Reply to I. Petropavlovskikh's interactive comment

We would like to thank the reviewer, Irina Petropavlovskikh, for her comprehensive review. The comments and recommendations in this review have lead to a great improvement of this paper.

Referee comments and statements are in bold. Author replies are in italics. Figures are located at the end.

Initial statements from referee

The paper by Stone at al. "A new Dobson Umkehr ozone profile retrieval method optimizing information content and resolution" describes the development of the new algorithm to process data from Dobson Umkehr measurements. In-depth analyses of measurements are presented with the aim to further improve the information content of the retrieval and to reduce noise in the conventional Umkehr ozone profiles. Authors exploit the measurements at three wavelength pairs vs. single standard C-pair approach, make use of all available measurements taken during the Umkehr observation in place of traditional interpolation to the nominal SZAs, extension of measurements beyond 90-degrees SZA (before the sunrise and after the sunset), and make adjustments to the a priori information. The method involves change to the forward model and development of the refraction correction code to accommodate the use of measurements at the variable SZAs. Among main limitations of the code is a lack of multiple scattering contributions in the zenith sky intensities simulated in the forward model and absence of the stray light corrections. The work is done on the Umkehr method that has been in use since 1930s. There are about 80 reporting Dobson instruments that monitor ozone around the world. Among these there are only about 19 stations that currently report Umkehr ozone profiles. Umkehr ozone profile data had been successfully used for long-term trend analysis and in validation of other ozone datasets. Special attention is paid to the data available prior to the onset of the ozone depletion and before the global satellite ozone profile retrievals became available (before 1978). In this case Australian Dobson data represent very important piece of puzzle that can be provided to the studies of the ozone in the Southern Hemisphere where the early period data are fairly scarce. The approach described in the paper shows potential improvements to the retrieved information, which can help to reduce retrieval noise in the profile (currently up to 25 % in the troposphere), separate ozone information into the vertical columns with less interference from adjacent layers and to extend retrieval sensitivity to 50 km (currently sensitivity to ozone changes above 42 km (2 hPa) is very limited in C-pair Umkehr retrievals). However, several concerns exist about the proposed approach. A demonstration of the stability of

the retrieval is needed, as well as implementation of the multiple-scattering and stray-light corrections in the forward model is of the importance for the high quality profile retrievals. This paper presents novel concepts and improvements to the existing Umkehr ozone profile retrieval algorithm, and thus is within the scope of AMT. The manuscript needs several clarifications and adjustments as indicated below before the paper can be published.

Comments

p. 8670, line 13, also on p.8672, line 11 and line 26, 27 (and more in the text) – I would not call these "raw" measurement since N-values are derived from R-values (which are raw measurements) by applying lamp and R-N table corrections. You can probably call it operational?

Instances of "raw" have been changed to "operational" throughout the text as suggested to avoid confusion between N and R-values as mentioned.

p. 8675, line 15. Please be aware that the errors were derived for C-pair only from the co-incident measurements between two co-located Dobsons. I expect that the noise in the A-pair might be higher and in the D-pair might be lower, but it needs to be confirmed.

Thank you for pointing this out. We have made the following change

p. 8675, line 17. Replaced the sentence: "As the majority of the measurements are taken manually, the values have been scaled to account for human error, for which we have used a factor of 10" with: "As the uncertainties were originally constructed for C-pair only measurements, and the majority of measurements at Australian sites are taken manually, the values have been scaled to account for human error and potentially larger errors present in A+C+D-pair measurements, for which we have used a factor of 10."

p. 8677, line 16. The lack of the stray light impact on the time series was shown only for the C-pair retrieval, it has not been tested for the three pair retrievals or the use of variable number of observations per Umkehr curve ("raw measurements" in your case). The Stray light is expected to have the largest effect on the A-pair at large SZAs, thus please change "the Umkehr" to "the traditional Umkehr" or (C-pair Umkehr) in the text. One point is worth to make about the stray light effect in the time series. It will likely remain the same if the same instrument is in operations for the entire record. However, if the instrument is replaced, then the step change can occur in the measurements and in the ozone record. It will be nice to mention in this section if there are any records of the instrument replacements for Australia stations. It could also be helpful to indicate these periods in figure 4.

We have updated all instances of when we referred to stray light and ozone cross section effects analysed by Petropavlovskikh et al. (2011), to specify that these where for C-pair retrievals only. Since this paper is only using Melbourne retrievals.

p. 8673, line 9. Changed "retrieval bias" to "retrieval bias for C-pair Umkehr retrievals"

p. 8675, line 12. Changed "Umkehr derived" to "C-pair Umkehr derived"

p. 8675, line 12. The following sentence has been appended to the end of the paragraph: "Stray light effects on Umkehr derived time-series are likely to be consistent if the same instrument is operation for the duration of the time-series. However, a step change may accompany instrument replacements."

We do not believe stating instrument replacements for all Australian sites are relevant for this paper. This will be documented in a future paper focusing on the Australian measurement datasets. We have included a statement about there being no instrument changes for the figure 4 (new figure 5) time series, however.

p. 8684, line 15. Added in the following statement "Also, during this period, the same Dobson instrument was used…"

p. 8677. Equations 6 and 7, not sure if they are needed here – I did not find a reference to these equations further in the text. If you want to keep it, please include figure with schematics of the light propagation to surface with single scattering event and define apparent and local SZA.

I have included figure 1 in the paper to show the refracted paths and single scattering event through the atmosphere, highlighting the apparent and local SZA. And appended the following sentence.

p. 8778, line 2. I have appended the text at the end of the paragraph: "An illustration of the propagation of light through the atmosphere with a single scattering event is shown in figure 2, highlighting r_g, θ_A, and ξ."

p. 8678, line 18. Add reference to Petropavlovskikh et al (2005) and to clarify that in the Umkehr algorithm ca. 2005 the a priori was modified from the algorithm ca. 1992 to remove constrain to the total column ozone.

p. 8672, line 4. Changed the sentence for clarity: "This variability in the a priori profiles was removed in the updated NOAA based algorithm by Petropavlovskikh et al. (2005), which was optimised for ozone layer trend analysis." To "Variability introduced by constraining the a priori profiles to total column ozone was removed in the updated NOAA based algorithm by Petropavlovskikh et al. (2005), which was optimized for ozone layer trend *analysis."*

p. 8678, line 19. Appended after "independent of the a priori": "consistent with Petropavlovskikh et al. (2005).

p. 8679, lines 4-7. Please clarify how the Sa is constructed. The way I read it now tells me that the new Sa is constructed for each profile with the adjustment (scaling) to the diagonal elements based on the L-curve analysis. It might create a better profile, while the time series might have an effect from the variability in the Sa matrix, which can impact the trend analysis. If this is the case, it needs to be mentioned in the text. It is also mentioned on the page 8683, line 21 "Sa changes slightly with season" – may be changes with total ozone as well? This can be checked by using trend analysis on Sa. By changing Sa you are changing the smoothing of ozone profiles and the AKs. How much does it impact the retrieval?

The L curve derived scaling was done for select C-pair and A+C+D pair retrievals and found that the same scaling was similar for all retrievals. The Sa matrix changes with month. It reads in the monthly climatological a priori standard deviation constructed from Hassler and uses them for the diagonal portion of Sa.

p. 8678, line 22. I have changed the sentence: "The a priori covariance matrix (Sa) in this work uses standard deviations" to "The a priori covariance matrix (Sa) in this work uses monthly standard deviations to" to help clarify that the standard deviations will change with month consistent with the a priori matrix.

p. 8679, line 4. I have changed the sentences describing the L-curve adjustment to: "This is done by constructing an L-curve on select C-pair and-A+C+D pair retrievals, where Sa is scaled from small to large values and plotted against the root mean square (RMS) of the retrievals to the measurements. The scaling value chosen is when there is no significant difference in the RMS as the scaling factor increases. The value derived from the selected retrievals is consistent for both the C and A+C+D-pair retrievals, and is used for all other retrievals."

p. 8680, section 5.1. Figure 1 shows Umkehr data for the clear sky day (very smooth curves). Can you show an example of the day with partial cloudiness and the model fit? One of the reasons for using nominal SZAs (or designated) in the traditional Umkehr retrieval is to be able to preserve some Umkehr measurement while removing cloud-affected points from the fit. Is it something that you are applying as well in the retrieval? It may not be a problem for the fit if outliers are dispersed through- out the entire range of data, but it would certainly be a concern if these points are located at either end of the SZA range (near 60 or near 90 degrees). The effect of the cloud on N-value often results in a constant offset of the Umkehr measurement. If you are using relatively

displaced N-value at 60-degrees SZA to normalize the entire curve you may introduce the offset in the retrieved profile. Have you looked into these cases?

This is a good point that we did not fully consider. Since the measurements that we are using are manually taken, we assume that evening based measurements will not begin when it is cloudy, so introducing an offset by normalizing to a cloudy measurement would be unlikely. However, this may not be the same for cloudy morning measurements, when the lowest SZA is taken last. We have included an example of a cloudy retrieval in the figure above (figure 2). The simulated fit still seems quite good. However, we agree that writing in a flag for cloudy day retrievals, especially for the normalization SZA is important, and thus will be implemented in the future. We have updated the text in the following places to state this:

p. 8675, line 3: In the case of cloudy days, an offset can be introduced into the measurement. If this offset is present at the normalisation angle, then the offset can be introduced into the retrieval, which is currently not taken into account.

p. 8681 line 2. It would be better to provide numbers to explain term "larger".

p. 8681, line 2. Have changed the sentences: " these differences consistently become larger (not shown). However, most of the differences are within the values assigned to the measured N values." to "the magnitude of these differences can increase significantly. For example, from the case at 13 April 1970, with a cut off SZA of 94°, the biggest difference in the N value is 30.4. This corresponds to only 13% of Se, but is outside the errors assigned to the measured N values."

p. 8681, line 4-5. When matching measurements at SZA larger than 90 degrees, how do you deal with the diurnal changes in ozone in the upper and even middle stratosphere. It takes between 90 to 180 minutes to perform full range of SZAs (depends on the station and season). By extending measurements over 90 degrees the profile retrieval has to account for a strong diurnal effect in order to derive a single ozone profile. Diurnal effect exists in the C-pair measurements, but it is likely more significant in the A-pair as it samples ozone at higher altitudes and thus encounters stronger photochemically driven changes in ozone.

This is a good point, and we thank the referee for bringing this to our attention. We do not account for diurnal changes in ozone. Figure 3 shows the diurnal concentration difference seasonally averaged over a year from the chemistry climate model: ACCESS-CCM (UM 7.3 and UKCA). The difference exceeds ten percent above around 50 km. This height range is relevant to all wavelength pairs when SZA exceed 90 degrees. This is highlighted in figure 4 for A-pair intensity curves. The ACCESS-CCM ozone profiles have also been *put into the forward model to show the difference in simulated N-value due to the diurnal cycle. This is shown in figure 5. There is around a 6 percent difference in the A-pair retrieval at 93˚. This is well within the measurement errors at that SZA. However, this may need to be taken into account in the future. We have added in the following statements to clarify this:*

p. 8675, line 12. The following Sentence was added: "Taken measurements beyond a SZA of 90˚ also introduces ozone diurnal effects that have consequences above 50 km. This is currently not taken into account, but is expected to be negligible due to the low ozone concentrations at these altitudes."

p. 8688, line 26. Included ozone diurnal cycle in future amendments.

p. 8681, lines 14-16. Can you please clarify what you mean by the statement - "the least smoothing occurs in layer 4" If you calculate the smoothing errors for all layers (Rodgers, 2000, eq. 3.29, or Bhartia et al., 2013, equation 9) – does it show the larges errors in layer 4? (WâA˘T˘I)*Sx*(W-I)ˆt, were Sx is the natural variability covariance matrix, it is different from Sa matrix that is used in the retrieval, W is the smoothing matrix W =ãA˘U˝SKãA˘U˚_nˆT ãA˘U˝(Kn SKnˆT+Se)ãA˘U˙ˆ(-1) Kn (Bhartia et al, 2013, equation 6)

Using Rodgers eq. 3.29. Ss does indeed show the largest errors in layer 4 (see figure 6). However, we agree that the statement "the least smoothing occurs in layer 4" needs clarification. We have made the following amendments.

p. 8681, line 14. The following sentence: "When looking at the designated C-pair measurement retrieval, it is seen that the least amount of smoothing occurs in layer 4, meaning this layer is influenced least by surrounding layers" has been changed to: "When looking at the designated C-pair measurement retrieval, it is seen that the least amount of influence from the a priori in the retrieval is seen in layer 4."

Discussion of Figure 3. The vertical distribution of the AKs seems to be unequally weighted – layers 5, 6, 7 and 8 have lower maximum while layers 2+3 and 4 show larger sensitivity in the profile retrieval. This may be caused by applying the "climatological" Sa which is constructed using time series of profiles from Hassler et al (2009) time series (representative of natural variability in the zonal mean approach). In Petropavlovskikh et al (2005) Figure 1 demonstrates the effect of using uniform vs. climatological Sa. May be this can be mentioned in the paper? Typically, one would prefer to have equally weighted AKs to get vertical information from profile, unless it is preferred to have enhanced sensitivity to the certain part of atmosphere.

The algorithm is also setup to also use Rodgers definition of Sa. This produces much more uniform AKs. But also changes the amount of *information retrieved, probably due to different scaling factor choices. The information gain when going from designated – measured – A+C+D measured is still similar using the full covariance uniform Sa. Therefore we decided to just show results from the first method.*

p. 8679, line 14. I have appended: "the effect of using these two methods to produce Sa has been discussed in Petropavlovskikh et al (2005).

It is important to make sure that the results presented in this paper as "C-pair retrieval: registered SZA" are not mistaken with the retrievals from Petropavlovskikh et al (2005). It would worth to add a sentence to clarify this point.

We have clarified that we are indeed using our algorithm for the designated SZA case data.

p. 8681, line 3. Changed: "Three cases are run, (1) running a spline through the operational C-pair measurements at designated SZAs (This is simulating the WMO reporting and fitting procedure),…" to "Three cases are run, (1) running a spline through the operational C-pair measurements at designated SZAs (This is simulating the WMO reporting and fitting procedure for use in this algorithm),…"

p. 8681 lines 21-22 – In comparisons between AKs for "raw" and "designated" measurements, and A+C+D vs. C -pairs do you use the same Sa in all retrievals?

Yes, the same Sa is used for all retrievals.

p. 8684, lines 17-19. I would not recommend adding the interpolated points to Figure 4 since it is misleading for the reader, even for the illustration purposes. Since, interpolation might create large differences between the two compared datasets, it is better to avoid it all together. The acknowledgement of the difference in sampling rates for the Umkehr and sounding records would point to the limitation of comparisons and would be appropriate in the discussion of results. I also find Figure 4 to be too busy to clearly see the points you are trying to demonstrate. I wonder if it is better to have an additional plot for the shorter time period when the 3-wavelength pair algorithm is applicable. Or replacing the entire time series with the shorter period when both types of algorithms are applicable (1970-1975) is another way to make comparisons more clear. Yet another approach is to replace a time series with a scatter plot. It would allow you to demonstrate the offset and correlation between several pairs of data in one plot. These results can also be captured in the table.

Figure 7 shows the reduced time series for figure 4 in the paper. Interpolation has also been removed as advised. This includes the update to the convolved case, taking into account considerations from Rodgers (2003). We have also *amended the following text to clarify that interpolation has no longer been performed, and a reduced time series is also now shown*

p. 8684, line 13. Changed: "1965-1982" to "1970-1975"

p. 8684, lines 16-18. Removed: "For visualisation purposes, any singular missing months in the Umkehr and ozonesondes time series have been interpolated from surrounding months"

p. 8685 lines 2-4. Following up on the discussion of Figure 4. It would be nice to demonstrate the result of application of the A+C+D pair AK smoothing on ozonesonde data vs. C-pair. It can be done on the data with even a shorter time scale (1970-1975 should have sufficient number of profiles for comparisons). This should be possible because when the A+C+D pair measurements are taken there is always the C-pair. The scatter plot of these results would be a nice demonstration of the advantage of the three pair retrieval. . . Otherwise it is hard to discern the difference in the Figure 4 format. One way to show the differences is to subtract the a priori profiles from the retrieved data (or to plot data as referenced to a priori, in percent). It will reduce seasonal cycle and allow for less noise in the results and thus would create cleaner comparisons.

Figure 8 shows a scatter plot of monthly average Umkehr C and A+C+D pair coincident retrievals compared to Ozonesonde AK convolved cases. We have not included this figure in the paper as the paper now shows similar information in the shorter time series as well as average percent differences provided in the text (see referee comment p. 8685, lines 17-21)

For layers 2+3 and 4. The C and A+C+D cases are quite similar and agree best with the ozonesonde data. For layer 5, the C and A+C+D pair data have consistently lower ozone compared to the ozonesondes, with the A+C+D pair case having slightly lower ozone compared to the C pair case. For layer 6, the A+C+D pair and C pair cases again show lower amounts of ozone compared to the ozonesonde data. The A+C+D pair is now showing slightly smaller amounts in layer 5 and slightly larger amounts in layer 6 compared to the C-pair but with a slightly worse R-value.

p. 8685, lines 9-10. Are you showing C-pair retrieval with all measured points in Figure 4? The AKs for layer 2+3 are more close between "C-pair: measured SZA" and A+C+D pair (0.7 vs. 0.8 at the maximum), as compared to "C-pair :designated SZA" (~0.55 vs 0.8) as shown in Figure 3. It should be clarified in the text by replacing "C-Pair" with "C-pair with measured SZA (CMS)" or "C-pair with designated SZAs (CDS)"

Yes, C pair retrievals with all measured points. Have changed to C-pair with measured SZA.

p. 8685, line 11, correct mistake in spelling Umkehr

Corrected, thank you.

p. 8685, lines 17-21. Please provide quantitative numbers of differences between two algorithms and in regards to the ozonesonde data. It is impossible to deduce results from the Figure 4.

The average percent difference between the two algorithms and the ozonesonde data for C-ACD coincident times only is for Umkehr C pair retrievals: Layer 2+3, -9.2%; Layer 4, -8.1%; Layer 5, 13.1%; Layer 6, 43.3%. Umkehr A+C+D pair retrievals: Layer 2+3, -12.5%; Layer 4, -8.6%; Layer 5, 16.2%; Layer 6, 41.3%. These numbers have been included in the text.

p. 8685, line 8. Included: "The average percent difference between ozonesondes and Umkehr data is -9.2% and -12.5% for C and A+C+D-pair data respectively for this layer."

p. 8685, line 14. Included: "The average percent difference between ozonesondes and Umkehr data is -8.1% and -8.6% for C and A+C+D-pair data respectively for this layer."

p. 8685, line 17. Included: "The average percent difference between ozonesondes and Umkehr data in layer 5 is 13.1% and 16.2% for C and A+C+D-pair data respectively for this layer, and 43.3% and 41.3% for layer 6."

p. 8685, lines 26-27. Add another reason to the list: "omission of the multiple scattering from the forward model".

Thank you, we have added the suggested reason to the list.

p. 8686, lines 1-3. I would rephrase it as "The algorithm presented here illustrates potential improvements to the ozone retrieval from the Umkehr technique. Results clearly demonstrate an increase in the informational content of the retrieved ozone profile when using measurements at three Umkehr wavelength pairs."

p. 8686, lines 1-3. Thank you, we have changed the sentence: "The algorithm presented here provides improvements to the most widely used Umkehr technique that are advantageous for obtaining a higher amount of information from the Umkehr measurements" to "The algorithm presented here illustrates potential improvements to the ozone retrieval from the Umkehr technique. Results clearly demonstrate an increase in the information content of the retrieved ozone profile when using measurements at three Umkehr wavelength pairs"

p. 8686, line 5. Please remove "accurately", replace "slight discrepancies" with numbers (residuals should be available from the retrieval of the data). Even if you can fit the profile to the observed Umkehr measurements by using only single scattering in the forward model, it may be for the wrong reason. For example, stray light effect

could be interfering with the Umkehr measurement fit that causes the offset in the retrieved profile.

p. 8686, line 5. Thank you. I have removed "accurately", and replaced slight discrepancies with the numbers from section 5.1 within the text.

p. 8686, lines 5-8. If you are outlining improvements in regards to the Umkehr retrieval ca 2005 algorithm (Petropavlovskikh et al., 2005), the removal of total ozone dependent a priori has been already applied (but not in algorithm ca 1992, Mateer and DeLuisi, 1992). It is better to state that "The presented algorithm uses a priori profiles and covariance matrix derived from the zonally and monthly averaged climatological dataset (Hassler et al., 2009). "

I have replaced the statement about not constraining to total column ozone with the suggested statement

p. 8686, lines 16-18. This is the first time that the spline extrapolation is discussed in the text. This should be moved into section 2.1. The estimates of errors for using spline extrapolation should be provided. It can be estimated from data that have data beyond 90-degrees SZA: applying extrapolation to the data with limit at 90-degrees SZA and comparing to the measurements between 90 and 94 degrees. In my experience the extrapolation at these SZAs does not work very well and should be avoided.

There must be some confusion, we did not perform any extrapolation, as noted by referee it would be very dangerous to do so. The text refers to the case where there was a limit at 90 degrees therefore no extrapolation was performed.

p. 8687, lines 8-7. Please provide the number to quantify "large". The number should be in %, which would allow for comparisons with the stray light effect (5-10 % in layers 7 and 8). (see above comments to p. 8685, lines 15-21).

Refer to referee comment p. 8685, lines 17-21

p. 8686, lines 26-28. You might add the following to the future work " assessment of the errors in the fitting of Umkehr curves in partial cloud conditions, assessment of errors in measurements at three wavelengths, assessment of the changes in a priori covariance on the ozone profile retrieval. And further validation work is needed to assess the improvement in operational ACD retrievals."

Thank you, we have included the appropriate statements to the future work.

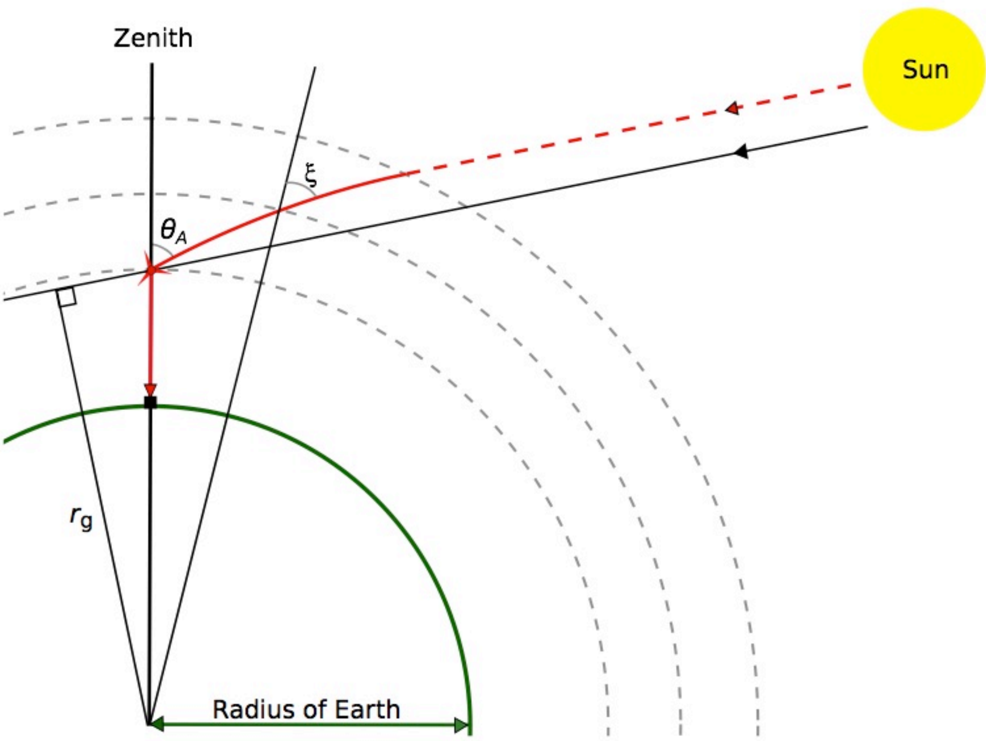

Figure 1. Refracted ray paths for a single scattering event. (For inclusion in the paper.)

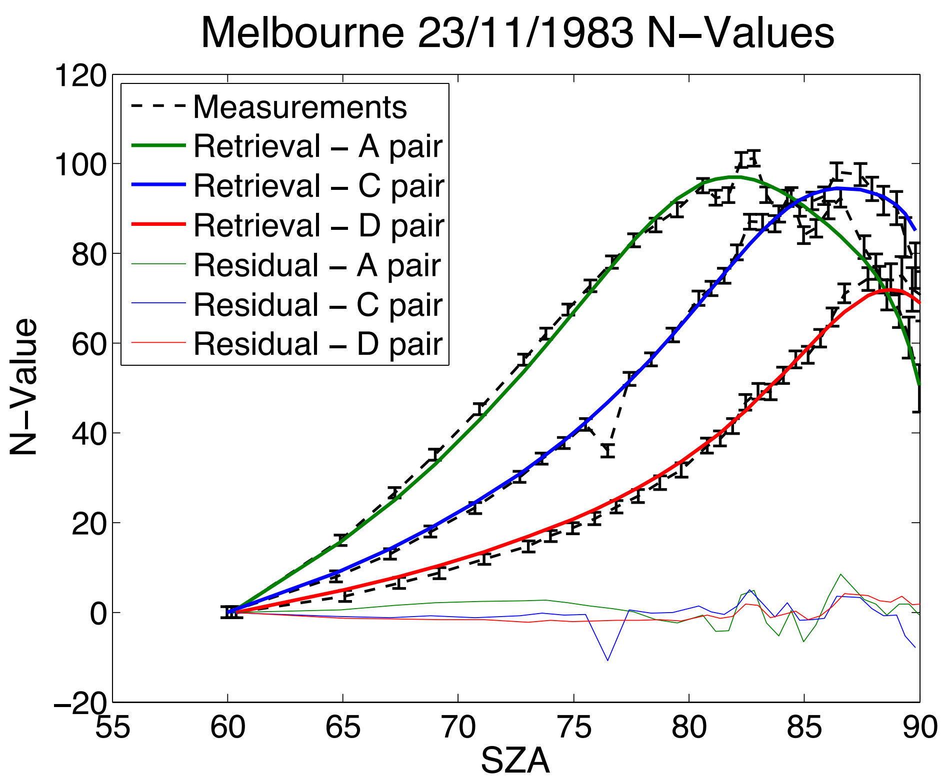

Figure 2. An example of a cloudy day retrieval.

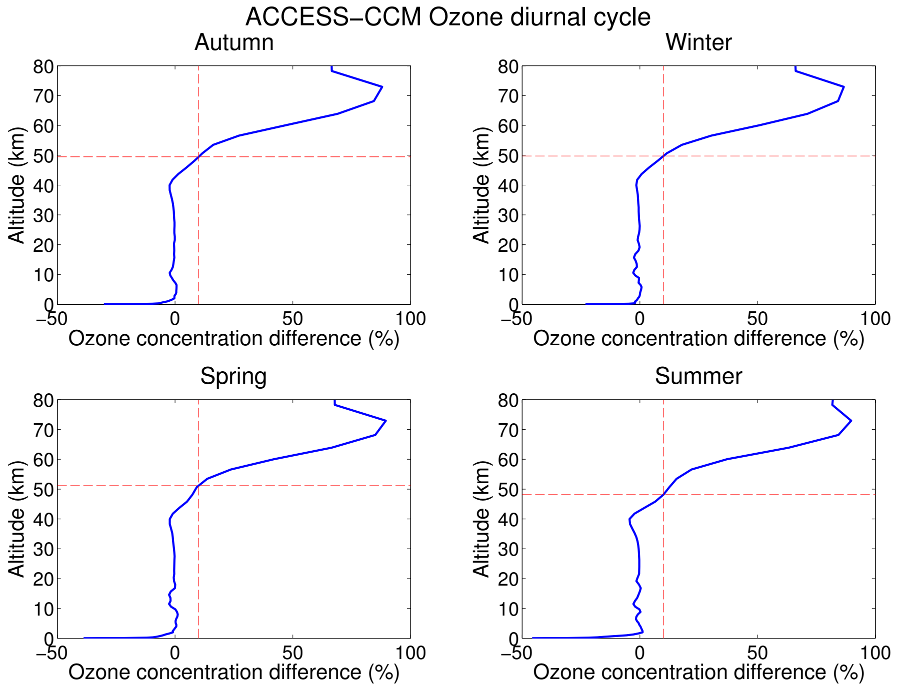

Figure 3. This figure uses 6 hourly averaged model output of ozone number density (molecules/cm^3), seasonally averaged over a year. The ozone concentration difference is 100(3pm-3am/3pm). Where 3pm and 3am are the center times of the 6 hourly averages. The red dashed lines show the ten percent mark.*

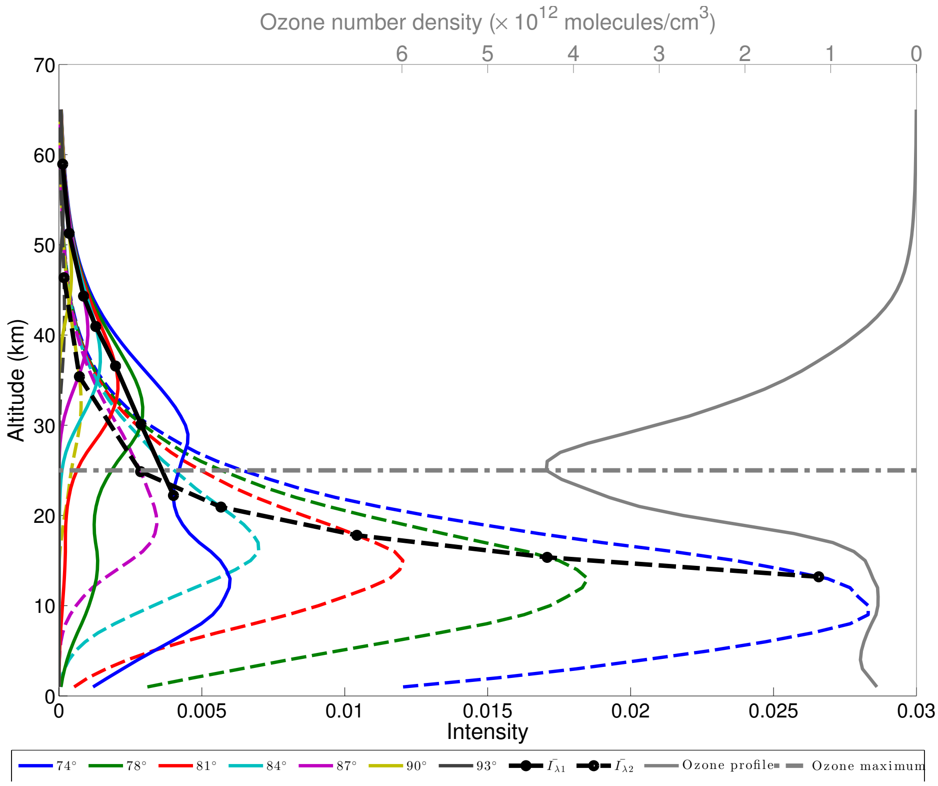

Figure 4. Zenith sky intensity contribution curves for A-pair wavelengths at selected SZAs for 13 April 1970. The weighted average height exceeds 50km after around 90 degrees for the A-pair short wavelength.

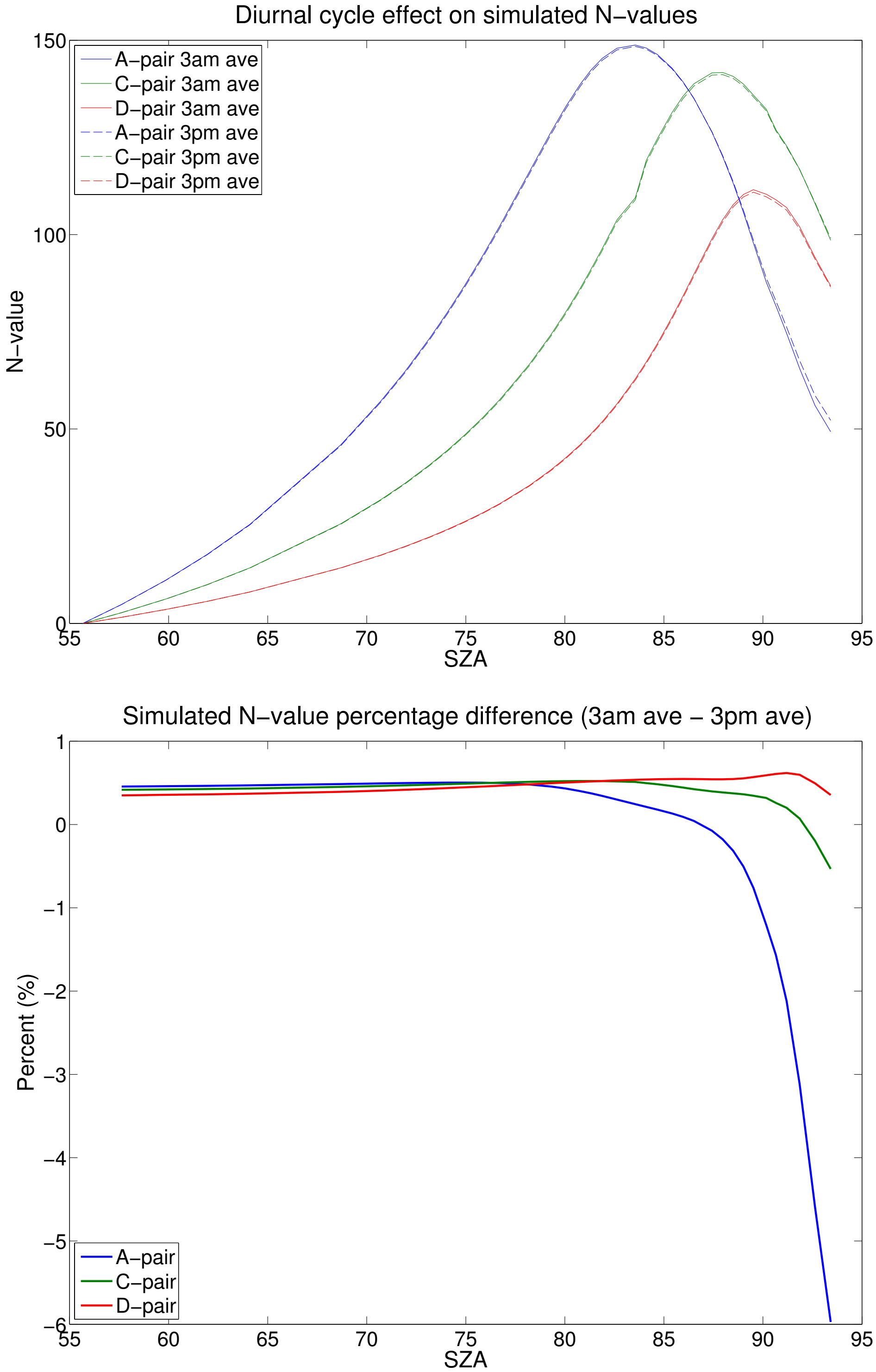

Figure 5. Difference in forward model simulated N-values when using the ACCESS-CCM 3pm and 3am 6-hourly ozone average profiles.

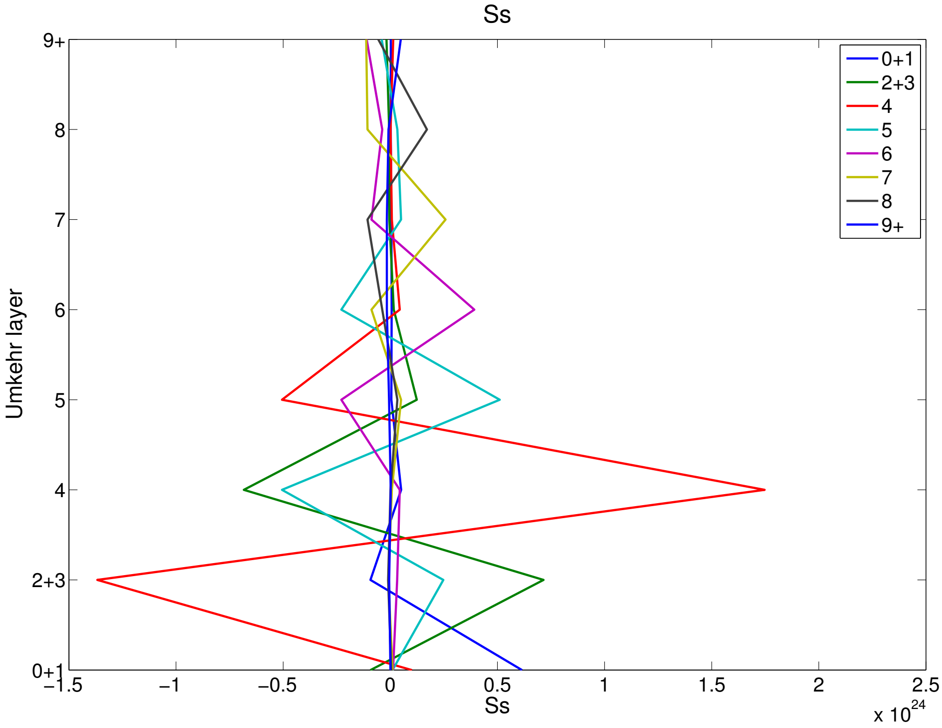

Figure 6. Ss for 29 January 1994 C-pair retrieval.

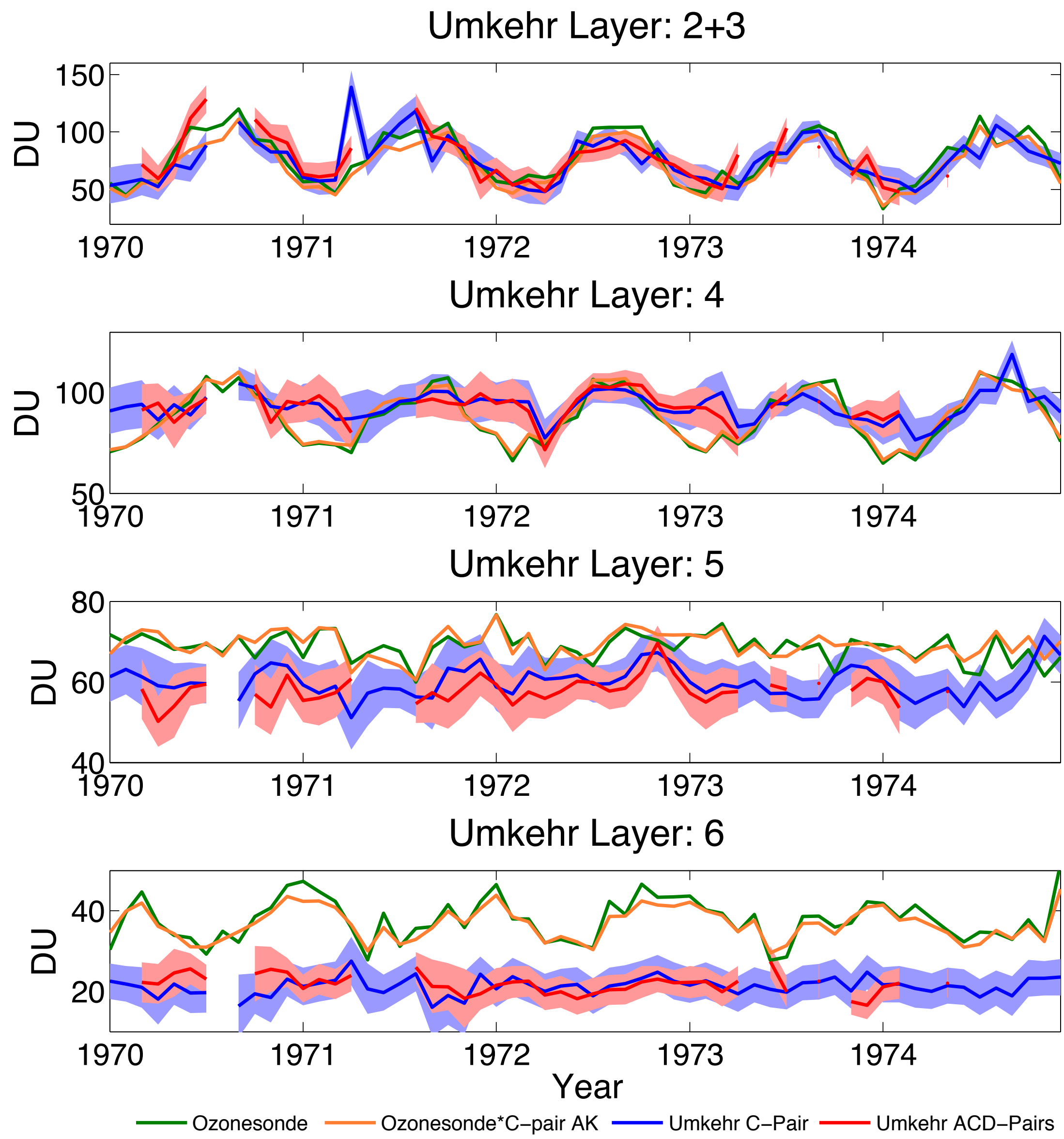

Figure 7. Update to paper figure 4.

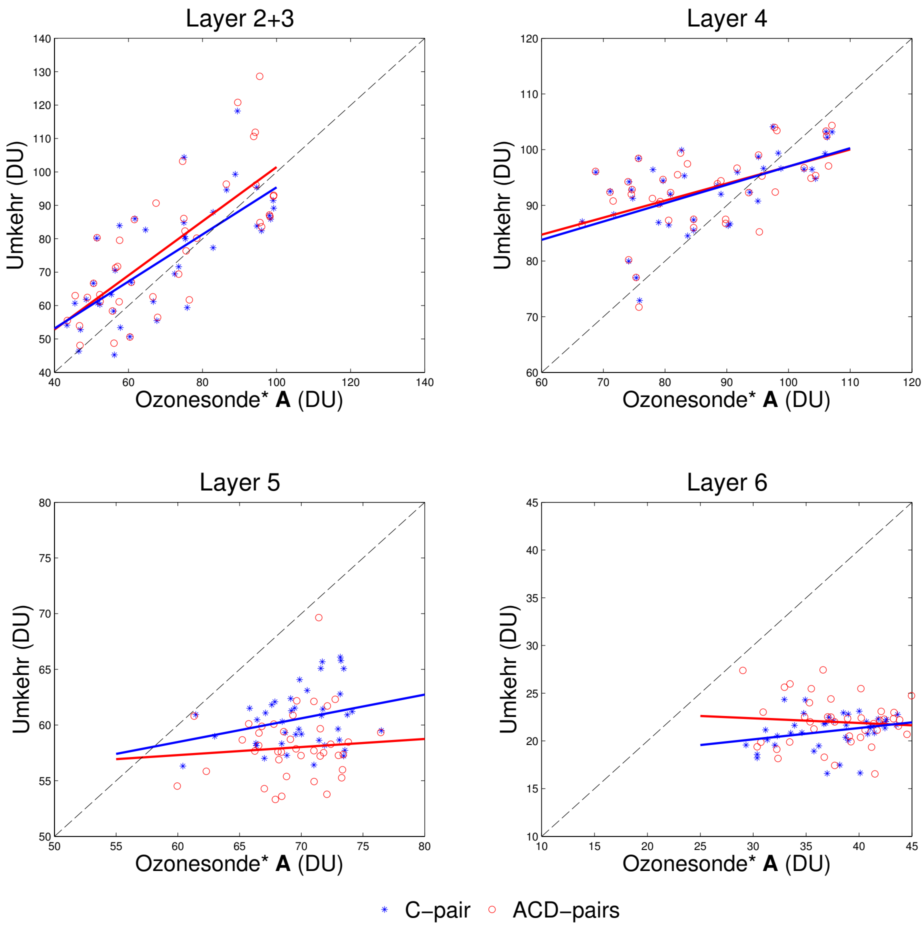

Figure 8. Scatter plots of monthly averaged Umkher C and A+C+D-pair retrievals against ozonesondes convolved against C and A+C+D-pair averaging kernels. Solid lines show least squares regression.