# Author's Response

Here I've listed the referee's notes and major concerns/comments, followed by my own response in blue. Minor comments have been handled and I'm thankful to the referee's for going through my manuscript with such attention, as it improves my work greatly.

First the referee responses are listed then the major comments and my responses for each referee are shown. Finally the minor corrections are listed with responses colocated.

# Referee Notes:

### Anonymous Referee #1

Received and published: 6 February 2017

The current study presents a method to identify stratosphere-to-troposphere transport (STT) events and estimate the associated ozone flux to the troposphere, based on ozonesonde profiles from three sites located in the Southern Hemisphere extratropics. Subsequently, the seasonality of STT events is determined, as well as the favorable synoptic conditions. Based on the stratospheric contribution to tropospheric ozone column estimated from the ozonesondes, the GEOS-Chem simulated tropospheric ozone columns are extrapolated to assess the stratospheric contribution over the Southern Ocean region. As the STT is of great importance for the tropospheric ozone budget and variability, and the number of relevant studies (both observational and modeling) for the examined region is limited, I find the topic of the paper within the scope of ACP. On the other side, there are several issues that need to be addressed before consideration for publication in ACP.

### Anonymous Referee #2

The authors present an observation-based method to estimate the total stratospheric ozone flux in the Southern Ocean. I think the approach is interesting and complement some model-based methods, and is also of interest to the readership of ACP. However, the method comes with some major uncertainties and I wonder whether an extrapolation to the whole Southern Ocean from only three measurement sites is reasonable. My major concerns are listed below, and based on them I only recommend the manuscript ready for publication in ACP if a carefully revised manuscript is provided.

### Anonymous Referee #3

Received and published: 23 February 2017

The paper by Greenslade and coauthors presents an analysis of ozone soundings from three locations between -31 ◦ S and -69 ◦ S over several years. The authors analyse the profiles for ozone enhancements in the troposphere, which they link to stratosphere-to-troposhpere exchange associated with cut-off lows and cyclones. Based on these enhancements they estimate the fraction of excess ozone from the stratosphere as percent of the tropospheric ozone column. They calculate a flux of ozone for the southern hemisphere by using tropospheric columns from GEOS-Chem times this fraction times the frequency of occurrence of STT events. This flux estimate is more than an order of magnitude lower than estimates from literature using various techniques. The authors conclude that these differences are due to conservative estimates of several thresholds and assumption regarding their method.

Observationally based estimates of ozone transport especially in the southern hemisphere are sparse and therefore valuable. However, the authors provide a flux estimate which is far from other studies, probably due to the sparse spatial and temporal coverage. If this however is the case, the method is simply not applicable in this case and the deduced flux does not mean anything. If the method is valid for the given data set I miss a careful analysis of the reasons for the discrepancy. Therefore I don’t see the paper as an ACP paper in the current form.

# Anonymous Referee 1

### Major Comments:

1) Calculating the 99th percentile from the perturbation profiles over that layer (2 to

1 Km below the tropopause) is a fairly strict criterion. Wouldn’t this threshold choice

avoid the selection of deeper stratospheric intrusion events as “STT events”?

This should read as 2 km above the surface to 1 Km below the tropopause since, as you point out, the one kilometre range would miss deeper intrusions. The text has been updated as follows on DoLast:page/line “... (2~km above the earth's surface to 1~km below the tropopause).” and DoLast:page/line “ … profiles between 2~km above the earth's surface and 1~km below the tropopause.”.

Have you consider modifying this criterion, and include others (e.g. significant negative O3 relative humidity correlation values above a threshold) to minimize false STT detection?

There is an issue with the ozonesondes humidity measurements which made me initially steer away from utilising them – TODO: Andrew knows about this?

2) The seasonality of STT events presented in Fig. 7 is not in line with the findings of

Škerlak et al. (2015) for the examined region. How are your results (STT seasonality)

compared with other modeling studies (Elbern et al., 1998; Sprenger et al., 2003)?

I have added comparisons between the seasonalities at my three sites and results from Wauben et al., 1998, Sprenger et al., 2003, and Skerlak et al., 2015.

The measurement sites are not in the regions which have a clear winter maximum seen in figure 1 of Sprenger et al., 2003, and the large scale winter maximum shown by all three studies seems to be dominated by the system in that region – this has now been noted in the text DOLAST: page/line “...This seasonality is not seen in the recent ERA-Interim tropopause fold analysis performed by Škerlak et al. (2015), where a winter maximum of ozone fold frequency (∼ 0.5% more folds in winter) over Australia can be seen. However their winter maximum is in the subtropics only - from around 20 ◦ S to 40 ◦ S, which can be seen as the prevalent feature over Australia in Fig. 5 of their publication. Wauben et al. (1998) look at modelled (CTM driven by ECMWF output) and measured ozone distributions and find more SH ozone in the lower troposphere during Austral winter, however they note that the ECMWF fields are uncertain here again due to lack of measurements. Their work shows a generally cleaner lower troposphere in the SH summer but can not be construed to suggest more or less STT folds in either season. Sprenger et al. (2003) examine modelled STT folds using ECMWF output over March 2000 - April 2001, and show that for this year there is a clear Austral winter maximum, again over the 20 ◦ S to 40 ◦ S band. The winter maximum does not include Melbourne, or the southern ocean, which may help explain why we see a seasonality which disagrees with these prior studies.”.

Is there any evidence from other studies that STT frequency over the examined region

exhibits a maximum during the austral summer (DJF) and not during the austral winter

(JJA) when the jet stream is strongest over the broader region?

I have not found any other literature which suggests a Summer peak over the Southern Ocean or Australia, however the prior work has not looked at the sites examined here (for instance see my previous response).

Have you tried to detect STT events from the model results? I guess this is strongly depended to the vertical resolution of the model, but it would be very interesting to see how the observed and

modeled STT seasonalities are compared.

Not only the vertical resolution but also the low horizontal resolution of GEOS-Chem output means an analysis would only perceive very large scale STT events, as they would need to span a large portion of 2 degrees latitude by 2.5 degrees longitude.

3) To my understanding, using the seasonality of STT events from the three sites to

extrapolate model results over the Southern Ocean region is a quite simplified and

coarse approach, especially when considering the previous comment.

This is correct, after seeing the reviews the SO extrapolation has been moved to a supplementary pdf as an example of one possible utility of the ozonesonde event detection. It is too simplified and uncertain to add any real substance to the paper.

Instead an estimate of STT ozone flux near the three sites has been performed and compared to Skerlak et al. 2014 on DOLAST:page/line “Škerlak et al. (2014) shows an estimate of roughly 40 to 150 kg km −2 month −1 in these regions, over all seasons, of which 0 to 10 kg km −2 month −1 enters the boundary layer (see Fig. 16, 17 in their publication) while we estimate 2 to 41 kg km −2 month −1 STT impact, of which the largest part is in Summer (DJF). Our calculated seasonal contributions, along with total

uncertainty are shown in Table 3.”

4) Overall, the presentation of the results can be further improved (please check my

suggestions further below), as well as the writing of the manuscript.

### Comments:

Škerlak et al. (2014) presented an STE climatology using the ERA-Interim data. This

study is important not only for the introduction, as it describes the STT climatology

for the SH, but for intercomparison of the results also. Similar climatologies can be

found in the modeling studies of Roelofs and Lelieveld (1997) and James et al. (2003).

Recently, Akritidis et al. (2016) explored the impact of stratospheric intrusions on tropospheric ozone and the associated stratospheric contribution over the eastern Mediterranean and the Middle East region, a task that is relevant with some of the purposes of this study.

Page 4, lines 3-4: Since the study is based on the ozonesondes launched from the three sites, it is important to present the location of the sites.

A very good idea, a brief description of the sites has been added: “... Melbourne, a major city in Victoria, Australia is in the far south eastern section of the Australian mainland, actual releases are north of the central business district in the Broadmeadows suburb. Macquarie island is isolated from the Australian mainland, situated in the remote southern ocean and unlikely to be affected by any local pollution events. Davis is on the coast of Antarctica and also unlikely to experience the effects of anthropogenic pollution.”

Page 4, line 22: “Figure 1 shows the monthly mean tropopause altitudes at ..”, while in Fig. 1 caption is stated “Multi-year monthly median tropopause altitude ..”. Is it the mean or the median? Please modify accordingly.

It is the median, Fig. 1 caption has been updated

Page 5, Figure 1: a) The shadings used to describe the 10th and 90th percentiles are

rather confusing. I suggest you replace the shadings with dashed lines (same color as

the solid lines).

b) Increase the range of the vertical axis to show the 10th percentile

value for February.

Done – including removal of colour shading.

c) Is it the case that tropopause drops below 4 km (10th percentile)

over Davis?

We looked into this and found a problem with the lapse rate tropopause picking up boundary layer temperature inversions – enough times that it showed up in the Davis 10th percentile.

We've now limited the lapse rate tropopause (LR TP) detection to 4km or greater (rather than 2km or greater), since every LR TP detected below that height was a false positive.

This is mentioned in the text at DoLast:page/line “... We slightly alter the lapse rate tropopause definition so as only to detect tropopauses above 4~km altitude, since at all three sites we saw several false positive lapse-rate detections due to temperature inversions near 2~km in altitude.”

That fix revealed two new events: one at Davis and one at Melbourne, and all plots are updated accordingly.

What is the minimum tropopause height value over Davis during February?

The lowest TP occurred at Macquarie: at 4.4 km, while Davis got as low as 6.14 km and Melbourne's lowest was 5.81 km.

Page 5, lines 5-6: “This seasonality at the high latitude sites is driven by a decrease in

photochemical destruction under the reduced radiation conditions around polar night.”

Please include a reference or information about the NOx levels at these sites (if avail-

able) to justify this statement.

TODO: ask Simon about this comment

Page 6, line 14: It is important to know the vertical resolution of the GEOS-Chem model

near the tropopause (although it can partially be seen from Fig. 13), as it is important

for the tropopause height detection and the tropospheric ozone column calculations

from the model results.

GEOS-Chem has roughlyRoughly 500m resolution near 10km altitude, I’ve now noted this in the text in the Model description secion DOLAST: page/line: “The vertical resolution is finer near the surface at ~60 m between levels, spreading out to ~500 m near 10~km altitude, and reaching ~1500 m near the top of the atmosphere.“

Page 7, lines 22-23: “The interpolated profiles … high frequency perturbations).” This is a rather brief description of the procedure. A more detailed description including a reference (if available) for the FT application would be necessary.

We've filled out the description a little, we've not come across this method for STT detection in any literature however the filter is quite common in sound and image filtering, and a reference to Numerical Recipes in C, by Press et al., Cambridge Uni. is now provided: DOLAST: page/line “... ”.

Page 7, lines 27-28: “We next use all the perturbation profiles at each site to calculate the 99th percentile perturbation value for the site”. How exactly is this cut-off threshold calculated?

Once the perturbation profiles are all created, the filtered interpolated values between 2km and TP-1km are used as the basis for this 99th percentile – which I've now noted in the text.

In Section 2.5, Page 9, the authors state that is calculated “between 2 km and 1 km below the tropopause”. This information should be provided earlier in the manuscript, at the point that the 99th percentile threshold is initially mentioned (Section 2.3).

As per your suggestion we've added this information TODO text/pageno/line: .

Page 8, Figure 3: Why the two panels have different units? Are the ozone units of

the left panel “1e+12 molecules cm-3”? Please change accordingly the Figure and the

Figure caption. mixing ratio -> number density

Thanks for spotting this, these units and caption are misleading/confusing – We've redrawn the plots using only ppbv.

Page 9, lines 1-2: “For this reason all detected STT events found near smoke plumes

are flagged”. How is “near” defined?

In this case near is defined as subjectively within 150km, This is now more clear in the text DOLAST: page/line “For this reason all detected STT events found near (within ~150 km of) smoke plumes are flagged, following visual inspection.”.

In my opinion, Figures 4, 5 and 6 are more supportive-descriptive without adding any-

thing new. Therefore, I suggest including them as a supplement. Moreover, Figures 5

and 6 can be merged into one.

Page 11, line 17: “We use the ERA-I 500 hPa data to subjectively classify the events

based on their likely meteorological cause.” Do the authors classify the events by visual

inspection of the 500 hPa maps for every STT event date?

Yes, we looked at each image, we were interested to see if there was any clearly discernible pattern or dominant weather system connected to the events. At DOLAST:page/line we've added “... likely meteorological cause, by visually examining each date where an event was detected” to the text.

Page 11, lines 20-21: “The stratospheric polar vortex may create ozone folds without

other sources of upper tropospheric turbulence”. Please include a reference for the

above statement.

The sentence now references Baray et al., 2000 and Sprenger et al., 2003nci

Page 14, lines 16-20: “The seasonal distributions . . . first half of the year”. To my

understanding Fig. 7 and Fig. 8 are not quite similar. Moreover, comparing Fig. 8

with Fig. 7 where fire influences are also included is somehow unfair. The fact that

ozonesondes are launched monthly at Davis from December to June is also the case

for Fig. 7, where high STT frequencies are found for the respective period.

Page 16: How is the modeled tropospheric column ozone calculated?

GEOS-Chem provides the ozone density (molecules/cm3), vertical column boxheights, and tropopause level. I am using the sum of the boxheight \* density for each box below the one containing the tropopause.

I've added to the text: “... Boltzmann constant. GEOS-Chem outputs ozone density (molecules cm$^{-3}$), and height of each simulated box, as well as which level contains the tropopause, allowing modelled $\Omega\_{O\_3}$ to be calculated as the product of density and height summed up to the box below the tropopause level. In both observations and model...”

How is the tropopause defined in the GEOS-Chem results?

That is a good question! GEOS-Chem uses the tropopause height provided by GEOS-5 met fields which are calculated using a lapse-rate definition using the first minimum above the surface in the function 0.03\*T(p) – log(p), with p in hPa.

This has now been added to the text

Page 17, lines 3-4: “Over Melbourne, ozone in the lower troposphere is well represented, but the model overestimates ozone from around 4 km to the tropopause”. This

is also seen for Macquarie and should be added to the discussion.

Over Macquarie Island the lower troposphere seems to be slightly underestimated, which is the same as seen over Davis, while ozone above 4 km does show similar overestimation as you point out.

The following has been added on DOLAST:pageno/lineno “ … Over Melbourne, ozone in the lower troposphere is well represented, but the model overestimates ozone from around 4~km to the tropopause. Similarly over Macquarie we see model overestimation of ozone above 4~km, suggesting that Macquarie may be influenced by processes seen at both of our other sites.”

Page 19: “Figure 14 shows the mean fraction of total tropospheric column ozone (cal-

culated from ozonesonde profiles) attributed to stratospheric ozone intrusions at each

site, averaged over days when an STT event occurred.” Please explain in more detail

how is this fraction calculated.

Thanks for pointing this out, the calculation was not well described. The following text has been added on DOLAST: page/line “ … ozone enhancements, based on a vertical integration of the ozone above baseline levels for each ozonesonde where an event was detected.

The area considered to be 'enhanced' ozone is outlined with yellow dashes on the left panel of Fig. DOLAST:figure number

…

First the tropospheric ozone column is calculated, then the enhanced ozone column amount is used to determine the relative increase.”

Page 19: “to the entire Southern Ocean region, defined here as 35\_ S-75\_ S to en-

compass”. What is the longitudinal range?

In our extrapolation we used the entire band from 35S to 75S (ie. 180W to 180E).

However this region has been replaced with two smaller regions covering the ozonesonde release sites, See figure DOLAST: image number for ComparisonRegion.png

Page 20: Fig. 14 and Fig.15 can be merged into one.

Good suggesion, these images have been combined into one with a top and bottom panel

Page 22: “If we we assume a fractional ozone impact due to each event STT event of

I=35% based on their results”. The 30-40% stratospheric contribution found by Terao

et al. (2008) is seen only during spring and at 500 hPa. Therefore, assuming a 35%

stratospheric contribution to the tropospheric column ozone seems a bit arbitrary.

We have updated how the calculation of flux is made, and are no longer using this change of I, the updated calculations are on TOLAST: text/pageno/line “... To determine the ozone column attributable to STT, ...”.

# Anonymous Referee 2

### Major Concerns:

1. **Extrapolation to Southern Ocean**: The authors look at three measurement sites (Davis,

Macqaurie, and Melbourne) in the Southern Ocean (SO), and then extrapolate their results to the

whole SO. I don’t think that this is valid. I think there ia quite a lot of spatial and temporal

variability that gets neglected in doing so. To make my point more clearly, I copy a figure (Fig. 16)

from Skerlak et al. (2014) here:

It shows the seasonally averaged STT ozone flux for the period 1979-2011. Evidently, there is a lot

of spatial and temporal variability. The next figure (Fig. 17) from Skerlak et al. (2014) shows the

estimated ozone flux into the PBL, which exhibits a still stronger variability. Hence, I think the

authors must be rather hesitating in extrapolating their results. I suggest to restrict the conclusions about the STT flux more to the regions around the three measurement sites. It will still be possible to compare the values, e.g., with the values in Skerlak et al. (2014).

As described for reviewer 1, comment 3, the SO extrapolation has been removed to a supplementary as it suffers from too many uncertainties/simplifications.

Your suggestion of using smaller region has been taken. We now examine small regions around each site shown in DOLAST: Fig. X. The comparison with Skerlak is also made on DOLAST: Page/lineno “...”

2. **Transport aspect**: An aspect that is not sufficiently discussed in the manuscript is the transport

of the ozone-rich air from its crossing to the measurement site. For instance, in Figure 5 the authors

show an STT event and the geopotential height field at 500 hPa. A nice cut-off low pressure system

is discernible in the geopotential. But it is not clear whether the STT event really occurred below

this cut-off. In fact, it could have happened quite a distance away from it and the be advected to this

place. I would argue that the transport aspect become more important if an STT event is detected at

middle or lower-tropospheric levels, i.e., when it is rather ‘detached’ from the tropopause above. As

an example, the following study shows that the crossing of the tropopause takes place in the western

North Atlantic but an ozone signal is discernible in the profile over western Europe:

Trickl, T. et al. "How stratospheric are deep stratospheric intrusions? LUAMI 2008." Atmospheric

Chemistry and Physics 16.14 (2016): 8791-8815.

I think the authors should more carefully discuss this aspect of STT event. Possibly, the do a short

literature review dealing with ozone transport and the long-range character of stratospheric

intrusions. It would also be interesting, and relevant to this manuscript, how long signals in

stratospheric ozone remain discernible in an atmospheric column after the air parcels have crossed

the tropopause.

This is an interesting question, TODO: Jenny or Ian could point me towards some literature?

3. **Uncertainty**: The method comes with quite a few uncertainties! I list some of them:

- P7,L30: “STT events at altitudes below 4 km are removed to avoid surface pollution, and

events within 0.5 km of the tropopause are removed to avoid false positives induced by the

sharp transition to stratospheric air.” → I see the problem with the near-surface STT events.

But still, even at this low altitude it could be due to a stratospheric intrusion.

This is one possible false negative which could occur, I have added a note at PX “...TODO...”.

Further, I expect quite some ozone flux to be across the tropopause without a very clear peak-like

structure in the profile. This could, e.g., be the case if the ozone flux is more related to a

continuous ‘diffusion’ of ozone across the tropopause in contrast to an ozone flux going

along with a coherent cross-tropopause air streams in distinct weather systems.

While this is true, the focus of this work is on discrete influx events, or folds, rather than continuous enhancement or reduction due to cross tropopause transport. This is due to the difficulty of detecting continuous influx with weekly ozonesondes. I do appreciate the detailed comments and have noted that we are likely ignoring this source of ozone enhancement at PX “...TODO...”.

- P7,L9-12: “This estimate is conservative because it does not take into account any ozone

enhancements outside of the detected peak that may have been caused by the STT, and also

ignores any enhanced ozone background amounts from synoptic-scale stratospheric mixing

into the troposphere.” → The ozone background is also enhanced in mixing across the

troposphere, or the background at any of the stations is enhanced by STT events taking place

outside its ‘range’.

This is true, thanks for pointing it out. We've added this possibility to the text: DOLAST:pageno/line“... increased the local background mixing ratio, and any influence from STT events nearby which may also increase the local background ozone.”.

- In section 5 (P19,L9) the overall ozone flux is determined as the product of the monthly

likelihoods of STT (f), the monthly mean fraction of an ozone column attributed to

stratospheric ozone (I) and the mean tropospheric ozone column (Omega). All these factors

come with a lot of uncertainty! Be it due to the method applied, or the spatial and temporal

variability.

As described to reviewer 1, comment 3, this extrapolation has been removed from the manuscript and included in a supplementary.

The monthly likelihoods of detection has been renamed to L, and a new term representing assumed event longevity has been added.

The uncertainties in the smaller extrapolations have now been somewhat quantified by using the multi-year standard deviations of each term in the multiplication, shown on DOLAST: Pageno/lineno “... The uncertainties in our two regions of extrapolated ozone are determined from the standard deviations in monthly values over the multi-year dataset range.

This ignores uncertainty due to the non-homogeneity of the regions extrapolated over, and would require more data, parameters, and analysis to apply to larger regions.

The overall uncertainty as a percentage is shown in parentheses in Table DOLAST: this reference.\ref{table:extrapolationResults}, these values are on the order of 200\%, largely due to uncertainty in the L and I factors which both sit near 100\% uncertainty for each month.”

- P9,L16: “While ozone production occurs in some biomass burning plumes, this is not

always the case; therefore ozone perturbations detected during transported smoke events

may or may not be caused by the plume. For this reason all detected STT events found near

smoke plumes are flagged.” → These events are not included in the calculation of the ozone

flux, but still they could be of relevance!- P9,L7-9: “ We use the 99th percentile because at this point the filter locates clear events with no obvious false positives. Event detection is highly sensitive to this choice; for example, using the 98.5th percentile instead increased detected events by 10 (22%) at Davis, 19 (40%) at Macquarie Island, and 24 (33%) at Melbourne.” → Does this mean that with a 98.5 th percentile, some of the events are clear false positives? Wo do you decide that this is the case? I am not sure whether this is obvious. In short, an additional uncertainty of the

method.

This is a good point, TODO: discuss this in paper

Given all these uncertainties, the estimate of the total STT flux based on the ozone profiles must be

rather conservative and going along with a big overall uncertainty! This is already discussed by the

authors, i.e., they are fully aware of it. What I would, however, suggest is a separate section (or

extended paragraph) where all uncertainties are presented and, if possible, quantified.

# Anonymous Referee 3

### Major comments:

1) Overall the manuscript leaves me a bit puzzled, since I’m not sure what to take out of this work. The authors state that their value of the ozone enhancement fraction might be largely underestimated by a factor of ten. If this is the case it is difficult to get the benefit of the study. Though the approach is reasonable, maybe the statistics and spatial coverage is to small to cover the full variability and frequency of occurrence of STT events for a quantitative flux calculation for the southern hemisphere. If the difference between observations and models is really between factors 30-200 depending on the reference (p.21, l.8-12), this needs more clarification than simply replacing observation with model results to find agreement with other studies. I found this approach at least very questionable.

We hope to present the ozonesonde dataset along with a new method of detecting STT ozone intrusions. STT flux estimation was included as a novel use for the ozonesonde dataset, although the extrapolation over the southern ocean was indeed far too simplified to be useful. This has been pointed out by all three reviewers in some form. We've now changed the calculations and are only extrapolating over regions near each site, with more analysis of uncertainties and a better comparison between our outputs and the literature.

2) Further: I missed a quantification of the uncertainties. This is partly done in section 2.5, but it is e.g. not clear why a threshold of 99% is the best choice nor which factor specifically leads to the very low flux estimate. I suggest the authors use a bunch of northern hemispheric sondes with higher spatial and temporal density to gauge their approach before applying it to the southern hemisphere.

# Minor Comments

All the minor comments have been implemented, any notes are added in blue

### Referee 1:

Page 1, line 4: seasonality -> seasonality of STT events

Page 1, line 9: 2.5 km, 3 km -> 2.5 km and 3 km

Page 1, line 14: these -> which

*Page 2, line 2: .Despite lingering -> . Despite the lingering*

Page 2, line 29: found STT -> found that STT

Page 2, line 31: challenging to accurately represent, and better model resolution → challenging to be accurately represented, and finer model resolution

Page 3, line 6: low -> lower

Page 3, lines 14-16: Add references.

**TODO: list of added refs for these notes: Skerlak,**

Page 3, line 16: characterized -> described

Page 8, line 12: transported -> transported over

Page 9, lines 22-23: (e.g., Sinha et al. (2004); Mari et al. (2008)). -> (e.g., Sinha et al., 2004; Mari et al., 2008). Please check the manuscript for similar instances.

Page 10, line 16: our three sites -> the three sites

Page 10, line 16: detected -> the detected

Page 11, line 23: profile -> vertical profile

Please replace all instances of “Brunt-Viäsälä” in the manuscript with “Brunt-Väisälä”.

Page 19, Figure 13: dash -> red dash, please also provide information about the black dashes.

Caption line has been altered to “... GEOS-Chem and ozonesonde pressure levels are marked with red and black dashes respectively”

Page 22, line 9: If we we assume -> If we assume

Page 22, line 10: impact due to each event STT event -> impact due to each STT event

Page 22: empirically-derived threshholds -> empirically-derived thresholds

Page 22: Comparison with ERA-Interim reanalysis data -> Analysis of the ERA-Interim reanalysis data

### Referee 2:

P2,L13-15: “While models show decreasing tropospheric ozone due to stratospheric ozone depletion propagated to the upper troposphere through vertical mixing (Stevenson et al., 2013), recent work based on the Southern Hemisphere ADditional OZonesonde (SHADOZ) network suggests increasing upper tropospheric ozone near southern Africa, most likely due to stratospheric mixing (Liu et al., 2015; Thompson et al., 2014)” → Simplify sentence structure! Rephrase.

This now reads: “Models show stratospheric ozone depletion has propogated to the upper troposphere (Stevenson et al., 2013). However, work based on the Southern Hemisphere Additional OZonesonde (SHADOZ) network suggests stratospheric mixing may be increasing upper tropospheric ozone near southern Africa (Liu et al., 2015; Thompson et al., 2014).”

- P2,L24: “ excedes” → “exceeds”

*- P2,L29: “STT is responsible” → “STT to be responsible”*

- P3, L4: “mixing across the tropopause mainly caused by the jet streams” → a little strange formulation. Mixing is not caused by the jet streams; maybe you can write that it is associated by the jet streams.

This now reads: “... mixing across the tropopause mainly associated with the jet streams over the ocean.”

- P3,L10-11: ”A big influence on the high surface ozone concentrations over the eastern Mediterranean is stratospheric mixing and anticyclonic subsidence (Zanis et al., 2014)” → ”A big influence on the high surface ozone concentrations over the eastern Mediterranean can be attributed stratospheric mixing and anticyclonic subsidence (Zanis et al., 2014)”

**TODO: read these + notes**

**- P3,L11-12: The authors might want to consider the following studies dealing with STT and ozone fluxes over the eastern Mediterranean:**

**Tyrlis, E., B. Škerlak, M. Sprenger, H. Wernli, G. Zittis, and J. Lelieveld (2014), On the linkage between the Asian summer monsoon and tropopause fold activity over the eastern Mediterranean and the Middle East, J. Geophys. Res. Atmos., 119, 3202–3221, doi:10.1002/2013JD021113. Akritidis, D. et al. "On the role of tropopause folds in summertime tropospheric ozone over the eastern Mediterranean and the Middle East." Atmospheric Chemistry and Physics 16.21 (2016): 14025-14039.**

- P3,L14-15: “The strength (ozone enhancement above background levels), horizontal scale, vertical depth, and longevity of these intruding ozone tongues vary with weather, topography, and season.” → This is a rather general statement. What do you mean with weather?

This line has been updated to : “... vary with wind direction and strength, topography, and season.”

- P3,L30-33: How relevant is it for the reader to know how the ozone mixing ratio is quantified? If not relevant, I would remove this sentence. It sounds rather technical!

The technical portion has been removed and the sentence now reads: “... Ozone mixing ratio is quantified with an electrochemical concentration cell, using standardised procedures when constructing, transporting, and releasing the ozonesondes http://www.ndsc.ncep.noaa.gov/organize/protocols/appendix5/.”

*- P4,L8-9: “Characterisation of STT events requires a clear definition of the tropopause. The two most common tropopause height definitions are the standard lapse rate tropopause (WMO, 1957) and the ozone tropopause (Bethan et al., 1996).” → I would mention already at this place the dynamical tropopause which is defined by means of a potential vorticity iso-surface. I would guess it to be rather similar to the ozone tropopause, but to differ from the WMO one.*

At the end of this paragraph DOLAST: page/line the following sentence has been added: “ Another commonly used tropopause definition (the dynamical tropopause) is determined from the ±2 PVU isosurface, which allows a 3D view of folds and other tropopause features in a sufficiently resolved model (Skerlak et al., 2014).”

- P4,L17: “The ozone tropopause can be less robust during stratosphere-troposphere exchange;” → What does ‘robust’ mean? What defines whether a tropopause is robust or not? The ozone tropopause certainly allows for much more details (and a much more complicated structure) than the WMO tropopause. But I would not say that it is less robust because of this!

Here we meant robust to mean 'less likely to misdiagnose the tropopause altitude'. This may be unclear so we have altered the text at DOLAST: page/line to read: “The ozone tropopause may misdiagnose the real tropopause altitude during stratosphere-troposphere exchange; however, it is more robust than the lapse rate tropopause at polar latitudes in winter and near jet streams in the lower stratosphere, which the lapse-rate definition may misdiagnose due to upper troposphere temperature inversions ...”

- P4,L19-21: “In this work, the lower of these two tropopause altitudes is used. This choice avoids occasional unrealistically high tropopause heights due to perturbed ozone or temperature measurements in the ozonesonde data.” → I feel a little uncomfortable by this definition! The two definitions of the tropopause are rather different, and by simply taking the lower one seems ‘dangerous’. The authors should motivate this approach more clearly. At least, I would like to know how often the ozone tropopause ‘wins’ and how often the WMO one. I would expect the ozone tropopause most often to be at lower heights than the WMO one! Correct?

- P7,L22-23: “The interpolated profiles are then bandpass-filtered using a Fourier transform to retain perturbations with vertical scales between 0.5 km and 5 km (removing low and high frequency perturbations)” → I see the 0.5-km threshold. What exactly is the aim of the low-pass filtering threshold (5 km). A more clear description would be helpful.

The low pass threshold 'flattens' the vertical profile, so that the general increase in ozone concentration from the surface up to the stratosphere is removed. DOLAST: page/line has been updated to read “... perturbations with vertical scales between 0.5~km and 5~km (removing low and high frequency perturbations). The high frequency perturbations are removed as they may represent noise in the measurements. The perturbations with scales longer than 5~km represent the vertical gradient of ozone concentration from the surface to the stratosphere.”

- P7,33-34: “The STT event is confirmed if the perturbation profile drops below zero between the ozone peak and the tropopause” → Why does have to drop below zero?

The drop represents a return to non-enhanced ozone concentrations, which suggests seperation between the ozone event and the tropopause. We've updated the text to read: “The STT event is confirmed if the perturbation profile drops below zero between the ozone peak and the tropopause, as this represents a return to non-enhanced ozone concentrations.”

- P8,Figure 3: Just for curiosity: In the ozone profile the Ozone mixing ratio (OMR) is rather low right above the identified STT event. The OMR is higher than immediately below the STT event. Is their a simple reason why the OMR is so low right above the STT peak?

It could be due to relatively clean free tropospheric air being advected over the event, or else the ozone rich air has been advected into the path of the ozonesonde while the free troposphere was particularly clean.

It's also worth noting that the x axis began at 5 molecules per cubic centimetre, and has since been updated to ppbv.

- P9,L16: “all detected STT events found near smoke plumes are flagged.” → What does ‘near’ mean?

This was also asked by Referee 1, in this case near is defined subjectively as within 150km.

- P 10,L15-16: “Data from the European Centre for Medium-range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-I) (Dee et al., 2011) product are used for synoptic-scale examination of weather patterns over our three sites on dates matching detected STT events” → Please rephrase! For instance: “Synoptic-scale weather patterns are examined based on the ERA-Interim dataset (Dee et al., 2011). More specifically, the ERA-I products over the three sites are used on dates matching detected STT events.

This sentence has been restructured as follows DOLAST:page/line : “Synoptic scale weather patterns are examined using data from the European Centre for Medium-range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-I) (Dee et al., 2011). This is done using the ERA-I data products over the three sites on dates matching the detected STT events.”

- P11,L17+26: Here, the STT event is subjectively linked to a meteorological feature, a cut-off low-

pressure system. The argument is not very ‘strong’. I don’t think that a lowering of the tropopause

itself can explain the flux of stratospheric ozone. It would be interesting to see a vertical cross

section the cut-off low, with tropopause height included. Is the cut-off low eroded away from below,

or how does the flux across the tropopause in the cut-off low really takes place? Some further

thoughts on this might be helpful. The following paper might be a starting point:

S**tohl, A., et al. "Stratosphere ‐ troposphere exchange: A review, and what we have learned from STACCATO." Journal of Geophysical Research: Atmospheres 108.D12 (2003).**

**Todo: read+notes**

### Referee 3:

p.1,l.5: Please add the period of observations

p.4, l. 9: At least mention the dynamical tropopause, it is more common than ozone...

This was noted by both the other reviewers, and added the following at DOLAST: page/line: “... Another commonly used tropopause definition (the dynamical tropopause) is determined from the ±2 PVU isosurface, which allows a 3D view of folds and other tropopause features in a sufficiently resolved model (Skerlak et al., 2014)..”

p.4, l.11: Correct definition of the thermal tropopause "... provided the lapse rate averaged between this altitude …"

C2p.4, l.20 (also Fig.1): The tropopause definitions are mixed here. Why do the authors not include the dynamical definition? The effect of the pure lapse rate criterion is misleading under specific synoptic conditions as correctly stated. This might explain the very low cases in Fig.1.

Well spotted, this is indeed the cause of the low tropopause detections: the lapse-rate definition has been fixed in the latest version to exclude detections below 4km, which were all due temperature inversions near the boundary layer.

As mentioned in response to another comment, the following text was added at DOLAST: page/line “... We slightly alter the lapse rate tropopause definition so as only to detect tropopauses above 4~km altitude, since at all three sites we saw several false positive lapse-rate detections due to temperature inversions near 2~km in altitude.”

Regarding the dynamical tropopause, using solely the sonde data we lacked sufficient information to determine the PV, and we wanted to keep the analysis of sonde records unmodified by other datasets (such as modelled PV).

p.6, l.15: How many model levels are between the sea level and 14 km? How many model levels are between 8 and 14 km and how are sonde and profile data compared? Pointwise or vertically averaged to fit the model levels?

Model and sonde datasets are only compared using the vertically summed tropospheric ozone columns [molecules / cm2].

Vertical model resolution is roughly 60 m near the surface, and around 500 m near 10 km altitude, which has been added to the text at DOLAST: page/line “The vertical resolution is finer near the surface at ~60 m between levels, spreading out to ~500 m near 10 km altitude, and reaching ~1500 m near the top of the atmosphere.”

p.6, l.15+: The sonde profiles are compared against model data of 2 x 2.5 degrees grid sizes (and the vertical model resolution). How well does the model resolve the soundings?

Generally not too well, but we do see an agreement between the datasets in terms of season and amplitude.

How do the authors estimate the fraction of ozone transport which is missed due to unresolved structures?

We assume that if the structure is unresolved, we cannot be certain that it is an STT event. At this time we have not examined the likelihood and frequency of false negatives, however it is an avenue for further research if the method becomes popular

Why do the authors don’t interpolate to the time window of the sounding (or at least use the according model time step)?

We actually did use the closest matching timesteps for comparison, although it is not obvious. The daily model time step over Davis is 0100, 0700, 1300, 1900, of which 0700 is generally closest to sonde release times. This is due to the model using a globally instantaneous (rather than local) snapshot.

p.11, l.17: Even if you did a subjective method, could you explain a bit more in detail in the manuscript, how you distinguished different potential situations? What are upper tropospheric "low pressure fronts"? Tropospheric intrusions (3D!) in the stratosphere or stratospheric cut-offs (fully detached)?

**TODO: UP TO HERE**

p.11,l.20: What are "ozone folds" without other sources of upper tropospheric turbulence and how are these related to the polar vortex?

p.11, l.25: Explain: "...ozone enhancements derived from dry stratospheric air..." didn’t you use the methods and criteria from sec.2?

Derived was poor word choice, we meant simply that the ozone enhancement was likely due to stratospheric influx. The sentence has been updated to “... suggesting the ozone enhancements **are due to** dry stratospheric air.”

p.12, Fig.6. and related discussion (shortly before section 3): Please show a cross section of PV since most likely the ozone peak is related to a tropopause fold.

p.12, last line: What is meant with increased winter activity? More tropopause folds, stronger tropospheric winds, cyclone activity, etc...? Please be more precise. How do you expect the vortex to affect the tropopause?

p.14, l.6-20: Why do you use N2 as indicator? The relation you found is interesting, but not necessarily valid since stability is not conserved. Why should it be ’retained’ when crossing the thermal tropopause?

In general the thermal tropopause is ill defined under these conditions. Why not simply taking PV for this excercise or humidity as a measurement based quantity? Fig.11 and related discussion: Couldn’t you provide scatter plots (or Taylor diagram) of the column ozone between sondes and model?

Fig.3 caption: Units: concentration or mixing ratio?

This image has been updated to use ozone ppbv for both panels.