Reappraisal of the climate impacts of ozone-depleting substances

Olaf Morgenstern[1], Fiona M. O'Connor[2], Ben T. Johnson[2], Guang Zeng[1], Jane P. Mulcahy[2], Jonny Williams[1], João Teixeira[2], Martine Michou[3], Pierre Nabat[3], Larry W. Horowitz[4], Vaishali Naik[4], Lori T. Sentman[4], Makoto Deushi[5], Susanne E. Bauer[6,7], Kostas Tsigaridis[7,6], Drew T. Shindell[8], Douglas E. Kinnison[9]

[1]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand
[2]Hadley Centre, Met Office, Exeter, UK
[3]Centre National de Recherches Météorologiques (CNRM), Université de Toulouse, Météo-France, CNRS, Toulouse, France
[4]Geophysical Fluid Dynamics Laboratory (GFDL), National Oceanic and Atmospheric Administration, Princeton, NJ, USA
[5]Meteorological Research Institute (MRI), Tsukuba, Japan
[6]NASA Goddard Institute for Space Studies (GISS), New York, NY, USA
[7]Center for Climate Systems Research, Columbia University, New York, NY, USA
[8]Earth & Ocean Sciences, Duke University, Durham, NC, USA
[9]National Center for Atmospheric Research (NCAR), Boulder, CO, USA

Key Points:

- Effective radiative forcing of ozone-depleting substances, as discerned from CMIP6 models, spans a large range.
- We use an Emergent Constraint technique, relating this efffective radiative forcing to ozone changes, to come up with a new range consistent with observational climatologies of ozone depletion.
- This range implies a larger impact of ozone depletion on the effective radiative forcing than the previous best estimate.

Corresponding author: Olaf Morgenstern, olaf.morgenstern@niwa.co.nz

Abstract

We assess the effective radiative forcing due to ozone-depleting substances using models participating in the Aerosols and Chemistry Model Intercomparison Project (AerChemMIP). A large inter-model spread in this globally averaged quantity necessitates an "emergent constraint" approach whereby we link the radiative forcing to the amount of ozone depletion simulated during 1979-2000, excluding two volcanically perturbed periods. During this period ozone-depleting substances were increasing, and several merged satellite-based climatologies document the ensuing decline of total-column ozone. We use these analyses to come up with effective radiative forcing magnitudes. For all of these satellite climatologies we find an effective radiative forcing outside or on the edge of the previously published "likely" range given by the 5$^{\text{th}}$ Assessment Report of IPCC, implying an offsetting effect of ozone depletion and/or other atmospheric feedbacks of -0.4 to -0.25 Wm^{-2}, which is in absolute terms is larger than the previous best estimate of -0.15 Wm^{-2}.

Plain Language Summary

Chloroflourocarbons and other compounds involved in ozone depletion are also powerful greenhouse gas, but their contribution to global warming is reduced due to the cooling effect of the ozone loss which they induce. Models informing an upcoming climate report disagree on the ozone loss and thus on the climate influence of these gases. Here we use observed ozone loss to reduce the resultant uncertainty in their overall climate influence and infer a larger cooling influence of ozone loss than was previously considered. The result implies a smaller benefit to climate of the phase-out of these ozone-depleting substances, mandated under the Montreal Protocol, than would have been the case under previous understanding.

1 Introduction

The Antarctic ozone hole remains arguably the most spectacular demonstration of human interference with the climate system. Within a few decades of starting to use chloro-fluorocarbons (CFCs) and other halocarbons on an industrial scale, humans had thinned the ozone layer above Antarctica in spring to a fraction of its prior thickness (WMO, 2018). This ozone depletion has had a substantial impact on the circulation and climate of the Southern Hemisphere (Velders et al., 2007; Myhre et al., 2013; Shindell et al., 2013). In 1987 the Montreal Protocol was enacted and subsequently strengthened which mandates a phase-out of these ozone-depleting substances (ODSs). It has been hailed as the most successful international treaty ever to protect the environment. Several of these ODSs also act as greenhouse gases, with global warming potentials many 1000's of times larger than that of carbon dioxide (WMO, 2018). By phasing them out, it is thought that the Montreal Protocol has averted more global warming than the Kyoto Protocol (Velders et al., 2007) despite not being designed for this purpose. The 5$^{\text{th}}$ Assessment Report of IPCC (AR5, Myhre et al., 2013) estimates that in 2011 the ODSs regulated by the Montreal Protocol exerted a direct radiative forcing of 0.34 Wm^{-2} and an indirect radiative forcing due to ozone depletion of -0.15 ± 0.15 Wm^{-2} relative to pre-industrial times. Accordingly, AR5 assesses that ODSs "very likely" have a positive radiative forcing as their direct radiative effects outweigh the indirect impact of ozone depletion.

The AR5 assessment was based essentially on one model (Shindell et al., 2013); also there is a large uncertainty associated with the negative radiative forcing due to ozone depletion. Furthermore, AR5 used the "stratospherically adjusted radiative forcing" concept, whereby any atmospheric adjustments other than a temperature adjustment of the stratosphere are not considered. AR5 stipulated that the effects of these were minor. These issues motivate a reassessment using new models and more developed methodologies. AerChemMIP (a subsidiary of the 6$^{\text{th}}$ Coupled Model Intercomparison Project, CMIP6; Collins et al., 2017) is a coordinated effort to re-assess these radiative forcing processes using

newer models and taking into account all atmospheric adjustments, arriving at the "effective radiative forcing" (ERF, Forster et al., 2016). In particular, here we will use simulations submitted under the "piClim-control" and the "piClim-HC" experiments. Both are atmosphere-only simulations driven with preindustrial sea-surface conditions and all atmospheric forcings at preindustrial levels, except that in the case of piClim-HC surface abundances of halogenated ODSs (short: halocarbons, HCs) compounds are elevated to their 2014 mean surface volume mixing ratios (Meinshausen et al., 2017). The difference in the global, multi-annual mean net top-of-the-atmosphere radiation between this pair of simulations defines the ERF due to ODSs. The problem, detailed below, is that this approach yields large inter-model differences for this ERF. This is a familiar situation in climate modelling. The "equilibrium climate sensitivity" (ECS, the equilibrium warming of the planet for a doubling of CO_2) is another well-studied example of similarly large inter-model spread (for a recent discussion see Zelinka et al., 2020). A technique to deal with such model disparities is known as an "emergent constraint" (e.g., Williamson & Sansom, 2019). This consists of relating a theoretical concept (such as the ECS or the ERF of ODSs, the topic of this paper) to a different, physically related quantity for which good-quality historical observational data exist. To constrain the ERF of ODSs, we evaluate here the choice of total-column ozone (TCO) measured comprehensively by satellite and ground-based instruments since the latter decades of the 20$^{\text{th}}$ century when ODSs were sharply increasing over time and ozone depletion was established. The remainder of this paper is devoted to applying this approach to AerChemMIP and "historical" (Eyring et al., 2016) simulations to arrive at a recommended value for the ERF of ODSs.

2 Models

We use here six different climate models, listed in table 1. The two GISS models differ in their coupled ocean models but have identical atmosphere and other sub-models.

Model name	Reference	Historical simulations
CESM2-WACCM	Gettelman et al. (2019)	1, 2, 3
CNRM-ESM2-1	Séférian et al. (2019)	1 to 5, 8, 9, 10
GFDL-ESM4	Dunne and et al. (2020)	1
GISS-E2-1-G/H	Kelley et al. (2019)	6, 8, 9, 10
MRI-ESM2-0	Yukimoto et al. (2019)	1 to 5
UKESM1-0-LL	Sellar et al. (2019)	1 to 4, 8 to 12, 16 to 19

Table 1. Models, key references, and "historical" simulations denoted by their run numbers.

The models are chosen because (a) all of them have explicit stratospheric chemistry, and (b) required data from these models is available for the piClim-control, piClim-HC, and historical experiments. Here we use total-column (or vertically resolved ozone, in the case of CESM2-WACCM) and outgoing short- and longwave radiation at the top of the atmosphere, all in monthly-means. The piClim-control and piClim-HC experiments all are one-member ensembles; the last 20 years of these simulations are used. We have established that the halocarbon increase in the piClim-HC simulation versus piClim-control has been fully communicated to the ozone layer and the ensuing ozone depletion is fully realized (figure S1).

In addition to these model data, we use five different ozone climatologies listed in table 2. Here we only consider data from the period 1979-2000 when ozone depletion was established. Four of these datasets are based on satellite measurements (instruments such as the Solar Backscatter Ultraviolet Radiometer (SBUV) series are used in all of them) but may use different versions and/or combine them in different ways with ground-based measurements to account for offsets between different overlapping satellite timeseries,

data gaps, drifts, and other instrumental artefacts. Details are in the references given (table 2). Furthermore, for reference we also consider total-column ozone derived from the recommended CMIP6 ozone forcing dataset (Checa-Garcia et al., 2018) used by CMIP6 models that do not compute their own ozone (i.e. not the six models listed in table 1). Referred to here as "CMIP6", this is not an observational dataset; rather it is derived from two historical chemistry-climate model simulations including one by CESM1-WACCM, an older version of the CESM2-WACCM model figuring in this study. In all references and in the "historical" simulations we remove two years each after the major eruptions of El Chichón (March 1982) and Mt Pinatubo (June 1991) although retaining these data would only have a small influence on our results.

Dataset	Coverage	Resolution	Reference
TOMS-SBUV v8	1978-2005	Zonal-mean, 5°	Frith et al. (2014)
SBUV v8.6	1970-2018	Zonal mean, 5°	Frith et al. (2014)
NIWA-BS (v3.4, unpatched)	1978-2016	$1° \times 1°$	Bodeker et al. (2005)
MSR-2	1979-2018	$0.5° \times 0.5°$	van der A et al. (2015)
CMIP6	1850-2014	$2.5° \times 1.5°$	Checa-Garcia et al. (2018)

Table 2. Four observational TCO climatologies and the reference ozone field used to force CMIP6 models without interactive ozone.

3 Results

We calculate trends in TCO for the period 1979-2000 for the five reference dataset and the six models (figure 1) excluding two years each after the El Chichón and Mt Pinatubo eruptions in 1982 and 1991. There are mostly relatively small differences between the observational datasets, as expected. The Multi-Sensor Reanalysis 2 (MSR-2) dataset uses data assimilation in a chemistry-transport model (van der A et al., 2015). As such it has a more comprehensive coverage than the other climatologies. It reveals substantial ozone loss in the Arctic peaking in March when the other three observational datasets have data gaps. It has slightly weaker Antarctic ozone depletion than the NIWA-Bodeker Scientific (NIWA-BS), SBUV, and TOMS-SBUV climatologies. MSR-2 has the winter polar observational gaps characteristic of satellite measurements largely or completely filled in (in the case of the Arctic), in contrast to NIWA-BS, SBUV, and TOMS-SBUV. The SBUV climatology does not have any data poleward of 80°N/S, i.e. it has the most restricted high-latitude coverage of the datasets considered here. TOMS-SBUV is very similar to the SBUV dataset but in summer does not have data gaps over both polar caps.

The CMIP6 ozone climatology compares well with the observations in the Southern Hemisphere but underestimates the decline in Northern-Hemisphere ozone. This is most evident in comparison to MSR-2 but can also be seen versus the other three datasets.

The six climate models exhibit highly variable Antarctic ozone trends in their ensemble-average historical simulations (figure 1), ranging from quite weak (MRI-ESM2-0) to extremely strong (GISS-E2-1-G and GISS-E2-1-H). Also in the Arctic, several models exhibit weak trends, with the exception of UKESM1 which exhibits excessive ozone depletion in both polar regions.

We next address the extent to which these historical ozone trends represent ozone loss attributable to ODSs. Considerations here include that at least in models, and likely in reality, some reductions in stratospheric ozone occurred before the onset of comprehensive satellite measurements in 1978 (figure S2, Dhomse et al., 2018), that the actual ozone depletion occurred in an atmosphere with the methane loading increased above

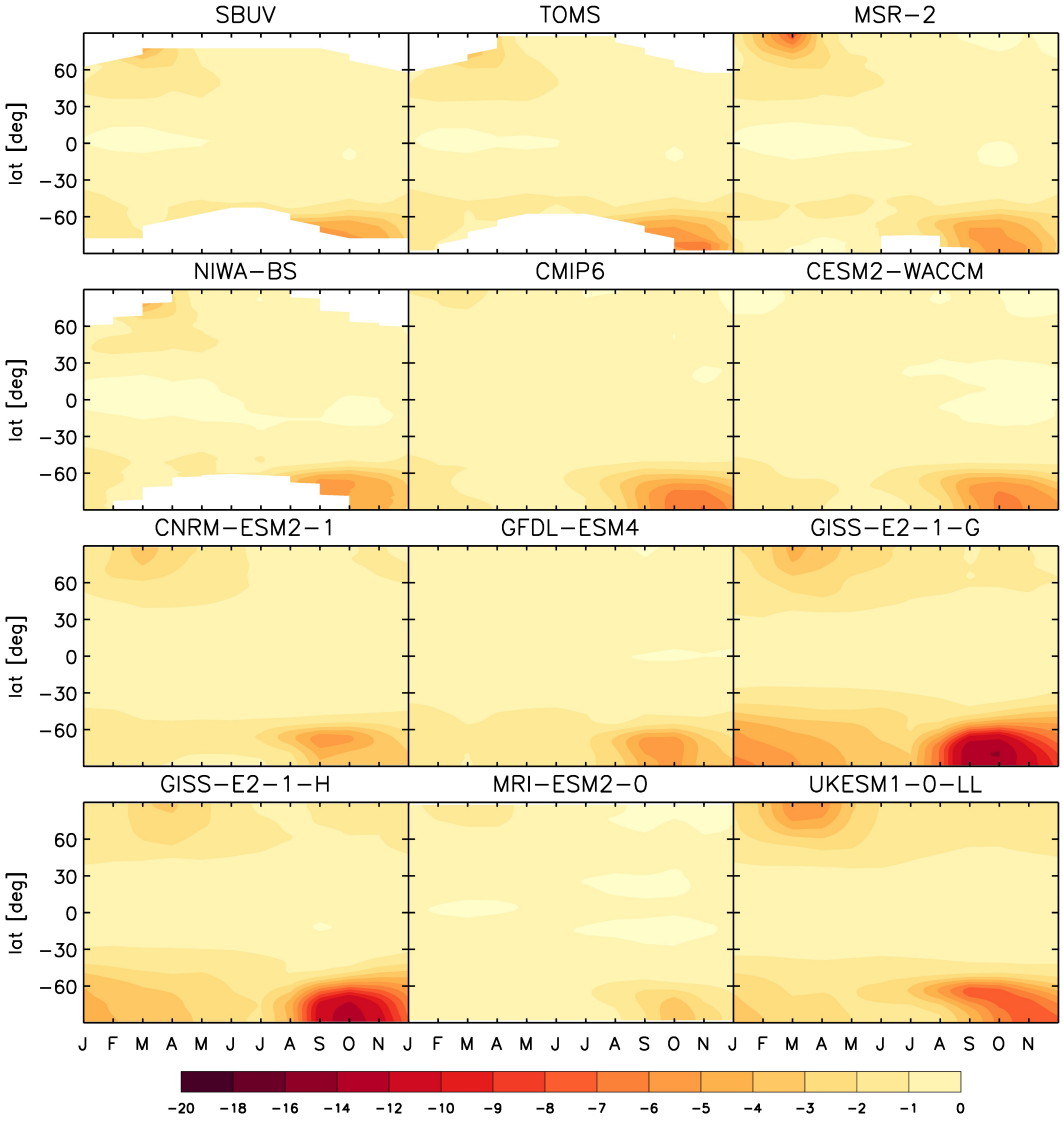

Figure 1. Zonal-mean TCO trends (Dobson Unit/year, DU a^{-1}) for 1979-2000, excluding two years each after the El Chichón and Pinatubo eruptions, in four observational datasets (SBUV, TOMS-SBUV, NIWA-BS, and MSR-2), the CMIP6 ozone climatology, and the historical simulations by the seven CMIP6 models considered here.

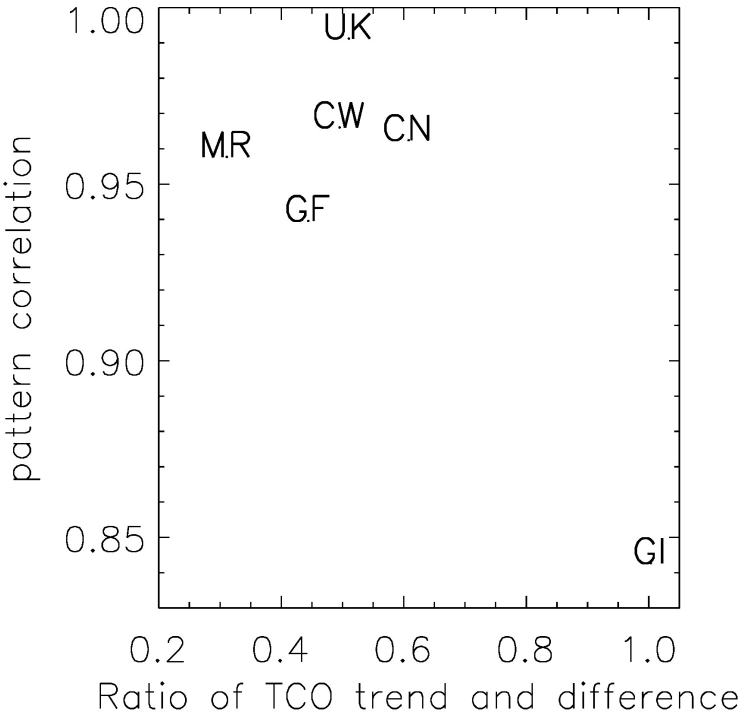

Figure 2. Abscissa: Ratio of area-weighted, global- and annual-mean ozone trend (DU a^{-1}) for 1979-2000 derived from the historical simulations, with two volcanically perturbed periods removed (figure 1) and the global- and annual-mean mean ozone difference (DU) between the piClim-HC and piClim-control experiments (figure S1) divided by 22 years. Ordinate: Pattern correlation between the zonal-mean ozone trends shown in figure 1 and the zonal-mean ozone differences between the piClim-HC and piClim-control experiments, for the six CMIP6 models. CN = CNRM-ESM2-1. CW- CESM2-WACCM. GF = GFDL-ESM4. GI = GISS-E2-1-G. MR = MRI-ESM2-0. UK = UKESM1-0-LL.

its preindustrial level which reduces the efficiency of chlorine at depleting ozone, and that stratospheric cooling since preindustrial times due to increasing greenhouse gases (GHGs) affects both polar and extrapolar ozone chemistry in different ways (Morgenstern et al., 2018). Single-forcing model simulations consistently show that the net effect of these GHGs increasing during the ozone-depletion period has been to mitigate some ozone depletion (Eyring et al., 2010; Dhomse et al., 2018). Figure S3 shows the difference of zonal-mean ozone between the piClim-control and piClim-HC experiments for the six models. A visual comparison to figure 1 indicates a high degree of pattern similarity for most models between the late-20$^{\text{th}}$ century ozone trends shown in figure 1 and the ozone loss attributable to ODSs. An exception to this is the GISS-E2-1-G model which has an anomalously large Southern-Hemisphere ozone trend associated with the formation of the ozone hole, but an ozone difference in figure S3 that is not anomalous relative to the other models.

Figure 2 displays a measure of the similarity between figures 1 and S3 for the six models. Five models cluster at a pattern correlation exceeding 0.94 and a trend to difference ratio of around 0.3 to 0.6 (i.e. the global ozone loss deduced from historical simulations between 1979 and 2000 is 30 to 60% of the difference in ozone between the piClim-control and piClim-HC simulations). GISS-E2-1-G has a somewhat lower correlation but a much larger trend ratio. Figure S2 shows that this model exhibits a much bigger response to volcanic eruptions than the other models (e.g. after the Krakatoa eruption of 1883) and enters into a volcanically disturbed phase of increased TCO in the 1970s. This implies that in this model, the 1979-2000 trend has a substantial contribution due to recovery from this volcanic perturbation. This manifests as an anomalous amplification of the influence of ODSs as the volcanic effect transitions from increasing TCO under low chlorine conditions to decreasing it during times of high chlorine loading. For this reason GISS-E2-1-G is not considered in the further analysis.

Figure 3 illustrates the "emergent constraint" relating simulated mean TCO loss for 1979-2000 (excluding volcanic periods) to the mean ERF of ODSs discerned from the piClim-control and piClim-HC experiments. The figure shows that this ERF closely relates to how much ozone depletion is simulated in these models. We capture this relationship through a least-squares linear regression line. Where this line intersects with the ozone depletion discerned from an observational climatology defines the ERF of ODSs that optimally corresponds to the ozone loss in that climatology. This process is illustrated in figure 3 for the example of the NIWA-BS climatology (similar plots for the other climatologies are in the supplement, figures S4 to S7). Here sampling the modeled TCO data in the same way as this climatology makes the modelled and observational climatologies directly comparable. Table 3 summarizes the ERFs discerned in this way using all TCO climatologies considered here as well as their means and standard deviations. A comparison with the straightforward multi-model mean (MMM) using the six models considered here (table 3) reveals that the multi-model mean is often outside the one-standard deviation uncertainty range spanned by the Emergent Constraint analyses, and that these analyses lead a considerably reduction of the uncertainty relative to forming the multi-model mean. A caveat in both cases is that the uncertainty analysis presented here does not capture some aspects of the total uncertainty such as sampling uncertainty (caused by the small number of models considered here) and model uncertainty (caused by structural and formulation problems inherent in models, which may be common to different models).

4 Discussion

Our analysis has shown that historical TCO trends for 1979-2000 simulated by CMIP6 full-chemistry models, with volcanic periods ignored, can serve as proxies for the total ozone loss caused by anthropogenic ODSs simulated by the same models. This enables us to conduct an "emergent constraint" analysis, relating this historical ozone loss to the

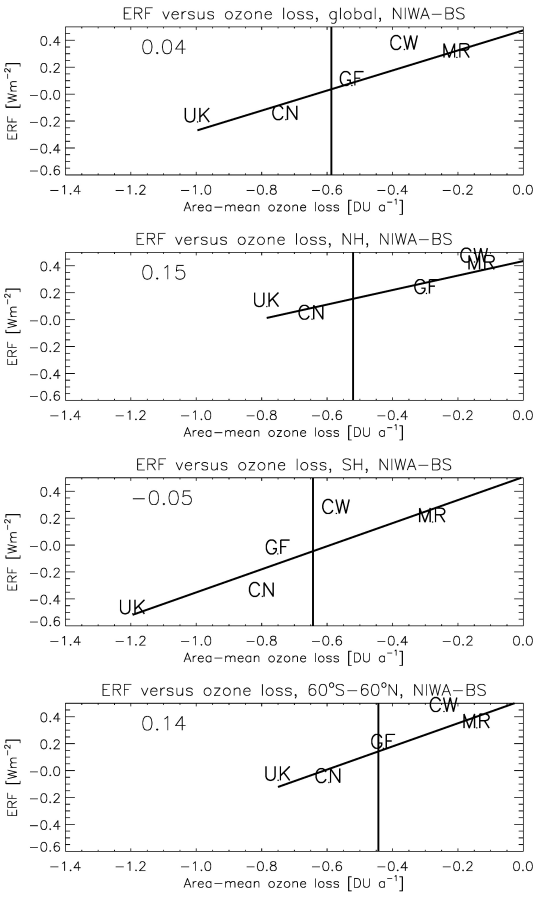

Figure 3. Abscissa: Area-weighted global- and annual-mean ozone loss (DU a^{-1}) for 1979-2000, discerned from the historical simulations of five CMIP6 models. Vertical line: Same, for the NIWA-Bodeker Scientific climatology. Ordinate: The ERF of ODSs as discerned from the piClim-HC and piClim-control simulations of these five models. Slanted line: linear regression. The four panels denote global, Southern-Hemisphere, Northern-Hemisphere, and 60°S to 60°N averages. The numbers inside the plots are the ERF values corresponding to the ozone loss discerned from the NIWA-BS climatology. CN = CNRM-ESM2-1. CW- CESM2-WACCM. GF = GFDL-ESM4. MR = MRI-ESM2-0. UK = UKESM1-0-LL. Equivalent plots for the other climatologies are in the supplement.

Climatology	90°S-90°N	0°-90°N	90°S-0°	60°S-60°N
CESM2-WACCM	0.31	0.43	0.19	0.41
CNRM-ESM2-1	−0.19	0.00	−0.38	−0.09
GFDL-ESM4	0.06	0.19	−0.07	0.16
GISS-E2-1-G	0.28	0.23	0.33	0.38
MRI-ESM2-0	0.30	0.40	0.20	0.34
UKESM1-0-LL	−0.19	0.12	−0.50	−0.06
MMM±σ	0.10±0.25	0.23±0.16	-0.04 ± 0.34	0.19±0.22
MSR-2	0.09	0.20	0.01	0.18
NIWA-BS	0.04	0.15	−0.05	0.14
SBUV	0.00	0.15	−0.12	0.09
TOMS-SBUV	−0.06	0.12	−0.20	0.03
MOM±σ	0.02 ± 0.06	0.16 ± 0.03	-0.09 ± 0.09	0.11 ± 0.06
CMIP6	0.13	0.32	−0.09	0.23

Table 3. ERFs of ODSs (Wm^{-2}). Top seven rows: Model results and multi-model mean (MMM) and standard deviation (σ). Bottom five rows: Results of the "emergent constraint" analysis and its mean ("multi-observations mean", MOM) and standard deviation.

ERF of ODSs simulated in five CMIP6 models. We find that indeed the ozone loss correlates with the ERF of ODSs, and that observed TCO loss provides the "emergent constraint" on an otherwise highly uncertain quantity. This analysis depends to some extent on the ozone depletion evident in the observational climatologies. The differences between them are small; they differ in their coverage of high-latitude and polar-night situations. However even if the analysis is restricted to extrapolar latitudes, removing the influence of such artefacts, some diffferences remain. Relative to the four observational climatologies, the ERF of ODSs falls into the range of −0.06 to +0.09 Wm^{-2}. Assuming now that the direct radiative forcing due to ODSs is 0.34 Wm^{-2} (Myhre et al., 2013), that means all other feedbacks (associated with ozone depletion, but also cloud and aerosol responses) cause an effective radiative forcing in the range of −0.4 to −0.25 Wm^{-2}. (As an aside, the CFC-12 equivalent mixing ratio of ODSs only dropped by 0.4% between 2011 and 2014; Meinshausen et al., 2017, meaning this direct radiative forcing is essentially unchanged during this period.)

Extrapolation of the regression line in figure 3 to zero ozone loss yields a global-mean ERF of ODSs, thus excluding the direct and indirect impacts of ozone depletion, of 0.45 to 0.55 Wm^{-2} in all cases. This is slightly larger than the AR5 estimate for the direct radiative forcing of ODSs of $0.34Wm^{-2}$ (Myhre et al., 2013), possibly because of additional feedbacks not included in the AR5 calculation, but also possibly highlighting that the "emergent constraint" analysis conducted here is based on only five models all using low spectral-resolution radiative transfer schemes, and hence is somewhat uncertain. (The uncertainty range stated above only includes the part of the uncertainty stemming from the different observational ozone climatologies, not any other, model-related uncertainty.)

For all ozone climatologies, the analysis suggests a near-zero or negative total ERF of ODSs in the Southern Hemisphere of between +0.01 and −0.20 Wm^{-2} (table 3), implying a cooling effect of ozone depletion and other feedback processes of between −0.33 and −0.54 Wm^{-2} (again based on the AR5 globally averaged direct radiative forcing estimate of 0.34 Wm^{-2} which we assume is also the Southern-Hemisphere average). For the Northern Hemisphere, in all cases the analysis yields positive ERFs, reflecting the

lesser role of ozone depletion here. If the analysis is restricted to the latitude range 60°S to 60°N, thus excluding the polar regions affected by variably large data gaps, the range of values (maximum − minimum) is the same (0.15 Wm^{-2}) as if the analysis is extended to all grid points with valid data in the observational references, suggesting that the mid- and low-latitude differences between the TCO climatologies play a significant role in driving the uncertainty in the global ERF.

Turning now to the CMIP6 ozone climatology (Checa-Garcia et al., 2018), the weak Northern-Hemisphere ozone depletion in this climatology, relative to observations, causes a smaller absolute ERF of ozone changes and hence a larger total global-mean ERF (0.13 Wm^{-2}) than would be consistent with the observations. In the Southern Hemisphere, ozone loss and thus the ERF of ODSs associated with this climatology (-0.09 Wm^{-2}) compare well with the observational estimates. The ERF associated with this climatology, for all four latitude regions, is close to but consistently smaller than the ERFs calculated for the CESM2-WACCM model (table 3). This behavior, paralleling CESM2-WACCM, is as expected as this model, in an older version, was one of two models used in the generation of the climatology (Checa-Garcia et al., 2018).

In summary, we find here a global-mean offsetting effect of ozone depletion on the ERF of ODSs which is only marginally overlapping the "likely" range given in AR5 of -0.3 to 0 Wm^{-2}. The best estimate of AR5, -0.15 Wm^{-2}, is outside the range derived here. Reasons for this difference may include (a) conceptual differences between the "stratospherically adjusted radiative forcing" evaluated in AR5 and the ERF evaluated here; (b) model differences, whereby the AR5 estimate was based on one model only (Shindell et al., 2013) – our analysis reveals a large model dependence of the result; and (c) methodological differences, whereby here we account for observational references using an novel "emergent constraint" approach. This approach turns large inter-model differences from a problem into an asset necessary for the regression analysis to become robust. A downside of the approach is that influences other than ODSs (such as increasing GHGs, tropospheric ozone pollution, or variations in solar output) have not been explicitly accounted for in the observational record (we have removed some volcanically affected periods), so there may be scope to further refine this analysis. In a follow-on publication we will endeavour to shed more light on the role of secondary feedbacks which in combination make up "effective radiative forcing" (namely chemical ozone depletion, cloud and temperature adjustments, and aerosol feedbacks). This analysis cannot be conducted based only on AerChemMIP simulations and hence is beyond the scope of this paper.

Acknowledgments

OM, FOC, BJ, GZ, JPM, JW, and JT acknowledge the UKESM1 team and other colleagues for their support, in particular, Luke Abraham, Paul Griffiths, Mohit Dalvi, James Manners, and Omar Jamil. OM and GZ were supported by the NZ Government's Strategic Science Investment Fund (SSIF) through the NIWA programme CACV. FOC, BJ, and JPM were supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra (Grant Number GA01101). JW acknowledges support by the Deep South National Science Challenge (DSNSC), funded by the New Zealand Ministry for Business, Innovation and Employment (MBIE). The authors acknowledge the contribution of NeSI high-performance computing facilities to the results of this research. New Zealand's national facilities are provided by the New Zealand eScience Infrastructure (NeSI) and funded jointly by NeSI's collaborator institutions and through MBIE's Research Infrastructure programme. MM and PN particularly acknowledge the support of the entire team in charge of the CNRM climate models, and especially that of Antoinette Alias and Laurent Franchisteguy for their technical assistance. Supercomputing time was provided by the Météo-France/DSI supercomputing center. Makoto Deushi was partly supported by JSPS KAKENHI grant no. JP20K04070. The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work sup-

ported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR.

We thank Bodeker Scientific, funded by the DSNSC, NASA, and KNMI for providing the total-column ozone databases. Data used in this study are all publicly available at `https://acd-ext.gsfc.nasa.gov/Data_services/merged/data/additional_files/toms_sbuv.v8.mod_v3.78-05.za.rev2.txt`, `https://acd-ext.gsfc.nasa.gov/Data_services/merged/data/sbuv_v86_mod.int_lyr.70-18.za.r7.txt`, `http://www.bodekerscientific.com/data/total-column-ozone`, `http://www.temis.nl/protocols/O3global.html`, and the CMIP6 portals (`https://esgf-node.llnl.gov/search/cmip6/` and `https://esgf-node.llnl.gov/search/input4mips/`).

References

Bodeker, G. E., Shiona, H., & Eskes, H. (2005). Indicators of Antarctic ozone depletion. *Atmospheric Chemistry and Physics*, *5*, 2603–2615. doi: 10.5194/acp-5-2603-2005

Checa-Garcia, R., Hegglin, M. I., Kinnison, D., Plummer, D. A., & Shine, K. P. (2018). Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophysical Research Letters*, *45*(7), 3264-3273. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076770` doi: 10.1002/2017GL076770

Collins, W. J., Lamarque, J.-F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., ... Smith, S. J. (2017). AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geoscientific Model Development*, *10*(2), 585–607. Retrieved from `https://www.geosci-model-dev.net/10/585/2017/` doi: 10.5194/gmd-10-585-2017

Dhomse, S. S., Kinnison, D., Chipperfield, M. P., Salawitch, R. J., Cionni, I., Hegglin, M. I., ... Zeng, G. (2018). Estimates of ozone return dates from Chemistry-Climate Model Initiative simulations. *Atmospheric Chemistry and Physics*, *18*, 8409–8438.

Dunne, J. P., & et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM4.1): Model description and simulation characteristics. *in prep.*.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved from `https://www.geosci-model-dev.net/9/1937/2016/` doi: 10.5194/gmd-9-1937-2016

Eyring, V., Cionni, I., Bodeker, G. E., Charlton-Perez, A. J., Kinnison, D. E., Scinocca, J. F., ... Yamashita, Y. (2010). Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models. *Atmospheric Chemistry Physics*, *10*, 9451–9472. Retrieved from `https://doi.org/10.5194/acp-10-9451-2010`

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., ... Schulz, M. (2016). Recommendations for diagnosing effective radiative forcing from climate models for cmip6. *Journal of Geophysical Research: Atmospheres*, *121*(20), 12,460-12,475. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025320` doi: 10.1002/2016JD025320

Frith, S. M., Kramarova, N. A., Stolarski, R. S., McPeters, R. D., Bhartia, P. K., & Labow, G. J. (2014). Recent changes in total column ozone based on the SBUV version 8.6 merged ozone data set. *Journal of Geophysical Research: Atmospheres*, *119*(16), 9735-9751. Retrieved from `https://`

agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD021889 doi: 10.1002/2014JD021889

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., ... Randel, W. J. (2019). The Whole Atmosphere Community Climate Model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12380-12403. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030943 doi: 10.1029/2019JD030943

Kelley, M., Schmidt, G. A., Nazarenko, L., Miller, R. L., Bauer, S. E., Ruedy, R., ... Yao, M. (2019). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modelling Earth Systems*. (Manuscript submitted for publication)

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., ... Weiss, R. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*, 2057–2116. Retrieved from https://doi.org/10.5194/gmd-10-2057-2017

Morgenstern, O., Stone, K. A., Schofield, R., Akiyoshi, H., Yamashita, Y., Kinnison, D. E., ... Chipperfield, M. P. (2018). Ozone sensitivity to varying greenhouse gases and ozone-depleting substances in CCMI-1 simulations. *Atmos. Chem. Phys.*, *18*, 1091–1114. Retrieved from https://doi.org/10.5194/acp-18-1091-2018

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., ... Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013 - The Physical Science Basis* (chap. 8). Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC).

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., ... Zerroukat, M. (2019). UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513-4558. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001739 doi: 10.1029/2019MS001739

Shindell, D., Faluvegi, G., Nazarenko, L., Bowman, K., Lamarque, J.-F., Voulgarakis, A., ... Ruedy, R. (2013). Attribution of historical ozone forcing to anthropogenic emissions. *Nature Climate Change*, *3*, 567.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., ... Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182-4227. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001791 doi: 10.1029/2019MS001791

van der A, R. J., Allaart, M. A. F., & Eskes, H. J. (2015). Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012. *Atmospheric Measurement Techniques*, *8*, 3021–3035. Retrieved from https://doi.org/10.5194/amt-8-3021-2015

Velders, G. J. M., Andersen, S. O., Daniel, J. S., Fahey, D. W., & McFarland, M. (2007). The importance of the Montreal Protocol in protecting climate. *Proceedings of the National Academy of Sciences*, *104*, 4814-4819.

Williamson, D. B., & Sansom, P. G. (2019). How are emergent constraints quantifying uncertainty and what do they leave behind? *Bulletin of the American Meteorological Society*, *100*(12), 2571-2588. Retrieved from https://doi.org/10.1175/BAMS-D-19-0131.1 doi: 10.1175/BAMS-D-19-0131.1

WMO. (2018). *Scientific Assessment of Ozone Depletion: 2018*. Geneva, Switzerland: World Meteorological Organization.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., ... ISHII, M. (2019). The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan Ser. II*, *advpub*.

396 doi: 10.2151/jmsj.2019-051
397 Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M.,
398 Ceppi, P., ... Taylor, K. E. (2020). Causes of higher climate sensitivity
399 in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782.
400 Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/`
401 `10.1029/2019GL085782` (e2019GL085782 10.1029/2019GL085782) doi:
402 10.1029/2019GL085782