The authors wish to thank reviewer #1 for their review.

General Comments:

The fact that the stratosphere is a key driver of tropospheric ozone variability in many locations stands in contrast to the emission centric viewpoint of tropospheric ozone variability that is often prevalent in the literature. As such we feel that this paper strengthens, extends and complements the arguments recently made in recent papers on the global scale (e.g., Hess and Zbinden, 2013; Neu et al., 2014) and the regional scale (e.g., Lin et al. (2012), Ordonez et al. (2007), Tarasick et al. (2005)).

Specific comments (reviewers comments in italics):

1. **Reply:** We agree that the proposed title change would be more appropriate.

Change: We will change the title as suggested by the reviewer: "Ensemble Simulations of the Role of the Stratosphere in the Attribution of Northern Extra-Tropical Tropospheric Ozone Variability".

2. **Reply:** In the revised version we will change the sentence: "Despite the simplicity of the tropospheric chemistry, the FR WACCM simulations capture the measured N. H. background interannual tropospheric ozone variability in many locations to a surprising extent, suggesting the importance of external forcing in driving interannual ozone variability".

Change: "Despite the simplicity of its tropospheric chemistry, at a number of N.H. locations the interannual ozone variability in the FR WACCM simulations is significantly correlated with the measured interannual variability, suggesting the importance of external forcing in driving interannual ozone variability"

3. **Reply:** In the revised version we will change the sentence: "The ensemble average tropospheric ozone record can largely be explained as a linear combination of the 30–90 area averaged 150hPa ozone flux and the global methane concentration. "

Change: "The ensemble average tropospheric ozone record can largely be explained as a linear combination of the 30–90^{o} area averaged flux of ozone across the 100 hPa surface and the global methane concentration. "

Note: We actually calculated the ozone flux across the 100 hPa surface, not the 150 hPa surface. This will be corrected here and throughout the text.

4. *"I think I largely agree with your caveats concerning the use of a simplified tropospheric chemistry scheme. I wonder if your simple scheme has a significantly different tropospheric ozone lifetime compared to a more complex scheme, and whether this may be important for the downwards propagation of ozone anomalies? For example, your simple CH4-NOx scheme presumably doesn't represent PAN, and thus misses some long-range transport of NOx. I find it hard to gauge how important this might be. Have you compared local ozone lifetimes in this model to your results in Hess and Zbinden (2013), with a more comprehensive tropospheric chemistry scheme? This may be instructive. I am slightly worried that by using a simple tropospheric chemistry you significantly change the lifetime of ozone in the troposphere, and thus either over- (or possibly even under-) emphasize the role of stratospheric ozone relative to in-situ production."*

Reply: The reviewer is correct in their assumption that the simple CH4-NOx scheme does not represent PAN, and thus misses some of the long-range transport of NOx (although NOx itself likely has a sufficiently long lifetime in the upper troposphere that even without the formation of PAN intercontinental transport of NOx is likely to be significant).

We would have liked to compare lifetimes with a more comprehensive model. However, the WACCM simulations we analyzed do not have the required output (i.e., the photochemical loss of ozone). These simulations were originally designed with a stratospheric focus.

No Change.

5. **Reply:** The ensemble members were initialized from different initial conditions.

Change: We will include this in the revised text.

6. *Just a grumble really: I do find figures etc. in a supplement annoying – I only printed out the main paper and not the supplement, so I have ignored them.*

Reply: Yes, I agree this can be annoying. The second reviewer also felt that we should include additional figures in the text.

Change: We will include the supplementary figures in the main text.

7. *It is interesting that ozone data over Japan cannot be successfully simulated. Is all the Japanese data unreliable? Similarly for the European data prior to 1990 (or 1998). It seems a bit convenient/sweeping to discount all this data. Could you expand on why this data is considered unreliable?*

Reply: One test for reliability might be how well the records of closely spaced stations are correlated. Over Europe there is a rather poor correlation between the measurement stations in the troposphere, even the ones that are quite close together. This raises some question about the accuracy of the measurement record. Hess and Zbinden (2013) pointed this out. Using a similar methodology of comparing the record between closely spaced stations Logan et al. (2012) also raises questions about the accuracy of European ozonesondes prior to 1998. In particular the ozone measurements from the MOZAIC program (available after 1994) and the measurements at the surface European alpine sites (e.g., Jungfraujoch and Zugspitze) are judged to be the most accurate. It is not clear why the earlier ozonesonde measurements are unreliable although Logan et al. (2012) lists a number of possible causes. Thus over Europe the earlier ozonesonde data should be treated with a great deal of care. The alpine site data is likely more reliable, although Logan et al. (2012) points out some possible glitches.

At 150 hPa the Japanese stratospheric ozonesonde measurements show very poor temporal correlation although the tropospheric measurements are better correlated (Hess and Zbinden, 2013). The poor correlation at 150 hPa may or may not be due to poor measurement reliability. We note that the latitudinal span of the Japanese stations is quite large (32^{o} -43^{o}N), possibly contributing to the poor correlation. We note that in both the stratosphere and the troposphere the ozone record for the Japanese stations shows rather poor temporal correlation with the record at many of the other more Northerly locations examined. Examining the EOF pattern, the first EOF has a comparatively small amplitude over the Japanese stations suggesting the processes that control Japanese ozone variability may be rather different from those that control much of the

larger region from 30 -90^{o}N. This is likely consistent with the more southerly location of the Japanese stations. At this point we don't know the cause of the poor correlation between the measurements and the simulation: it is possible the station record is unreliable, or it is possible the model does not accurately capture the interannual variability over Japan.

Change: We will change the text to better reflect the fact that only the 150 hPa interannual ozone record over Japan shows little correlation between measurement sites.

8. *The sentence beginning 'To' is unclear to me. Do you mean: 'If the model variability arose purely due to internal model dynamics, we would expect the ozone records from the different ensemble members to be uncorrelated with each other and uncorrelated with the measurements.'?*

Reply: Yes. This is what we mean.

Change: We will reword the sentence to: "To the extent that the ozone record is driven by internal model dynamics versus interannually varying external forcing we would expect the ozone records from the different ensemble members to be uncorrelated with each other and uncorrelated with the measurements"

9. **Reply:** Agreed.

Change: We will write "externally forced" here.

10. *The lower correlations nearer the surface could be related to the shorter ozone lifetime closer to the surface.I think you say as much later it may be worth mentioning here.*

Reply: Agreed.

Change: In the revised version we will include: "These lower correlations likely reflect the shorter lifetime of ozone near the surface and thus an increased importance of local and regional processes."

11. *Comments on Figure 7: I wonder if the trend in wstar could (at least partly) reflect an expansion in the latitudinal width of tropical upwelling (and shrinkage in extratropical downwelling)? I.e. is 30-90N all downwelling at 150 hPa, or does it include some times/regions of upwelling? (e.g., during NH summer.)*

Reply: There are indeed times when there is tropical upwelling north of 30^{o} N. During the course of the simulation both the rate of downwelling increases (averaged over those regions where there is downwelling; Figure 1 below) and the area of downwelling increases (Figure 2 below). Other modeling studies have found a narrowing in the upwelling branch of the Brewer-Dobson circulation (Li et al., 2010; Hardiman et al., 2014) , consistent with a widening in the downwelling branch.

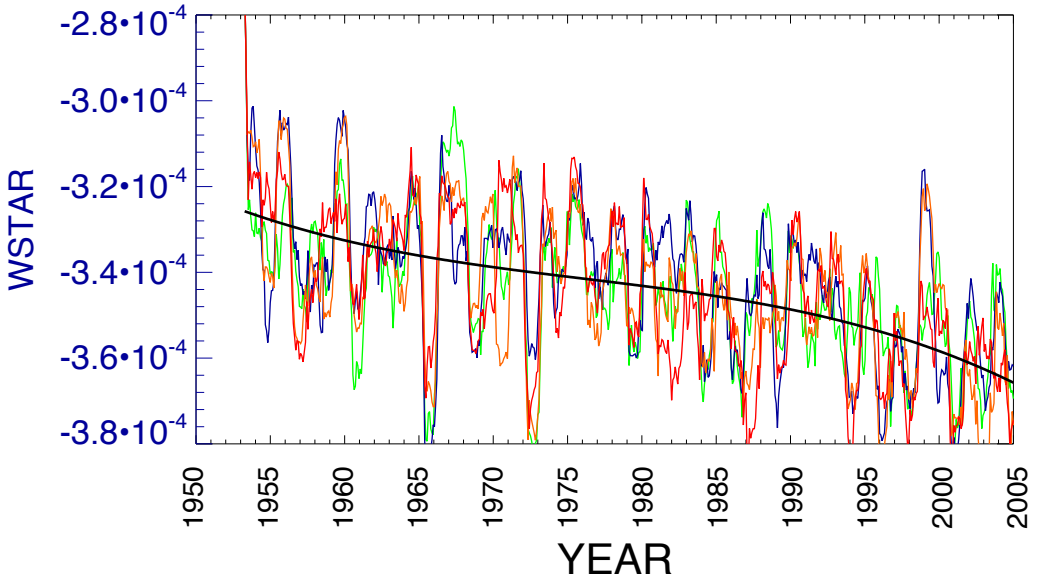

Figure 1: Zonally averaged w* (m/s) weighted only over those regions where it is downwards from 30-90° N for the four WACCM ensemble simulations. Monthly averaged values are zonally averaged and smoothed over 12 months.

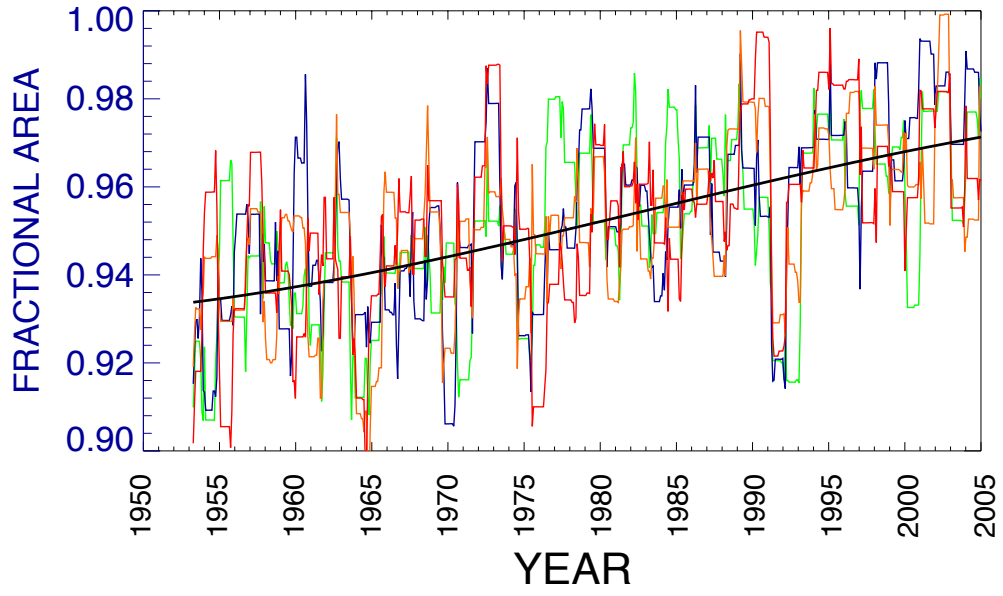

Figure 2: Fraction of area that zonally averaged w* is downwards from 30-90° N for the four WACCM ensemble simulations. Monthly averaged values are zonally averaged and smoothed over 12 months.

Change: We will include in the text that the simulation change in w* is due to both strength in w* over the downwelling region and an increase in area of the downwelling region.

12. *– is the cubic fit with respect to time or CH4?*

Reply: The fit is against methane.

Change: We will clarify this in the following sentence: "We find that a cubic fit of ozone *against methane* captures some of the subtleties of the long-term record better than a quadratic fit."

13. *'The long-term cubic fit to simulated ozone is not linear…' This seems a bit oxymoronic? If the cubic fit was linear, I'd hope you would have just used a linear fit, not a cubic?*

Reply: Good point.

Change: We will delete this sentence.

14. *Not consistent with what? Presumably with emissions changes, but clarify.*

Change: We will reword this sentence to: "After 1985 the shape of the various measured curves is not consistent with the emission changes"

15. *I think you are missing the word 'flux', which makes this confusing – i.e. I think you mean 150 hPa ozone flux.*

Change: Yes it should be ozone flux. We will correct in the revised version.

16. *I wondered what the 5-6 months lag meant physically. Is this the ozone lifetime at 150 hPa? Or the (mean?) transport timescale from 150 hPa to the surface? I would like to understand the significance of this lag length.*

Reply: Using a scale height of 7 km, 150 hPa corresponds to approximately 13 km and 300 hPa corresponds to a height of approximately 8.5 km. We assume 300 hPa corresponds to the height of the tropopause. Given an average downward velocity of approximately 3.4 10^{-4} m/s (see figure above) in the lower stratosphere this gives a timescale of: (13-8.5) (km) x 1000 (m/km) /3.4 10-^{4} (m/s) ~ 153 days. Below the tropopause (e.g., ~300 hPa) the mixing is likely to be more rapid. Thus the timescale of 5-6 months is consistent with a lower stratospheric advective timescale. Reviewer #2 also raised additional questions about this timescale and we have replied in some more depth there.

Change: We will include in the text that the lag is consistent with lower stratospheric advective velocities.

17. *Suggest insert partly or largely before 'ascribed'*

Change: Thank you.

18. *The values are 0.84 and 0.73 in Table 3 (i.e. different to those in text).*

Change: The values in the text 0.83 and 0.74 should be 0.84 and 0.73 (consistent with the table). I inadvertently switched the second digit and will correct.

19. *'physically deep' – do you mean they span from 150 hPa to the surface?*

Chnage: Yes, we will clarify this by qualifying physical deep as spanning from at least 150 hPa to the surface.

20. *By large you mean ~0.4 ppb?*

Change: We will change "large" to "comparatively large"

21. *'…high alpine sites over Europe…where amplitude of first EOF is also large.' I can't see high values over the Alps, either at the surface or 500 hPa in Figure 11. Clarify what you mean.*

Reply: These measurements are usually made in free tropospheric air. Thus it is appropriate to compare the surface ozone concentrations measured at these mountain sites with simulated ozone concentrations at the altitude of the measurements, above the model surface. While the surface EOF values at the surface are small over Europe, at 500 hPa (Figure 11) the alpine sites show a sizeable EOF response of close to 1 ppbv.

Change: This will be clarified in the revised version.

22. *I think these locations can also be usefully described as storm tracks?*

Reply: They are certainly associated with the storm tracks, but I would be hesitant to equate the two.

No change.

23. *This suggests the external forcing is sometimes important, sometimes not?*

Reply: It suggests the response is likely sensitive to the particular kind of forcing. This may have to do with the particular type of "El Nino" event or the conjunction between El Nino and the QBO.

Change: We will clarify this in the revised paper.

24. *Figure 3 (and subsequent related figures): Is the model line all stations, all of the*

time, or does it mimic the observations used? I guess it must be all + all, as it goes back beyond the first measurement data. Is this subtlety important?

Reply: We show the monthly averaged simulated ozone at all the stations within a region. Going backwards in time there are fewer measurements made at fewer stations. This makes it increasingly likely that measurements will not represent the regional variability. We do not have the option of sampling at the measurement sampling times as only monthly averaged output is available from the model simulation. Would our correlations improve if we sampled only at the points where ozone is measured geographically? It is possible, but it seems unlikely. (1) The correlation between the measurement sites within each region is generally high in the simulations. It seems unlikely that sampling the model at only those locations where ozone was actually measured during any given month would improve the model's correlation with the measurements. (2) Since these are free running simulations it seems unlikely the simulations will have much skill in capturing finer scale regional features, e.g., the geographic differences in the interannual variability of ozone within a region. So as to best represent simulated interannual ozone variability as sampled over current sampling network we choose to retain the current sampling methodology.

Change: We will clarify that the simulations are not sampled according to the temporal sampling variability at the individual measurement sites.

24. Technical corrections

Change: Thank you. We will fix these in the revised version.

References:

Hardiman, S. C., N. Butchart, and N. Calvo (2014), The morphology of the Brewer-Dobson circulation and its response to climate change in CMIP5 simulations, Q.J.R. Meteorol. Soc., 140: 1958–1965. doi: 10.1002/qj.2258

Hess, P. G. and Zbinden, R.: Stratospheric impact on tropospheric ozone variability and trends: 1990–2009, Atmospheric Chemistry and Physics, 13(2), 649–674, doi:10.5194/acp-13-649-2013, 2013.

Li, F., R. S. Stolarski, S. Pawson, P. a. Newman, and D. W. Waugh, 2010: Narrowing of the upwelling branch of the Brewer-Dobson circulation and Hadley cell in chemistry-climate model simulations of the 21st century. Geophys. Res. Lett., 37, 1-5, doi:10.1029/2010GL043718.

Lin, M., Fiore, A. M., Cooper, O. R., Horowitz, L. W., Langford, A. O., Levy, H., Johnson, B. J., Naik, V., Oltmans, S. J. and Senff, C. J.: Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions, Journal of Geophysical Research: Atmospheres, 117(D21), D00V22, doi:10.1029/2012JD018151, 2012.

Logan, J. A., Staehelin, J., Megretskaia, I. A., Cammas, J.-P., Thouret, V., Claude, H., De Backer, H., Steinbacher, M., Scheel, H.-E., Stübi, R., Fröhlich, M. and Derwent, R.: Changes in ozone over Europe: Analysis of ozone measurements from sondes, regular aircraft (MOZAIC) and alpine surface sites, Journal of Geophysical Research, 117(D9), D09301, doi:10.1029/2011JD016952, 2012.

Neu, J. L., Flury, T., Manney, G. L., Santee, M. L., Livesey, N. J. and Worden, J.: Tropospheric ozone variations governed by changes in stratospheric circulation, , 7(May), 340–344, doi:10.1038/NGEO2138, 2014.

Ordóñez, C., Brunner, D., Staehelin, J., Hadjinicolaou, P., Pyle, J. a., Jonas, M., Wernli, H. and Prévôt, a. S. H.: Strong influence of lowermost stratospheric ozone on lower tropospheric background ozone changes over Europe, Geophysical Research Letters, 34(7), 1–5, doi:10.1029/2006GL029113, 2007.

Tarasick, D. W.: Changes in the vertical distribution of ozone over Canada from ozonesondes: 1980–2001, Journal of Geophysical Research, 110(D2), D02304, doi:10.1029/2004JD004643, 2005.