Response to comments of Referee #2

We thank the referee for his or her thoughtful comments, which we have addressed as follows:

Referee comment

1. It is good to provide more details about how to derive a priori profiles from MLS and ozonesonde data. Are MLS data collocated with ozonesonde data around the Summit station? How MLS and ozonesonde data are merged as they cover different altitude ranges? Have other ozone profile climatologies such as McPeters et al. (2007) and McPeters and Labow (2012) been considered?

Response

MLS data are "overpass" data provided by the Aura Validation Data Center at https://avdc.gsfc.nasa.gov/pub/data/satellite/Aura/MLS/V04/L2GPOVP_Prof/O3/Summit/ as indicated in the manuscript. Data files provide the distance between the locations of Summit and the MLS profile. On average, the distance is 160 km. This was added to the manuscript.

As stated in the manuscript, "*A priori* state vectors $\mathbf{x}_a$ were constructed by combining balloon sonde profiles for altitudes below 10 km and profiles measured by the Microwave Limb Sounder (MLS) on NASA's Aura satellite for altitudes above 10 km […] Profiles for both seasons were constructed by calculating the median of a large number of sonde and MLS profiles measured during the two periods using data from the years 2004 to 2014." We believe that this description is sufficiently clear to indicate how the profiles were constructed.

We have not considered the ozone profile climatologies such as McPeters et al. (2007) and McPeters and Labow (2012) because we felt that *a priory* profiles constructed specifically for the location at Summit, separately for the spring and fall periods and using only data from the time period of relevance (2004 - 2014) are the most appropriate profiles.

Referee comment

2. Instead of using fixed a priori error of 0.1 and 0.4, you mentioned the use of altitude-dependent a priori errors in the discussion (P21 L18), which is likely more appropriate as the ozone variability is relatively small in most of the stratosphere, ~10% based on your analysis, but increases significantly in the lower stratosphere and upper troposphere to ~40%. You can modify Eq 4 to be more generic, allowing for altitude-dependent a priori errors: [Sa]mn = sigma_amˆ2 * [Xa]m * sigma_anˆ2 * [Xa]n * exp(- |m-n|/d)

Response

We agree that the altitude-dependence of the ozone variability could be considered when setting up the covariance matrix $\mathbf{S}_a$. For this reason, we mentioned this possibility in the Discussion as one of the options to optimize the Global-Umkehr method further. However, considering that the results obtained with σ_a = 0.1 and 0.4 are fairly similar (see Table 2), we feel that the minor effect does not warrant a recalculation of all results.

Referee comment
3. P5, L8, one of the most important diagnostics is averaging kernels A, which is described in section 2.4. I suggested moving section 2.4 to in front of L8 as ds, is typically derived from A, as the trace of A. The diagonal elements of A are the ds at each layer.

Response
The contents of Sect. 2.4 were moved to the place suggested by the referee. We are aware that there are different methods to calculate d_S, all resulting in the same value for d_S. We describe the method suggested by Goering et al. (2005) because this is the method that was actually used in our calculations.

Referee comment
4. P6, Equation 8 is confusing. Looks like Dc(theta(t)) is not based on actual measurement, but based on the parameterization of clearly sky measurement as a function of SZA. You may change "Dc(theta(t)) is the measurement : " to "Dc(theta(t)) is the modeled photodiode measurement at time t that would be observed during clear skies, parameterized a function of SZA after filtering cloudy measurements." Also what criteria are used to filter cloudy measurements?

Response
The sentence was replaced with:
"$D_C(\theta(t))$ is the hypothetical clear-sky photodiode measurement at time t. The function was parameterized as a function of SZA using measurement of the photodiode obtained during clear skies. Clear-sky periods were determined based on temporal variability using the method described by Bernhard et al. (2008)."

Referee comment
5. It is better to switch sections "Retrieval method" and "Measurements" as the section of retrieval method depends on the description of measurements.

Response
We prefer to keep the original sequence of the manuscript, which starts with the fundamental equation of the optimal estimation approach (Gauss-Newton method) developed by Rodgers (2000). If we were to move the "Measurement" section to the front, we would also have to move the "Forward Model" section because the retrieval method also depends on the model. We feel that it is better to present the principle of the method first before discussing the various parameters that go into Eq. (1).

Referee comment
6. P6, L18-21, what is the main motivation of interpolating measurements to a common SZA grid that has 8 SZAs other than reducing the computation time. What is the typical number of spectra during the collection period (SZA change form 70 to 90)? Looks like it is much larger than 8, so interpolating it to 8 SZAs only while keeping the same measurement error can reduce the available information content and increase the measurement error. Have retrievals been conducted using the measurements at individual SZAs and compared with retrievals interpolated to 8 common SZAs?

Response
The interpolation of measurements to a common SZA grid is a common procedure for any Umkehr technique. For example, Petropavlovskikh et al. (2000) uses 14 fixed SZA between 60° and 90°. Using a common grid simplyfies calculations greatly. Note that our data always have to be interpolated because measurements at 310 and 337 nm are not performed at the same time and are therefore measured at slightly different SZAs. When developing our method, we also tried retrievals with up to eleven fixed SZAs. We found that there is virtually no benefit by increasing the number of SZAs beyond eight that would justify the greater computation time. This is not surprising because the "number of degrees of freedom for signal" d_S is typically less than 3.1, suggesting that eight SZAs are more than sufficient to characterize the information content provided by the observations.

The number of spectra recorded during the observation time varied between 17 and 52, so there are enough spectra for accurate interpolations. Because we use approximating (smoothing) splines for interpolations, random errors are reduced, so even though retrievals are only based eight SZAs, we take advantage of the much larger number of spectra.

We have not conducted retrievals using the measurements at individual SZAs, but as noted above, results did not change significantly by using more than eight SZAs.

We added the following sentence to the manuscript:
"Tests indicated that retrieval results do not change significantly by adding measurements at additional SZAs."

Referee comment
7. P6, L30, why not using more recent ozone cross sections based on the activities of ACSO (Absorption Cross-Sections of Ozone) summarized in Orphal et al. (2016), which recommends that the BP data should not be used. Is this for consistency with the OMI TOC retrieval, which also used the BP data?

Response
Note that Referee 1 had a similar comment.
We are aware that more accurate ozone absorption cross sections than those published by Bass and Paur (1985) are now available and recommended. Nonetheless, we decided to use the Bass and Paur (1985) data because OMI total ozone data are based on B&P. Using a different cross section would have complicated the validation of our results with OMI data. We added the following to the manuscript:
"While more accurate ozone absorption cross sections are now available (Gorshelev et al., 2014; Orphal et al., 2016), we used Bass and Paur (1985) data to facilitate validation with OMI total ozone column measurements, which are also based on Bass and Paur (1985)."

The following references were added:

Gorshelev V., Serdyuchenko A., Weber M., Chehade W., and Burrows J.P.: High spectral resolution ozone absorption cross-sections–Part 1: Measurements, data analysis and comparison with previous measurements around 293K, Atmos. Meas. Tech., 7, 609-624, doi:10.5194/amt-7-609-2014, 2014.
Orphal J., Staehelin J., Tamminen J., Braathen G., De Backer M. R., Bais A., Balis D., Barbe A., Bhartia P. K., Birk M., and Burkholder J. B.: Absorption cross-sections of ozone in the

ultraviolet and visible spectral regions: Status report 2015, J. Mol. Spectrosc., 327, 105-121, doi:10.1016/j.jms.2016.07.007, 2016.

Referee comment

8. Are both SDISORT and MYSTIC RTMs based on scalar (rather than vector) radiative transfer models? If so, this is another source of forward model bias. What are the impacts of neglecting polarization (i.e., assuming scalar) on the calculated radiances? Just check if any such analysis has been done for either SIDOSRT and MYSTIC RTM.

Response

Yes, both SDISORT and MYSTIC were run in scalar mode. We agree with the referee that the MYSTIC results may be biases relative to the "truth" by neglecting polarization (i.e., by not using a vector model). Errors arising from neglecting polarization in radiative transfer calculations have been quantified by Lacis et al. (1998). The authors conclude that errors for radiances can be as large as 10% for a purely Rayleigh scattering atmosphere. However, most of these errors cancel when integrating over viewing angles to calculate global spectral irradiance. Fig. 1 below is reproduced from the top-right panel of Fig. 3 of Lacis et al. (1998) and shows error in irradiance for a Rayleigh atmosphere that result from the omission of polarization.

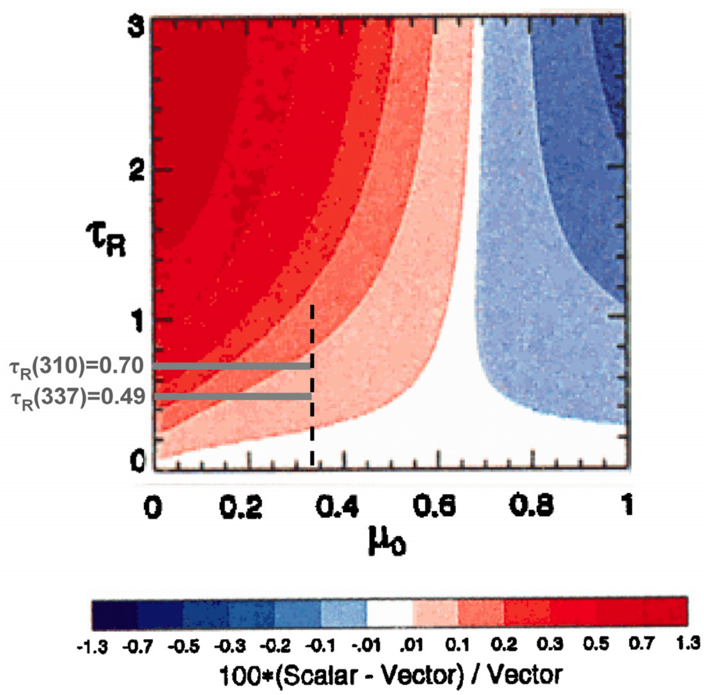

Fig. 1. (Scalar – Vector)/Vector errors for a Rayleigh atmosphere adapted from Fig. 3 of Lacis et al. (1998). Rayleigh optical depths of relevance for Umkehr retrievals are indicated by grey horizontal lines.

Errors are plotted as a function of the cosine of the SZA (abscissa), μ_0, and the Rayleigh optical depth, τ_R, (ordinate). Relative errors range between –1.3 and 1.3%. For our Umkehr retrievals, only the difference in errors between measurements at 310 and 337 nm is relevant. For the altitude of Summit, $\tau_0(310)$ is 0.7 and $\tau_0(337)$ is 0.49. These optical depths are indicated by grey horizontal lines in the figure below. Further, SZAs range only between 90° and 70°, corresponding to $0 \leq \mu_0 \leq 0.34$ (vertical broken line in the Fig. 1). It is apparent from Fig. 1 that the difference in irradiance errors for 310 and 337 nm is about 0.1% for all SZAs of relevance. According to Lacis et al. (1998), these errors are further reduced by a Lambertian surface, which is the case for Summit (snow surface with albedo of about 0.97).

Based on this analysis, we added the following to the Discussion:

"Finally, the MYSTIC Monte Carlo model, which was used to calculate the correction function $R(\theta)$ (see Fig. 1b), was run with a scalar radiative transfer solver, which did not take polarization into account. Lacis et al. (1998) calculated that modelling errors for irradiance resulting from the omission of polarization in these calculations can be as large as 1.3% for a Rayleigh atmosphere. However, errors for 310 and 337 nm (i.e., the wavelengths used in the Global-Umkehr method) agree to within 0.1%. We therefore conclude that the omission of polarization is not an import error source in our calculations."

Reference:

Lacis, A.A., Chowdhary, J., Mishchenko, M.I., and Cairns, B.: Modeling errors in diffuse-sky radiation: Vector vs. scalar treatment, Geophys. Res. Lett, 25(2), 135-138, 1998.

Referee comment

9. P10, L12, how is this threshold of 20 DU be determined?

Response

We assumed that changes in ozone occur linearly over time, so a 20 DU change over one day (as measured by OMI) translates to about 5 DU change over the course of the ~6 hours required for the Umkehr measurements. At Summit, the total ozone column varies between 250 and 350 DU during the spring period and between 320 and 480 DU during the fall period. So 5 DU make up about 1.6% of the column in spring and 1.3% in fall. Considering that day-to-day variations in the ozone profile occur mostly in the troposphere and lower stratosphere, relative ozone variations in these levels may exceed the percentages calculated for the column by a factor of about 2 to 3, resulting in relative variations of about 4% in these levels. Thus, by choosing a threshold of 20 DU, we ensure that variations in ozone over the course of the Umkehr observations are not a important uncertainty when comparing our results with MLS measurements. However, 20 DU is not a "magic number", and the criterion could be relaxed for operational processing.

We added the following to the manuscript:
"This criterion ensures that changes in the ozone profile remain below about 4% for all Umkehr layers."

Referee comment

10. P11, MLS measurements from consecutive days are used to quantify the temporal variation of ozone. It should be noted that the MLS measurements from consecutive days will be measured at different locations, maybe ~100 km apart. So some of the MLS1/2 difference is due to spatial variability. What is the average distance between MLS 1 and 2?

Response

We are aware of this problem and the following sentence had therefore been included in the Discussion of the original manuscript:

“However, a portion of the change in the MLS profile from one day to the next may be caused by the relatively poor horizontal resolution of MLS profiles of about 200 km. For example, some variability in the MLS overpass dataset can be attributed to the slightly different geolocation of two consecutive overpass profiles.”

In response to the referee’s comment, we have added the following:
“For example, the average horizontal distance between the locations of Summit and the MLS overpass is 160 km.”

The number was calculated from the “Distance to the station” field provided in the MLS data files.

Of note, in addition to distance, the different viewing geometry of MLS 1 and 2 data and diurnal variations of the ozone profiles in the upper stratosphere are also important sources of variability. This point was raised by Referee #1. Please see our reply in response to the remark of Referee #1.

Referee comment
11. P12, L3-15, a lot of the description can be reduced as this has been described in the figure caption.

Response
The description was slightly reduced in length (see annotated manuscript).

Referee comment
12. P21 L20, you may consider using some recent cross sections as suggested in Orphal et al. (2016), and use meteorological data (e.g., temperature profiles) to account for the temperature dependence of the ozone absorption cross section. To reduce the impact of Ring effect, you may consider optimizing not only wavelengths, but also the magnitude of bandpass (currently 2 nm) used to degrade the spectral resolution. In addition, you can also mention the correction of forward model errors due to the neglect of polarization as commented earlier.

Response
The following was added:
“More current ozone absorption cross section data could be used (e.g., Orphal et al., 2016) than the Bass and Paur (1985) data implemented in this work and by OMI. If temperature profile data are available, these could be utilized to account for the temperature dependence of the ozone absorption cross section.”
and
“For example, by degrading the spectral resolution (currently set to 2 nm), the impact of the Ring effect could be reduced. Finally, the MYSTIC Monte Carlo model, which was used to calculate the correction function $R(\theta)$ (see Fig. 1b), was run with a scalar radiative transfer solver, which did not take polarization into account. Lacis et al. (1998) calculated that modelling errors for irradiance resulting from the omission of polarization in these calculations can be as large as 1.3% for a Rayleigh atmosphere. However, errors for 310 and 337 nm (i.e., the wavelengths used in the Global-Umkehr method) agree to within 0.1%. We therefore conclude that the omission of polarization is not an import error source in our calculations.”

Referee comment
13. P22 L8, multiple scattering effect is also important for zenith sky measurements. You may say multiple scatterings effects become more important and the sphericity of the viewing geometry should be taken into account.

Response
Sentence changed to:
"Compared to the standard zenith-sky Umkehr method, multiple scattering effects become more important when exploiting global irradiance measurements, which also include contributions from photons received from directions close to the horizon. Therefore, the sphericity of the viewing geometry needs to be taken into account."

Referee comment
14. P24, L21, The poor sensitivity of the Umkehr method to ozone retrieval at layer 0 & 1 was mentioned here. Because only 2 wavelengths are used in the retrievals, measurements at other wavelengths especially the global irradiance spectrum can be used to improve the retrieval sensitivity in the first few layers as shown in Liu et al. (2005). You may add a few sentences about the possibility of exploring this for future studies.

Response
This comments seems to refer to P22, L21.
When we started working on the Global-Umkehr method, we were aware that tropospheric ozone profiles can be retrieved from off-axis radiance measurement by using more than two wavelengths in the UV. In addition to Liu et al. (2005), this method has for example also been used by Tzortziou et al. (2008). When developing our method we therefore explored retrievals with additional UV wavelengths, specifically combinations of E(305)/E(337); E(310)/E(337); E(325)/E(337), see P3, L5 of manuscript. We were hopeful that the tropospheric resolution of the Global-Umkehr method could be improved by including this additional spectral information. Unfortunately, adding these additional wavelength pairs did not lead to a significant improvement. We are therefore not hopeful that tropospheric ozone profiles can be retrieve from global spectral irradiance spectra without adding additional viewing angles, which is not possible for our instrument. We therefore do not think that it is appropriate to raise hopes. No change to manuscript.

Technical comments

1. P2, L20, change "a.s.l" to "a.s.l."
Period added.

2. The section of "1 Method" should be "2 Method" and "1.1 Retrieval method" should be "2.1 Retrieval method"
Corrected.

3. P2, L25, change "depends" to "depend"
Changed. (Typo was on P4, L25).

4. P5, L17, "and is part of… :"

No change because the sentence is an enumeration with the phrase in question being the second of three phrases.

5. P5, L27, change “wavelengths shifts” to “wavelength shifts”
Extra “s” removed.

6. P6, L17, this sentence can be grouped to the above paragraph.
Paragraph break removed.
7. P6, L28, change “result are” to “results are”
“s” inserted.

8. P7, L6, change “reference” to “references”
“s” inserted.

9. P7, L24, change “considered” to “considered as”
“as” inserted.

10. P15, L18, add “,” before “interquartile”
Comma inserted.

11. P19, L28, change “decreased” to “has decreased” or “decreases”
“has” inserted (but on P18, L28)

12. P19, L18, change “is varies” to “varies”
“is” deleted.

13. P19 L26ïijˇN change to “compared”
“McElroy and Kerr (1995) compare Umkehr profiles” changed to “McElroy and Kerr (1995) compared Umkehr profiles”

14. P20, L1, L2, L5, L6, change to “compared”, “found”, “concluded”, “compared”
Changed to past tense as suggested.

15. P23 L10, change to “on a weekly basis”
“a” inserted.

16. P24, L17, add “,” after “Phys.”
Comma inserted.

17. P24, L18, add “,” after “Res.”
Comma inserted.

18. P26, last line, use normal font for the journal title.
Italic formatting removed.