Reviewer 1

The authors evaluate, on a species-by-species basis, the Effective Radiative Forcings associated with ozone-depleting substances (ODSs). They find that while ODSs as a group exhibit a positive ERF, in agreement with previous evaluations, feedbacks due to ozone depletiong and impacts on tropospheric methane loss both substantially offset this aggregate ERF. However, many individual ODSs exhibit negative ERFs when these feedbacks are taken into account, including CFC-11, CCl4, and all halons. This is important new information. Species with positive ERFs include CFC-12 and the HCFCs.

We thank the reviewer for their thoughtful review. We agree this is important new information.

The authors present a well-developed methodological framework to arrive at this conclusion that is mostly based in previous work including Pyle at et 2022. My comments, detailed below, are all minor. I am sure the authors can address these in a revision. I recommend publication subject to providing this revised version for a second round of review.

Details:

Figure 1: Perhaps you want to write a sentence or two about the outliers in Fig. 1 (CH3Cl, CH3Br, CCl4). Why are these substances relatively far from the best linear fit?

It is beyond the scope of this paper to explore this in depth, but we will add comments on this and the possible implications for indirect GWPs.

L179: Replace "is" with "are".

This has been replaced

L202: Please rewrite. This seems to be grammatically wrong.

This has been rewritten as: “The chemical scheme considers several family species. For halogens these are the ClOx, BrOx, Bry, and Cly families.”

L248: I think this is incorrect. Stratospheric ozone depletion in UKESM1 outweighs the positive direct RF from the ODSs not any positive trend in tropospheric ozone. The negative stratospheric ozone trend also substantially, via STE, reduces the tropospheric ozone abundance but I'm unsure about the sign of the tropospheric ozone trend in this model during the period od increasing ODSs.

We have clarified the sentence to reaffirm that the ozone depletion in UKESM1 outweighs the forcing from tropospheric ozone.

L388: Do you really need to make this assumption? I understand that this has some history, but the models that you use as a basis of analysis do not include this assumption.

No, we don’t make this assumption and have clarified the text to explicitly state this: “do not include this effect” -> “do not assume this saturation of polar depletion”

L449-450: Similarly I don't understand why this assumption needs to be made. Is that really a requirement of the method?

We now explain that this polar/extra-polar split is necessary because the fraction release values differ between the polar and non-polar regions.

Table 3: I don't believe that the errors are simply additive. Usually if you assume normally distributed uncertainties and uncorrelated errors, the uncertainties add in quadrature. This would yield an uncertainty of +-1685 not +-2053. Similarly the uncertainties for "O3" and "CH4" can be combined in quadrature, yielding a smaller uncertainty. Please comment.

The uncertainties in two O3 components (extra-polar and polar) are correlated as the uncertainty in each is dominated by the uncertainty in the overall ozone ERF. All other errors are added in quadrature.

L483: What is "FDH"?

This is expanded as “fixed dynamical heating” on the line above. A reference to fixed dynamical heating (Fels et al. 1980) was provided in the introduction on line 85 and we have now provided the abbreviation (FDH) there.

Reviewer 2

General: This is a very nice multi-model experiment on ERFs from halocarbons. The effort that has gone into this manuscript is commendable. Could the authors include some evidence that the models are all capable of representing key parameters such as stratospheric distributions of O3 and ODSs (e.g., by comparisons to observations)? Considering the large spread in ozone depletion that these models produce (e.g., from looking at Fig. 2 and 4), one gets the impression that they might however not be very suitable for assessing any effects from ODSs. Should that perhaps be the main message of the paper? Certainly, a better description of, and justification for, the calculation of uncertainties is needed. Here a few more detailed comments:

Thank you for the comments and for the thoughtful review. We agree that the models have a large spread. This is why the observational constraint from Morgenstern et al. is useful as it allows us to combine an observable quantity (historical rate of depletion) with a modelled quantity (ERF). This means that we do not rely on the models to correctly model the magnitude of the ozone depletion, but we do assume that the models represent the relationships between rate of depletion and ERF. We have added this explanation to the Discussion and Conclusions.

l51-54: Not sure why halons are not mentioned here. Why mention two substance classes that are not covered but not all that are part of the focus.

Accepted, these lines are removed and halons added.

l145-146/fig 1: This slope is probably dependent on the location in the stratosphere, most commonly approximated by the mean age of air. Can the authors provide evidence that it (i.e., the slope) can be generalised throughout the extra-polar stratosphere? Also, what is the uncertainty range of this slope, and how was that taken into account in the analysis?

Further discussion of the slope of this figure is beyond the scope of this paper. This figure is purely to illustrate an approximate linearity between f and 1/tau_strat, and hence to show that although the IOD and ODP metrics appear conceptually different, their behaviour is similar. The

25:1 slope is to guide the eye and is not used at all in the analysis. It is not a fit through the points and as it is a purely graphical construct it doesn't have an uncertainty.

l152 pp: This is an outdated definition of EESC. In the two most recent WMO Ozone Assessments, improved EESC calculations have been used that include, e.g., the shape of the age spectrum, as well as time-independent fractional release. It is unclear why the authors have chosen to ignore these developments here (although they appear to be aware of some of them as Engel et al. (2018) is cited and partly used later).

Thank you, we agree that it will be more appropriate to use the Engel et al definition here. This does not affect the analysis in this section since here we are describing constant concentrations. Using an age spectrum will affect the GWP-100 calculations in Section 5 and these will be updated.

l159 pp: Have the authors assessed the uncertainties or biases introduced if concentrations are not constant (as they are in reality)?

The formulations of the IOD metric and EESC were designed for different purposes. IOD is the integrated depletion from a pulse emission, EESC is the transient evolution of reactive chlorine following a historical or future emission trajectory. In this paragraph we are simply illustrating that if instead of their original purposes, they were both applied to an idealised scenario of a step change in concentration their response would be approximately proportional. The step change scenario is a useful way of determining the GWP as explained in the text. This relationship is not used explicitly in our analysis and therefore the uncertainties and biases are not assessed. The reason for demonstrating that there is an approximate proportionality is to explain the results in Section 5 where the GWPs using the IOD and EESC methods are found to be similar.

l180-293 Clearly, the model-setups are already very different, which might explain the massively different levels of ozone depletion in the runs (as shown in Fig. 2 and 4). A table might enable the reader to get a better overview of the differences in Cl and Br concentrations and how they evolve over time. For instance, the Cl and Br catalytic ozone depletion cycles are known to enhance each other, so getting both of them right is crucial (as are many other things, e.g., heterogenous chemistry, vortex dynamics,..). Looking, e.g., at line 344, the chemical mechanism could therefore be very similar but still produce different O3 depletion results. More generally, the current descriptions are lacking an overview/comparison of such key parameters which prevents a proper assessment of the performance of the models.

We will add further details on the chemistry, but a full evaluation of the models is outside the scope of this paper. We already refer to Thornhill et al. (2021), Griffiths et al. (2021), and Keeble et al. (2021) where more details are available. It is important to note (and we accept we didn't make this sufficiently clear) that we do not directly use the ozone depletion results from individual models in this paper, but rather the empirically constrained result from Morgenstern et al. (2021). This is now made clearer in the text.

l333-334: Using a "fill factor" from a 1997 publication seems highly inappropriate here. These factors will have changed drastically over time, depending on the temporal evolution of the tropospheric concentrations of the individual halocarbons. They therefore cannot be applied as a blanket solution. This oversimplified approach is all the more puzzling since one of the great advantages of models is the wealth of data they provide. It should well be possible (if a bit onerous) to derive time-dependent global-scale fill factors from these data, which could serve as a quality check of the models as well if compared with existing observational data sets such as the one from Volk et al. (1997).

We have added extra points from Hodnebrog et al. 2020 which align with those from Volk et al. showing that they have not “changed drastically over time”. These diagnostics were not archived in CMIP6. It should be noted that this correction factor is between 5% and 10% (a common assumption is 7% for CFCs e.g. Daniel et al. (2007)). Any uncertainties in this factor will be of the order of 1 or 2% and hence are substantially smaller than those in the radiative efficiencies which are of the order of 30%.

l401 pp.: Could the observed O3 column difference (including uncertainty envelope) be added to this figure? It would enable a better perspective on which model might be closest to reality (as clearly not all of them can be very close).

This figure (4) is an idealised model construct of with and without halocarbons, so we don’t have observations from an alternative history in which all other changes occurred (including CO2 and climate change) but halocarbons remained constant. We will add a reference to Keeble et al. (2021) where comparison between modelled and observed polar ozone is shown in their figure 6.

l440-442: This is partly misleading. For example, while many HCFCs appear to have negative GWP100 (at least if one believes the results and especially uncertainties of this model experiment), two of the three main ones (22 and 142b) have substantially positive ones. A total contribution from each substance class weighted by some (current) atmospheric concentration or EESC contribution would help shedding a more balanced light on the matter.

This is a good point, we have clarified the text to say that the most important ones warm.

l449-450: How does this assumption from 30 years ago compare to reality?

This is still the ratio used in the most recent WMO (2022) assessment that has been published. We have added a reference to this. We are not aware of any other updates.

l468: How are the uncertainty ranges in this table calculated? Looking at Fig. 4, where the polar O3 depletion varies by more than a factor of 2 between models, this does not seem to be reflected in these GWP100 uncertainties.

Thank you, we realise that we did not explain fully that we use the Morgenstern et al. ERFs in the GWP calculation. We have now explicitly stated this in Section 4.2. The uncertainty range comes from Morgenstern et al. We have added further explanation in the Discussion and Conclusions as to why we do not use the full model range. The Morgenstern et al. study uses a correlation between the historical ozone trend and the ERF. Although there is a more than a factor of 2 in the modelled ozone trend the constraint methodology can be considered (in an approximate sense) to scale down the ERFs from models with a higher trend than observed, and scale up those from models with a lower trend than observed, hence the range is reduced.

l472-473 Are there any other references to support this statement?

We have added a reference to WMO 2022 here, as the most recent published ozone assessment, which combines the observed EESC with modelled forcing.

l490 Could you elucidate why you used the emergent constraint of Morgenstern et al. (2021)? Using results from a different study if your own ones have too large uncertainties does not seem very convincing.

Thank you, we realise that we did not explain that the model data here is exactly the same as that used by Morgenstern et al. (2021) – i.e. the piClim-HC simulations from the same models from the CMIP6 archive. This means that the constraint from Morgenstern et al. is valid as a constraint on our data. We now explain this in Section 4.2.

l492-494: The main author of this study is one of the three authors of Western et al., which is apparently coming to a different conclusion (although this cannot be confirmed by the reader as the latter paper is still in review). What does this mean? Which study is right or at least better? Or is the lead author of this study really questioning his own work?

Thank you, we understand the confusion. The Western et al. paper uses the SARF metric for the forcing from ozone depletion and discusses that this is around half the ERF metric. In this study we explicitly state that our use of ERF increases the radiative effect of ozone depletion by a factor of two compared to previous WMO reports and Western et al. We have added further comments in the Discussion and Conclusion to explain this.