Review #1.

We thank the referee for the thoughtful review and comments that helped us to improve the manuscript. To address some of the referees' comments we decided to attach a Supplement, where some additional results and discussion are presented.

Please, find our responses to the referee's comments below.

Referee: p 2723 1st par: Is there any paper available where the most recent retrieval algorithm is described in more detail? If so, a reference to this would be appropriate. If not, it might be helpful to add some more details on the retrieval, e.g. with respect to clouds, surface albedo, further fit variables except ozone partial columns constraints and a priori used, etc. (c.f. comment on p2734 l22)

Authors: Thank you for the comment. We added section 2.3.1 that briefly describes the main features of the new version 8.6 algorithm that are relevant to the problems discussed in this paper. We also put a reference to the manuscript that fully describes the SBUV retrieval algorithm. We added:

"**2.3.1 SBUV version 8.6 algorithm**

In this section we outline the main features of the SBUV v8.6 retrieval algorithm, fully described by Bhartia et al. (2012), that are relevant to the present study. In v8.6 the Optimal Estimation technique (Rodgers, 2000) is used to retrieve ozone profiles as partial ozone columns (DU per layer) at 80 pressure layers plus a top layer above 0.1 hPa. The seasonal ozone climatology, derived from Aura MLS and ozone sonde observations (McPeters and Labow, 2012), is used by the retrieval algorithm as the a priori information. The a priori covariance matrix $\mathbf{S_a}$ is constructed assuming that the variance at each layer is equal to a constant fraction of the a priori, and that adjacent layers are highly correlated: $\mathbf{S_a}(\mathrm{i,j}) = \sigma^2 \boldsymbol{x_a}(i)\boldsymbol{x_a}(j)e^{-|i-j|/N_c}$, where $\mathbf{x_a}$ is the SBUV a priori; *i* and *j* are layer indices; σ^2 is the fractional ozone variance, and N_c is a number of adjacent layers that are highly correlated. We set σ=0.5 and N_c=12 (~10 km) in the v8.6 algorithm. The algorithm uses the same a priori covariance matrix for all latitude bins and seasons. The measurement error covariance matrix $\mathbf{S_e}$ is constructed as a diagonal matrix with the diagonal elements σ_e=0.43 N-value, where N-value is the logarithm of the backscattered radiance to solar irradiance ratio: $N_{value} = -100\log_{10} {I}/{I_0}$ (see Bhartia et al., 2012).

In v8.6 ozone profiles are reported as partial ozone columns (DU per layer) at 20 pressure layers (plus a top layer above 0.1 hPa) by combining ozone in every 4 retrieved layers. The 81 layers (80 plus a top layer) are needed to increase the accuracy of the forward model calculations, but the vertical resolution of the SBUV measurement system is much coarser, thus it is reasonable to report data at thicker layers. All correlative quantities, such as a priori, Jacobian, **A** matrix etc., are reported at the same 20 layers. The total ozone columns are calculated as sums of the partial ozone columns at all 21 layers.

For the first time, v8.6 SBUV mzm ozone profiles have been released as a primary product for use in the long-term time series analysis. The mzm profiles are calculated in 5º latitudinal bins with midpoints starting at 87.5º S by simply averaging individual profiles in the specific month and latitude bin. The smoothing errors are calculated for the mzm profiles."

Referee: p 2724 l6: The smoothing error represents differences not only due to vertical smoothing but also due to biasing by the a priori (particularly if the sums over the averaging kernels are not unity). I suggest to write "...due to vertical smoothing OR ANY OTHER EFFECT OF THE CONSTRAINT TERM by the retrieval algorithm."

<u>Authors:</u> We agree that the smoothing error depends on the a priori and we do say that in the next sentence. However, we believe that our statement is accurate since the a priori constraint is the part of the retrieval algorithm.

<u>Referee:</u> p 2725 l19: "an ensemble of true states" sound a bit vague to me. Care must be taken that C is representative for the sampling of the instrument under investigation.

<u>Authors:</u> Thank you for the comment. We agree, and we removed the word "true".

<u>Referee:</u> p2726 l3: "on the magnitudes AND INTER-ALTITUDE CORRELATIONS of the..."

<u>Authors:</u> Thank you, we added.

<u>Referee:</u> p2726 l17: Why are ideal AKs Gaussian? Why not rectangular of triangular?

<u>Authors:</u> We re-phrased: " An idealized AK for a defined layer would have a δ-function shape with an integrated value of about one, ..."

<u>Referee:</u> p2727 l6: Some more motivation why the normalization is done would be helpful. Is the Rodgers formalism still applicable to normalized smoothing errors? This issue deserves a more thorough discussion. Does this imply that there is somewhere a hidden transition between two representation systems (e.g. concentrations vs. partial columns)? I.e. is the retrieval performed in concentrations but the data are finally represented in partial columns? I am a bit confused about this issue. I trust that it is correct what the authors are doing but I am afraid that I miss an important piece of information to understand the rationale behind this. And please make clear at any point in the manuscript where you use the original averaging kernels and where the normalized ones are used.

<u>Authors:</u> Thank you for the comment. We realized that we didn't explain it clearly in the text. The SBUV algorithm retrieves ozone as partial columns at 81 pressure layers, and reports profile at 21 layers. All correlative data (**A** matrices, Jacobians etc.) are also reported for the ozone partial columns at 21 layers. The smoothing errors have been calculated in the same units as the retrieved ozone profiles (DU per layer) using **A** matrices.
We show normalized AK only in Figure 2. This was done because the shape of Averaging Kernels, associated with the partial column profiles, is very different from the familiar bell-like shape. This issue is fully discussed by Bhartia et al. (2012). We considered normalized AK to simplify a visual analysis.
To avoid any confusion in the revised version of the manuscript we define a term "**A** matrix" (or **A** in equations) to refer to the matrix associated with the partial column profiles and used in all calculations. We use a term "normalized **A** matrix" (or $\mathbf{A_n}$ in equations) to refer to the normalized matrix shown in Fig. 2 to simplify the visual analysis (and $\mathbf{A_n}$ not used for the calculations). We put some relevant clarification in new section 2.3.1. We also added an explanation in section 2.3.2:

" The **A** matrix is relevant to profiles of partial ozone columns in units of DU per layer, but the shapes of Averaging Kernels (rows of **A** matrix) are different from well known bell-shape. To simplify visual analysis, we show rows of normalized $\mathbf{A_n}$ matrices in figure 2. The normalization is done as follows:

$$\mathbf{A_n}(\mathbf{i},\mathbf{j}) = \mathbf{A}(\mathbf{i},\mathbf{j}) * \boldsymbol{x_a}(\boldsymbol{j})/\boldsymbol{x_a}(\boldsymbol{i}) \quad (3)$$

where $\boldsymbol{x_a}$ is the SBUV a priori profile, and *i* and *j* are layer indices from 1 to 20. This normalized $\mathbf{A_n}$ matrix is applicable to the ozone profiles expressed as a fraction from the a priori. We need to emphasize that for the smoothing error calculation the original **A** matrices have been used. "

Referee: p2728 l9: "... is the A PRIORI covariance matrix." (because in the retrieval there is also the measurement error covariance matrix).

Authors: The covariance matrices **C** we use for the smoothing error estimation are different from the a priori covariance matrices $\mathbf{S_a}$ used in the retrieval (see new section 2.3.1).

Referee: p2728 l12: Why only "year-to-year" variability? Don't you lose variability if you restrict yourself to year-to-year variability? (c.f. comment on p2734 l22)

Referee: p2734 l22: Here it is stated that seasonal prior is used. Without this information some of the earlier contents of the paper cannot be understood, e.g. why this study focuses only on year-to-year variability (c.f. comment on p2728 l12)

Authors: The SBUV measurements are very often used for long-term trend analysis. For trend analysis ozone data are usually averaged by month and latitude bin (monthly zonal means). Thus our goal was to estimate the smoothing error for mzm profiles. Since in the retrieval algorithm we use seasonal a priori profiles, we estimated year-to-year ozone variability by constructing covariance matrices from the 6-year record of mzm profiles.

Referee: p2729-2730: You have made tests that the use of only the off-diagonal elements of C is a justified simplification in your case. This is a consequence of the limited number of altitude grid points of the retrieval. For other applications the off-diagonal elements might be essential. I suggest to add "in our case" here and there in order to avoid that an unexperienced reader understands this as a universal statement.

Authors: Thank you, we added in the text "in our case" where appropriate.

Referee: Sect 3.1. It is not clear to me if full C or diagonal C is used here. Further, it remains unclear to me why the investigation is not made on the basis of the full C.

Authors: In all calculations we use full covariance matrices C and full A matrices (all elements: diagonal and off-diagonal). We did two test runs where we set the off diagonal elements of A and C equal to zero. Whenever we discuss test runs we always mention that in the text. All results presented in section 3 are for the case where we use full C and A matrices. To emphasize that we added the following text in the beginning of section 3:

" For each SBUV mzm profile the smoothing error covariance matrix $\mathbf{S_{serr}}$ was calculated using Eq. (1). All elements (diagonal and off-diagonal) of the **C** and **A** matrices are included in the computation of the smoothing error. The diagonal elements of $\mathbf{S_{serr}}$ represent the error variances of the elements of the SBUV mzm profile $\hat{\boldsymbol{x}}$, and the off-diagonal elements of $\mathbf{S_{serr}}$ indicate the inter-level error correlations (Rodgers, 1990). When the off-diagonal elements of $\mathbf{S_{serr}}$ indicate that the errors are highly correlated, then we have more information about $\boldsymbol{x}$ (Rodgers, 1990) and the errors are expected to be smaller."

Referee: p2730 l15 "caused" seems more appropriate to me than "defined".

Authors: we replaced "defined" by "caused".

Referee: p2731 l8-11: With diagonal A with diagonal elements lower than one, A does not smooth the profile but just tells how large the weight of the measurement is with respective to the weight of the a priori. This leads to another question: Are the mean values of SBUV and MLS the same? The MLS covariance matrix tells you how the MLS values vary around the MLS mean value. What you need, however, is a covariance matrix which tells you how the true values vary around the a priori actually used. Is there a "mean smoothing error" to be considered, which would result in a bias? I do not want to urge you to do the study again, but if these considerations are not included, this should be clearly stated. This issue may deserve some discussion. Another issue: Have you subtracted the MLS measurement error covariance matrix to get the covariance matrix representing only the natural variability?

Authors: Thank you for the comment. We added in sect 2.4 par. 4: " The calculated covariance matrices **C** represent the variability of the merged MLS/sonde data about their mean. We assume that the MLS and sonde measurement error covariance matrices $\mathbf{S_e}$ are small compared to **C**, thus the **C** matrix represents natural ozone variability. However, to compute the actual smoothing error we need to know the variability of the 'true' ozone profiles about the SBUV a priori. The difference between the estimated and 'true' variability will add additional errors in smoothing error calculations, but since the 'true' state is not known these errors cannot be estimated."

Referee: p2731 l15/16: The recommendation "to convolve ... with the SBUV AK (or integrated kernels)" is a bit vague. Are indeed both these approaches correct? What would be the difference?

Authors: Bhartia et al. (2012) discuss how to normalize reported **A** matrixes to apply them to convolve mixing ratio profiles. To avoid any confusion we re-phrased our statement (also see response on comment "p2727 l6" above) :
" One approach to such comparisons is to convolve a highly resolved profile with the SBUV AK as shown in Fig. 1. The profile with finer vertical resolution should be degraded first onto the SBUV vertical scale and then convolved using the SBUV **A** matrix (Rodgers and Connor, 2003)".

Referee: p2731: Not sure if the Rodgers book is the appropriate reference. I think C. D. Rodgers and B. J. Connor, Intercomparison of remote sounding instruments, J. Geophys. Res., 108, D3, doi10.1029/2002JD002299, 2003 is the appropriate reference. To my knowledge, this particular concept of intercomparison is not yet covered by the book.

Authors: Thank you, we added the reference.

Referee: Eq. 5: It would be helpful to derive Eq 5 from generaslized Gaussian error estimation: The partial column of a merged layer is the sum of the partial columns of the individual layers. The 2-layer merging operator thus is (1,1). The resulting error is thus (1,1) S_serr(1,1)ˆT where ˆT denotes the transposed and where S_serr is the submatrix of S_err with elements related to the layers to be merged. This results in Eq 5. The reader might appreciate some guidance how Eq 5 is obtained.

Authors: We included an explanation to the Supplement:

"Derivation of equation 5.

To calculate the smoothing error for the thick, combined layer we use the following equations:

$$S_{\text{merged}} = \boldsymbol{L^T S_{serr} L} \quad \text{(S1)}$$

Where $\boldsymbol{L}$ is a state vector with 1 for the layers to be merged and 0 elsewhere.

If, for example, we want to merge layers 2 and 3 then the state vector $\boldsymbol{L}$ will be $\boldsymbol{L} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$$S_{2+3} = [0 \quad 1 \quad 1] * \begin{bmatrix} S_{11} & S_{21} & S_{31} \\ S_{12} & S_{22} & S_{32} \\ S_{13} & S_{23} & S_{33} \end{bmatrix} * \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} =$$

$$[S_{12} + S_{13} \quad S_{22} + S_{23} \quad S_{32} + S_{33}] * \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = S_{22} + S_{23} + S_{32} + S_{33} = \sum_{i=2}^{3} \sum_{j=2}^{3} \mathbf{S}(i,j)$$

(S2)

Thus the smoothing error for the thick merged layer can be calculated using the expression:

$$S_{serr}^{k_o,k_n} = \sqrt{\sum_{i=k_0}^{i=k_n} \sum_{j=k_0}^{j=k_n} \mathbf{S_{serr}}(\mathbf{i},\mathbf{j})} \quad \text{(S3)"}$$

Referee: p2732 l21: It is not at all clear to me how the DFS can increase by merging.
Referee: p2735 l20: Is the DFS of the thick layer really larger than the sum of the DFS of its "parent layers"? How can this be? Merging into one layer is a kind of "hard constraint" which should reduce the total DFS of the profile. Or do you mean larger than the DFS of each of the parent layers?
Referee: p2735 l22: I do not understand how it is possible to get more information by merging layers. I agree that the DFS of the thick layer can be larger than any of its parent layers but this does not maximise information. If it is argued with the term "information", the Shannon information content of the entire profile needs to be formally evaluated.

Authors: Thank you for the comments. You are absolutely right, the DFS of the thick layer cannot be larger than the sum of the DFS of its parent layers, and information content is not increased by merging layers. What we meant in the text was that the DFS of the thick layer is larger than the DFS of any parent layer. We re-phrased our statements in section 3.2 and in Conclusions to clarify that. We added: "If the thickness of the combined layer is close to the vertical resolution of the measured signal, then the smoothing error for the combined layer will decrease. The DFS of the combined layer is equal to the sum of the 'parent' layer DFS. "

Referee: p2733 l8/9: This is certainly true for concentrations but I doubt that it is true also for partial columns. A more precise wording is needed here. From Eq. 5 at least it is not obvious how the error becomes less, unless there are large negative correlations.

Authors: The main factor in reducing the smoothing error for the thick layer is a large inter-level correlation. We assume that the diagonal elements of the smoothing covariance matrix represent the layer smoothing error. This approach provides a good estimation of the range of the smoothing errors. However, for the precise calculations the off-diagonal elements should be considered as well. We include the following text: " The high negative inter-level correlation of errors (off-diagonal elements of $\mathbf{S_{serr}}$) plays a significant role in reducing the merged layer error (see Supplement Fig. S3). "

We also included figures in the Supplement that shows inter-level error correlations.

Referee: p2733 l23: what do these standard deviations refer to? Standard deviations of the differences between SBUV and MLS? Please specify.

Authors: We added: " deviations of the differences between SBUV and MLS mzm measurements."

Referee: p2734 l7: "2-sigma range" is a bit ambiguous. Do you mean plus minus 1 sigma (which gives a span of 2 sigma) or do you mean plus minus 2 sigma?

Authors: In Fig. 8 we showed ±2σ range, because in some layers ±1σ range is too small and cannot be visually distinguished from the line itself. We specified that in the text: "... indicate the ±2-σ range of the calculated SBUV layer smoothing error. "

Referee: p2734 l19: Is this really a limitation of the SBUV algorithm, or a limitation of the measurement system. The wording "limitation of the algorithm" suggests that with another algorithm better information could be retrieved from the same measurement data. I doubt that this is what you intend to say.

Authors: Thank you for the comment. We changed:" ... the limitation of the SBUV measuring system. "

Referee: Fig 3: It is misleading to represent the DFS by a continuous line because these are discrete numbers referring to the layers. The total DFS is not the integral over the line but the sum of the dircrete DFS. Thus, symbols should be plotted. If need be, these can be connected by a faint line to guide the eye but the information is contained in the symbols, not the line.

Authors: We agree. We updated Fig.3.

Referee: Fig 4 vs. Fig 5: According to the plots, the smoothing error can even be larger than sqrt of the diagonal of C, (particularly near 100 hPa) i.e. after the retrieval there seems to be less information than before. How can this be? Does the retrieval destroy information?

Authors: Thank you for the observation. Indeed, in a few cases the smoothing error can be larger than the square roots of the diagonal elements of C. In particular, this is true for layers 6 and 10 in the tropics. This is due to the assumption we made assigning the layer smoothing errors to be equal to the square roots of the diagonal elements of the smoothing error covariance matrix $\mathbf{S}_{\mathbf{serr}}$. Also, some portion of the differences between Fig. 4 and Fig. 5 is due to different references that we used to convert ozone columns in percent. In Fig. 4 we normalized the diagonal elements by the a priori, while in Fig. 5 errors were calculated as % from the retrieved profiles. We updated Fig. 5 using the same a priori profiles (as we used for Fig. 4) to normalize errors.
We added: " In very few cases, for example in layers 6 (100-63 hPa) and 10 (16-10 hPa) in the tropics 10° S-10° N, the smoothing errors are larger than the estimated ozone standard deviations (square roots of the digonal elements of **C,** Fig. 4). This is a limitation of our approach considering only the diagonal elements of $\mathbf{S}_{\mathbf{serr}}$ and ignoring inter-level error correlations to estimate the layer smoothing errors. "

Technical comments:

Referee: p 2723 l23: The abbreviation mzm is defined only in the abstract but not in the body of text. However, since both the abstract and the body of the text must stand alone, a definition in the body of the text is necessary, and the definition in the abstract might be obsolete.

Referee: p2734 l1: Shouldn't the caption read "QBO"?

Authors: Thank you for the technical comments. We accepted all technical corrections pointed by the referee and made appropriate changes.