fine-tuned to each instrument (spectral quality, response curve,..) and site (think of surface albedo, station height, ...) which all will have an effect on the CI, which at this point cannot just be calibrated out.

Minor observations: Line 11: Here, in the introduction, is mentioned we focus on 90 elevation observations since simulations show these are the most sensitive to the sky conditions. Please indicate or introduce a reference pointing this out. **Added to text** Sec 2.2, p5888 Change Rayleigh and Mie scattering for scattering due to molecules and particles **Adjusted** The viewing angles (azimuth angles) are missing in the description of the Jungfraujoch and Brussels instruments. **Adjusted for JFJ, Brussels was already present in text.**

Please elaborate more in the DOAS analysis of O4. Do the UV and visible analysis carry less/more/the same information for the cloud screening approach described here?. The UV spectra range might be sensitive to air mass closer to the instrument, while the distance range in the visible may be larger.
Our cloud screening in itself does not make a distinction on UV or VIS spectra, only the AOD model retrievals are determined for both if data is present in both the UV and VIS. This means the exact same cloud-screening flags are applied to retrievals in the UV and the VIS. For the flags based on the colour index we use a fixed wavelength ratio for each site, so we do not make a distinction on flags derived from UV or VIS spectra. The same is true for the multiple-scattering flag, where we for all sites use O4 DSCDs which are derived from VIS spectra, unless no such data is available (e.g. Brussels) where UV-based O4 DSCDs are used.

Also, it has been noted a correction factor in Clemer et al of 0.8, is this correction factor applied here?. Please elaborate more in the use of this correction and for which elevation angles is important and how this factor could potentially affect the CI.
This correction factor was also used for the AOD modelling at Xianghe, but not for Brussels and JFJ, where it did not improve the modeling. This factor has no effect on the CI, it is only a parameter used in the bePRO modelling and AOD calculation. Values of the correction factor was added to the text.

This method is applied using only zenith sky spectra, however is restricted only to SZA > 85. . . what do the authors recommend for SZA > 85. This might be important for the zenith sky DOAS where stratospheric trace gases are aimed, particularly for SZA> 85
For high SZA values the cloud-screening based on the colour index will no longer contain information, as it will become increasingly difficult to separate between different sky conditions. One might extrapolate information from the flags from smaller SZA (70-80deg), under the assumption that the meteorological conditions didn't change drastically. However, this could introduce large uncertainties, as some sites for example often experience an increase in clouds during twilight. Other information on the sky quality will have to be derived from parameters such as the O4 or the Ring effect (Wagner2009).

It is noted here that the main sensitivity in the simulations of CI was using a cloud base height at 1km, but is not clear what was the actual altitude of the cloud.
Adjusted in 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. We also tested the effect of varying the cloud base height, ranging from 1 km to 8 km, but found very little influence on the derived CI values, especially for higher elevation angles.

Please expand more in the sensitivity studies of the CI, especially in the type of clouds, altitude (distance from the ground and vertical extent - cloud thickness), and for different elevation angles.
I have added some extra description of the simulations and added a figure to describe in more detail the effect of the elevation angle and the wavelength ratio: see added text above and Figs 4 and 5, with the accompanying text: ***Fig. 5, where simulations of the CI are presented for the three different wavelength ratios used for the different measurement sites, shows that the CI derived from spectra with low elevation angles shows a much more narrow spread regarding different aerosol settings, making it difficult to distinguish between the different parameters. These simulations furthermore show that the same problem of overlapping simulations occurs for observations taken at SZA*** $> 85 degree$***. For this reason we exclude these data from our study. The simulations at 90degree elevation show a more narrow spread for lower SZA values (***$\lesssim 40$***degree), compared to the 15 and 30degree elevation angle. However, at larger SZA (***$\gtrsim 55$***degree) the situation is reversed. The same result is found for simulations made under different cloud optical depth settings.***
As we only have little observations made at low SZA (< 40***degree), we therefore choose the 90degree as the best elevation for our further study. In principal, the method can be extended in a similar way to include CI (and*** O_4 ***DSCD) information from multiple elevation angles, with the realisation that the higher elevation angles will give the best constraints. This is discussed briefly in Sect. 6.1.***
We restrict ourselves to only one elevation angle for the sake of simplicity and to show that the method already works well with this restriction. The zenith elevation further has the advantage of not depending on the viewing azimuth of the instrument, which simplifies the computational effort for the CI simulations if data sets from instrument with different pointing direction are used.

The description of figure 1 on the main text refers to figures 1a-d, however the labels (a,b,c, and d) are missing in the actual figures. The color chosen for the Elev=2,8, and 90 are difficult of distinguish, please change it accordingly. **Adjusted**

The AERONET component is very important along this manuscript, for instance, in fig.1 the aerosol load is essential in order to makes sense of the CI and O4 dSCDs.
This is not completely true: the AERONET data is presented to show that it corroborates our expected CI and O4 DSCD behaviour, namely that a decrease in CI corresponds to an increase in AOD or clouds. Or that a that a non-varying CI and erratic O4 profile corresponds to thick clouds.

For the fractional day 35 in figure 1 the AOD increases in the early morning, however there is not data after 35.25. The AERONET data used in this work is level 1.5, which is cloud screened data, so presumably when there is not AOD data there are clouds present, could you please expand a description when there is not AOD in terms of O4 dSCDs and CI for this figure?. **We have adjusted the analysis and figures to show both the unscreened and cloud-screened AERONET data. This indeed shows that the second half of that day is flagged as cloudy by AERONET. However, we note that this might be a misflagging by the AERONET cloud-screening method in the text:** ***In the supplementary material we also show the correlation between our AOD retrievals and co-located measurements, but now using cloud-screened AERONET level 15 and Brewer data. For Jungfraujoch no such cloud-screened data are available. We find that the AERONET cloud-screening procedure (Smirnov2000), based on the stability of a measured AOD triplet over a 30 second interval and temporal AOD hourly and diurnal variability, removes more data compared to our cloud screening, leaving around 28% for Xianghe and 10% for Brussels. This results in better correlation and slope values for both Xianghe and Brussels, compared to the correlation with the non-screened level 10 data, with***

improvement on average of the order of $0.05 - 0.1$ ***for both*** R ***and*** S***. As the AERONET cloud screening is based only on temporal variability of the AOD, stable uniform clouds and aerosol plumes can be misidentified. This could account for differences between our cloud-screening method and the AERONET screening, as for example seen in the first plot of Fig. 1. For this day with a strong rise in aerosol load, the second half of the day is flagged as mainly cloudy by AERONET, whereas we do not.***

Also, the fractional day in fig 1 is confusing. The measurements started at 35.0, is this midnight?.
Fractional day is given in UT, adjusted in the caption.