Response to Reviewer 1's comments

We would like to thank the reviewer for their time and comments to improve the current manuscript. We have addressed the reviewer's comments below, with our responses indicated in blue.

Review of “Future volcanic eruptions may delay the recovery of lower stratospheric O3 over Antarctica and Southern Hemisphere mid-latitudes” by Chim et al.

This manuscript used the model simulation output from Chim et al. (2023) to investigate how sporadic volcanic eruptions impact the O3 recovery in the coming decades. The topic is of interest to the research community and ACP readers. The use of stochastic volcanic eruption scenarios is an advantage of this study over previous studies using a constant background volcanic emission. Also, the analysis based on four different O3 recovery indicators are very useful to understand this problem from different perspectives. However, current model evaluation results do not present a very solidity of using this model version on O3 studies, thus casting doubts on the validity of the results and conclusions. Besides, the structure of the manuscript does not look reasonable. Figure 2 to 7 are all used for model evaluations. Especially, the main conclusion of the manuscript is on volcanic impacts on O3 recovery, but the conclusion draws mostly based on table 2, plus figure 8-9, from which the results does not look that convincing to corroborate the main conclusion of the paper. Some of the texts describing the figures and table 2 are not accurate also cast doubt on the validity of the conclusion. The mechanism investigation is quite limit only based on figure 10. Based on these concerns, at least major revisions are needed to improve the quality of the manuscript.

Below lists the detailed major and specific comments:

Major comments

1. The introduction section should be largely improved. L49-62, L69-86 and figure 1 can be merged and shortened, these are well known basic knowledge on O3 chemistry. Instead, L64-L67 only very shortly mention the advance of the study field, but volcanic halogen and water vapour lead to strong O3 depletion, thus have a large impact on O3 recovery. Many modelling studies already investigated these but are completely missing in this manuscript.

   We thank the reviewer for their suggestions and have rearranged and revised the introduction section accordingly. We initially did not elaborate much on the effects of volcanic halogen and water vapour on ozone because the focus of our study is on volcanic sulfur dioxide emissions and we did not co-inject halogens or water vapour in our simulations. However, we agree with the reviewer that these emissions are important for understanding ozone recovery. We have revised the introduction as follows:

   L49-60, "Stratospheric volcanic sulfate aerosols provide surfaces facilitating

heterogeneous chemical reactions that catalyse the release of reactive chlorine and bromine species from their respective reservoir species (Eqn. 1 to 3). At present, volcanic eruptions that inject sulfur into the stratosphere result in a net decrease in Antarctic column ozone. This is because volcanic sulfate aerosols enhance ozone loss via the HOx, ClOx, and BrOx catalytic cycles, which dominate over the suppression of ozone loss driven by the NOx cycle (Eqn. 4). Observational and modelling studies have provided evidence of the reduction in NOx and enhancement in halogen radicals at mid-latitudes after the 1991 Mt. Pinatubo eruption (Fahey et al., 1993). The ozone response to volcanic sulfate aerosols is greater over Antarctica than other latitudes due to the extreme cold temperatures and the presence of polar stratospheric clouds inside the Antarctic polar vortex. As anthropogenic ODSs continue to decline in this century and assuming no injection of volcanic halogen or unexpected rise in CFC emissions, future volcanic eruptions with the same stratospheric SO_2 injection are expected to cause less ozone loss via the ClOx and BrOx cycles. Therefore, future enhancements in stratospheric aerosol loading are anticipated to lead to a net increase in Antarctic column ozone towards the middle or the end of this century (Klobas et al., 2017)."

Line 67-143, "Volcanic eruptions may also inject water vapour and volcanic halogen species into the stratosphere in addition to volcanic SO_2 and cause additional chemical ozone loss (Bobrowski et al., 2003; Pyle and Mather, 2009; Evan et al., 2023; Santee et al., 2024). Recent modelling studies demonstrate that the co-injection of volcanic sulfur and halogens into the stratosphere can lead to greater and prolonged ozone depletion compared to sulfur injections only (Klobas et al., 2017; Brenna et al., 2020; Ming et al, 2020; Staunton-Sykes et al., 2021). Large amounts of volcanic water vapour injection, as demonstrated by the 2022 Hunga Tonga-Hunga Ha'apai eruption, can also perturb stratospheric ozone for 4-7 years (Zhu et al., 2022; Fleming et al., 2024; Zhou et al., 2024; Zhuo et al., 2025). While volcanic halogen and water vapor emissions are important for ozone recovery projections, strong stratospheric water vapor and halogen injections are rare and highly variable. Volcanic SO_2 is the most commonly emitted species with comprehensive ice-core and satellite emission inventories (Carn et al., 2016; Sigl et al., 2021), whereas volcanic halogen and water vapor injections lack comprehensive records, particularly prior to the satellite era. We therefore focus only on SO_2 emissions in this study."

Added references:

Santee, M. L., Manney, G. L., Lambert, A., Millán, L. F., Livesey, N. J., Pitts, M. C., Froidevaux, L., Read, W. G., and Fuller, R. A.: The Influence of Stratospheric Hydration From the Hunga Eruption on Chemical Processing in the 2023 Antarctic Vortex, JGR Atmospheres, 129, e2023JD040687, https://doi.org/10.1029/2023JD040687, 2024.

Brenna, H., Kutterolf, S., Mills, M. J., and Krüger, K.: The potential impacts of

a sulfur- and halogen-rich supereruption such as Los Chocoyos on the atmosphere and climate, Atmos. Chem. Phys., 20, 6521–6539, https://doi.org/10.5194/acp-20-6521-2020, 2020.

Zhu, Y., Bardeen, C. G., Tilmes, S., Mills, M. J., Wang, X., Harvey, V. L., Taha, G., Kinnison, D., Portmann, R. W., Yu, P., and others: Perturbations in stratospheric aerosol evolution due to the water-rich plume of the 2022 Hunga-Tonga eruption, Communications Earth & Environment, 3, 248, 2022.

Fleming, E. L., Newman, P. A., Liang, Q., and Oman, L. D.: Stratospheric Temperature and Ozone Impacts of the Hunga Tonga-Hunga Ha'apai Water Vapor Injection, JGR Atmospheres, 129, e2023JD039298, https://doi.org/10.1029/2023JD039298, 2024.

Zhou, X., Dhomse, S. S., Feng, W., Mann, G., Heddell, S., Pumphrey, H., Kerridge, B. J., Latter, B., Siddans, R., Ventress, L., Querel, R., Smale, P., Asher, E., Hall, E. G., Bekki, S., and Chipperfield, M. P.: Antarctic Vortex Dehydration in 2023 as a Substantial Removal Pathway for Hunga Tonga-Hunga Ha'apai Water Vapor, Geophysical Research Letters, 51, e2023GL107630, https://doi.org/10.1029/2023GL107630, 2024.

Zhuo, Z., Wang, X., Zhu, Y., Yu, W., Bednarz, E. M., Fleming, E., Colarco, P. R., Watanabe, S., Plummer, D., Stenchikov, G., Randel, W., Bourassa, A., Aquila, V., Sekiya, T., Schoeberl, M. R., Tilmes, S., Zhang, J., Kushner, P. J., and Pausata, F. S. R.: Comparing multi-model ensemble simulations with observations and decadal projections of upper atmospheric variations following the Hunga eruption, Atmos. Chem. Phys., 25, 13161–13176, https://doi.org/10.5194/acp-25-13161-2025, 2025.

2. two median scenarios with small-magnitude eruptions only (VOLC50-1s and VOLC50-2s) are mentioned, but none of the figure showed any results of these experiments except for table 2, why? And no results and discussions addressed the question of "evaluate the effects of small-magnitude eruptions".

We have revised the Figures 5 and 6 (previously Figure 8 and 9) to include the results of small-magnitude eruptions, and revised the results and discussion sections accordingly.

L569-572, "The median return years for ozone mass deficit with a 220 DU threshold are the same for the VOLC50 scenarios and their respective runs with small-magnitude eruptions only (VOLC50-1s and VOLC50-2s). For the 175 DU threshold, VOLC50-2s shows an earlier return by 5 years for both the median and 5th to 95th percentile ranges as compared to VOLC50-2, while VOLC-501s shows a 1-year delay in return year as compared to VOLC50-1."

L595-605, "One of the median scenarios with small-magnitude eruptions only (VOLC50-1s) has a 4-year delay in the median return date of total column

ozone compared to VOLC50-1, while the other median scenario (VOLC50-2) has the same return year as the scenario with small-magnitude eruptions only (VOLC50-2s)."

L617-618, "Small-magnitude eruptions (VOLC50-1s and VOLC50-2s) have no impact on the median return dates of the Antarctic ozone hole area."

L749-755, "We also show that future scenarios with small-magnitude eruptions only (VOLC50-1s and VOLC50-2s) have no effect on the return years of Antarctic ozone mass deficit with a 220 DU threshold, but have a mixed effect on the recovery of deeper ozone holes at the 175 DU threshold. One of the scenarios with small-magnitude eruptions only (VOLC50-2s), which has the lowest annual SO_2 flux (0.57 Tg SO_2 yr^{-1}), shows earlier recovery by 5 years for the 175 DU threshold metric, while the other small-magnitude scenario (VOLC50-1s) shows a 1-year delay compared to VOLC50-1 (Table 1). This suggests that ozone recovery timing is primarily governed by large-magnitude eruptions, with small-magnitude eruptions playing a secondary role."

Conclusion, L968-969, "Our results also showed that small-magnitude eruptions have little effect on the recovery of Antarctic and SH mid-latitudes stratospheric ozone."

3. The Pinatubo eruption erupted at a specific latitude-longitude location, but it needs to inject across 13 latitude bands to allow aerosols to be distributed correctly. This cast doubts on the validity of the model ability to simulate the eruptions. As for other eruptions, did you also use wider latitude bands? And different eruptions need different latitude bands? As The distribution of the aerosol is one of the key factors that affecting the regional O3 responses.

For the UKESM1.1 Pinatubo-only simulations, we included 13 injection latitudes between 0° and 15°N. This approach is consistent with previous modelling studies using UM-UKCA and CESM, including Dhomse et al. (2014) and Mills et al. (2016), and is designed to replicate the observed initial southward spread of the aerosol cloud. We acknowledge that the injection spreading required by some models to reproduce the hemispheric aerosol distribution might not only be caused by the specific meteorological conditions in 1991 that resulted in a southward cloud spreading, but also be a consequence of model biases in stratospheric circulation. Such biases are not specific to UM-UKCA (e.g., Mills et al. 2016, Clyne et al., 2021).

We did not use a wider latitude band for eruptions in the stochastic volcanic eruption scenarios. For the stochastic volcanic eruption scenarios, each volcanic eruption is assigned with an eruption latitude during the resampling process. If the eruption is drawn from the bipolar ice core array, according to the eruption hemisphere indicated in the ice core dataset, we resample the eruption latitude from the geological dataset (Smithsonian Global Volcanic Programme Holocene eruption database) of the respective hemisphere. For

instance, if a Northern Hemisphere eruption is selected from the ice core dataset, then we resample from the Northern Hemisphere eruptions in the geological dataset for the eruption latitude. Therefore, each eruption in the stochastic scenarios is assigned with a specific latitude and longitude for injection.

L322-324, "This injection approach is consistent with previous modelling studies using UM-UKCA and CESM (Dhomse et al., 2014; Mills et al., 2016)."

Added references:

Dhomse, S. S., Emmerson, K. M., Mann, G. W., Bellouin, N., Carslaw, K. S., Chipperfield, M. P., Hommel, R., Abraham, N. L., Telford, P., Braesicke, P., Dalvi, M., Johnson, C. E., O'Connor, F., Morgenstern, O., Pyle, J. A., Deshler, T., Zawodny, J. M., and Thomason, L. W.: Aerosol microphysics simulations of the Mt. Pinatubo eruption with the UM-UKCA composition-climate model, Atmos. Chem. Phys., 14, 11221–11246, https://doi.org/10.5194/acp-14-11221-2014, 2014.

Mills, M. J., Schmidt, A., Easter, R., Solomon, S., Kinnison, D. E., Ghan, S. J., Neely, R. R., Marsh, D. R., Conley, A., Bardeen, C. G., and Gettelman, A.: Global volcanic aerosol properties derived from emissions, 1990–2014, using CESM1 (WACCM), J. Geophys. Res.-Atmos., 121, 2332–2348, https://doi.org/10.1002/2015JD024290, 2016.

4. Why VSL chlorine and bromine compounds are mentioned and discussed even more than the volcanic halogen emissions, which lead to stratosphere halogen injections more than background emissions and have a larger post-eruption impact on stratospheric O3. Is this reasonable?

We agree with the reviewer that the discussion on volcanic halogen and VSL chlorine and bromine seem to be imbalanced in the introduction. We have revised the introduction accordingly (see response to major comment #1), and moved some of the lines about VSL chlorine and bromine compounds to the discussion, see below.

L893-899, "Klobas et al. (2017) emphasised the sensitivity of future ozone changes to a Pinatubo-like eruption with VSL bromine injection between 0 to 8 pptv, which can lead to changes in total column ozone over Antarctica between -3% and 3% under one future representative concentration pathway (i.e., RCP6.0). Although VSL chlorine has a small contribution to total stratospheric chlorine (about 3%) between 2010 to 2019 (Bednarz et al., 2022), the increasing emission of VSL chlorine highlights the potential importance of its impact on future ozone changes. By not accounting for VSL species in our simulations, we likely underestimate the full magnitude of volcanic impacts on ozone recovery. The lack of comprehensive historical records on volcanic halogen emissions, the variability in halogen injections across different eruptions, and the uncertainties in future VSL halogenated

compounds emissions make it challenging to project future stratospheric halogen loadings using our stochastic approach."

5. The analysis methods can lead to biased understand on evaluating the model and showing O3 impacts. Why different number of months are used for calculating cumulative O3 loss between Antarctica and SH mid-latitude.

   We agree with the reviewer that using consistent time periods provides a fairer comparison of ozone responses between Antarctica and SH mid-latitudes. We initially used October to March for Antarctica to capture the peak ozone loss during the Antarctic spring and summer. However, this approach introduces inconsistency in the comparison with the 12-month calculation for SH mid-latitudes and could lead to biased interpretation of the results in Figure 7 (previously Figure 10).

   We have revised Figure 7 to calculate cumulative ozone loss over a 12-month period for both regions. For Antarctica, we integrate from August to the following July to capture the full seasonal cycle beginning with the onset of ozone depletion in August. For SH mid-latitudes, we calculate cumulative ozone loss for the 12 months following each selected volcanic eruption, as the effects of volcanic aerosols at the SH mid-latitudes are not seasonally confined.

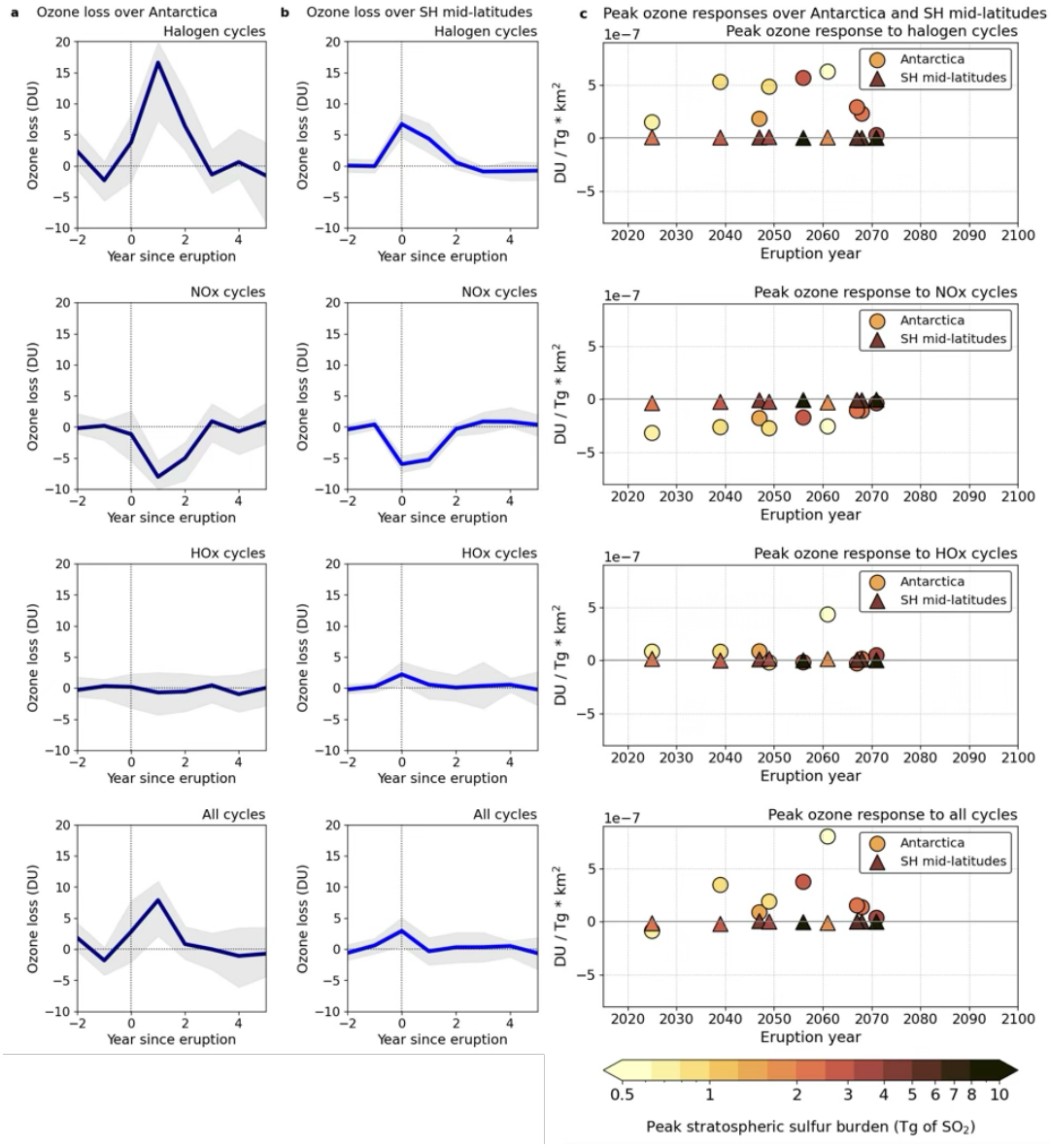


Figure 7. Composite analysis of the cumulative stratospheric ozone loss (up to 25 km) relative to NOVOLC for the 9 selected eruptions (a) over Antarctica, integrated by time from August to the following July, and (b) SH mid-latitudes, cumulative loss for 12 months after the onset of each eruption. Year 0 denotes the year of eruption. The shaded region shows the range of the first and third quartiles. (c) Peak ozone response of the selected eruptions to halogen cycles, NOx cycle, HOx cycle, and all cycles (halogen + NOx + HOx) over Antarctica (circles) and SH mid-latitudes (triangles). The values are normalised with the respective peak stratospheric S burden for each eruption and total area over Antarctica and SH mid-latitudes, respectively.

Why model evaluations are against with other models or CMIP6 multi model mean? There are observations, even used in the cited reference in Keeble et al. 2021, like SWOOSH. Results also show large difference between models, what are these differences mean? What impact do they have on the O3 recovery? There should be important aspects for understanding the validity of using the model on this study and shown results of O3 responses, but not discussed. Figure 7 also shows the timing of the modelled Antarctic O3 hole deviate largely from historical observations, and it was mentioned in the paper that another version of UKESM simulate O3 correctly. Does these

mean this model version cannot simulate O3 reasonably? The conclusion the texts based on these figures does not read that convincing for the conclusion on the validity of the model.

Since UKESM1.1 is a recent model version not included in the previous CMIP6 multi-model intercomparison studies, we compared its results against the CMIP6 multi-model mean to evaluate the model performance and aid our interpretation of the results. Ozone return years in CCMI simulations are typically calculated by taking the multi-model mean, so understanding where UKESM1.1 sits relative to other CMIP6 models helps contextualise our projections of estimated ozone recovery dates. Our primary objective here is to evaluate the model's performance within the CMIP6 framework. ML-TOMCAT, which we also reference and compare, is a machine learning-trained dataset based on the TOMCAT 3-D chemical transport model and satellite observations, which has a good agreement with independent satellite-based dataset including GOZCARDS and SWOOSH (for non-MLS periods, as the MLS period SWOOSH is used as training data).

We acknowledge that UKESM1.1 exhibits known biases that affect its simulation of Antarctic ozone. The prolonged ozone depletion over Antarctica in UKESM is primarily attributed to the model's stratospheric cold bias and excessively strong polar vortex, which are known issues in UKESM (Sellar et al., 2019; Hall et al., 2021). As shown in the comparison with ML-TOMCAT (Figure S1), UKESM overestimates ozone loss over Antarctic between 10 and 30 hPa, and underestimates ozone loss over the Antarctic lower stratosphere. This stratospheric cold bias persists across all UKESM versions, and future model development is needed to address this limitation. Despite these biases, our 1991 Mt. Pinatubo simulation using UKESM1.1 shows a comparable magnitude of Antarctic ozone loss to the NIWA-BS dataset, and the Antarctic October mean total column ozone simulated by the model agrees well with the satellite observation. UKESM remains suitable for investigating the relative impacts of different volcanic forcing scenarios on ozone recovery. However, due to the prolonged Antarctic ozone hole in the model, we expect our results overestimate the duration and cumulative ozone loss over Antarctica (see L1079-1091).

We also agree with the reviewer's comment on the structure of the manuscript, and decided to move Figures 4 and 5 (now Figure S1 and S2) to the supplementary information. We have revised the discussion section as follows,

L485-490, "UKESM1.1 simulates higher ozone mass mixing ratio over the tropics and extratropical regions, lower ozone over Antarctic stratosphere from 1-10hPa, and higher ozone in the Antarctic lower stratosphere between 10-30 hPa compared to both ML-TOMCAT and the CMIP6 multi-model mean (Fig. S1d and e). This suggests that UKESM1.1 may overestimate lower stratospheric (10-30 hPa) ozone concentration over Antarctica relative to other CMIP6 models and ML-TOMCAT. UKESM1.1 also simulates higher

total column ozone between 30°S and 60°N, and a deeper, more persistent Antarctic ozone hole compared to the CMIP6 multi-model mean (Fig. S2)."

L421-424, "Although our simulations capture the magnitude of the Antarctic total column ozone response, the timing of ozone loss in 1991 does not match NIWA-BS. This discrepancy likely reflects differences between our free-running ensemble climatology and the 1991 atmospheric conditions, as our Pinatubo simulations are not nudged to observations."

L517-519, "The magnitude of Antarctic total column ozone response in our 1991 Mt. Pinatubo simulations also show a good agreement with observational datasets, suggesting that our estimates of relative effects across different volcanic scenarios are robust."

L887-899, "We acknowledge that UKESM1.1 exhibits a stratospheric cold bias and excessively strong polar vortex that leads to prolonged ozone depletion over Antarctica (Fig. 4). The comparison of the UKESM1.1 simulation with ML-TOMCAT shows that the climatological mean of UKESM1.1 overestimates lower stratospheric ozone loss over Antarctica and SH mid-latitudes. Due to these model biases, our results likely overestimate the cumulative ozone loss over Antarctica and SH mid-latitudes. However, the relative volcanic effects on ozone in our simulated scenarios remain robust. In addition, our stochastic scenarios include stratospheric volcanic SO_2 emissions only, but not volcanic halogen species, water vapour and VSL chlorine and bromine compounds, which affect stratospheric ozone recovery."

L901-902, "Therefore, our model-simulated effects on ozone represent a lower bound of the potential effects of future volcanic eruptions on ozone depletion."

6. Table 2 shows the metrics to evaluate O3 recovery, but only VOLC98 shows consistent delay based on all the metrics, other eruption clusters sometimes show an earlier recovery. For VOLC98, there are way more strong volcanic eruptions in 2050-2075 with high Antarctica SO2 loadings as shown in Fig. 2, while the O3 recovery metric is also mostly fall in this range 2058-2066, then how can you exclude that the delay of O3 recovery metric is just due to these short episodic large eruptions, especially the huge tropical eruption in 2056 with 114 Tg of SO2 emission?

We agree with the reviewer and do not exclude that the delay in ozone recovery is related to large-magnitude eruptions. In fact, we would like to highlight that the occurrence of large-magnitude eruptions close to the timing of stratospheric ozone recovery can lead to delays in ozone return years. The delay in return year is sensitive to the timing of large-magnitude eruptions in the stochastic scenarios. VOLC2.5 has a large-magnitude eruption in 2061, while VOLC98 has clusters of large-magnitude eruptions between 2035 and 2060, which contribute to the delays in ozone recovery metrics. We have

made changes in the results section (see the end of the response to major comment #6).

VOLC50-1 and VOLC50-2 is used, but it's not clear whether large or small eruptions affect the results. For O3 mass deficit to 220 DU, VOLC2.5 and VOLC98 show a delay, but other scenarios all show an earlier recovery, why? For O3 mass deficit to 175 DU, VOLC2.5 show an even longer delay compared to VOLC50-1S and VOLC50-2s and even VOLC98, while VOLC50-1 even show an earlier recovery.

The apparent inconsistencies across scenarios arise from the different sensitivities and recovery timelines of the two ozone mass deficit metrics. The ozone mass deficit with a 175 DU threshold recovers faster (with a median return year of 2038 for NOVOLC) than the ozone mass deficit with a 220 DU threshold (with a median return year of 2062 for NOVOLC). These metrics are most sensitive to aerosol perturbations during the years before they return to their respective thresholds.

For the ozone mass deficit with 220 DU threshold, VOLC2.5 and VOLC98 are characterized by large-magnitude eruptions with Southern Hemisphere aerosol distribution after 2050, which explains why both scenarios show later return years. The earlier return year in VOLC50-1 is unexpected; however, the uncertainty range for VOLC50-1 (2050-2069) is much wider than NOVOLC (2054-2065), suggesting this result may be related to internal variability.

For the 175 DU threshold, VOLC50-2 shows the longest delay (6 years), while VOLC2.5 and VOLC98 both show 3-year delays, and VOLC50-1 has the same return year as NOVOLC. This pattern can be explained by the timing of eruptions: VOLC50-2 has higher Southern Hemisphere SAOD than VOLC2.5 and VOLC98 before 2038, which affects the recovery of the 175DU threshold ozone mass deficit (see SAOD panel in Figure 6). In contrast, the first large-magnitude eruption in VOLC50-1 occurred in 2047, which is after the return year of the 175 DU threshold for ozone mass deficit.

We have also added discussion of the small-magnitude eruption scenarios (VOLC50-1s and VOLC50-2s) to address whether large or small eruptions dominate these effects (see the end of the response to major comment #6).

For O3 hole area to 220 DU (unit also in DU?), why 2058 (2058-2059) for NOVOLC, but 2059 (2058-2059) for VOLC50-2S, considering this uncertainty, are three members enough for this? These questions/inconsistencies make it not convincing on the main conclusion of the paper. The table is not that intuitive for range comparison, can be changed into a more illustrative figure.

We calculated the uncertainty range using Monte Carlo bootstrap sampling and report the median of the return year. Although the return years for Antarctic ozone hole area are close to each other, the median return year for VOLC50-2s (2059) is slightly later than NOVOLC (2058). We acknowledge

that the limited ensemble size (three members) contributes to the large uncertainty ranges, and this limitation is now discussed in the revised manuscript. The overlapping uncertainty ranges demonstrate the substantial natural variability in ozone recovery timing between ensemble members, even under stochastic scenarios with the same forcing magnitude. This variability highlights that projecting recovery dates remains challenging, and our results should be interpreted as ranges rather than exact predictions.

Regarding the table format, we prefer to keep Table 2 as is for easier direct comparison of median and uncertainty ranges between scenarios.

We have revised the discussion section as follows,

L739-759, "For the 175 DU threshold, VOLC50-2 shows the longest delay (6 years), while VOLC2.5 and VOLC98 both show 3-year delays, and VOLC50-1 has the same return year as NOVOLC. The apparent inconsistencies in the delays of recovery across scenarios arise from the different recovery timelines, which is due to differences in the timing of volcanic eruptions across scenarios, as well as varying sensitivities of the two ozone mass deficit thresholds. The recovery for the 175 DU threshold is much faster (median year of 2038 in NOVOLC) than the 220 DU threshold (median year of 2062 in NOVOLC), making each metric sensitive to eruptions occurring at different times. For instance, VOLC2.5 and VOLC98 both have large-magnitude eruptions with SH aerosol distribution before 2062, thus showing greater delay in return years for the 220 DU threshold. VOLC50-2 shows the longest delay for the 175DU threshold because it has higher SH SAOD than VOLC2.5 and VOLC98 before 2038 (Fig. 6). In contrast, the first large-magnitude eruption in VOLC50-1 occurs in 2047, after the 175 DU threshold recovery year, explaining why it shows no delay for this metric.

We also show that future scenarios with small-magnitude eruptions only (VOLC50-1s and VOLC50-2s) have no effect on the return years of Antarctic ozone mass deficit with a 220 DU threshold, but have a mixed effect on the recovery of deeper ozone holes at the 175 DU threshold. One of the scenarios with small-magnitude eruptions only (VOLC50-2s), which has the lowest annual SO_2 flux (0.57 Tg SO_2 yr^{-1}), shows earlier recovery by 5 years for the 175 DU threshold metric, while the other small-magnitude scenario (VOLC50-1s) shows a 1-year delay compared to VOLC50-1 (Table 1). This suggests that ozone recovery timing is primarily governed by large-magnitude eruptions, with small-magnitude eruptions playing a secondary role. However, the large uncertainty ranges and overlapping values in the return years between VOLC and NOVOLC scenarios reflect substantial internal variability between ensemble members; this highlights that recovery projections should be interpreted as ranges rather than median values alone. Our simulations are also limited by the number of ensemble members. Future modelling experiments with larger ensemble sizes would further quantify the uncertainty in volcanic effects on ozone recovery and distinguish between volcanic forcing effects and internal variability."

Specific comments

1. Figure 1 does not add important information to the text and the impact of volcanic eruptions on O3, not that needed.

   We have decided to remove Figure 1 from the main text and revised the introduction.

2. Table 1 is not that necessary, the Tg of SO2 per year can be just mentioned in the text. Need to mention more information regarding volcanic distributions and location etc. can be found in which figure/table in Chim et al., 2023. Then readers can easily know where to find this details that are useful for understanding the method.

   We have removed Table 1 and included the information of SO_2 flux for each scenario in the text.

   L281-289, "We consider four scenarios as input for our future climate simulations: the low-end scenario (VOLC2.5, with 0.64 Tg of SO_2 yr^{-1}), the two median scenarios (VOLC50-1 and VOLC50-2, with 1.44 Tg of SO_2 yr^{-1}) and the high-end scenario (VOLC98, with 5.23 Tg of SO_2 yr^{-1}), each correspond to a scenario sampled near to the 2.5th, 50th, and 97.5th percentiles, respectively, of the ranked total sulfur mass of the 1000 stochastic scenarios. We perform simulations of the two median scenarios with small-magnitude eruptions only (VOLC50-1s and VOLC50-2s, with 0.64 and 0.57 Tg of SO_2 yr^{-1}) to evaluate the effects of small-magnitude eruptions. The results of the VOLC runs are compared against control simulations with constant volcanic forcing as in CMIP6 ScenarioMIP (CONST, with 0.77 to 0.86 Tg of SO_2 yr^{-1} according to current volcanic emission records) and without explosive volcanic emissions (NOVOLC). All runs used the SSP3-7.0 scenario for anthropogenic forcings with simulation years between 2015 and 2100, with three ensemble members each."

3. Figure 2 suggest adding latitude range also in the figure, then readers don't need to find it in the text.

   Thank you for the suggestion. We have added the latitude ranges and the injection latitude of the selected eruptions in Figure 2 (now Figure 1), also attached below.

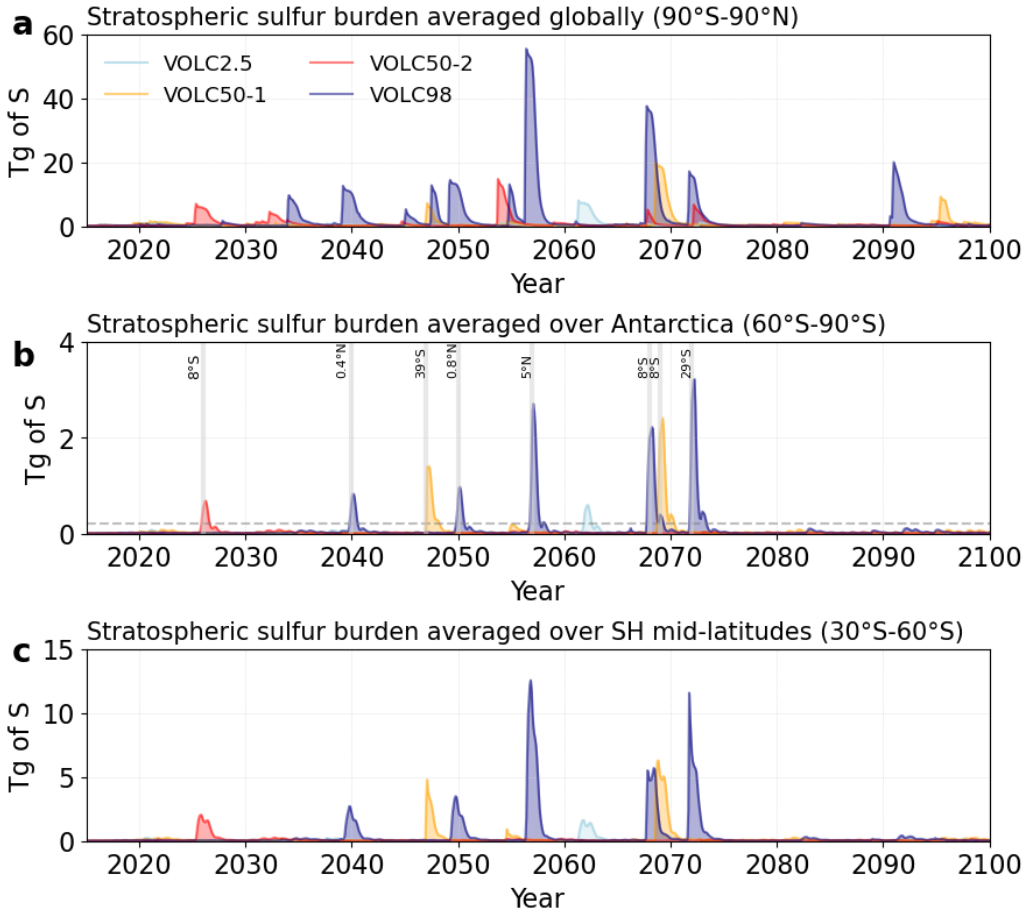


Figure 1. Stratospheric sulfur burden (in Tg of S) averaged (a) globally, (b) over Antarctica, and (c) over SH mid-latitudes from 2015 to 2100 for the four stochastic scenarios. The dotted line in panel (b) shows the stratospheric sulfur burden threshold of 0.2 Tg of S, which is used to select eruptions for composite analysis. The injection latitudes of the selected eruptions are shown in panel (b).

4. L175: access -> assess.

   Corrected in the text.

5. L190: confusing “October-mean”.

   Corrected in the text as "October Antarctic ozone hole area".

6. L194-196: different font.

   Revised in the text.

7. L196-200: “October to March” for Antarctica but 12 months for SH mid-latitude cumulative O3 loss calculation, any O3 increase in any months that counteract the O3 loss effect in mid-latitude? Then it’s not apple-to-apple comparison, will this contribute to a biased understanding on results shown in Fig. 10?

We agree with the reviewer that using consistent time periods provides a fairer comparison between Antarctica and SH mid-latitudes. We initially used October to March for Antarctica to capture the peak ozone loss during the Antarctic spring and summer. However, we acknowledge that this approach is inconsistent with the 12-month calculation for SH mid-latitudes and could lead to biased interpretation of the results in Figure 7 (previously Figure 10).

We have revised Figure 7 to calculate cumulative ozone loss over a 12-month period for both regions. For Antarctica, we integrate from August to the following July to capture the full seasonal cycle beginning with the onset of ozone depletion in August. For SH mid-latitudes, we calculate cumulative ozone loss for the 12 months following each selected volcanic eruption, as the effects of volcanic aerosols at the SH mid-latitudes are not seasonally confined.

8. L207-208 “2-year window prior to the eruption” Any double eruptions within several years? How did you deal with this? As the former eruption can elevate the pre-eruption background conditions.

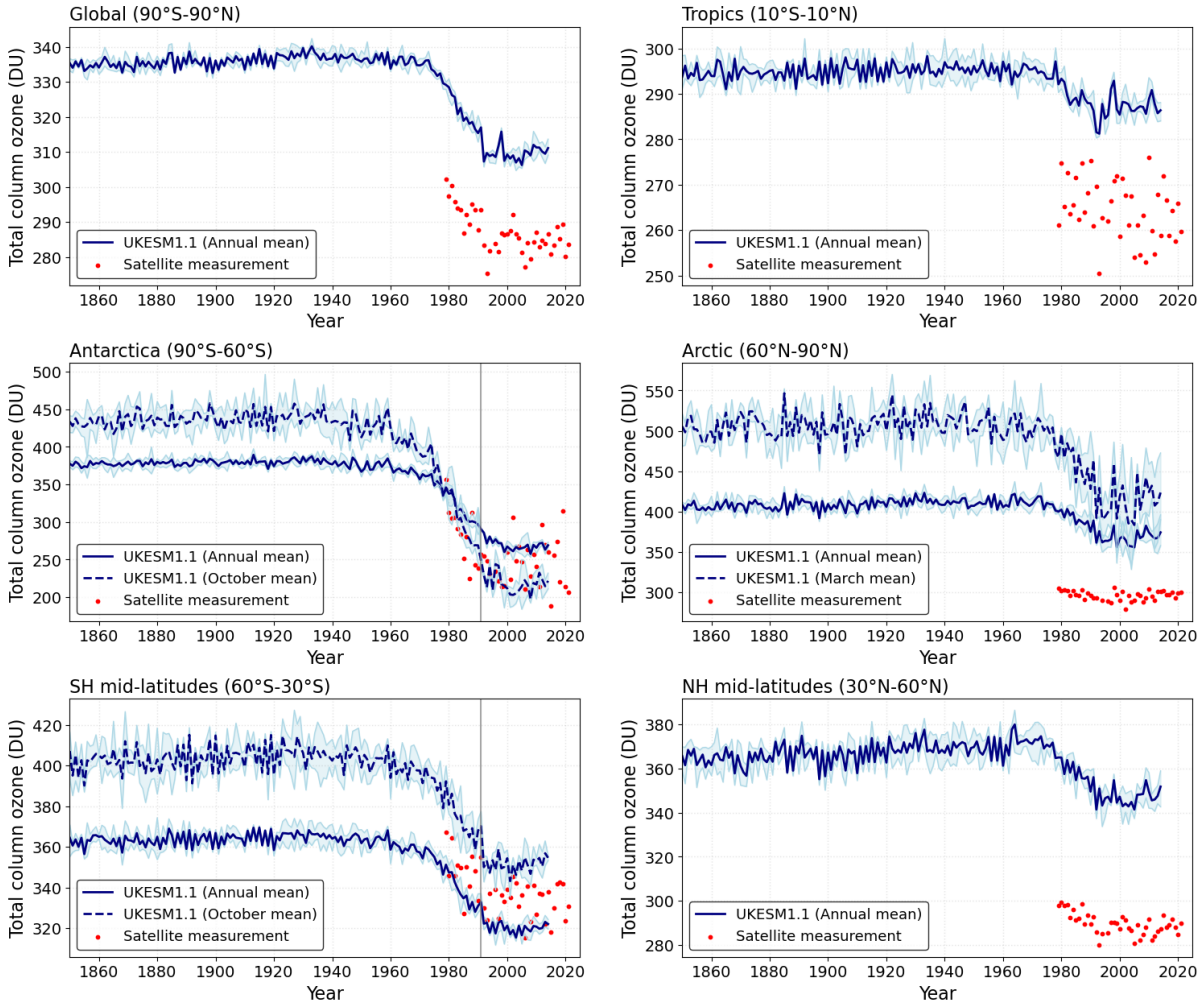
   The high-end volcanic scenario (VOLC98) has clusters of large-magnitude eruptions that occurred very close to each other (i.e., in year 2045, 2047, 2049, 2054, 2057, 2067, and 2071) (see Figure 1a). Of these events, only the eruptions in 2067 and 2071 produced stratospheric sulfur burdens exceeding 0.2 Tg of S. Given that these two eruptions occurred just 4 years apart, we selected a 2-year pre-eruption window for the composite analysis to minimize the effects from the earlier event. During this two-year pre-eruption period before the 2071 eruption, the stratospheric sulfur burden over Antarctica remained below 0.1 Tg of S (Figure 1b).

9. Fig. 3: why do you only show comparisons for Antarctica in October? How different are they for annual mean Antarctica? How other regions look like in October? Besides, observed annual mean SH mid-latitudes O3 is higher compared to modelled one, does it have a stronger SH mid-latitude O3 depletion (whether in Oct.?) due to hinder of aerosols transport to Antarctica as written in L276-278? These can be used to understand the inconsistency between observations and model.

   In the below plot, we have added the Antarctic annual mean total column ozone (solid line) for comparison with the Antarctic October mean total column ozone (dotted line). The October mean Antarctic total column ozone has a greater total column ozone loss than the annual mean from around 1980 onwards.

   We have also included the October mean SH-mid latitude total column ozone (in dotted line) for comparison with the annual mean SH-mid latitude values. The model-simulated October mean SH mid-latitude total column ozone is greater than the satellite observed values. The October mean Antarctic and SH mid-latitude total column ozone after the 1991 Mt. Pinatubo eruption (the grey vertical line denotes year 1991) shows a great reduction of 19-20 DU in year 1992. However, the UKESM1.1 simulation results shown in the figure below use

the prescribed volcanic forcing from CMIP6 rather than interactive volcanic sulfur injections. The prescribed forcing approach does not explicitly simulate the transport of volcanic aerosols, and therefore we cannot directly assess whether differences in aerosol transport between Antarctica and SH mid-latitudes contribute to the observed model-observation discrepancies. We have revised the figure (now Figure 2) in the main text to show the annual mean Antarctic and Arctic total column ozone instead of the October and March mean values.



Annual mean total column ozone (in DU) averaged (a) globally (90°S to 90°N), over the (b) tropics (20°S to 20°N), (c) Antarctica (60°S to 90°S, annual mean in solid line, October mean in dotted line), (d) Arctic (60°N to 90°N, annual mean in solid line, March mean in dotted line), (e) SH mid-latitudes (30°S to 60°S, annual mean in solid line, October mean in dotted line), and (f) NH mid-latitudes (30°N to 60°N). The grey vertical line denotes the year of the 1991 Mt. Pinatubo eruption. The blue lines represent the UKESM historical simulations from 1850 to 2014, and the red markers represent satellite measurements from 1978 to 2023. The blue shading shows the maximum and minimum values of the ensemble members.

10. Fig. 4 and 5: Here only shows model inter-comparisons, why not comparing with observations, like that shown in Keeble et l., 2021? And what these model differences mean?

We compared our results with the ML-TOMCAT dataset, which provides a long-term chemically-consistent ozone profile data with a good agreement with satellite observations. We have elaborated on the interpretation of the model differences in the text.

L407-411, "UKESM1.1 simulates higher ozone mass mixing ratio over the tropics and extratropical regions, lower ozone over Antarctic stratosphere from 1-10hPa, and higher ozone in the Antarctic lower stratosphere between 10-30 hPa compared to both ML-TOMCAT and the CMIP6 multi-model mean (Fig. S1d and e). This suggests that UKESM1.1 may overestimate ozone concentration over Antarctic lower stratosphere (10-30 hPa) relative to other CMIP6 models and ML-TOMCAT."

11. L253-254: what impacts does this modelled prolonged Antarctic O3 loss have on the O3 recovery results?

    The prolonged ozone loss will lead to a higher cumulative ozone loss for earlier decades. We have added this in the discussion section.

    L889-890, "Due to these model biases, our results likely overestimate the cumulative ozone loss over Antarctica and SH mid-latitudes."

12. Fig. 6: c) why TOC near 1992 is so different between NIWA-BS and UKESM?

    The NIWA-BS total column ozone anomaly presented in Figure 3 (previously Figure 6) is relative to the 1986-1990 climatology. Apart from the June 1991 Mt. Pinatubo eruption, another volcano in the Southern Hemisphere (Chile), Cerro Hudson, erupted in August 1991. The anomaly shown by GloSSAC and NIWA-BS reflects the responses of both eruptions. In Figure 3c, we see a negative anomaly of Antarctic TCO for NIWA-BS after June 1991, which changed to a positive anomaly towards the end of 1991. This shows that the onset of Antarctic ozone loss started much earlier in year 1991 than the 1986-1990 climatology. The TCO anomaly in UKESM simulations are relative to the control simulation without volcanic eruption. The TCO reduction occurs at the end of 1991 as the volcanic aerosols have a slower transportation to Antarctica than GloSSAC (Figure 3a and 3b). The magnitude of the peak Antarctic TCO loss between NIWA-BS and UKESM in 1991 is comparable.

13. Fig. 6 f) there is a huge positive anomaly between UKESM and NIWA-BS in NH mid to polar latitudes, does this mean a shortened impact of volcanic eruptions on NH O3? Above, a modelled prolonged Antarctic O3 loss was mentioned, why? Understand these might be quite useful to understand the ability of model simulation on O3. Besides, how to understand their impacts on the interpretation of the modelled O3 recovery results?

    Our 1991 Mt. Pinatubo simulations represent the ensemble mean of free-running

simulations, which are not nudged to observations and therefore do not replicate the specific meteorological conditions from 1991 to 1993. The ozone response to volcanic eruptions in polar regions is highly sensitive to stratospheric aerosol transport, which depends on atmospheric circulation patterns that vary between our model climatology and the actual conditions.

The prolonged ozone hole in Antarctica is attributed to the stratospheric cold bias and strong polar vortex in UKESM. A stronger and more persistent polar vortex in UKESM extends the ozone loss period over Antarctica and thus leads to a prolonged ozone hole. These model biases likely leads to an overestimation of cumulative ozone loss over Antarctica and SH mid-latitudes. We have elaborated the potential impact of the model biases on the ozone recovery results in the discussion.

L887-891, "We acknowledge that UKESM1.1 exhibits a stratospheric cold bias and excessively strong polar vortex that leads to prolonged ozone depletion over Antarctica (Fig. 4). The comparison of the UKESM1.1 simulation with ML-TOMCAT shows that the climatological mean of UKESM1.1 overestimates lower stratospheric ozone loss over Antarctica and SH mid-latitudes. Due to these model biases, our results likely overestimate the cumulative ozone loss over Antarctica and SH mid-latitudes. However, the relative volcanic effects on ozone in our simulated scenarios remain robust."

14. L276-278: may hinder the transport? Where is the assessment based on the data used?

    UKESM is known to have a stratospheric cold bias and strong polar vortices (Sellar et al., 2019; Hall et al., 2021), which likely limits the transport of volcanic aerosols into the Antarctic stratosphere. We have clarified the sentence in the main text.

    L519-521, "Since the stratospheric cold bias and strong winter polar vortex in UKESM (Sellar et al., 2019; Hall et al., 2021) likely hinder the transport of volcanic aerosols into the Antarctic stratosphere, we assess the ozone responses over both Antarctica and SH mid-latitudes."

    Added reference:

    Hall, R. J., Mitchell, D. M., Seviour, W. J. M., and Wright, C. J.: Persistent Model Biases in the CMIP6 Representation of Stratospheric Polar Vortex Variability, JGR Atmospheres, 126, e2021JD034759, https://doi.org/10.1029/2021JD034759, 2021.

15. L281-285: Many boxes do not show a delayed recovery and even an advanced recovery, thus looks quite different to the texts written here. It seems all experiments show a delayed recovery based on O3 mass deficit (175 DU) except for VOLC50-1, but O3 mass deficit (220 DU) show quite contradictory results for

different experiments, why it's so different when using different threshold?

The delay of return years for ozone mass deficit is most sensitive to aerosol perturbations during the years before the ozone mass deficit return to their respective thresholds. Please see the response to comment 6 for details. We have elaborated in the discussion section as follows,

L739-748, "For the 175 DU threshold, VOLC50-2 shows the longest delay (6 years), while VOLC2.5 and VOLC98 both show 3-year delays, and VOLC50-1 has the same return year as NOVOLC. The apparent inconsistencies in the delays of recovery across scenarios arise from the different recovery timelines, which is due to differences in the timing of volcanic eruptions across scenarios, as well as varying sensitivities of the two ozone mass deficit thresholds. The recovery for the 175 DU threshold is much faster (median year of 2038 in NOVOLC) than the 220 DU threshold (median year of 2062 in NOVOLC), making each metric sensitive to eruptions occurring at different times. For instance, VOLC2.5 and VOLC98 both have large-magnitude eruptions with SH aerosol distribution before 2062, thus showing greater delay in return years for the 220 DU threshold. VOLC50-2 shows the longest delay for the 175 DU threshold because it has higher SH SAOD than VOLC2.5 and VOLC98 before 2038 (Fig. 6). In contrast, the first large-magnitude eruption in VOLC50-1 occurs in 2047, after the 175 DU threshold recovery year, and thus showing no delay for this metric."

16. Fig. 8: 3-year moving mean plots are shown, but none of the texts mentioned these subplots. The difference between control and experiments are not that clear with 3-year moving mean. What does these subplots mean when compared to the 30-year moving mean results?

We present both 3-year and 30-year moving means to show the interannual variability from individual volcanic events and the long-term trends in volcanic effects on ozone recovery. We have revised the results section as follows,

L562-566, "Figure 5a and 5b shows the October-mean ozone mass deficit over Antarctica with 220 DU and 175 DU thresholds respectively, shown as 30-year and 3-year moving means. The 3-year moving mean reveals substantial interannual variability in ozone mass deficit across scenarios, driven by individual eruptions and internal variability. When smoothed with a 30-year moving mean to show longer-term trends, all stochastic scenarios consistently exhibit higher ozone mass deficit compared to NOVOLC prior to their respective return years (Fig. 5a and 5b)."

17. L306-307: any model data analysis to confirm this?

CFC concentrations in the model follow the SSP3-7.0 scenario from ScenarioMIP (Meinshausen et al., 2020), which projects continued decline in atmospheric CFCs consistent with the Montreal Protocol. The reduction in stratospheric ClOx over Antarctica (Figure S6) reflects the corresponding

decrease in stratospheric chlorine loading from declining CFCs.

18. L311-312: but O3 mass deficit (220 DU and 175 DU) in table 2 also shows later years for VOLC2.5 compared to NOVOLC.

    We have elaborated the responses of the ozone return years in the results and discussion. Please see the response to comment #15.

19. L317-318: October Antarctica (Fig. 8) vs annual mean global and mid-latitude (Fig. S1), Is the comparison reasonable?

    We agree with the reviewer that comparing October mean Antarctic TCO values with annual mean values for other regions is not ideal. We have revised Figure S3 (previously Figure S1) to show Antarctic total column ozone as an annual mean (with 30-year moving average applied) for consistent comparison with the global and SH mid-latitude annual mean total column ozone. The text has been revised as follows:

    L610-612, "The magnitude of response in global and SH mid-latitude total column ozone is consistently negative throughout this century and comparable to or slightly greater than that over Antarctica (+0.1% to -0.5%, Fig. S3d), where the response shows greater variability."

20. L324-325: is this event dominated the O3 impact and the delayed O3 recovery result?

    The clusters of large-magnitude eruptions in VOLC98 between 2035 and 2060 collectively contribute to the delayed ozone recovery. The delay in ozone mass deficit with a 175 DU threshold is driven by the large-magnitude eruptions occurring before 2038 (the median return year of NOVOLC), rather than by the single event in year 2056.

21. L327-328 the Antarctic O3 hole area is highly variable between ensemble members, then more members are need to study this?

    We agree with the reviewer that the results of Antarctic ozone hole area would benefit from more ensemble members to account for its variability. Our model simulations are limited by the availability of computing resources since it is a fully-coupled Earth System Model. We suggest future studies can design experiments with additional ensemble members to quantify the uncertainty of Antarctic ozone chemistry and the ozone return years. We have revised the discussion section as follows,

    L758-759, "Future modelling experiments with larger ensemble sizes would further quantify the uncertainty in volcanic effects on ozone recovery and distinguish between volcanic forcing effects and internal variability."

22. L320-331: this paragraph describes Fig. 9, but only a few sentences are based on Fig. 9. It looks like Fig. 9 does not show a clear difference between VOLC and NOVOLC, except for VOLC98. However, VOLC98 has several large eruptions in 2040-2060, the delayed recovery can be just due to these eruptions especially the 2056 eruption with 114 Tg SO2 emissions.

    We agree with the reviewer's comment that the delayed recovery of the Antarctic ozone hole area largely depends on the timing of the large-magnitude eruptions. Our simulations show that the delay in ozone recovery is largely due to the occurrence of large-magnitude eruptions before the ozone recovery metrics recover to their respective threshold and baseline. We have revised the discussion section (L739-748) accordingly to highlight this point.

23. L348: refer to Fig. 2b.

    We have revised in the text (now Figure 1b).

    L659-660, "To assess the volcanic effects on the chemical loss of ozone, we select 9 large-magnitude eruptions with stratospheric sulfur burden > 0.2 Tg of S over Antarctica for composite analysis (Figure 1b)."

24. Section 3.3 and Fig. 10: it's not clear what's the connection of these content to the main conclusion of the delayed O3 recovery and comparison between Antarctica and mid-latitude changes. Does different response along time due to background O3 and halogen or other chemical family changes in different latitudes?

    Figure 10 (now Figure 7) shows the changes in Antarctic and SH mid-latitude stratospheric chemistry following the large-magnitude eruptions with over 0.2 Tg of sulfur over Antarctica. As the atmospheric chlorine decreases in the century, the ozone loss associated with the halogen cycles is expected to decrease as well (Figure 7c). We show that the ozone loss response per Tg of sulfur injection over the respective latitudes is stronger in Antarctica than SH mid-latitudes. We expect that eruptions occurring after year 2070 will eventually lead to a net ozone gain instead of ozone loss. Such shifting in the chemistry will affect the future ozone recovery trend.

    L855-857, "We expect that eruptions occurring after year 2070 will eventually lead to a net ozone gain instead of ozone loss. Such shifting in the chemistry will affect the future ozone recovery trend."

25. L378: "depends on the eruption timing, latitude and aerosol distribution in the stochastic scenarios", the figures do not provide support on this conclusion.

    We have elaborated further regarding this point in the discussion section.

    L735-748, "We find that the extent of the delay in Antarctic ozone return dates

depends on the eruption timing, latitude and aerosol distribution in the stochastic scenarios, with two scenarios showing delays in ozone mass deficit of 2 years (VOLC2.5) and up to 5 years (VOLC98) with a 220 DU ozone threshold (Table 1). If we use a lower ozone threshold of 175 DU, the relative effect of volcanic eruptions on ozone mass deficit can be up to 23% in 2030 for a high-end future scenario (Fig. 5b). For the 175 DU threshold, VOLC50-2 shows the longest delay (6 years), while VOLC2.5 and VOLC98 both show 3-year delays, and VOLC50-1 has the same return year as NOVOLC. The apparent inconsistencies in the delays of recovery across scenarios arise from the different recovery timelines, which is due to differences in the timing of volcanic eruptions across scenarios, as well as varying sensitivities of the two ozone mass deficit thresholds. The recovery for the 175 DU threshold is much faster (median year of 2038 in NOVOLC) than the 220 DU threshold (median year of 2062 in NOVOLC), making each metric sensitive to eruptions occurring at different times. For instance, VOLC2.5 and VOLC98 both have large-magnitude eruptions with SH aerosol distribution before 2062, thus showing greater delay in return years for the 220 DU threshold. VOLC50-2 shows the longest delay for the 175 DU threshold because it has higher SH SAOD than VOLC2.5 and VOLC98 before 2038 (Fig. 6). In contrast, the first large-magnitude eruption in VOLC50-1 occurs in 2047, after the 175 DU threshold recovery year, and thus showing no delay for this metric."

26. L387: why loading so small but comparable O3 hole size?

    The background CFC concentration is much higher in 2015 than in 2056. With higher CFC concentration in the atmosphere, there are more reactive halogen radicals available for ozone-depleting reactions on volcanic aerosol surfaces. Therefore, even relatively small stratospheric sulfur injections can lead to significant ozone loss. By 2056, atmospheric CFC levels have declined substantially following the Montreal Protocol, limiting the availability of reactive halogens. Consequently, even a large-magnitude eruption with 114 Tg SO_2 injection (in year 2056) produces less ozone depletion than comparable eruptions during the high-CFC period, as the reduced halogen loading in the atmosphere leads to less ozone loss via halogen cycles.

    Line 830-832, "The comparable ozone hole size is attributable to the substantially higher stratospheric chlorine loading in 2015 compared to that projected for the mid-21st century."

27. L391: is it accurate to say “contrast”, as the settings are very different. In this study, the delayed O3 recovery can be dominated by a huge eruption at that time, but Naik et al., 2017 used constant emissions.

    We agree with the reviewer that the wording might be misleading. We have changed the wording in the text.

    L834, "Our results differ from the findings of Naik et al. (2017), ..."

28. L392-393: how much loading in the Antarctica? This is important for the O3 change in Antarctica.

    The future volcanic forcing is prescribed with the stratospheric aerosol optical properties derived from the historical mean in Naik et al. (2017). They did not implement direct volcanic SO_2 emission in their model design.

29. L395-396: “leads to an earlier recovery of global stratospheric O3”, this seems to be the case in your VOLC2.5 scenario.

    For the VOLC2.5 scenario, it shows a 1-year earlier return in Antarctic October mean total column ozone than NOVOLC. For the ozone mass deficit with 175 DU and 200 DU thresholds, the VOLC2.5 scenario shows 2 to 3 years delay in the return year. The return year is sensitive to the eruption timing and aerosol distribution.

30. L396-397: can be wrong, as the difference can be dominated by a large difference in Antarctica aerosol loading mass.

    Naik et al. (2017) used a prescribed historical-mean volcanic forcing in their model simulation. We agree that the differences between our result and Naik et al. (2017) can arise from the differences in forcing magnitudes. We have revised the text as follows,

    L839-840, "The disagreement between Naik et al. (2017) and our study potentially arises from differences in forcing magnitude, model responses and chemistry representations between GFDL-CM3 and UKESM."

31. L400-402: then is this model version appropriate for this study?

    At the time this study was initiated, we used UKESM1.1 with the standard heterogenous chemistry scheme for the simulations. Our 1991 Mt. Pinatubo simulations in Section 3 demonstrates that UKESM1.1 adequately captures the magnitude of volcanic-induced total column ozone response over Antarctica, which is appropriate for our comparative analysis of different volcanic forcing scenarios. We acknowledge that implementing improved heterogeneous chemistry could enhance the simulated magnitude of volcanic ozone depletion, and future work incorporating these improvements in fully coupled simulations would help reduce uncertainties in ozone recovery projections.

32. L424-425: different sign in the text and in Fig. S10, makes it harder to understand.

    We have reversed the signs in the main text to avoid confusion.

    L876-879, "Composite analysis of 15 selected eruptions (with a stratospheric

sulfur burden exceeding 0.5 Tg of S over the NH mid-latitudes) reveals substantial net ozone loss of 9 DU over the Arctic (interquartile range: 18 DU to -2 DU) and 2 DU loss over the NH mid-latitudes (interquartile range: 6 DU to -3 DU) (Figures S12)."

33. L436-437: already written in the introduction. Volcanic halogen emissions can lead to significant O3 depletion and delayed O3 recovery, there are already modelling studies simulated co-emission of halogen with sulfur, which should be discussed more in depth. Besides, modelling studies on Hunga eruption (like Fleming et al., 2024, JGRA; Zhu et al., 2022, ACP; Zhuo et al., 2025, ACP) showed significant O3 impact from water vapor emissions, this is also very important for understanding the delayed O3 recovery but is completely missing in the discussion.

    We have revised the introduction and discussion to include the discussion on volcanic halogen and water vapour (see response to major comment).

34. L461-463: not clear where does it show in the paper?

    The weakening of peak ozone loss response following volcanic eruptions over time is shown in Figure 7c, which demonstrates that the magnitude of peak Antarctic ozone response decreases until year 2070. We have added the reference to the figure, and revised the sentence for clarity.

    L1179-1181, "Our results also showed that the peak ozone loss due to large-magnitude volcanic eruptions in our stochastic scenarios weakens from 2015 to 2070 (Fig. 7c) as a result of declining CFC concentrations."