This paper presents an interesting analysis of ozone variability above the observation site of Lauder in New Zealand. The objective is to identify and quantify the main drivers of ozone variability and trend at different altitudes as monitored with ozone sondes. The attribution is carried out using multivariate regression analysis and sensitivity simulations from a chemistry-climate model. A large part of the ozone variability is found to be driven by dynamical/climate variability. Some of it is also linked to changes in O3 precursors emissions. The results suggest that ozone long-term monitoring at specific sites contains valuable information in terms of the causes of ozone changes. This study illustrates how to extract information about atmospheric composition changes from observational time series. The results will be useful to scientists running monitoring sites. The paper is reasonably clear and well written. Its scope fits perfectly with those of ACP. Therefore, I recommend publication with minor corrections that the authors may consider.

Thanks very much for the reviewer's positive comments.

l19, p2 : 'where climate change accelerates the export of O3 to higher latitudes, reducing O3' I think the impact of climate change on tropical ozone in the stratosphere is more about enhanced tropical uplift and hence reduced time for ozone production in rising air.

We have revised this sentence to "…but there may be no sustained recovery in the tropical lower stratosphere where climate change increases tropical upwelling, leading to less time for O$_3$ production and hence reducing O$_3$ in this region (Eyring et al., 2010)." (Page 2, lines 18-20)

l5, p3: for more clarity, I would suggest to have: 2.1 ozone record 2.2 statistical method 2.3 chemistry-climate model simulations

Thanks for the suggestion. We have now added subsections in Section 2.

l21, p3: 'Tropopause height is calculated using the 150 ppbv O3 chemical tropopause'. Why 150 pbbv ? Prather et al., 2011, recommended the 100 ppbv O3 contour. Prather, M. J., et al. (2011), An atmospheric chemist in search of the tropopause, J. Geophys. Res., 116, D04306, doi:10.1029/2010JD014939.

We have now adopted the definition of the tropopause based on vertical gradient of ozone, which is greater than 60 ppbv/km and remains so for a further 200 meters, constrained by ozone mixing ratio greater than 80 ppbv and exceeds 110 ppbv immediate above the tropopause (Bethan et al., 1996). This ozone tropopause definition has the advantage of being consistent with the PV definition (Beekmann et al., 1994, Bethan et al., 1996)). Consequently, we have recalculated the regression function and the trend of the observed and simulated tropopause height based on revised tropopause height, and revised figures (2, 3, and 4), tables (2 and 3), and the relevant text accordingly. Resultant changes are small. (Page 3, line 24-28)

l24, p3: The different model simulations are forced by different scenarios in CH4 and other O3 precursors, greenhouse gases (GHGs), and ozone depleting substances (ODSs) over the period of the time series. There does not seem to be any forcing that is characteristic of the effect of tropospheric O3 precursors.

We do not include short-lived ozone precursors in the regression as this would be difficult to formulate in terms of a usable, univariate regression function. Moreover, CH$_4$ is the most effective ozone precursor in the remote Southern Hemisphere background air, but it doesn't impact ozone variability due to its well-mixed nature.

l24, p3: The effect of surface temperature is discussed in the text and appears in Table 3 (correlation coefficient). It seems to be the dominant factor in O3 variability at the surface. Why wasn't it included in the regression?

We have now included surface temperature as a regression function, and made changes accordingly. As a result, we have revised Figures 2 and 3, and Tables 2 and 3 accordingly, and have added related discussions in Section 3.1. (Page 3, Eqn. (1) and Sec. 3.1)

l20, p3: What is Cly? 'effective equivalent chlorine loading (Cly)'. and 2 lines down : 'Cly is the total chlorine loading'.

Thank you. We have now clarified that "Cly is the total chlorine loading". (Page 3, line 23)

l27, p3 : 'We note that interdependencies and correlations, e.g., between stratospheric temperature, relative humidity at surface, and tropopause height as regression functions, cannot be excluded'. There is no need to speculate here. For example, the effective equivalent chlorine loading, a regression function, impacts O3 and stratospheric temperature, one of the regression function. Furthermore, correlations between regression functions can be calculated. It is a significant source of uncertainties and should be discussed.

We have removed this sentence and have added a discussion of the correlations between regression functions in Section 3.1. (Page 6, lines 32-34 & page 7, lines 1-2)

l24, p4: 'CH4 mixing ratios are prescribed at the surface, and the same CH4 scenario is used in both chemistry and radiation.' The wording is a bit confusing. Isn't the model-calculated CH4 fed into the radiation scheme, like O3?

Yes, model-calculated CH$_4$ (i.e. prescribed at the surface but subjected to transport and chemical removal) feeds into the radiation scheme. We have revised the text to make this clearer. (Page 5, lines 2-3)

l32, p4: This paragraph would greatly gain in clarity and usefulness if the different simulations were described and contrasted instead of just referring to the Table 1.

We have now modified the texts to make the description of this second set of CCMI simulations clearer. (Page 5, lines 10-12, 16-18)

l10, p5: 'Fig. 2 shows the deseasonalised O3 anomalies at the eight layers from the surface to the lower stratosphere, and the respective regressed O3 anomalies'. Please, could you indicate when a figure is about observed or model-calculated variables? Also, is it monthly means for all the analysis (as indicated in one figure)? Give more information about the time resolution and data processing/filtering of the different analyses.

We have added more details in Section 3.1 to achieve greater clarity. (Page 5, line 24-30)

l11, p5: Can be quantitative about the amount of variability captured by the regression? just provide the values of the determination coefficient R2.

We now present R^2 in the new Table 2. We have also added some related discussions in Section 3.1. (Page 5 line 30 to page 6 line 3)

l19, p5: "project onto" ? replace by "are mostly driven by"

Changed.

l26, p5: se comment above about Tsurf missing from the regression.

Surface temperature (Tsurf) has now been added in the regression model (Eqn. (1)), and the text has been modified accordingly (Section 3.1).

l29, p6: it's very unlikely that temperature is the driver of O3 variability in the lower stratosphere. The slowing down of O3 destruction by stratospheric cooling (via Chapman cycle) occurs in the upper stratosphere. The ozone budget and variability in the mid-latitude lower stratosphere is dominating by dynamics. I think temperature changes simply reflect dynamical changes that drive O3 variability.

Agree. We have modified the text accordingly. Now it reads "In the stratosphere above 15 km, negative O_3 trends become weaker and are not significant at the 95% confidence level. Long-term changes in O_3 in this region are influenced by changes in dynamics (which are reflected in changes in stratospheric temperatures linked to O_3 changes), and O_3 depletion/recovery that is governed by changes in O_3 depleting substances; such changes seem to have cancelled out the decrease in O_3 resulting from the tropopause height increase in this region." (Page 7 lines 31-34 to page 8 line 1)

l9, p8: Difficult to expect a CCM to reproduce specific short-term O3 anomalies. Those anomalies are often driven by specific dynamical variations. Comparisons between CCM simulations and observational time series make more sense when considering long term trends. I would suggest to add the analysis of a REF-C1SD simulation (wind/temperature forced by meteorological analyses instead of being calculated by the model) and compare to REF-C1. It would give an estimate about the effect of model-calculated dynamics/meteorology and biases (yes, there are some) on O3 variability, including trend, for the considered site. Even on long-time series, this could be significant. The authors would be in a stronger position in their attribution analysis. It is less of a problem when considering large-scale averages or multiple sites but here the analysis is limited to a specific site.

We agree with the reviewer that a free-running CCM has limitations w.r.t. reproducing short-term O3 anomalies. Unfortunately, we have not performed specified dynamics (SD) simulations for comparison with our RED-C1 simulation presented here. However, we did assess simulations by other CCMI models (REF-C1 and REF-C1SD) w.r.t. how their ozone compares to the Lauder ozone record. These other REF-C1 simulations often compare worse to the record than our own simulation. The REF-C1SD simulations usually better reproduce the O_3 variability in the stratosphere than REF-C1, but with mixed performance in the troposphere.

Therefore, for brevity and clarity, we have decided not to include SD simulations from different models here. The focus of this paper is on attribution of anthropogenic forcings to O_3 trends at Lauder, and we believe that this is best achieved by contrasting sensitivity simulations to the reference simulation in a consistent model setup. Moreover, SD simulations are not suitable for use in the attribution study, due to externally imposed meteorological forcing, and in some cases are not self-consistent (r/f Hardiman et al. 2017). Free-running simulations have the merit being more internally self-consistent, taking into account both dynamical and chemical feedbacks, and are more suitable for assessing long-term impact. We have now noted in the text that free-running CCMs have limitations in re-producing O_3 variability and trends at one specific site. Furthermore, a comparison between free-running and SD simulations would be more meaningful in a multi-model and a multi-site context that will avoid any potentially biased conclusion, which is outside the scope of this study. We have added some discussions in the text. We hope this is satisfactory. (Page 9 lines 22-29)

We very much concur with the reviewer's point that it is important to compare free-running and SD simulations in reproducing O_3 trends and variabilities in general, and we will pursue a following-up study to more comprehensively analyse multi-model results for more locations.

Figures : Can the authors add in all the figure captions which curves are what?

We have now added the description of each curve in respective figure captions.