To the editor

It is apparent that Reviewer #2 has fundamentally misunderstood the focus of this Letter. We added clarifying text to avoid the potential for misunderstanding in our first revision and have added further clarifications in this second revision. We note that the focus seemed perfectly clear to Reviewer #1 who was "*convinced [the letter] is technically correct*"..

Reviewer #2 interprets our Letter as a redefining of the relationship between surface observations, tropospheric mole fractions and whole-atmosphere mole fractions. This would be a fundamental change and their comments could be relevant if that were the case; however the text of the Letter makes no suggestion of addressing this topic. It is quite clear that this Letter focuses solely on the need to convert existing reported radiative efficiencies when using them to derive climate metrics (such as GWP). Existing text: "*This ratio should strictly be the ratio of the changes as used in the radiative efficiency calculations which is not necessarily that observed in the atmosphere or derived from models*". Any comments on the treatment of AGAGE or NOAA observations and the expertise or otherwise of the authors in analysing observations are not pertinent to this Letter. Indeed, we note that in Reviewer 2's first review, they noted that "GWP calculations since the TAR [IPCC's 2001 Third Assessment] have assumed that all GHG are well mixed to the top of the atmosphere and now need to be corrected … and [we should] make sure the AR6 GWPs are updated on /ipcc.ch".

This Letter in no way seeks to redefine radiative forcing, hence the comments on the best means to characterise this are not relevant. In this Letter we note that the radiative efficiency diagnosed by Radiative Transfer Modelling (RTM) studies is the change in radiative forcing divided by the change in surface mole fraction in their RTM (e.g. Etminan et al. 2016, Hodnebrog et al. 2020, Freckleton et al., 1998; Hodnebrog et al., 2013; Jain et al., 2000; Myhre et al., 1998; Myhre & Stordal, 1997; Naik et al., 2000; and presumably future studies to come). We then identify that this quantity needs a correction factor to be converted to a radiative efficiency per mass (as required by climate metric calculations, such as the Global Warming Potential). Comments on whether RTMs should diagnose the tropospheric mole fraction rather than surface, or the relationship between mole fraction at the "surface" in the RTMs and "surface" as constructed from AGAGE/NOAA observations have no bearing on the corrections calculated in this Letter.

Publication of this Letter is essential for the IPCC and WMO assessments as the uncorrected calculations are being widely used in industry and policy.

Response to Michael Prather

GWPs need to be corrected, but not in this way.

The GWP corrections introduced here may be needed, but they are not well founded, nor adequately documented. Thus, they fail to provide a traceable or reproducible chain of uncertainty, which is critical for their use in high level policy as intended for IPCC or WMO O3 assessments.

We agree with the reviewer that the corrections certainly are needed, it will be crucial for IPCC AR7 to apply these corrections when calculating GWP/GTP/GWP*/CGTP from radiative efficiencies derived from WMO 2026 or Etminan et al. 2016 (and hence also Meinshausen et al. 2020). For some species the exact profiles used in the underlying studies were available, hence there are no uncertainties in the corrections. For other species we have added calculations of the uncertainty involved in using the parameterised equations. Added statements:

"The root mean squared error (RMSE) in the ratio using the fit is less than 0.008. "

"For species with a lifetime greater than 1 year (right hand panel of Fig 1c) the RMSE in the ratio using the fit is less than 0.002. For species with a lifetime less than 1 year (left hand panel of Fig 1c) we calculate a fractional uncertainty of 5% of the ratio."

These uncertainties have been added to Table 1,

"Uncertainties arising from the use of the equations are provided for these species."

"Note uncertainties in GWP100 are around 40% for most of these species (Smith et al., 2021)."

Moreover, the work here continues to commit errors in deriving metrics for the short-lived gases based on use of a standard lifetime without recognizing that the lifetime depends on the spatial-temporal emission patterns (this is now noted briefly but not addressed in the calculations here). The author's rebuttal to the review highlights the errors in the approach used here that would need to be corrected before it is accepted as a standard for IPCC or WMO O3 assessments. Some of these failings are highlighted below.

In this letter we do not generate new metrics for the very short-lived species, but note that these metrics already exist in published literature (notably IPCC AR6, WMO 2022 and 2026) and it is well recognised that they will vary with the spatial-temporal emission pattern. Whether or not differences in spatial-temporal emission patterns should have been taken into account in these assessments (and underlying literature) is not a matter for this Letter. The fact on the ground is that these radiative efficiencies have been published and the derived GWPs are used widely. Given their existence it is imperative

that they are not further used without the corrections provided in this letter. We have added text to clarify that Table 1 is to illustrate the impact of the correction, not to produce new definitive GWP100 values. New text: *"Although we use here GWP100 values from Forster et al. (2021) and WMO (2022) for illustration, the correction factors are appropriate for all climate metrics at all timescales"*. We have also reordered the columns in Table 1 to make it clear that the key calculation is the correction factor.

(1) The authors have an obsession with using the global atmospheric mean mole fraction, x_atm, to describe the radiative properties. This makes no sense, they should be using the tropospheric mean mole fraction, x_trp, as used in the IPCC for observations and scenarios since 1994. The value of x_atm is the key factor in determining the burden (Tg) of a gas which is critical when connecting emissions, lifetime, and burden. The problem is that the x_trp values more accurately reflects the RF of a gas as stated in their Line 54-55. For the well mixed GHGs it is x_trp that proportional to the RF and also directly predicts the mole fraction of Cl or Br entering the stratosphere. Also, the value x_trp is readily measured by the remote marine surface sites or transect campaigns like NASA's ATom. The values of x_atm cannot be simply measured without measuring the falloff of x_ in the stratosphere. Likewise, the average surface value x_sfc is not really measurable since we cannot average over the large spatial-temporal heterogeneity of boundary layer mole fractions over the large emission regions. Similarly, x_sfc is very hard to model consistently across our different models.

We think that this might be one of the key misunderstandings of the Letter by the reviewer. This Letter does not in any way redefine radiative efficiencies or their use to calculate radiative forcing. This Letter instead draws attention to the fact that the definitions of GWP (and similar metric) calculations require the derivation of a radiative efficiency per atmospheric mean mole fraction so that they can be converted to a mass basis. This Letter does not redefine the surface average mole fractions or tropospheric average mole fraction. Instead, it requires the "surface" mole fraction from the underlying RTMs (here Etminan et al. 2016, Hodnebrog et al. 2020; but also Freckleton et al., 1998; Hodnebrog et al., 2013; Jain et al., 2000; Myhre et al., 1998; Myhre & Stordal, 1997; Naik et al., 2000) . We have added text to clarify that "surface" mole fraction is that used in the RTMs to normalise the radiative forcing and does not necessarily correspond to a strict averaged surface concentration. New text: *"although for species with a steep vertical gradient (lifetime < 1 year) the definitions of "surface" in the RTM may not correspond exactly to the definition of "surface" in the observations"*.

What is vexing to this reviewer is that the authors claim that all the modern RF calculations used a surface value that was different from the mean tropospheric value, but they cannot document this, nor even recover what was used. Using their prime

example of Etminan 2016, I tried to recover the GHG profiles used but could find no reference to profiles or surface values anywhere in the main text or supplemental data. In fact, the supplemental tables imply that a constant GHG profile is used since only a single value of x_ is given. Given their participation in these papers, they must have inside information that was never documented. This speaks to the need to redo these RF calculations from scratch with proper documentation, rather than trying to correct something (x_sfc/x_atm) that you are unsure of exactly what was used in the original calculation.

We already provided a citation for the Etminan profiles in the original text *"The profiles of methane and nitrous oxide used in that study were from Myhre & Stordal (1997)"* so it is incorrect for the reviewer to say "they cannot document this, nor even recover what was used", it is also incorrect for the reviewer to say "they must have inside information that was never documented" as this information is available in the reference we cited. For additional transparency we have now provided the profiles as a Zenodo link and cite this under "Data Availability".

Modified text: *"The profiles of methane and nitrous oxide used in that study were from Myhre & Stordal (1997) (see Data Availability)"*.

New text under Data Availability: *"The methane and nitrous oxide profiles used in Etminan et al. (2016) come originally from Myhre & Stordal (1997) and are provided as text files with the DOI 10.5281/zenodo.21791079 (W. Collins et al., 2026)."*

Note, that the Letter does not comment on whether the RTM calculation used a surface value that was different from the mean tropospheric (as opposed to total atmospheric) value, rather that it is the surface value from the RTM that is used in the denominator in their radiative efficiencies and therefore that is the value that should be corrected.

As a general point it is not appropriate for us to respond to the reviewer's criticism of other papers. Only one of the current authors was involved in the Etminan study, and hence the reviewer's generalisation to all the current authors is incorrect, and it is inappropriate to assert that there is a need to "redo the calculations from scratch". The Etminan study and before that Myhre et al. 1998 are widely used and if the reviewer has criticisms of them this is not the right forum to raise them. In this Letter we provide a transparent calculation to show the literature values of radiative efficiency should be corrected when calculating GWPs.

(2) The authors assert that this work is important for future assessments, but without proper uncertainty assessment it really cannot be a mainstay of any assessments. Furthermore, without documentation of the testable intermediate steps in the calculation as requested in the first review, no one can use or test the results using different modeling tools. Calculating "quick & dirty" correction factors to climate

metrics that were calculated incorrectly (or were undocumented) in the first place is not a way to provide traceable metrics. It is not clear what the RF radiation models used for profiles; same problem applies to the CTM used for 'correcting' them?

The RTMs are documented in the Etminan et al. 2016 and Hodnebrog et al. 2020 papers and papers referenced within these. As stated above, we explicitly did provide a reference to Myhre and Stordal (1997) which contains the GHG profiles used in Etminan (the reviewer seems to have missed this), and now include a link to the data for these. Data from the Hodnebrog et al. paper are explicitly shown in figure 1 and uploaded to Zenodo. Rather than a "quick & dirty" correction factor, it is explicitly stated in the text that the appropriate factor to use is that taken from the RTM used to generate the original radiative efficiencies, hence the future recommendation for radiative transfer modellers to publish radiative efficiencies per atmospheric mean mole fraction alongside (not instead of) their standard values.

(3) The author's rebuttal states that the correct balance between greenhouse gas content (burden) and mean tropospheric abundance (dry-air mole fraction, ppb) has not been used by the IPCC previously in "climate metrics". This assertion is highly misleading and wrong. In the 2001 IPCC TAR, the many chapters dealing with greenhouse gas emissions, burdens, observations, and future scenarios all used a near-correct conversion factor relating these (see details in my review). I would call the calculation of the mean tropospheric abundance of greenhouse gases a climate metric. Their use in the TAR chapter 4 was clear, but not well documented as to how they were calculated. Thus, I wrote up a 2007 Phil. Trans. Roy. Soc. Monograph on lifetimes ("Lifetimes and time scales in atmospheric chemistry", Prather) and then we prepared a detailed analysis of the factors and uncertainties in the 2012 GRL article ("Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry", Prather, Holmes, Hsu). I know these studies and their uncertainties were carried through to the 2017 IPCC AR5, since I edited the Annex II summary tables. That 2012 paper needs updating. All the greenhouse gas abundances in the IPCC are calculated as the tropospheric mean, not "atmospheric" mean; the latter requires one estimate how much the gas fills the stratosphere and this gives the standard conversion factor (CF, Tg/ppb) used since 2001. It appears that only the chapter calculating the damage metrics (GWPs, ODPs) was incorrect. It is interesting to note that chemistry-climate models calculating RF from their own GHG distributions were correct. Please correct the manuscript with an accurate history and avoid misleading the readers, i.e., that we did not know how to calculate the abundance anddistribution of the greenhouse gases.

This Letter does not claim to make a more general quantification of tropospheric abundance and does not have implications for the calculations in Prather et al. 2012. This Letter does not discuss the history of atmospheric burden calculations, and it

would not be appropriate to add this as it does not relate to the calculation of radiative efficiencies.

(4) The obsession here with the "surface" abundance is confusing and misleading. Regarding 'surface' values, the author's rebuttal shows a lack of experience in analyzing NOAA or AGAGE surface measurements. Yes, these are surface sites (for the most part) but they have very strict protocols for taking flasks or excluding polluted data. They are meant to represent the remote marine troposphere and thus their reported surface measurements are systematically biased to represent the free troposphere. One cannot compare these observations with a CTM averaged surface value because the surface data representing emissions has been removed or avoided. Cape Grim Observatory is an obvious example. I know this because I had to ask ALE/GAGE to provide the full, unbiased data set for my 1985 Nature paper based on the polluted surface layer ("Continental sources of halocarbons and nitrous oxide"). The "remote marine clean air" observations, when latitudinally averaged, are typical of the tropospheric mean. Since we cannot really observe the surface abundance with any accuracy, we must also admit that we cannot model it very well either since the results will depend on the height of the nighttime boundary layer and hence on the diurnal cycle of emissions. Thus, the ratio of "surface" to "atmospheric" mole fraction developed here using one set of emissions and resolution with OsloCTM2 (and used as a correction factor) is irreproducible in other models. Further, the curve fitting on a log scale does not help address uncertainties. On the other hand, the mean tropospheric abundance for a GHG like CH4 is easy to estimate or observe – even the Atom mission transects provide a good estimate.

As with previous comments, the reviewer may have misinterpreted this Letter. The Letter does not at any point redefine "surface", hence the experience or otherwise of the authors in analysing NOAA or AGAGE measurements is not a pertinent criticism of a Letter on radiative efficiencies. The key quantity is the mole fraction used in the original RTM calculations (e.g. Etminan et al. 2016, Hodnebrog et al. 2020, Freckleton et al., 1998; Hodnebrog et al., 2013; Jain et al., 2000; Myhre et al., 1998; Myhre & Stordal, 1997; Naik et al., 2000; and presumably future studies to come) to derive radiative efficiencies from radiative forcing calculations . We have added text to clarify that "surface" mole fraction is that used in the RTMs to normalise the radiative forcing and does not necessarily correspond to a strict averaged surface concentration. New text: *"although for species with a steep vertical gradient (lifetime < 1 year) the definitions of "surface" in the RTM may not correspond exactly to the definition of "surface" in the observations"*. We also have added clarification that the ratios defined here are exact only for the radiative efficiencies in the current literature and assessments. Text added : *"For studies other than Etminan et al. (2016) and Hodnebrog et al. (2013, 2020) the $(\Delta\chi_sfc)/(\Delta\chi_atm)$ correction should ideally be taken from the profiles used in the RTM,*

however if these are not available our uncertainty analysis shows that Equations 4 and 5 are able to estimate the ratio to within less than 1%."

(5) A much broader issue that could be addressed here - and really should be addressed by the community as a whole - is that we must stop reporting these damage metrics (e.g., GWP, ODP) for short-lived gases without also documenting how they were derived. Defining short-lived as <1/2 yr (i.e., not well mixed within a hemisphere) we must recognize that their lifetimes – a critical factor in these metrics – will vary depending on where and when it is emitted. A recent quick-study of CO, showed lifetimes ranging from 55 to 85 days. While the lifetime for a well mixed GHG is about the same for any emissions within the troposphere, for stratospheric emissions (e.g., some proposed mechanisms for photochemical production of N2O) it is much, much shorter (1998 Science paper, "Time scales in atmospheric chemistry: coupled perturbations to N2O, NOy, and O3"). More applicable to this problem, we can look in our 2024 Elementa paper ("Lifetimes and timescales of tropospheric ozone", Prather & Zhu) where we calculate the lifetime of a tropospheric O3 emission to vary from 6 days (E. Asian pollution source) to 26 days (aviation or strat-trop exchange). Thus, if we calculate the RE of an ozone source we have a large factor-of-four range, but fortunately we usually calculate the ozone RE from a modeled change in burden, and the lifetime is factored into the burden. This same range, or potential error in GWP, could apply to short-lived HFCs. The authors feel that giving a "best estimate" for HFC-152 is appropriate since people are using it, but I would contend that they do not necessarily know what "best" is, and they certainly have not documented that choice. I find the revised wording to be inadequate since their correction of 10% is probably dwarfed by the uncertainty they should report on the GWP.

As stated above, the facts on the ground are that these radiative efficiencies have been published and the derived GWPs are used widely. The issue is not what the authors "feel" is best, but rather that these values already exist and are being used. Given their existence it is imperative that they are not further used without the corrections provided in this Letter. We have added clarification that the focus of this letter is to provide correction factors to those in the literature and that the GWPs are shown for illustration. We have also reordered the columns in Table 1 to make it clear that the key calculation is the correction factor. New text: *"Although we use here GWP100 values from Forster et al. (2021) and WMO (2022) for illustration, the correction factors are appropriate for all climate metrics at all timescales"*. We have added text on the uncertainties in the GWP100: *"Note uncertainties in GWP100 are around 40% for most of these species (Smith et al., 2021)"*.

I find the authors show very little understanding of what the "lifetime" of a gas depends

on, they seem to think it can be scaled from knowledge of chemical kinetics alone. With these short-lived gases, a lot of the build up of the burden and the chemical rates of destruction depend on tropospheric processes such as boundary layer venting, scavenging, clouds and convection, which are not consistent across models and introduce unknown uncertainties. The range of emission location/timing determines lifetime and is not a scientific uncertainty but a societal choice. All we can do is to calculate the scientific uncertainty for a specific emission and then include the possible range from emission choices.

The authors' understanding or otherwise of "lifetime" is not pertinent. This Letter addresses a correction to the use of RTM-generated radiative efficiencies in GWP calculations.

The publication of this article will not help the IPCC or WMO O3 assessments, it will only confuse. Approximate correction factors, parameterized by a curve fit on a log scale, for an incorrect on unknown original value with unclear documentation does not lend confidence, but erodes it.

On the contrary, publication of this article is essential for IPCC and WMO assessments. These assessments have been using established methods that at the time were well documented. However, this letter has addressed a deficiency in those methods. The reviewer is correct that there is uncertainty in the correction that is being presented in this letter. We have addressed this uncertainty in response to comments above.

Response to Chris Smith

It is necessary that this correction to calculating policy-relevant metrics is available in the scientific domain.

We agree that publication of this article is essential for the IPCC and WMO assessments as the uncorrected calculations are being widely used in industry and policy.