

## Author's Response

We thank the three reviewers for their detailed, thoughtful and insightful reviews which have improved the quality of our manuscript. In particular, we value their critique on our technique to determine a whole hemisphere flux, which has led to an adjustment of our thresholds to incorporate more data and improve robustness in this data sparse Southern Hemisphere region. We write our detailed responses to each comment in blue text (and quoted text in magenta), directly below the specific reviewer comment.

## Anonymous Referee 1

### Referee Notes:

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.

### 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 presently written, it implies a one kilometre range which would miss deeper intrusions. The text has been updated as follows on page 8, line 7: "... (2 km above the earth's surface to 1 km below the tropopause)." and on page 24, line 3: "... 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?

We investigated modifying this 99<sup>th</sup> percentile threshold in detail and found that lowering the limit to the 95<sup>th</sup> percentile is the optimal balance between including as many detections as possible while minimising false positives. Thus, we use the 95<sup>th</sup> percentile limit in our revised manuscript. Following an inspection of the parsed data, we found that lowering the threshold further resulted in too many clearly incorrect "O3 events" being incorporated into the results. We prefer to only include events which are clear STE events, which we accept could result in an underestimate of STE flux, than including data which are clearly spurious. Regarding use of humidity, this parameter is known to be biased low in the upper troposphere. To some extent this can be corrected for with the type of radiosonde (RS-90) which we used. However we have not performed this correction because it is difficult to apply uniformly across all sites due to seasonal variation in bias that is driven by the seasonal variation in tropopause temperature.

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)?

We have added discussions between the seasonalities reported at Melbourne, Macquarie Island and Davis in this manuscript and results from Wauben et al., 1998, Sprenger et al., 2003, and Škerlak et al., 2014 and 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 as follows on page 11, line 16: “...The SH summer maximum we see for STT ozone flux can also be seen in Fig. 16 of Škerlak et al. (2014), which shows seasonal flux over the southern ocean, although this is less clear over Melbourne. This seasonality is not clear 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 to the north of Melbourne. Their work seems to show slightly higher fold frequencies over Melbourne in summer (Škerlak et al., 2015, Fig. 5), which agrees with our ozonesonde measurements. Their winter maximum is in the subtropics only - from around 20S to 40S, which can be seen as the prevalent feature over Australia in their Fig. 5. 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 this 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 20S to 40S band. The winter maximum does not include Melbourne, or the southern ocean, which explains 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?

Fig. 16 in Škerlak et al., 2014 shows the seasonal STT ozone flux, and a summer maximum is apparent over the Southern Ocean. Fig. 5 of Škerlak et al., 2015 may also agree with our ozonesondes, as there appears to be slightly higher summer fold frequencies over Melbourne.

Please refer to our response immediately above for the additions to the text which also address this point.

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. Therefore we are unable to use GEOS-Chem to detect the STT events which our ozonesondes are capable of detecting.

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.

We agree with this comment. After considering the reviews, we have removed the SO extrapolation and replaced it with a smaller scale examination of STT events in regions surrounding the three release sites. Please refer to the new Figure 1 for the location of these regions.

STT ozone flux near each site has been calculated and compared against Škerlak et al. 2014 on page 20, line 25: “Škerlak et al. (2014) show an estimate of roughly 40 to 150 kg km<sup>-2</sup> month<sup>-1</sup> in these regions, over all seasons, (see Fig. 16, 17 in their publication) while we estimate from 0 to 180 kg km<sup>-2</sup> month<sup>-1</sup> STT impact, following a seasonal cycle with the maximum in austral summer. We

estimate higher maximum flux over Melbourne, (178, and 150 kg km<sup>-2</sup> month<sup>-1</sup> in January and February) than in either Davis (89 kg km<sup>-2</sup> month<sup>-1</sup> in March) or Macquarie Island (68 kg km<sup>-2</sup> month<sup>-1</sup> in January). 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.

We thank the reviewer for the comments below which have helped us improved presentation 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.

We have now referenced Akritidis, and included specific mention of Roelofs and Lelieveld as follows: “... other studies which have found STT ozone fluxes in the SH extra-tropics are largest from autumn or winter to early spring. Roelofs and Lelieveld (1997) use a model with a stratospheric ozone tracer to estimate STT impacts, they see higher SH tropospheric ozone concentrations, as well as STT flux, in the SH winter.”

Full comparison of our results against Škerlak et al. (2014, 2015) has been added where possible, including a conversion of our output to Kg per month within each season and the uncertainties as shown in Table 3, and in the added text quoted in response to Major comment 3.

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 brief description of the sites has been added to the revision: “... Melbourne, a major city in the south east of Australia. 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 (and please note this is now Figure 2 in the revised manuscript)

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).

We have made this change (now Figure 2 in the revision)

b) Increase the range of the vertical axis to show the 10th percentile value for February.

[please refer to our reply directly below]

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 – frequently enough that it showed up in the Davis 10<sup>th</sup> percentile.

We set a lower limit in our algorithm requiring the lapse rate tropopause (LR TP) to be a minimum

of 4km altitude (rather than the 2km minimum previously), since every LR TP detected below that height was a false positive.

This is mentioned in the text at page 5 line 17: “... We require lapse rate tropopauses to be at a minimum of 4 km altitude”

We are now using the ozone defined tropopause in our event detection algorithm, which means the LR TP no longer affects event detection. The ozone tropopause climatology is now displayed in the revised Figure 2.

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 NO<sub>x</sub> levels at these sites (if available) to justify this statement.

The seasonality shown in Figure 3 is consistent with remote free tropospheric photochemistry determined by solar radiation availability and temperature, resulting in higher ozone in winter [Lelieveld and Dentener, 2000]. NO<sub>2</sub> stratospheric observations have been conducted in the Southern hemisphere at Lauder, Macquarie island and Arrival Heights i.e. [Struthers et al., 2004] which displays a winter minima in seasonality consistent with an ozone maxima. We add these references to the text.

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 roughly 500m resolution near 10km altitude. We have now noted this in the text in the Model description section at page 7, line 14: “The vertical resolution is finer near the surface at ~60 m between levels, increasing to ~500 m near 10 km altitude.”

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 have added a reference to Press et al., 1992, along with the following improved description on page 8 line 25: “...to 14 km altitude. Small vertical-scale fluctuations in ozone, which are captured by the high-resolution ozonesondes, can be regarded as sinusoidal waves superimposed on the large vertical scale background tropospheric ozone. As such, the interpolated profiles are bandpass-filtered using a fast Fourier transform (Press et al., 1992) to retain these small vertical scales, between 0.5 km and 5km...”

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? (Please note that in our revision, we now use the 95<sup>th</sup> percentile as our threshold cut-off as we found this the optimum balance between maximise data acceptance and minimising spurious false positives). Once the perturbation profiles are all created, the filtered interpolated values between 2km and TP-1km are used as the basis for this percentile. This has been added to the text at page 8 line 4: “... For an event to qualify as STT, a clear increase above the background ozone level is needed, as a bandpass filter leaves us with enhancements minus any noise or seasonal scale vertical profile effects. The 0.5 km scale limit is set in order to remove any spikes of ozone which could be considered noise. We next use all the perturbation profiles at each site to calculate the 95th percentile perturbation value for the site. The threshold is calculated from all the interpolated filtered values between 2 km above the surface and 1 km below the tropopause. This is our threshold for tropospheric ozone perturbations, and any profiles with perturbations exceeding this

value in individual ozonesondes are classified as STT events.”

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 our response to Major Point 1, we have rewritten this sentence in the revision because it was not clearly detailing our analysis procedure.

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

We have redrawn the plots (now Figure 4) with consistent units.

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 at page 10, line 32: “For this reason all detected STT events found near (within ~150 km of) smoke plumes are flagged, following visual inspection. Removal of these detections reduces the yearly estimated ozone flux by ~15% at Macquarie Island and ~20% at Melbourne.”

In my opinion, Figures 4, 5 and 6 are more supportive-descriptive without adding anything new. Therefore, I suggest including them as a supplement. Moreover, Figures 5 and 6 can be merged into one.

As suggested, these figures and accompanying text have been moved into a supplementary document.

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, because we wanted to determine whether any clearly discernible pattern or dominant weather system connected to the events. At page 10 line 19 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., 2003.

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.

This earlier wording was reflecting that a lower proportion of measurements were made during these months compared with the remainder of the year. In particular, discounting 2011 and 2012 where ozonesondes were launched weekly throughout the year, and 2013 when not flights were made in the early part of the year, the remaining months in summer and autumn have only about 5 flights per month in total. More importantly, we have reconsidered the analysis of STT based on use of the stability criterion, where we now require the square of the Brunt-Väisälä frequency ( $N^2$ ) to be higher at the lapse rate tropopause than the ozone tropopause. This situation allows intruded air to reduce its stability through thermal processes, and discriminates against a small number of cases where intruded air is more stable than the air at the thermal tropopause (which could be the case



when convective air parcels not associated with the STT are near the thermal tropopause). The correspondence between the seasonal distributions between the STT proxy and Fourier analysis are now in closer agreement for Macquarie Island and Melbourne. For Davis, we still have a lack of events for the STT proxy in summer and autumn, and have revised the final sentence of the section that introduces the STT proxy as follows on page 13, line 11:

“... For our STT proxy, we only detect intrusions where the lowest altitude of the intrusion satisfies the ozone tropopause definition. During summer and autumn, the vertical ozone gradients at Davis are weaker compared with the other seasons, and the detected ozone tropopause tends to lie above the lapse rate tropopause potentially reducing the ability to identify STT events based on the definition of our proxy.”

Additionally, we clarify the resolution of the data used. The radiosonde data are binned at 100 m resolution (not 500 m as originally stated), and “ $N^2$  is evaluated using 250 m resolution data (to smooth variability in the vertical gradient of potential temperature that is due to small temperature fluctuations likely associated with gravity waves).” is now included in the same paragraph.

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

GEOS-Chem provides the ozone density (molecules/cm<sup>3</sup>), vertical column boxheights, and tropopause level. We are using the sum of the boxheight \* density for each box below the one containing the tropopause.

We’ve added to the text at page 15, line 3: “Figure 9 compares the time series of tropospheric ozone columns ( $\Omega_{O_3}$ ) in molecules cm<sup>-2</sup> simulated by GEOS-Chem (red) to the measured tropospheric ozone columns (black). GEOS-Chem outputs ozone density (molecules cm<sup>-3</sup>), 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?

We have added this description to the text at page 7, line 18. : “GEOS-Chem uses the tropopause height provided by GEOS-5 meteorological fields, which are calculated using a lapse-rate tropopause definition using the first minimum above the surface in the function  $0.03 \times T(p) - \log(p)$ , with  $p$  in hPa (Rienecker, 2007).”

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

The following has been added on page 16, line 5“ ... The model generally underestimates ozone in the lower troposphere (up to 6 km) over Davis, although this bias is less pronounced during summer. Over Melbourne, ozone in the lower troposphere is well represented, but the model overestimates ozone from around 4 km to the tropopause. Over Macquarie Island we see model overestimation of ozone above 4 km, as well as underestimated ozone in the lower troposphere, suggesting that this region is influenced by processes seen at both of our other sites. Also shown is the mean tropopause height simulated by the model (horizontal dashed red line), which is always higher than the observed average, although this difference is not statistically significant. The effect of local pollution over Melbourne during austral summer (DJF) can be seen from the increased mean mixing ratios and enhanced variance near the surface. The gradient of the O<sub>3</sub> profiles is steeper in the measurements than the model, at all sites during all seasons. Recently Hu et al. (2017) examined GEOS-Chem ozone simulations and saw a similar overestimation of upper troposphere ozone in the mid southern latitudes when running with GEOS5 meteorological fields.”

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.

In order to clarify how we perform this calculation, we have added the following text on page 20, line 8: “ ... ozone enhancements, based on a vertical integration of the ozone above baseline levels for each ozonesonde where an event was detected. The top panel of Fig. 12 shows the STT 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. 4. ...

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 encompass”. What is the longitudinal range?

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

However, following comments from other reviewers, we have replaced this entire Southern Ocean region with three smaller regions each covering the ozonesonde release sites, please see the new Figure 1.

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

These images have been merged into one as suggested

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 agree with the reviewer that this was arbitrary and so we have removed from the revision. We have updated how the calculation of flux is made, and are no longer using this change of  $I$ . The updated calculations are on page 19, line 20: “... To determine the ozone column attributable to STT, we determine monthly averaged STT impact ( $I$ ; fraction of tropospheric ozone sourced from the stratosphere as shown above) and the monthly mean tropospheric ozone column (from the GEOS-Chem multi-year mean,  $\Omega_{O_3}$ ) over the regions described above. This can be expressed simply as the STT flux per event (flux  $I$  in each month:  $\text{flux}_I = \Omega_{O_3} \times I$ . Next we determine how many events are occurring per month by assuming only one event can occur at one time, and that no event is measured twice. ...”.

### Minor comments:

We have implemented the following comments as suggested, or else noted here what we have done.

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.

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 thresholds -> empirically-derived thresholds

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

## Anonymous Referee 2

### Notes

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.

### Major Concerns:

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

We agree with this comment and as such, in the revision we have removed the Southern Ocean extrapolation from the manuscript. We adopt the reviewer's suggestion of using smaller, more local regions: we examine three regions surrounding each ozonesonde launch site as shown in a new Figure 1. Text has been added at page 19, line 15 to reflect these changes:

“... Three regions are used to examine possible STT flux over a larger area using modeled tropospheric ozone concentrations. The regions are shown in Fig. 1. The regions are centred at each site, plus or minus ten degrees latitude, and plus or minus 25, 16, and 11 degrees longitude for Davis, Macquarie Island, and Melbourne respectively. These boundaries approximate a rectangle centred at each site with ~2000 km side lengths, covering ~ 4.4, 4.6, and 4.8 million square km, for Davis, Macquarie Island, and Melbourne respectively.”

**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 then 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, they 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.

Transport is hard to quantitatively analyse using sparse ozonesonde measurements. We have updated the flux calculation with an assumed STT event lifetime of ~ 2 days, which makes the transport question even more pertinent as we see in Trickl et al. that long range transport is possible in certain conditions within 48 hours. We have now included more literature in the introduction, dealing with transport at page 3 line 11:

"Stratospheric ozone intrusions undergo transport and mixing, with up to half of the ozone diffusing within 12 hours following descension from the upper troposphere (Trickl et al., 2014).

The long range transport of enhanced ozone can be facilitated by shear upper tropospheric winds, with remarkably little convective mixing, as shown by Trickl et al., 2014, who measure STT airmasses two days and thousands of kilometres from their source. Cooper et al., 2004, also shows how STT advection can transport stratospheric air over long distances, with a modelled STT event spreading from the northern Pacific to the East coast of the USA over a few days."

Additionally we discuss the effects of uncertainty to do with advection upon our flux estimation:

"... Transport uncertainty is very difficult to estimate with the disparate point measurements; it's possible that detected events are (at least partially) advected out of the analysis regions, which would mean we overestimate the influx into the region, and it is also possible that we are influenced by STT events outside the regions of analysis."

**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 page 25, line 12:

"However, STT events which reach below 4~km are possible and we may have some false negative detections due to the altitude restricted detections."

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.

The reviewer raises a very important point here, in that some of what we are defining as our 'background' tropospheric ozone may in fact be a diffuse ozone of recent stratospheric origin, with vertical scales exceeding our bandpass filtering limits. While our work focuses on the strengths of ozonesonde data (namely, their very high vertical resolution), we recognise that we are ill-placed to capture these type of STT ozone flux events noted by the reviewer. This is, in part, due to the low temporal resolution (weekly ozone flights). To make it clear that we are likely missing some of these type of STT events, we have added text at page DOLAST: 8, line 10 which states: "*We note that this ozone detection methodology detailed above does not allow us to resolve STT events where the ozone flux is spread diffusely across the troposphere without a peak-like structure in the ozonesonde profile. In other words, STT events which might have occurred some distance and time away from the location of the ozonesonde profiles may not be readily detected using the high*

*vertical resolution, but infrequent, ozonesonde launches.”*

- 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’.

We have noted this possibility in our revision at page 19, line 9: “... increased the local background mixing ratio, and any influence from STT events nearby which may also increase the local background ozone.”. Also we mention this aspect in our response to the comment above.

- 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.

Following advice from another reviewer, we have removed this extrapolation from the revised manuscript. Instead, we focus on three smaller regions which are shown in a new Figure 1, located around the sites of the ozonesonde launches. The monthly probabilities of detection have been renamed to  $P$ , 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 page 25, line 3: “Flux is calculated as  $I \times P \times M \times \Omega_{O_3}$ , with each term calculated as described in Sect. 5.1. The uncertainty is determined using the standard deviation of the product, with variance calculated using the variance of a product formula, assuming that each of our terms is independent:

...

Uncertainty in assumed event lifetime is set at 50%, as we believe it is reasonable to expect events to last 1-3 days.  $P$  is the probability of any ozonesonde detecting an event, and is assumed to be constant (for any month). The overall uncertainty as a percentage is shown in parentheses in Table 3, these values are on the order of 100%, largely due to relative uncertainty in the  $I$  factor which ranges from 50-120% for each month.”

We also mention the shortfalls of our uncertainty calculations in section 6.2 at page 25, line 27:

“...Other possibly important uncertainties in our calculation of STT flux which we don’t cover are listed here. Filtering events which occur within 500 m of the tropopause may also lead to more false negatives. This could also cause lower impact estimates due to only measuring ozone enhancements which have descended and potentially slightly dissipated. On the other hand we have no measure of how often or likely the detached ozone intrusion reascends into the stratosphere, which would lead to a reduced stratospheric impact. The estimated tropospheric ozone columns modelled by GEOS-Chem may be biased, for instance Hu et al. (2017) suggest that in general GEOS-Chem (with GEOS-5 met. fields) underestimates STT, with  $\sim 360 \text{ Tg a}^{-1}$  simulated globally, compared to  $\sim 550 \text{ Tg a}^{-1}$  observationally constrained. Transport uncertainty is very difficult to estimate with the disparate point measurements; it’s possible that detected events are (at least partially) advected out of the analysis regions, which would mean we overestimate the influx into the region, and it is also possible that we are influenced by STT events outside the regions of analysis. Uncertainty in event longevity is set to 50%, however this implies a very simplistic model of event lifetimes, a great deal of work could be done to properly model the regional event lifetimes however this is beyond the scope of our work.”

- 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!

This is now noted at page 25, line 21: “Our STT event impact estimates have some sensitivity to our biomass burning filter: including smoke-influenced days increases the mean per area flux by 15-20%. Although events which are detected near fire smoke plumes are removed, some portion of these could be actual STTs. The change in our P parameter when we include potentially smoke influenced events leads to a yearly estimated STT of  $11 \times 10^{17}$  molecules  $\text{cm}^{-2} \text{yr}^{-1}$  over Melbourne, which suggests that up to  $2.1 \times 10^{17}$  molecules  $\text{cm}^{-2} \text{yr}^{-1}$  ozone enhancement could be caused by smoke plume transported precursors. This is a potential area for improvement, as a better method of determining smoke influenced columns would improve confidence in our estimate..”  
which also shows the calculated affect of removing potential smoke events.

- 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.

The sensitivity to this threshold is challenging, we were seeing some false positives when we relaxed to the 98.5<sup>th</sup> percentile however with the other changes made to our STT detection algorithm as part of addressing the reviewers’ comments (such as using the ozone tropopause), we find that the 95<sup>th</sup> percentile is the optimal balance between maximising data acceptance while minimising false positives (from visual inspection, less than 5%) 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.

We have added a Sensitivities and limitations section (Sect. 6) with the uncertainty calculations for both detections and flux estimation. Overall uncertainty in flux is calculated to be around 100%, and we discuss both how we determine this number and the potential uncertainties which we are not able to quantify.

### Minor comments:

We have implemented the following comments as suggested, or else noted here what we have done.

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 propagated 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)”

- 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, at page 5, line 18, the following sentence has been added: “ Another commonly used tropopause definition (the dynamical tropopause) is determined from the  $\pm 2$  PVU isosurface, which allows a 3D view of folds and other tropopause features in a sufficiently resolved model (Škerlak 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'. As this is unclear, we have changed the text at page 5, line 15: “The ozone tropopause may misdiagnose the real tropopause altitude during stratosphere-troposphere exchange; however, it is useful at polar latitudes in winter, where the lapse-rate definition may result in artificially high tropopause values (Bethan et al., 1996; Tomikawa et al., 2009; Alexander et al., 2013)”

- 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?

We now use the ozone tropopause exclusively. We saw several misdiagnosed tropopause heights when using the lapse-rate definition due to low lying temperature inversions.

Actually we tend to see the ozone tropopause at higher altitudes (higher median can be seen in Fig. 3).

- 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.

We select a low-pass limit to remove any background ozone which might, for instance, be slowly increasing with height and could otherwise result in a false STT flag. As noted by other reviewers, our ozonesonde method therefore cannot determine STT which might have occurred some time and space away from the ozonesonde flights because the resultant ozone becomes smeared out through (part of) the troposphere and does not in these cases produce a clear perturbation signal.

- 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 separation 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 did not begin at zero 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?

‘Near’ is defined subjectively as within 150km, which has been added to the text.

- 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 the reviewer has suggested page 10, line 16 : “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:

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

We would like to examine the meteorological event relationships in more detail however we feel that this could best be addressed in a future publication and stress that the relationships noted here are subjective. The case studies in this work and some of the discussion about them has been moved into a supplementary document.

## Anonymous Referee 3

### Notes

Received and published: 23 February 2017



The paper by Greenslade and coauthors presents an analysis of ozone soundings from three locations between  $-31^{\circ}$  S and  $-69^{\circ}$  S over several years. The authors analyse the profiles for ozone enhancements in the troposphere, which they link to stratosphere-to-troposphere 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.

### 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 too 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 aim to present the ozonesonde dataset along with a method of detecting STT ozone intrusions which takes advantage of the very high vertical resolution nature of the ozonesondes. STT flux estimation was included as a novel use for the ozonesonde dataset, although we agree with reviewers' comments that the extrapolation over the southern ocean was too simplified. As such, we have restricted the calculations to regions adjacent to each ozonesonde launch site, as illustrated in the new Figure 1. In the revised manuscript, we provide more analysis of uncertainties and a better comparison between our results 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.

Section 6 has been added which discusses in detail both the uncertainties we calculate and those we are unable to quantify. The 99<sup>th</sup> percentile is replaced by the 95<sup>th</sup>, which, following improvements to our algorithm upon reading the reviewers comments, has fewer obvious false positives (less than

5%) detections, which is compensated for by the increased number of detections. We find the 95<sup>th</sup> percentile level to be the optimal balance between minimising false positives and accepting more ozone events. This is added to the text at page 24, line 2: “The cut-off threshold (defined separately for each site) is determined from the 95th percentile of the ozone perturbation profiles between 2 km above the earth’s surface and 1 km below the tropopause. We use the 95th percentile because at this point the filter locates clear events with fewer than 5% obvious false positive detections. Event detection is sensitive to this choice; for example, using the 96th, and 97th percentile instead decreased detected events by 2, 9 (2,10%) at Davis, 13, 31 (11, 28%) at Macquarie Island, and 8, 24 (6, 18%) at Melbourne.”

This updated and improved analysis procedure is more robust than that described in our submitted manuscript. We think that the reviewer’s suggestion to use a dense Northern Hemisphere network of ozonesondes is valuable and we will consider pursuing this in a separate study.

### Minor comments:

We have implemented the following comments as suggested, or else noted here what we have done.

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...

We have added the following at page 5, line 18 “... Another commonly used tropopause definition (the dynamical tropopause) is determined from the  $\pm 2$  PVU isosurface, which allows a 3D view of folds and other tropopause features in a sufficiently resolved model (Škerlak 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.

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. The revised Figure 2 now shows the multi-year ozone tropopause altitude climatology.

The following text was added at page 5, line 17: “... We require lapse-rate tropopauses to be at a minimum of 4 km 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 / cm<sup>2</sup>].

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 page 7, line 14: “The vertical resolution is finer near the surface at ~60 m between levels, spreading out to ~500 m near 10 km altitude.”

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. We show some examples of the comparison between model and ozonesondes in Figure 14: clearly the model struggles to reproduce the short vertical features in ozone recorded by the ozonesondes.

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.

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

We do use the closest matching time steps for comparison, although we did not make this point clearly in the manuscript. 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)?

The case studies showing two of these events has been moved into a supplementary document following a suggestion from Referee one, and the following has been added to the text: page 11, line 28: "... To detect cut-off low pressure systems we look for cyclonic winds and a detached area of low pressure within ~500 km of a site on days of event detection. For low pressure fronts we look for low pressure troughs within ~500 km."

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

We meant to say tropopause folds, as this tropopause folding and increased stratospheric mixing is seen near many jet streams. The sentence has been changed to "... The stratospheric polar vortex may create tropopause folds without other sources of upper tropospheric turbulence such as low pressure fronts or cyclones (e.g. Baray et al., 2000; Sprenger et al., 2003; Tyrlis et al., 2014) ."

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 changed 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.

We have removed this figure and discussion in the revision.

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?

We were commenting on the relative increase in STT event detections over Davis compared to the other sites, the text has been altered to read: "... At Davis, there are more STT detections during winter relative to our other sites, which may be due to the polar vortex and its associated lowered tropopause and increased turbulence."

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?

We use Brunt-Väisälä frequency as an indicator of transport as it has characteristically different behaviour in the stratosphere and troposphere and can be directly evaluated from the radiosonde measurements. We agree that static stability is not conserved when crossing the tropopause. As discussed in the manuscript, for this particular analysis we require the transport time to be short (hours to 1-2 days) compared with the radiative time. This is generally the case in the extratropical tropopause (Gettleman et al., 2011 - <http://onlinelibrary.wiley.com/doi/10.1029/2011RG000355/pdf>), although we are necessarily biased identifying STT that satisfies this case.

In general the thermal tropopause is ill defined under these conditions. Why not simply taking PV for this exercise or humidity as a measurement based quantity?

As we were seeing problems with the lapse-rate tropopause, and in response to several comments regarding the tropopause definitions, we now use the ozone defined tropopause exclusively for event detection.

Regarding PV: PV is necessarily only available at lower vertical resolution from reanalyses. The radiosonde humidity is biased low near the tropopause, and the performance of the humidity sensor degrades with lower temperatures, giving poorer performance in winter compared with summer.

Fig.11 and related discussion: Couldn't you provide scatter plots (or Taylor diagram) of the column ozone between sondes and model?

We prefer to display the model and ozonesonde as time-series (now Figure 9 in the revision) to allow us to comment on the seasonality.

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

This image has been updated to use ozone ppbv for both panels, with the caption updated.