The manuscript makes a useful contribution to the literature by exploring explicitly how different time scales for GWP relate to GHG equivalence ratios based on damage costs and different discount rates. It is clearly written and highly readable. I have no fundamental concern with the technical approach and quantitative results, but I feel the manuscript needs to work a bit harder to develop its value proposition, discussion of results including sensitivity analysis, and finally the conclusions, before it is fit for publication.

We would like to thank the referee for their comments. For responses comment by comment, see below. All author replies are in red. There is also a summary of new sensitivity analyses that is included at the end of the reply to William Collins.

I'm comfortable with and largely endorse the comments already posted by Bill Collins and anonymous referee #2, and will try not to repeat the specific points they made.

My main concerns where I feel the manuscript needs to work harder are as follows:

1) value proposition: it is mainly in the SI that the authors acknowledge prior work that linked GHG equivalencies based on damage costs and discount rates to GWP. I believe this needs to be brought into the main paper up-front, and the authors need to do a better job explaining where their study adds value to those existing studies. For example, one could argue that their approach is simply a reverse reading of Boucher (2012). I don't think that accusation would be justified, but neither is it justifiable for the manuscript not to recognise the fact that a range of studies have already found that discount rates around 2-3% give the same GHG equivalence between CH4 and CO2 as GWP100. In this context, in the discussion, I would have liked to see a better explanation why their GWP100-equivalent discount rate of 3.3% is higher than that derived by both Boucher and Fuglestvedt et al.

We will bring the SI discussion into the main text, particularly with the comparison to Boucher & Fuglestvedt. See our response to Collins. Better understanding the differences between Boucher & Fuglestvedt is still in progress.

2) discussion of results including sensitivity analysis: in my view, the authors should include an explicit simulation of results if climate-carbon cycle feedbacks following a pulse emission of CH4 are included. The IPCC AR5 and subsequent studies demonstrated that including this results in a significant increase in the GWP100. This is flagged (p7 of the manuscript) but appears not to have been included in the actual sensitivity analysis. It should be fairly easy to modify the radiative forcing calculations to simulate climate-carbon cycle feedbacks and it doesn't have to change the study design at all. There really is no good justification in my view not to include this, other than this is not how the GWP has been defined historically – but from a scientific consistency perspective, it makes no sense to include an effect for one gas (CO2) but not for the other. Including this in the sensitivity analysis (perhaps as a special case, since this is a binary choice rather than something that can be expressed via a pdf) would at least tell us how important this is when we are concerned about choosing GHG equivalencies based on damage functions and discount rates. I could even live with the authors running this only for a central estimate for all other parameters so we can get an order-of-magnitude sense.

See our description of sensitivity analyses at the end of the response to William Collins for a description of the sensitivity analysis we performed in response to this comment, showing that sensitivity to exclusion of the climate-carbon feedback from CO_2 had only a small effect. (as well as discussing why we chose that approach rather than inclusion of the climate-carbon feedback in the CH_4 effect). We agree that this was an important analysis to do. The effect of the exclusion was small, due to cancellation when the GWP and the damage ratio are both calculated using consistent assumptions about gas lifetimes and radiative efficiencies.

Related to this, but more difficult to do (hence I would not insist that this is done quantitatively) is consideration of the rate of change as a source of damages. Again this could be parameterised and quantified, but there is a large degree of arbitrariness how much weight to place on rate of change vs amount of change. The manuscript would be much stronger though if it could demonstrate under what circumstances including the rate of change might affect the conclusions, or whether the conclusions might be robust even if rate of change damages have been incorporated within reasonable bounds.

We have implemented a crude rate of change analysis within our framework (see discussion in the sensitivity analysis at the end of the reply to William Collins for details), and determined that under RCP6 inclusion of even extreme damage estimates due to rate of change have little effect on implicit timescales. However, under the RCP3PD scenario, the incorporation of rate of change has a larger effect, reducing the implicit timescale by almost half under the rate of change damage parameters that may be more realistic in magnitude. Our approach is not sophisticated enough to become a major component of the paper, but the results would be worth noting in a sentence or two, along with reference to Bowerman (2013) which provides a good explanation for the reasons why we see this result.

3) interpretation and conclusions: I would endorse some of the comments made by anonymous reviewer #2, that the authors are effectively beating up a strawman. Yes some people have argued that we should simply ‘switch’ to GWP20, but the more intelligent arguments are all for considering the effect of multiple alternative time horizons to inform abatement decisions and policy choices. See e.g. the conclusions in Levasseur et al 2016 (doi: 10.1016/j.ecolind.2016.06.049) regarding the use of multiple time horizons and metrics in lifecycle assessment. The discussion and conclusions need to add quite a bit of nuance to reflect what those studies actually say, and hence the degree to which this manuscript challenges their conclusions or simply adds another dimension that can help choosing the right metric for the right purpose.

We have added more nuance to our description (see responses to other referees), as well as additional context to our results.

There are two additional points that the discussion and conclusion needs to address:

(a) one is that a key context in which GHG metrics are used are in emission trading schemes, and to help governments evaluate policy choices that directly affect near- term commercial decisions, i.e. policy that would “alter the use of capital in the private sector”. So there are very different contexts in which GHG metrics are actually used in climate policy and where different discount rates are commonly applied, and the paper would be stronger and more relevant if it recognised and addressed these explicitly.

We recognize that there is common justification (e.g., OMB Circular A-4) for the use of one discount rate for social problems and another for policies that “alter the use of capital in the private sector”. We would argue, however, that comparing the relative impacts of CO_2 v. CH_4 should always been considered a social problem in this context, regardless of whether it is being used in decisions regarding capital or by governments within emission trading programs.

We will consider how to address this within the paper, but it is a somewhat tricky topic, as it can pose consistency challenges that go well beyond the implications of this paper. For example, a decisionmaker deciding between investing in a CNG vehicle and a gasoline vehicle might want to look at vehicle costs, maintenance, and fuel prices under a high discount rate to reflect the opportunity cost of that investment, but, because the benefits of the GHG abatement are not received by the decisionmaker but by society as a whole, it could be argued that the latter issue should be considered with a societal discount rate. It is unclear how to bring those two monetization streams into a single analysis given the difference in discount rates.

(b) The second is a recognition that IAMs used to design cost-minimising emission pathways often use a discount rate of 5%. Given that a(nother) key use of GHG metrics is to help IAMs make trade-offs between different gases with different mitigation costs, this should enter into the discussion in this paper. I don't think this materially changes the conclusions since we know that different GHG metrics don't have a massive effect on total mitigation costs (although there is a systematic effect especially when moving towards GWP20), but the issue is not trivial especially for countries or sectors with non-negligible non-CO_2/SLCF sources. Some discussion on this is needed.

We plan to add more discussion regarding some of the IAM-based tradeoff work such as van den Berg et al. (2015), Reisinger et al. (2013), and Smith et al. (2013). The common use of a 5% discount rate might relevant to that discussion.

I believe that all the above points (with the exception of quantifying the effect of including climate-carbon cycle feedbacks for CH_4) can be addressed by a careful revision of the text itself. The manuscript needs to avoid what currently appears as the too- simplistic conclusion that "actually, GWP100 is largely ok, let's move on" (which is how I read P8L22). The fundamental finding from virtually all metrics papers is that the right metric depends on the application, and hence it is rather jarring to read a conclusion that continued use of GWP100 is 'reasonable' without any caveat.

We will endeavor to add more nuance to the conclusion, though the authors do believe that this analysis is fairly strong support for a 100 year GWP in many contexts, given acceptance of a few key assumptions. These key assumptions can be made more clear, and might be as follows (but with some improvement in phrasing):

1) The context of comparing pulses of emissions. This may not hold true within a framework of, for example, an absolute temperature target or other global decision-making process.

2) A valuing of time similar to a 3% discount rate

3) Given the assumptions about damage functions, future scenarios, and other parameters made within this paper.

4) Not considering additional impacts such as CH_4-O_3 or acidification (which could be addressed with post-hoc adjustments of relative weights)

5) Not considering how the metric functions within the context of the broader economic system (which could be better addressed within an IAM)

We do recognize that there are other uses of metrics such as the AGTP to produce a mechanism by which future temperatures changes resulting from emissions pulses can be quickly estimated, but that is a separate issue from the relative impacts addressed by this paper.

I am not repeating the above points in my specific comments below and would be happy for the authors to decide how they can best address them.

Specific comments:

P1L22: insert 'emission' after gases – we're talking about emission metrics here

Will do.

P2L3: 'endorsed' is too strong in my view for the UNFCCC – 'used' is more factual, I cannot recall an explicit endorsement in the sense that the UNFCCC would have explained and justified its choice.

We propose language such as:

> The 100-year time horizon of the GWP (GWP_{100}) is the has been time horizon most commonly used in many venues, for example in trading regimes such as under the Kyoto Protocol, perhaps in part because it was the middle value of the three time horizons (20, 100, and 500 years) analyzed in the IPCC First Assessment Report.

P2L11: I believe the correct term for GTP is Global Temperature CHANGE Potential

In our defense, AR4 and other sources also refer to the GTP as the Global Temperature Potential: e.g., "2.10.4.2 The Global Temperature Potential"). But "change" seems to be more standard, including in AR5, and so the updated manuscript will reflect that.

P2L12: the reason why GTP downplays SLCFs is not primarily that it is temperature based but that it is a point metric. iGTP is very similar to GWP.

We are updating language to better differentiate integrated versus endpoint metrics. We note, however, that Brazil and New Zealand specifically suggest the GTP, not the iGTP, and justify it based on the temperature argument.

P2L22-27: editorial only: I prefer if introductions don't include the conclusions but rather focus on making the point of why the conclusions are worth having.

We will edit accordingly.

P3L15: shouldn't the N2O effect on CH4 forcing depend on the RCP pathway? Perhaps this was done but this isn't clear to me from the text.

The particular adjustment for N_2O cited here is based on 8.SM.11.3.3, which estimates that emissions of 100 molecules of N_2O will lead to a destruction of 36 molecules of CH4. For this effect, the RCP pathway does not matter.

This is in contrast to the overlap between N_2O and CH_4 radiative absorption bands, which does depend on the RCP pathway, and is captured in the radiative forcing equations used for this paper (and for the GWP).

P4L9: 'future years are cooler than present': helpful if you could indicate what years we are talking about (presumably you mean after 2200 or thereabouts, depending on the reference period/warming – meaning that much of those will be heavily discounted anyway).

Under RCP3PD, with climate sensitivity of 3.92, and a forcing imbalance of 0.84, temperatures drop below the starting temperature only 458 years into the analysis. Therefore, in this case, when the damage function is relying on temperature change since present (rather than since 1950-1980 or earlier), these years would need to be set to zero. It is possible that under different climate sensitivities or forcing imbalances, this could occur somewhat earlier, but we believe that the referee's intuition that this effect will be negligible due to discounting is correct.

P4L21: here and later, please clarify where you truncate your damage calculations (when I read this sentence, I thought you truncate at 2300, but later (P5L14) it seems you truncate at 2500). You note below that this may matter for very low discount rates. Can you quantify/illustrate this?

We will clarify that the graphs are to 2300, but the calculations do go to 2500. We also include text regarding the size of the effect as follows:

> Even at a 2% discount rate, 95% of the CO_2 damages come in the first 287 years. At discount rates lower than 2%, however, truncation effects can account for errors in damage ratio estimates of greater than a percent, indicating that longer calculation timeframes may be necessary to capture the full effect of the emissions pulse.

P6L4: I think the entire sensitivity discussion should note that projecting damages multiple centuries into the future is increasingly fraught with difficulties. The AR5 chose not to evaluate GWP500 because the authors felt that (deep) uncertainties were simply too large – but here you evaluate damages from temperature responses from forcing 500 years into the future? At least a comment on this is needed here – the discussion of what percentage of total damages occurs up to a given year for CH4 and CO2 is useful in this context and could be linked to this point about uncertainty.

See the response above regarding the low percentage of damages that occur after the first 290 years even under a 2% discount rate. But we can also include a note about the challenges of projecting damages many decades into the future. In particular, the GDP projections are a key component of the analysis, and therefore the dramatic uncertainties about future economic growth are particularly important. We also propose to include a sentence noting that uncertainties as shown in Figure 2 are larger at low discount rates in part for this reason.

P8L2: I feel the statement "We note no metric designed to tradeoff emission pulses is consistent with stabilization" is too strong. Of course, no metric in itself delivers stabilisation, but almost any metric can be used wisely enough to help countries achieve stabilisation.

We propose to move some of the following discussion into the main text to make a more in-depth discussion of this particular point, hopefully in a more nuanced fashion. We particularly wanted to highlight the challenge of using, for example, a GTPX metric to achieve a cost-optimal approach to achieving a temperature target in a given year, but which when used in that way would lead to increasing challenges of maintaining temperatures at that target in future years (in part because the short-term GTP would lead to more SLCF abatement relative to CO_2 abatement than would be indicated by, for example, a GWP100). This is in addition to the fact that any SCLF/CO_2 trading using whatever pulse-based metric post-stabilization will lead to moving away from stabilization. Allen et al. 2016 has a particularly elegant approach to this question. Though all of this is perhaps not central to the main analysis in the paper...

> A number of authors have recognized that the GWP is not designed to achieve stabilization goals (Sarofim et al. 2005, Smith et al. 2012, Allen et al. 2016). Some actors (Brazil INDC, 2015) have claimed that certain metrics such as the Global Temperature Potential (Shine et al. 2005) or the Climate Tipping Potential (Jorgensen et al. 2014) are more compatible with a stabilization target such as 2 degrees C because they are temperature based. However, these metrics are also not designed to achieve stabilization goals, but rather to achieve a temperature target in a single given year. The challenge is that in any year after stabilization, any trading between emission pulses of carbon dioxide and a shorter-lived gas will cause a deviation from stabilization. For example, trading a reduction in methane emissions for a pulse of CO_2 emissions will lead to a near term decrease in temperature, but also a long-term increase in temperature above the original stabilization level.
>
> One solution to the problem is a physically-based one. Allen et al. (2016) suggest trading an emission pulse of carbon dioxide against a sustained change in the emissions of a short-lived climate forcer. This resolves the issue of trading off what is effectively a permanent temperature change against a transient one. However, the challenge becomes one of implementation, as current policy structures are not designed for addressing indefinite sustained mitigation. Alternatively, a number of researchers (Daniel 2012, Jackson 2009, Smith et al. 2012) suggest addressing CO_2 mitigation separately from short-lived gases. Such a separation recognizes the value of the cumulative carbon concept in setting GHG mitigation policy (Zickfeld et al. 2009).