We appreciate comments from Reviewer 1 for the improvement of the manuscript. We color code this document as follows:

Black: Comments from Reviewer 1.

Blue: Our response to Reviewer 1.

Green: Changes made in the manuscript reflective of Reviewer 1's comments.

General Comments

The manuscript entitled „Troposphere – stratosphere integrated BrO profile retrieval over the central Pacific Ocean" by Koenig et al. presents a novel retrieval algorithm for the simultaneous determination of tropospheric and stratospheric trace gases and aerosols from MAX-DOAS measurements performed on mountain tops.

Knowledge on the composition of the atmosphere, in particular in the upper troposphere and stratosphere, is crucial for the understanding of the atmospheric chemistry and its impact on climate. MAX-DOAS measurements provide a simple and cost-effective way to retrieve information on the vertical distribution of atmospheric trace gases. The mountain-top based MAX-DOAS system presented here allows for gaining enhanced information on the UT/LS region. The authors describe novel and improved retrieval methods, and the subject of the manuscript fits well into the scope of AMT.

The level of agreement of modelled and measured diurnal variability of dSCDs is impressive, indicating that the retrieval is modelling the atmospheric radiative transfer realistically. However, I find it difficult to understand the principal approach of the so-called "Time-dependent retrieval", and I feel that the respective section 2.5.3. requires substantial revision as detailed in the specific comments.

The abstract is too long and should, as an AMT contribution, focus more on the retrieval algorithm itself than on the chemistry of the UT/LS region, for which far too many details are provided.

We thank Reviewer 1 for highlighting the need for clearer explanation of the methods, especially the time-dependent retrieval also identified by Reviewer 2.

We have shortened the abstract in the revised manuscript and improved its readability. A certain length is needed to descript the key points the paper, incl. key features of the retrieval methods, figures of merit, atmospheric state changes in BrO columns and concentration profiles that are the result of time dependent optimal estimation, and the evaluation of the MT-DOAS profiles using aircraft measurements. The revised abstract is not longer than published in AMT articles elsewhere.

Specific Comments

L13: The term "trace element" is not appropriate for bromine since it has an impact on atmospheric chemistry not only in its elemental form.

With the reworking of the abstract this sentence and phrasing no longer occurs.

L76: It would be good if you would be more specific regarding the exploitation of the "sun motion" (i.e., increase in light path with increasing SZA).

We have reworded and further specified: …motion of the sun varying atmospheric path length and scattering attitude …

L149: It is possible to fit certain parameters or state/measurement vector elements, but it not possible to fit "degrees of freedom". Please specify what is fitted here.

We have changed degrees of freedom to width parameters

L161 and thereafter: What do you mean with the term "principal program"? Are the RTMs not just programs?

This wording was chosen so that it can be understood at the end of the paragraph that we ran both RTMs for all the data for purposes of comparison but only utilize one for the retrieval. On rereading the fact of comparison is likely understood regardless and we have edited it for clarity as follows:

Two radiative transfer codes were used for this study. For ZS-DOAS measurements, Discrete Ordinate Method Radiative Transfer (DISORT) was used and for OA measurements the Monte Carlo Atmospheric Radiative Transfer Inversion Model (McArtim) was used.

L177: Can you give an estimate how large the errors are if the atmosphere below the instrument is not considered in the DISORT model? Can this simplification be justified?

The effect is most clear in O_4-based aerosol retrievals which have better signal to noise. We examine this in section B of the supplement focusing on the impacts on O_4. The O_4 concentration profile is such that it presents a worst case scenario for BrO. The key effect is increased signal contribution from lower altitudes which can be partly reproduced by increasing albedo in DISORT. If DISORT were used for OA geometries, the effect could be as large as ~10% but for ZS data this agrees with McArtim results including lower altitudes to better than ~1% differences.

Section 2.3.: Please explain why you use two different RTMs in a single retrieval algorithm. This is a quite unusual approach. McArtim is capable of modelling both tropospheric and stratospheric radiative transfer also during twilight, so it is not clear what the advantage of using DISORT is.

We agree that McArtim is capable of modeling stratospheric radiative transfer at twilight, however, it presents technical challenges. We found some differences between the models even in 1D implementation around SZA = 90° (this included running McArtim with and without forcing twilight), but these could likely be resolved. The larger problem was finding a suitable 2D implementation of McArtim to use in conjunction with UVspec. In DISORT the 2D geometry is defined along the solar azimuth which is ideal for accounting for photochemical effects along the principle line of sight. The documentation of 2D and 3D implementations of McArtim is relatively sparse, but as far as we could find we are limited to lattice in altitude, latitude, and/or longitude. We successfully ran altitude-longitude, and altitude-longitude-latitude runs, which had seemingly sensible results but are much less facile to combine with the photochemical data. Furthermore, these McArtim runs were particularly computationally and time inefficient in comparison to DISORT. From the documentation available it appears that other geometries could be defined (at least in some versions of McArtim) but attempts to implement this failed. In addition, for all the McArtim versions tested we found that attempting to combine 2-D and 3-D runs with certain surface altitude and surface albedo settings had a hard requirement of supplying surface leaving radiance the required format of which we could not find documented.

The following text was added to the revised manuscript: "While McArtim is in principle capable of modeling stratospheric radiative transfer at twilight, a suitable 2D implementation of McArtim to use in conjunction with a photochemical model (i.e., UVspec) was not straightforward. We use DISORT instead since the 2D geometry is defined along the solar azimuth, which is ideal for accounting for photochemical effects along the principle line of sight."

L195: It is not clear what you mean with "layering approach".

We have added further detail. Aerosol conditions for the data in this work had significant extinction below the instrument but sub-Rayleigh extinction near and above the instrument. For these conditions, each EA was given an altitude sensitivity mapping and the extinction profile adjusted starting from lower EA and lower altitudes. This bottom-up as opposed to top-down approach was chosen because the albedo effects from lower altitude aerosol were needed to reproduce observations at higher EA. Initial aerosol profiles were found by manually testing different boundary layer heights and AOD for a box aerosol layer below the instrument above which extinction decreased exponentially. Thereafter the sequence of O_4 comparisons and adjustments was run six times.

We have also added significantly more detail on the aerosol retrievals in response to Reviewer 2.

Sections 2.5.1 and 3.3.3: Is there a specific reason why SCD_ref is not simply retrieved as part of the state vector in the optimal estimation algorithm, instead of using Langley plots as an extra step?

The results in 3.3.3 hint at the reason for this choice although the reviewer is correct that other choices could have been made. When supplying $SCD_{Ref} = 2.50\times10^{13}$ molec. cm^{-2} it recovers itself despite other deficiencies in the solution, but even more importantly it diverges away from the better solution. This suggests that the solution space has local minima and the choice and assessment of the a priori SCD_{Ref} requires assessment outside a single optimal estimation in any case. The reviewer is correct that SCD_{Ref} could be handled instead as an element of the state vector in principle. The relevant Jacobian elements are reasonably simple to define, but the definition of relevant a priori covariance terms is not to us obvious.

Section 2.5.3 requires substantial revision as it is not possible to understand what the actual approach is. What is the basic idea behind your approach? What is the difference between time-dependent and time-independent retrievals? The term “time-independent formulation” occurs in L291, but it is not explained anywhere what this means. What is the exact meaning of the mathematical objects in Equation 3, which of these are scalars, vectors or matrices, and what are their shapes/dimensions? Does the vector x contain profiles at a single time or are BrO profiles over a period of time which are retrieved simultaneously? What exactly is x_0 and how is it determined? Please specify in detail the individual components of the measurement vector and the state vector.

We have added a paragraph leading into and motivating the time-dependent dependent retrieval which is addressed in the response to reviewer 2. Regarding the time-independent retrieval, this is explained there also but for ease of understanding here is the integrated retrieval just outlined in Section 2.5.2. We have reworked and expanded on the definition of terms around Eq. 3 as follows.

The conventional time independent retrieval assumes constant Br_y (a static atmosphere), with the only changes in BrO being those predicted by the photochemical model (Br_y repartitions as a function of SZA). This assumption was ultimately found to be invalid for one day where dynamical changes in Br_y were observed, in addition to chemical repartitioning (see Sect. 3.3 and 3.4 for details). We addressed this by augmenting the optimal estimation with time-dependent variables. To our knowledge such an approach has not been employed for DOAS optimal estimation before, so we describe it here in detail.

We define the time evolution of the BrO profile at time t ($\mathbf{x}$) in terms of L altitude regions (here L = 4) consisting of a weighted set of related atmospheric grid layers ($\mathbf{W}^{\mathrm{L}}$) where Br_y is expected to evolve consistently such that:

$$\mathbf{x} = \mathbf{x}_0 \left(1 + f(\mathrm{t}) \sum_{\mathrm{L}} \mathbf{W}^{\mathrm{L}} \mathbf{C}^{\mathrm{L}} \right) \qquad (3)$$

Where $\mathbf{x_0}$ is a vector, i.e., the BrO profile at some reference time (t_0) with dimensions of 1×36 for this work, and $f(\mathrm{t}) : \mathrm{t} \rightarrow (-1, 1]$ is a scalar time evolution function such that $f(t_0) = 0$. For convenience we choose t_0 to match SZA = 70° so that photochemical and dynamical effects vary on a common time axis. In principle, $f(\mathrm{t})$ could be indexed to the layers L and folded into the sum. However, for this work there was insufficient external information to constrain more than one choice; a single $f(\mathrm{t})$ function describes the relative time variation in all four altitude regions. In practice, only a linear and a ramp function form – which mirrors the stratospheric O_3 column on Apr. 29 were tested for $f(\mathrm{t})$. The choice of the codomain (-1, 1] is more generally important for reasons outlined below.

$\mathbf{W}^L \rightarrow (0, 1]$ is a matrix of altitude weights defining the mapping of altitude regions L onto the altitude grid that $\mathbf{x_0}$ is defined on constructed such that $\Sigma_L \mathbf{W}^L \leq 1$ for all altitudes with dimensions of 36×4 for this work. For this work the stricter condition that $\Sigma_L \mathbf{W}^L = 1$ for all altitudes is fulfilled. Finally, $\mathbf{C}^L$ are scaling factors describing the proportional change in the BrO profile within altitude region L, with dimensions of 1×4 for this work e.g. if an element of $\mathbf{C}^L$ has a value of 1.1, BrO increased by 10% in that layer when $f(t) = 1$.

We choose to fully constrain $f(t)$ and $\mathbf{W}^L$ as such the combination of $\mathbf{x_0}$ and $\mathbf{C}^L$ fully specifies the state vector $\mathbf{x}$. We seek to retrieve $\mathbf{C}^L$ in addition to $\mathbf{x_0}$ given a choice of $f(t)$ and $\mathbf{W}^L$.

Eq. 4: It is stated that H represents a Jacobian, but H = K_0*x is not a Jacobian Matrix but a vector in measurement space.

The language gets ahead of what is done here, we have changed the introduction text to:

We seek to derive a time-augmented Jacobian starting from the transformation H in terms of a left-side transform matrix $\mathbf{K_0}$:

We have added more detail and reworked the further detailing of the time-dependent retrieval:

A close observer may notice a potential challenge posed by these equations; the weighting functions for $\mathbf{x_0}$ require knowledge of $\mathbf{C}^L$ while the weighing functions for $\mathbf{C}^L$ require knowledge of $\mathbf{x_0}$. To resolve this challenge we take advantage of the fact that the extended definition for $\mathbf{x_0}$ still contains a term which is fully independent of $\mathbf{C}^L$ and approaches the time-independent formulation as $t \rightarrow t_0$. We therefore leverage the time-independent retrieval (Section 2.5.2; already photochemically indexed to t_0) to gain imperfect knowledge of $\mathbf{x_0}$. This retrieval already averages over the time dependence and should get us close to the true state as such we supply it as the a priori for $\mathbf{x_0}$ to compute the weighting functions. If the solution is too far from this a priori however, the computed partial derivatives might no longer be locally valid. To limit this effect we reduce the a priori covariance for the spatial variables ($\mathbf{x_0}$) by a factor of ten. In reporting results we use the spatial variables ($\mathbf{x_0}$) retrieved using the procedure in Section 2.5.2 including their corresponding AVK. The values and AVK for the time dependence vector ($\mathbf{C}^L$) are reported for the results of this second stage.

It remains to choose $f(t)$ and $\mathbf{W}^L$. Examination of Eq. 5a reveals the rationale for setting the codomain of $f(t)$ to (-1, 1], for solutions in which the entire retrieved column entirely disappears or doubles (considered reasonable bounding cases) this bounds the time dependent term to (-L, L] which for the small values of (L = 4) considered here is comparable to the time independent weight fixed at 1, hence ensuring the relative importance of measurements is at least partly preserved. In addition, by defining all $f(t)$ to have a maximum value of 1, the profile which is maximally different from that at t_0 is readily computed and compared. For this work we considered only linear and ramp functions, ultimately using a ramp function defined to be zero prior to 70° SZA and increasing to 1 which matches an observed trend in O_3 VCDs (see Supplement for details). For $\mathbf{W}^L$ logistic curves were chosen as the functional form with a logistic steepness in all cases of 2 km^{-1}. The atmosphere was first divided at the tropopause at 17.5 km. Then at 6 km and 10 km based on the results of retrievals using single scans and modeled behavior in CAM-chem (see Sect. 3.3 for details).

The solutions to the inversion of the time-dependent retrieval were found to be highly sensitive to the a priori supplied for $\mathbf{C}^L$ including non-physical results. We suspect that this might be because the assumptions for the staged retrieval only work where the partial derivatives are sufficiently flat, or perhaps at least smooth. This necessitated systematic sensitivity studies to find solutions which were physical as well as categorizing solutions based on minimizing any time-dependent trend in the a posteriori residuals. This methodology identified a family of solutions with similar values that met stringent criteria, from which the solution with lowest overall residual was selected (see Sect. 3.3 for details).

L317: I suppose the “high concentrations” of BrO are expected in the FT. Please specify.

We have reworded this to “increased concentrations”. Prior to this study it could only be inferred that BrO should increase. What baseline concentrations are over the central Pacific was uncertain in the absence of measurements and given the variability observed elsewhere.

L354: Cross-sections do not have an optical density. Please rephrase.

We’ve specified that the cross-sections are “multiplied by ZS dSCDs”

L357: To what is the Aliwell fit window insensitive?

“to the choice of O_3 cross-section”

L363: Is a wavelength shift of 3 pm leading to any noticeable difference in the fit if the instrument has a spectral resolution of about 0.5 nm?

The difference is not statistically significant, however, it is noticeable in the third digit of fitted BrO dSCDs especially in ZS data. Even this is likely because the O_3 absorption is so great compared to BrO. We have included this as we believe it represents the current best practice.

L376: Please explain why the O4 scaling factor should scale with lambda^4 like Rayleigh extinction. I do not see an immediate physical reason for this.

Because $[O_4] \propto [O_2]^2$, O_4 signal in dSCDs overwhelmingly comes from lower altitudes typically after the final scattering event. In the single-scattering approximation the path-length from the final scattering event is inversely proportional to extinction. For low aerosol conditions extinction is well approximated by Rayleigh extinction. We have rephrased the sentence and added a reference: “This value is similar to what one might expect for a Rayleigh comparison of optical depth and pathlength of the 360 nm and 344 nm bands $(344/360)^4 = 0.83$ (Wagner et al., 2004)”

L382: Please explain what you mean with “intensity effects”. Could this be instrumental non-linearity? If you suspect that NO2 is affected by such effects, then why not other trace gases, in particular if they have lower optical density?

Intensity effects is used in the context of DOAS to refer to effects which are linearly proportional to intensity rather than proportional to the ratio of intensity. Instrumental non-linearity is indeed a possible explanation, however, it is relatively unlikely as the exposure is dynamically adjusted for consistent average saturation. We’ve made the possibilty more explicit by specifying: Those effects could be accurately fitted as offsets from changing straylight; notably, the exposure of spectra measured at different EA is dynamically adjusted for consistent average saturation in our setup (Coburn et al., 2011).

For the UV fitting window and the measurement conditions, the differential optical density of NO_2 is comparably low, and we’ve added that it: is relevant to the small NO_2 signals measured in the free troposphere for which OA NO_2 dSCDs differ from nearby zeniths by less than three times the fit uncertainty. We suspect that the HCHO-BrO cross-talk specifically may have similar effects as in individual scans it shows similar patterns, but the “drift” effect obscures attempts to determine this more clearly.

L395ff: Here it is not clear what you mean with the terms “component” and “signal”. Do they refer to the retrieved dSCDs or to the fit residuals? What exactly are replicate measurements? Do you mean subsequent measurements along the same line of sight?

We've reworded the sentence to: After optimization of the fitting window we believe that this spectral cross-talk is handled by the DOAS fit with the exception of a fast-changing anticorrelation identifiable as opposing changes in BrO and HCHO for sequential measurements of the same viewing geometry, and slow-changing opposing "drifts" in both HCHO and BrO dSCDs.

The "replicate measurements" refers to sequential repeats of EA = 0°, 30°, 90° for which the fast change is most easily identified. We've reworded to "sequential measurements of the same viewing geometry" for clarity.

L564: A "change" has no DoF.

This change does have DoF. To be more precise the time-dependent, spatially-constrained scaling factors we retrieve have information content independent of the a priori information provided which can be quantified as DoF. We took this comment as further motivation to improve the revised section 2.5.3.

L635: Here it would be good to cite Rodgers and Connor [2003].

Agreed and added

L640: In what respect is the analysis limited by the RTM calculations? In terms of accuracy? Computational time?

Specified computational time.

L643: To my knowledge, McArtim already fulfils the required capabilities listed here – see Deutschmann et al. [2011].

We agree in principle, and address this above.

Please add a "Code Availability" section stating the availability of the retrieval algorithm presented here.

Added and archived at the same locations as the data.

Technical Corrections

L30: near -> nearly

Changed.

L56: "BrOx adds radical species to oxidative capacity" does not make much sense. Suggestion: "BrOx increases the oxidative capacity"

Suggestion adopted

L106: I guess you mean the azimuth angle when you talk about "primary viewing direction"?

Clarified to specify azimuthal viewing direction.

L123: What do you mean with the dagger symbol as prefix for the elevation angles?

We have clarified the existing explanation at the start of the sentence which now reads:

"…where angles preceded by † are collected in the reverse azimuthal direction (+130°±2):"

L175: Stratospheric aerosol WAS modelled…

Changed

L189: Tropospheric aerosol WAS assumed…

Changed

L190: Approximated by an approximation?

Modified sentence to remove both instances of approximation as the sentence is already discussing model assumptions.

"…non-absorbing Henyey-Greenstein aerosol phase function with asymmetry parameter …"

L194: Add "The retrieval of" to the beginning of the sentence

Changed

L393: The term "method-based anticorrelation" is not clear to me.

We have rephrased the sentence for further clarity:

This similarity in measurement leads to an empirical anticorrelation, the characterization of which is further confounded by chemical coupling of BrO and HCHO via reactions of Br atom with HCHO and other aldehydes, which often correlate with the latter, suppressing BrO formation and creating a chemical anticorrelation."

L404: The part of the sentence after the semicolon is without any context.

We have replaced the semicolon with a period and combined the following clause with the following sentence for which it provides context.

L409: I suggest to add that the additional HCHO absorption feature is at 330 nm.

We have specified that the feature is "peaking between 329 nm and 330 nm" to ensure it is understood that windows starting at 330.0 nm do not capture the peak.

L497: comparison -> difference

Changed

L506: "While it is clear the retrieval … can be further improved" is grammatically incorrect. Is a "that" missing?

We have add "that" for clarity.

L585: Explain abbreviation "KOA".

Changed to "Kona airport (KOA)"

We wish to note that when conducting sensitivity studies in response to comments from Reviewer 2, we discovered that the DoF reported for April 29 had been erroneously entered for different retrieval settings than those used. We reviewed all other results and confirmed that this error was limited to only the DoF. The reviewer may wish to check the revised numbers.

References

1. Deutschmann et al., “The Monte Carlo atmospheric radiative transfer model McArtim: Introduction and validation of Jacobians and 3D features,” J. Quant. Spec. Rad. Trans., vol. 112, pp. 1119–1137, 2011, doi: 10.1016/j.jqsrt.2010.12.009.
2. D. Rodgers and B. J. Connor, “Intercomparison of remote sounding instruments,” J. Geophys. Res, vol. 108, no. D3, pp. 4116–4229, 2003, doi: 10.1029/2002JD002299.