The authors wish to the reviewer for their review.

General Comments (reviewers comments are in italics)

Primary concerns:

1. *Several of the results should be compared/contrasted in more detail with previous studies. In particular, the magnitude of the stratospheric impact on the tropospheric ozone found in this work appears to be much greater than what is found by Neu et al. (2014) in their observational based study.*

Reply: This is a good suggestion. Neu et al. (2014) found the correlation between 40-50^o N 150 hPa O3 and 30-50^o N 510 hPa O3 to be 0.40. As we state in the paper: "The correlation between the 150hPa area averaged ozone and the detrended tropospheric area average at 500hPa reaches 0.80 with a lag of 3months; the correlation between 150hPa and surface ozone reaches 0.75 with a lag of 4months."

Reasons for these differences are unclear but might include the length of the sample compared, model biases, measurement biases, and the somewhat different quantities compared.

In the conclusions we state: Extrapolating these changes, a 30% increase in the ozone flux by 2100 (Hegglin and Shepherd, 2009) would result in 3% increase in surface ozone and a 6% increase in 500hPa ozone. The analysis of Neu et al. (2014) gives a 40% increase in the ozone flux would change tropospheric in Northern mid-latitudes by 2%. While Neu et al. (2014) does not specify the exact level that this change will occur at it is clear our paper is at least a factor of two higher.

Change: Results in the revised paper will be explicitly compared with Neu et al (2014).

2. *Also the ozone flux across the 150 hPa surface (1.1 x 10ˆ4 kg/yr; Figure 7) is significantly less than current estimates of stratosphere-troposphere exchange of ozone (For example, the model study by Hsu and Prather (2009) and the observation/model study by Olsen et al. (2013)). I first assumed these results were per unit area and just missing the units, such as km-2. Even if so, the reported values are still much less than current estimates.*

Reply: Thanks for catching this. The figure is correct but with the units kg/sec (the caption says kg/year which is wrong and axis simply gives kg, which is also not correct). The revised version will give Tg/year in the figure.

Note also that the actual calculation used in the paper is the ozone flux across the 100 hPa surface, not the 150 hPa surface as stated in the text. The global flux of ozone across the 100 hPa surface in these simulations is 513 Tg/year. The mean ozone flux from 30-90^{o}N is 352 Tg/year. These estimates are consistent with current estimates.

Change: The revised paper will show the figure in Tg/year and will note the global total. In addition the revised paper will state the vertical ozone flux is calculated at 100 hPa.

3. *Much more discussion needs to be made of the lag times used in the correlations. The lag times were selected by minimizing chi squared. However, these considerable lags (e.g. Table 3) do not seem to have any physical justification. The tropospheric lifetime of ozone is thought to be on the order of a month (e.g. Stevenson et al., 2006). If a mass of ozone descends into the troposphere, I would expect most of those ozone molecules to be lost by 5-6 months later. Later in the paper (p20486), the first EOF is associated with regions of known stratosphere-troposphere exchange, particularly deep exchange. The works cited do demonstrate these preferred regions but they also demonstrate that the exchange is relatively rapid. This does not support the justification of 6-9 month lags used for the correlations in Table 4.*

Reply: We have reviewed the analysis used to create Table 4 (see below). A correction in the analysis gives qualitatively the same results, but with the lags slightly reduced.

LEVEL	Explained Variance[1]	Correlation 150 hPa PC[2]	Correlation O3 FLUX 150 hPa[3]	Lag[4]
150 hPa	79-85%	NA	0.85 (0.43)	2(2)
500 hPa	71-77%	0.78	0.85 (0.50)	5(5)
1000 hPa	40-48%	0.66	0.73 (0.34)	6(6)

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 the tropopause is located at ~300 hPa. The average downward velocity is approximately 3.4 10^{-4} m/s (see figure in paper). This gives a timescale of: (13-8.5) (km) x1000 (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. A timescale of 5-6 months is consistent with a lower stratospheric advective timescale, roughly consistent with the lag-times obtained in the analysis.

The lag-time within the troposphere is much shorter. We obtain a lag on the order of 1 month between 500 hPa and the surface as indicated in the discussion of Figure 10.

A one-month timescale is consistent with the tropospheric lifetime of ozone. However, this timescale increases with altitude; a timescale of closer to a year is probably appropriate for the lowermost extratropical stratosphere.

We have calculated the lag for *interannual* ozone anomalies between different heights. As pointed out by the reviewer the advective timescale for ozone exchange within the troposphere can indeed be measured in days. However, the ozone anomalies associated with these individual events arre of short duration and occur on small scales. The timescale for the transport of annually averaged zonally averaged ozone anomalies within the troposphere probably corresponds to a tropospheric mixing timescale. For this a timescale of a month for transport within the troposphere seems reasonable (as suggested by the discussion in association with Fig. 10). The timescale for transport from 150 hPa is much longer as suggested by the calculation above.

In addition, the correlations shown in Table 4 have a broad and smooth peak with lag-time. For example, while the correlation between the 500 hPa PC – and the 100 hPa O3 flux is 0.85 at a lag 5 months, the lag at 3 months is 0.81. Thus the results are not terribly sensitive to the lag time.

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

4. *Also, a 3-month lag for the 150 hPa EOF correlation with ozone flux at 150 hPa seems counterintuitive. However, this could be justified if the air mass flux was significantly out of phase with the seasonality of ozone.*

Reply: The revised calculation suggests the lag is two months. The maximum response of ozone may be expected to lag the ozone flux as the time change of ozone is proportional to the ozone flux. For example if the ozone flux corresponds to a sin-wave the ozone response would be expected to be pi/2 out of phase with it.

In addition, the ozone flux is actually calculated at 100 hPa.

Change: No change to this part.

Other comments:

5. *I question why many of the figures are placed in the supplementary material. Many of these are presented and discussed on par with the rest of the material in the paper. These would not be "supplemental" and only make it more difficult on the reader to have to jump back and forth between two different places to look at the figures discussed! Most of these should be placed into the body of the paper.*

Change: Reviewer #1 also commented on this. We will move the supplementary figures to the main body of the paper.

6. *The title of the paper should reflect that the study considers only the Northern Hemisphere extratropics.*

Change: We will change the title to: "Ensemble Simulations of the Role of the Stratosphere in the Attribution of Northern Extra-Tropical Tropospheric Ozone Variability" as suggested by Reviewer #1.

7. *It would be helpful if it were explicitly stated earlier that these are free-running simulations. I currently see that in the beginning of the conclusion section. Also, this section could additionally describe how the ensemble members were created (differences in the initializations).*

Change: We will explicitly make this point at the beginning of section 2.1. The four ensembles were initialized using different initial conditions. The revised paper will explicitly state this.

8. *This really doesn't suggest a long-term ozone decrease, especially with the large standard deviation. It appears rather flat over the long-term.*

Reply: The measured record over Japan is rather spotty and noisy. However, the measured values from the late 1990s to at least 2000 are on average less than the measured values during the 1970s and early to mid 1980s. While it is true we haven't evaluated this statistically the paper only states the data suggests a long-term ozone decrease, not a terribly strong statement.

No Change.

9. *Given this statement, should the Northern Europe value in the Table be in bold?*

Reply: The correlation over Northern Europe is significant at the 98% level, not the 99% significance level required to be bold in the table.

Change: We will give the significance levels in the text to remove the seeming contradiction.

10. *As I understand it, the model values are averaged over the region and the observations in each region are averaged together. If so, I am not surprised that the 500 hPa measurements have a standard deviation much larger than the model but they are much more comparable at 150 hPa. The spatial variability of ozone in the troposphere is much greater than in the stratosphere. An average of a small sample of points in the troposphere (the observations) is likely to have greater variability than the average over that continental-scale tropospheric region. Figure 4 also provides supporting evidence of this. The standard deviations are quite large (and time series look completely different) during the earlier record when there are far fewer sites and measurements. After about 1980 in Canada and the early 90s in Europe, the number of sites increase and observation frequencies become greater. This corresponds to the time when the observations and model results begin to agree much better.*

Reply: This is an interesting observation. However, there is likely to be more to it, at least in some locations. (1) An extensive analysis by Logan et al. (2012) shows that in fact many of the ozonesonde observations are unreliable over Europe prior to 1998. This is based partly on the correlation between individual station records. (2) The model simulations (as well as the measurements) suggest that to a large extent the annually averaged signal is highly correlated on a regional basis. While the spatial variability of the troposphere may be large, it is not as clear that the annually averaged interannual variability of the troposphere occurs on small (subregional) spatial scales.

In relation to the difference between the 500 hPa and 150 hPa records the current paper states: "This suggests a comparative degradation in the measurement accuracy at 500hPa compared to 150hPa and/or geographical variability not simulated." This, we believe, captures the reviewer's sense that differences between 500 and 150 hPa are related to the scale of spatial variability.

Change: We will also mention that the decrease in the standard deviation between measurement sites may relate to decreases in the number of sites and a decrease in observation frequencies.

11. *P20478: It appears that the increase after 1990 could be due to the impact of Pinatubo. Thus, it appears that the 1960-2005 trend would be fairly linear if 1990-1995 were removed from the time series.*

Reply: We presume the reviewer is referring to the ozone flux here. We agree that if 1990-1995 were removed from the record it may be hard to discern an increase in ozone between the late 1980 and 2005. However, there seems to be a clear decrease in the ozone flux between mid-1960s to the late 1980s (prior to Mt Pinatubo eruption).

No Change.

12**.** *Table 4, note 3: I don't understand exactly what you mean by "The correlation in parenthesis is computed individually for each simulation; however, the correlation coefficient comprises the overall relationship for all ensembles."*

Reply: We computed an overall correlation coefficient by correlating the overall series consisting of [S1,S2,S3,S4] with [S1*,S2*,S3*,S4*] where S1 and S1* consist of different components of the simulations (e.g., the ozone flux and the first component of the EOF).

Change: We agree that our phrasing here is rather opaque. We will change it in the revised version.

13. *P20488, L6 (referencing Fig. 13): The individual titles in each panel of Figure 13 labels Mace Head, Lassen, and Alpine as the surface rather than 500 hPa. I assume it should be 500 hPa.*

Reply: The measurements at the Mace Head, Lassen and the European Alpine stations are indeed measurements at the surface.

Change: The phrasing used here may have caused some confusion: "(500hPa Canadian sites, Mace Head, Lassen and the European alpine sites)". We will change this to "(500hPa Canadian sites, and the surface sites at Mace Head, Lassen and the European Alps)"

14. *P20488, L28: And the most minor comment: the words "entire" and "the" should be reversed!*

Change: Thank you. We will change.

References:

Hegglin, M. I. and Shepherd, T. G.: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux, Nature Geoscience, 2, 687–691, doi:10.1038/ngeo604, 2009.

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.