This paper usefully links the well-used GWP metric with more economic comparisons of the ratios of damages. It is suitable for publication and I only have minor comments.

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 this comment reply.

The relationship to the Paris goals are only briefly alluded to. I suggest including a longer discussion of the differences between: a temperature limit, a long-term temperature goal, and least economic cost. Presumably neither a temperature limit nor a long-term temperature goal are optimal economically using the damage function here? Is this a problem with the Paris agreement? The choice of metric depends entirely on the choice of target, and the authors here are implicitly assuming that least economic damage is the most important target. The authors dismiss short-term GTPs as implying unrealistic discount rates, which of course they do for an economic damage target. However, if the Paris agreement is taken to imply that temperatures should not be allowed to exceed 2 degrees, then a GTP with a time horizon ending at the time of peak warming (20-30 years) is an entirely appropriate metric. Similarly, if the Paris agreement is taken to mean a long-term goal to stabilise at 2 degrees then GWP* is the appropriate metric.

The reviewer brings up an interesting question regarding the goals of global climate agreements. To some extent, the optimal design of a global goal is beyond the scope of this manuscript, but given that one important use of climate metrics is within cap and trade programs that are meant to help achieve these global goals, some discussion would be appropriate. To that end, we propose to bring some of the text that was originally in the SI into the main manuscript, with some updating and editing. Such language might look like the following:

> This paper does not address the construction of global climate goals (such as the 2 °C goal in the Paris agreement). However, because one use of metrics is to enable trading between gases in cap and trade programs, and because cap and trade programs are one of the measures used to help achieve these long-term targets, a discussion of the relationship between metrics and temperature stabilization is relevant. 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 what is effectively a permanent temperature change for 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) have suggested 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).

We note as an aside that for a purely economic approach to recommend an absolute target such as 2 degrees would require a world in which damages were infinite above 2 degrees, and zero below two degrees (see e.g., the discussion of “cost-effectiveness” in Deuber et al. 2013, Physico-economic evaluation of climate metrics: A conceptual framework).

The overall formula for the damage function needs to be shown as a function of temperature, discount rate, GDP etc.

We propose to add text such as

> Damages: Damages as a percent of GDP were calculated by multiplying a constant times the temperature squared. E.g., $D(2050) = a*\Delta T(2050)^2*GDP$, and the net present value is calculated using the discount rate such that $NPV(D(t)) = D(t)/(1+r)^{t-2010}$.

Page 1, line 26: Maybe a different word other than “endpoint” could be used so as to avoid confusion with the later discussion of integrated and endpoint metrics.

We will replace endpoint here with “measure of impact”

Page 2, first paragraph: The main difference between GTP and GWP is the difference between endpoint and integrated metrics. This should be brought out more in this paragraph. The iGTP could be mentioned as it is more similar to GWP than GTP.

We propose to add a discussion along the lines of the following:

> While we argue that quantitative justifications for timescales within the GWP are rare, as reflected by the judgment of the IPCC authors that no scientific arguments exist for selecting given timescales, there is a rich literature addressing many aspects of climate metrics. Deuber et al. (2013) presents a conceptual framework for evaluating climate metrics, laying out the different choices involved in choosing the measure of impact of radiative forcing, temperature, or damages, and temporal weighting functions that can be integrative (whether discounted or time-horizon based) or based on single future time points. Deuber et al. conclude that the Global Damage Potential (GDP) could be considered a “first-best benchmark metric”, but recognize that the time-horizon based GWP has advantages based on limited value-based judgments to a choice of time horizon, reducing scientific uncertainty by limiting the calculations of atmospheric effects to radiative forcing, and eliminating scenario uncertainty by assuming constant background concentrations. Mallapragada and Mignone (2017) present a similar framework which also notes that metrics can consider a single pulse of a stream of pulses over multiple years. Several authors have noted that under certain simplifying assumptions, the GWP is equivalent to the integrated GTP, and therefore any timescale arguments that apply to analyses of one metric would also apply to the other (Sarofim GTP paper and related references).

Page 2, second paragraph: Boucher ESD 2012 should also be discussed for economically-based equivalences.

We propose to bring a discussion of Boucher 2012, Fuglesvedt 2003, Reilly 2001, and other relevant papers into a paragraph in the introduction. There is some relevant language that was originally in the SI that, with some modifications, might serve:

> We acknowledge that a number of authors have considered the use of relative damages as a potential metric, and in doing so have often compared the resulting global damage potential to the GWP. However, we believe that none of these papers has done so with the explicit goal of evaluating GWP timescales, and determining the discount rate implied by any given timescale choice. Boucher (2012) in particular compares a global damage potential to the GWP_{100} with an uncertainty analysis, and finds, similar to our manuscript, that the median values of each approach are consistent with one another. Fuglestvedt et al. (2003), in an assessment of a number of different approaches to metrics, performed a similar calculation, and also found that a 100 year GWP was consistent with a discount rate on the order of 2%. De Cara et al. (2005), in an unpublished manuscript, calculated the relationship between discount rates and time horizon, though they assumed linear damages.

Page 4, line 1. These GDP pathways should be shown (maybe in the supplement).

We will include a graph of the GDP pathways in the SI.

Page 4, line 6. It is not obvious why 1951-80 should be chosen as a baseline. A problem with damage functions that are non-linear functions of temperature is that a point needs to be chosen when temperatures were optimal.

We agree that there is no clear best baseline: that is why we did a sensitivity analysis using a baseline of 0 (effectively, that damages are a function of temperature change relative to 2010) and a baseline of 0.8 degrees (assuming that damages are a function of temperature change since preindustrial times) were used. Table 1 indicates that this choice can have a difference of up to 26% in terms of damages. We propose to add a sentence on this sensitivity analysis to the paper.

Figure 1: I was surprised by the shape of 1 (c). Why does the damage from CH4 keep increasing? Is the damage an integral quantity, or is this increase purely due to an exponential increase in GDP? In 1(d) the damage decreases. Is this because the discount rate is larger than the GDP growth? With the GDP growth of 2.06% would a discount rate of less than 2% give an increasing damage for a gas like CO2 with a non-decaying component?

We propose to add a sentence discussing this effect:

> Fig. 1c demonstrates the dramatic increase in damage over time due to the relationship of damage to economic growth. In the case of CH_4, damage peaks in 2032 and declines until 2080 as a result of the short lifetime of the gas. The increase in damages after 2080 is due to the component of the temperature response function that includes a 409 year timescale decay rate,

such that after 100 years the decrease in the ΔT^2 is about 0.5%, and because that decay rate is slower than the rate of consumption growth, the net damages grow.

Page 6, line 13: I don't think "exponential function of temperature" is the right term for temperature raised to a power.

Will change to "polynomial function".

Table 2: The ranges (either 25%-75% or 10%-90%) need to be shown as well as the central value. These are quite large for the timescales and may well include 1.0 for many of the damage ratios.

We propose to update the table as follows, including uncertainty ranges for CH_4: Uncertainty ranges for the longer-lived gases are more challenging, as the non-monotonicity adds complications.

Table 1: Parameter sensitivity analysis: Examining the sensitivity of the GWP-discount rate equivalency as shown in the uncertainty ranges in Fig. 2 as a function of the individual parameters of the calculation. The ratio is calculated as the ratio of the median of the estimated GWPs given the highest and lowest value of each parameter. Results in this table are derived assuming as discount rate of 3%.

Gas	Lifetime	Optimal timescale	GWP_{100}/damage ratio	GWP_{20}/damage ratio
CH_4	12.4	120 (84-172)	1.15 (1.52-0.87)	3.4 (4.49-2.57)
N_2O	121	*52	0.85	0.84
HFC-134a	13.4	115	1.11	3.2
HFC-23	222	*105	0.71	0.62
PFC-14	50000	>400	0.62	0.45

Table 2: Optimal timescale of non-CO_2 gases. Implicit timescale evaluated for non-CO_2 gases with the GWP to damage ratio for the two most common GWP timescales. Asterisks indicate no exact match between GWP ratio and damage ratio, closest value is given instead. The third and fourth columns show the ratio of the GWP for a given gas to the calculated damage ratio. Interquartile uncertainty ranges are presented for the timescale and damage ratios for CH_4. Results in this table are derived assuming as discount rate of 3%.

Page 6, line 25: GWP100 seems to agree very well with the 3% discount rate within the uncertainty rather than overvaluing or undervaluing.

We propose to modify the text "However, the analysis shows that the 100-year timescale is consistent, within the interquartile range, with the 3% discount rate that is commonly used for climate change analysis."

Page 6, line 25-29: I didn't understand this sentence. Are you saying that the uncertainty in GWP100 is such that it covers agreement with the 3% discount? If so, that seems to contradict the previous sentence which suggested a under/overvaluing

We chose to delete the sentence in question. Upon reflection, the uncertainty in the GWP100 is due to factors like uncertainty in radiative efficiency, but (as shown in a sensitivity calculation) the timescale is not sensitive to radiative efficiency uncertainty.

Additional Sensitivity Analyses in response to various Referee Comments:
The four referees raised a number of points, several of which were inter-related. In response to these points, we have performed a number of additional sensitivity analyses, which will be detailed here. One lesson from these analyses: the analysis is robust to a surprising number of changes. This relates to the fact that any change in the analysis which changes both the GWP and the damage ratio generally cancels out in terms of calculating the relationship between discount rate and timescale: e.g., most factors in the causal chain from emissions through radiative forcing. That includes the size of the emissions pulse, the radiative efficiency of the gas, and the lifetime of the gas (within certain limits). Note that changes in these parameters may well change the best estimate of the relative importance of reducing methane compared to CO_2, but they do not change the implicit timescale that is the focus of this paper.

In contrast, anything which changes only the damage ratio can have a larger impact on the implicit timescale. This includes the factors which have the largest influence on the uncertainty, as reflected in Table 1: the rate of GDP growth, the damage exponent, the scenario (because the GWP assumes constant concentrations), the baseline temperature from which damages are calculated, and the climate sensitivity.

A third category is influences which are specific to a given gas, whether the efficacy of that gas (e.g., Modak et al. noted by Referee #3), or the health effects of methane-related ozone productions. Like changes to radiative efficiency or lifetime, these influences can also change the best estimate of the relative importance of reducing methane compared to CO_2, but because they are specific to a single substance and not generalizable between substances, they would not be appropriate for calculating implicit timescales.

Herein, we summarize the additional sensitivity analyses:

Pulse Size:

In response to a comment by Referee #2, we performed a sensitivity analysis on the size of the emissions pulse. Using 373 MMT (one year's emissions according to Saunois 2016, http://iopscience.iop.org/article/10.1088/1748-9326/11/12/120207, where global emissions are 559 MT of which 2/3rds are anthropogenic), the relative damage ratios of CH_4/CO_2 are the following:

quartile	25%	50%	75%
Damage Ratio	18.80	25.05	32.69

A similar analysis for the 28.3 MMT used in the paper, performed in response to a comment by reviewer 1, yields

quartile	25%	50%	75%
Damage Ratio	18.84	25.12	32.86

The differences between the two analyses are less than 1%. We will note this sensitivity analysis in the paper. Arguments can be made for either the larger quantity or the smaller quantity depending on the purpose of the analysis. When GWPs are used to inform decision making, these decisions are often regarding mitigation actions at the national or sub-national scale, which is closer to the 28 MMT scale than the 370 MMT scale. When GWPs are used to compare CO_2equivalents for global scenarios, then the global annual number might be more appropriate. But in any case, this appears to be an uncertainty much smaller than any of the others considered in this paper.

Radiative Efficiency:

Referee #2 also raised a question regarding the use of the IPCC AR5 radiative efficiencies rather than the more recent Etminan et al. 2016 article. In order to test the general sensitivity of this approach to changes in the radiative efficiency of methane, a sensitivity analysis was performed, wherein forcing from CH_4 was doubled compared to the standard calculation. Relative damages of CH_4 and CO_2, as might be expected, also doubled:

quartile	25%	50%	75%
Damage Ratio	37.98	50.35	65.50

However, calculated GWPs based on the updated radiative forcing also double: the new GWP100 being 56.79, and the new GWP20 being 166.92. The optimal timescale using the central parameters ends up being 119.79 years, under 1/10th of a percent different than the original calculated timescale of 119.85 years.

We can also mention this in the paper.

Consistent Climate Carbon Feedbacks:

Referee #4 points out, with good cause, that it would be a good sensitivity analysis to calculate an implicit timescale using consistent assumptions about climate-carbon feedbacks for both carbon and methane lifetimes (whereas, in the main body of the paper, we used a CO_2 lifetime that included climate-carbon feedbacks, and a methane lifetime that did not, consistent with the approach of the IPCC AR4 and the GWP values reported in the main body of chapter 8 in IPCC AR5). Using the "no-climate-carbon-feedback" CO_2 equation from Gasser et al. 2017, we can calculate CH_4/CO_2 damage ratios:

quartile	25%	50%	75%
Damage Ratio	20.57	27.16	35.23

This can be compared to the original damage ratios from the paper:

quartile	25%	50%	75%
Damage Ratio	18.84	25.12	32.86

The damage ratios are about 8% larger under the consistent no-CC-feedback case than in the CO_2-feedback/CH_4-nofeedback case, due to what is now a shorter CO_2 lifetime. But a GWP calculated with consistent assumptions regarding climate-carbon feedbacks also changes. The no-feedback GWP100 is now 30.74 rather than 28.64, a 7% increase. These two effects largely cancel out, leaving the effect on implicit timescales smaller, again, than most of the other sensitivities examined in the paper.

quartile	25%	50%	75%
consistent no cc feedback timescales	170.56	117.9	83.38
original analysis	170.87	118.30	83.81

So while the damage-ratio differences are on the order of 8%, the implicit timescale differences are less than 1%. Similar to the impact of changing methane's radiative forcing, a large impact on damage ratios has a small impact on implicit timescales because the update happens in both the calculation of the damage-ratio AND in the calculation of the GWP, and therefore cancels out to a large extent.

We show here a calculation that was consistent in not including climate carbon feedbacks in either damage calculation. While it would be possible to do a similar consistent calculation with climate carbon feedbacks in both damage calculations, implementation of the climate-carbon feedback for methane is not without challenges. Such a calculation requires first calculating of the temperature impact resulting from an emissions pulse over the full time period (as done in the original), and then, for the temperature change in each year, requires a calculation of the carbon perturbation resulting from that temperature change over the remainder of the time period, and then calculating the additional forcing and temperature change resulting from that carbon perturbation. Therefore, we decline to do this calculation at this time (absent an already available methane lifetime equation with the CC feedback built in, which the authors do not have on hand). The authors are already on record as preferring a GWP metric that does not include CC feedbacks in either the CO_2 or the CH_4 response rather than inclusion of CC feedbacks in both CO_2 and CH_4 responses – see the response by Sarofim, Giordano, and Crimmins to the Gasser paper for a more detailed explanation (https://www.earth-syst-dynam-discuss.net/esd-2016-55/esd-2016-55-SC1.pdf)

Ramsey Discounting:

Of all the sensitivity analyses (other than discount rate), the calculation of implicit timescales was most sensitive to changes in the assumption about future GDP growth rates. The original draft raised the question as to what extent this sensitivity to GDP growth might diminish were the discount rate to be a function of GDP, as in a Ramsey discounting framework. In this case, a high GDP growth rate (which implies large future damages, and therefore a low methane/CO2 damage ratio) would be counteracted by a high discount rate which would be expected to lead to a high methane/CO2 damage ratio.

For this sensitivity analysis, the Ramsey discounting approach was calibrated to yield an average discount rate for the first 30 years of the analysis of 5% for the reference GDP growth rate, with a pure rate of time preference of 0.01% (a very low pure rate of time preference being consistent with the assumption that, holding consumption constant, all generations should be given equal weight). That required an elasticity of marginal utility of consumption of 1.53.

Interestingly, the ratio of the median damage estimate from the low GDP scenario to the high GDP scenario was still about a factor of 2, except that now the low GDP scenario leads to the lowest damage ratio and the high GDP scenario leads to the highest damage ratio: effectively, in the Ramsey case, the discount rate effect overwhelms the GDP growth rate effect. Note that the equivalent timescale in the median GDPref case under this set of assumptions is 135 years.

GDP scenario		GDPlow	GDPref	GDPhigh
Discount rate, first 30 years->		2.5%	5%	7.5%
Discount rate, full run->		0.8%	1.5%	2.3%
Ramsey discounting approach	25%	11.28	15.71	21.78
	50%	16.08	22.72	31.05
	75%	18.42	26.40	35.16
		3%	3%	3%
Original constant discount rate approach	25%	29.96	21.11	14.46
	50%	38.69	27.58	18.70
	75%	42.52	30.29	19.81

Rate of Change Calculation:

Referee #4 also brought up the idea of adding in damages from rate of change of temperature as well as from absolute temperature change. This was an important addition in, for example, Manne & Richels (2001) where it turned a dynamic optimization metric from one that didn’t value methane until a decade or two before threshold temperatures were reached to a metric which had fairly constant value over time.

In order to test this, we added code that included damages from rate of change. The damage is calculated by taking the square of the rate of change and multiplying by the GDP and a constant – similar to the damage calculation for absolute temperature.

For the first test, peak rate of change damages under the central scenario were calibrated to be equal to the damages in 50 years (2060), where there is 1 degree of temperature change (above the temperature offset) under central parameter estimates. This led to an increase in damage ratio of 2.4%.

A second test, with peak rate of change damages 10 times as large, led to an increase in the damage ratio of 5.1%. That yields a timescale change from 120 years to 112 years.

The small impact of the rate of change damage is likely due to timing of peak rates of change under RCP6. In this analysis, there is an initial high rate of change for the first few years, followed by a secondary (though smaller) peak about 60 years in. Because of the GDP growth in the interim, damages at the 60 year secondary peak are higher than damages in the initial years. Therefore, reducing the rate of temperature growth 60 years into the run is more important than reducing the rate of growth at the beginning of the run, and for this reason, reducing short-lived forcer emissions does not have large advantages over reductions in CO_2 emissions.

However, this raises a question about the importance of the scenario assumption. Therefore, we also tried the same exercise with the RCP3PD scenario as the baseline. In RCP3PD, in this analysis, the rate of temperature change is at its highest in the first years of the scenario, which means that the increase in rate of change due to additional radiative forcing in those early years can have a disproportionately large impact, favoring short-lived climate forcers. Note that because RCP3PD cools in later years, we chose to set the damages resulting from a negative rate of change to zero.

In this case, the first rate of change analysis yields an increase in the damage ratio of 53%, and the second analysis yields a damage ratio increase of 85%. This yields a timescale change from 94 years (the timescale for RCP3PD without rate of change damages) to 54 years to 42 years. This is consistent with some other literature such as Bowerman et al. (2013), which suggests that under stringent CO_2 mitigation scenarios, the rate of change peaks in the near future, and therefore reduction of short-lived climate forcers can be particularly valuable – but that in scenarios where CO_2 emissions continue to rise in the near term, the rate of change peaks further in the future, and therefore delay in short lived climate forcer mitigation will lead to the greatest reductions in peak rate of change.

However, this particular sensitivity analysis was rather crude. In order to improve it, we would have to develop a better reasoning for choosing the parameters involved (e.g., the damage exponent for rate of change, and the damage constant), but also investigate how well our approach models near-term rate of change compared to more complex models, as that is of importance for this calculation.