Climatic Change (2016) 136:203–216
DOI 10.1007/s10584-016-1603-7

How climate metrics affect global mitigation strategies and costs: a multi-model study

Mathijs J. H. M. Harmsen[1,2] · **Maarten van den Berg**[2] · **Volker Krey**[3] · **Gunnar Luderer**[4] · **Adriana Marcucci**[5] · **Jessica Strefler**[4] · **Detlef P. Van Vuuren**[1,2]

Received: 4 January 2015 / Accepted: 6 January 2016 / Published online: 19 January 2016

Abstract In climate policy, substitutions metrics are used to determine exchange ratios for different greenhouse gases as part of a multi-gas strategy. The suitability of the metric depends on the policy goals and considerations regarding its practical use. Here, we present a multi-model comparison study to look at the impact of different metrics on the mitigation strategies and global climate policy costs. The study looks into different Global Warming Potentials (GWP) and the Global Temperature change Potential (GTP). The study shows that for all the models, varying between GWPs - from different IPCC reports, with different integration periods: 20 or 100 years - has a relatively small influence on policy costs (< 2.2 % spread across scenarios with a 2.8 W/m^2 target) and climate outcomes. Metrics with a constant low substitution value for methane (effectively reducing its abatement), in contrast, lead to higher-cost mitigation pathways (with an average cost increase of 32.8 % in a 2.8 W/m^2 scenario). If implemented efficiently, a time-varying GTP leads to a limited cost reduction compared to GWP. However, under imperfect foresight in combination with inertia of CH_4 abatement options, or if implemented sub-optimally, time-varying GTP can result in higher costs than a 100-year GWP. At the same time, given a long-term radiative forcing target, a time-varying GTP results in slightly higher maximum global temperature change rates.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-016-1603-7) contains supplementary material, which is available to authorized users.

✉ Mathijs J. H. M. Harmsen
mathijs.harmsen@pbl.nl

[1] Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, The Netherlands

[2] PBL Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands

[3] International Institute for Applied Systems Analysis (IIASA), Vienna, Austria

[4] Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

[5] Center of Economic Research, ETH, Zurich, Switzerland

1 Introduction

In addition to CO_2, several other greenhouse gases (GHGs) contribute to climate change. To express the contribution of each of these gases in a common indicator, several emission and climate equivalent metrics have been developed. Such metrics are necessary in tracking total emission trends and to compare the contribution of different countries or sectors to climate change. Yet, arguably the main application of metrics is to allow for substitution of gases in multi-gas emission trading schemes (O'Neill 2003). It has been shown repeatedly that multi-gas mitigation strategies, which allow substitution across different gases based on marginal costs, are able to achieve targets at lower costs than a CO_2 only approach (van Vuuren et al. 2006b; Weyant et al. 2006).

However, developing a common metric for different climate forcers is far from straightforward, because of the large differences in physical properties of GHGs, such as atmospheric lifetime and radiative potency (Myhre et al. 2013). This is further complicated by the fact that a metric implicitly or explicitly includes value judgments concerning the overall goal of specific climate policy (Deuber et al. 2013; IPCC 2009). Different climate policy goals may be pursued, such as limiting long-term or short-term temperature change.

In 2011, the UNFCCC requested further research on the impact of the choice of global warming potentials (GWPs) on climate policy strategies (UNFCCC 2011). In response, several single-model studies were carried out to analyse different metrics (See section 2). These studies generated some robust results, but also led to varying conclusions. Notably, they do not always agree on the relation between metrics and policy costs in mitigation scenarios. As a result, different most cost-effective metrics have been proposed. Related to this, some studies show large cost differences between metrics, while others indicate a small spread. The problem with these study results is that it is not directly possible to trace differences in outcomes to underlying model assumptions, to input data or to the setup of the experiment. These are all factors that are relevant in a policy context.

With this study, we assessed the use of common metrics in a multi-model comparison study with Integrated Asessment Models (IAMs) to provide insights into key uncertainties and the difference in outcomes of earlier studies and to identify robust results across all IAMs. This analysis was part of the EU FP7 AMPERE project (Kriegler et al. 2014), using the following four IAMs: IMAGE, MERGE_ETL, MESSAGE and REMIND. IAMs are particularly suitable to use for such an inter-disciplinary analysis, because they simulate the interplay of atmospheric-chemical and socio-economic mechanisms.

Our study specifically focussed on assessing the influence of different metrics on mitigation strategies and costs, and how differences in results can be explained by different modelling approaches. For specific metrics, models were compared on GHG emission pathways, policy costs and global mean temperature profiles in achieving the same climate target in 2100, while using the same scenario setup. This study shows the impact of using other metrics than the 100 year GWPs from the IPCC assessment reports for climate policy and related fields (see section 2).

2 Earlier assessments of the influence of different metrics

2.1 Different types of climate metrics

A large number of metrics to convert GHGs to a common unit have been proposed, based on very different principles. One class of metrics are based exclusively on physical parameters.

The best known examples are the global warming potential (GWP) and the global temperature change potential (GTP). The GWP is based on the GHG induced integrated total radiative forcing (RF) over a certain timespan. The GTP, one of the most suggested possible alternatives to the GWP, is aimed at optimally reducing temperature change in a specific target year (Shine et al. 2007, 2005). Other metrics focus more on climate change related damages, such as the global damage potential (Kandlikar 1995; Tol 1999) or the global cost potential that accounts for the contribution of each gas to overall mitigation costs (Johansson 2011; Manne and Richels 2001). Already in 1990, the IPCC stressed that there is no unambiguous methodology for combining all factors in one metric (IPCC 1990). The clear advantage of physical metrics is that they can be derived in a relatively transparent way. This has the additional advantage that socio-economic uncertainties can be treated separately (Deuber et al. 2013; O'Neill 2000). Therefore, for actual 'real world' purposes so-far only physical metrics have been discussed. Here, we also focus on a selected set of physical metrics. Economic considerations, such as social discounting and GHG emission mitigation costs, are included in the IAMs that are involved in this study and the combined effect is analysed.

The 100 year GWP from the Second Assessment Report (SAR) is without question the most widely used metric in climate policy. It is used in most climate policies to-date, including for the first commitment period under the Kyoto Protocol as well as in different assessment reports (e.g. the Fourth Assessment Report (AR4)). More recently, it was replaced by the values of 100 year GWPs of AR4. At the same time, GWPs have also been extensively criticised by natural scientists and economists (Fuglestvedt et al. 2003; Manne and Richels 2001; O'Neill 2000; Shine 2009). Main points of critique have been the arbitrary timespan used as a basis for the metric (20, 100 or 500 years), the lack of rooting in economic theory, and the metrics' inability to reflect damages caused by (the temperature increase due to) climate change.

The time varying global temperature potential (GTP(t)) has been proposed as a suitable candidate for cost-optimal climate policy (Shine et al. 2007). It differs with the GWP in two ways: 1) It compares gases on the basis of the induced temperature change instead of radiative forcing, 2) It focuses on a certain target year (a so-called "snapshot" approach). The GTP(t) does not have a single numerical value for a specific GHG but its value varies over time. A short-lived GHG such as methane thus has a low value (normalized by CO_2) early on, but a very high one when the target year is approached (here: 2100). Depending on the overall goal of climate policy, the GTP(t) can be more cost-efficient as it provides the largest incentives to reduce emissions when it really matters most.

2.2 Effects on policy costs and strategies

Climate metrics are used both to facilitate comparison of past and future emission trends of individual gases as well as to facilitate substitution as part of actual climate policy. In this paper we concentrate on the latter use, as only this influences actual emission reduction strategies. The choice of the metric can influence the overall costs, the emission reduction strategies of individual gases as well as the overall timing of emissions, and thus temperature.

The current literature gives some common conclusions, but also some clearly different messages. In most studies the cost differences as a result of the choice of different metrics are found to be relatively small, when considering the same prescribed climate target (Ekholm et al. 2013; Reisinger et al. 2013; Strefler et al. 2014; van den Berg et al. 2015). However, if relatively long time horizons are assumed, such as with the 500 year GWP or the 100 year

GTP, policy costs are likely to increase considerably (in the order of 5 % to 20 %) (Ekholm et al. 2013; Reisinger et al. 2013; van den Berg et al. 2015). The reason is that with long timescales, methane reduction becomes unattractive because of a low metric value, and other gases have to be abated more at a higher cost. The use of time-dependent, GTP(t) metric, leads to different results: while some studies have reported a cost reduction (Johansson et al. 2006; Reisinger et al. 2013), others have reported equal costs (Strefler et al. 2014) or even higher costs (van den Berg et al. 2015). This study aims to understand the underlying reasons for this diversity in cost estimates resulting from GTP (t) and to conclude the ongoing debate about this topic.

Although global differences in policy costs can be small, the choice of a metric has large implications for the timing and amount of methane emissions (Reisinger et al. 2013; Smith et al. 2013; Strefler et al. 2014; van den Berg et al. 2015; van Vuuren et al. 2006a). Next to emission reduction profiles, another strategic consideration for the choice of a metric is the induced temperature profile. Temperatures could potentially overshoot unacceptably high before a target year or change too rapidly (Ekholm et al. 2013). By comparing several metrics, used to reach the same two degree climate target, Strefler et al. (2014) found very small differences in maximum transient temperatures (<= 0.05 °C), with the slightly lower temperatures generally corresponding with slightly higher policy costs.

The insights discussed above mostly result from single model studies that often use somewhat different assumptions. By comparing multiple models following exactly the same approach it is possible to thoroughly check these results and to consolidate the understanding of the impact of metric choice on transformation pathways. A model intercomparison approach, as the one adopted in this paper, can be a powerful tool in deriving robust conclusions required for informing decision makers.

3 Research methods

For this study, results have been used from several IAMs that were involved in the EU FP7 AMPERE project: IMAGE, MERGE_ETL, MESSAGE and REMIND (see Table 1)(Kriegler et al. 2014). These represent a range of different models. One distinction between models is how they describe relevant economic processes. These can be based on the concept of economic equilibrium, when aiming for a minimum overall cost from a centralized perspective, taking into account price-elasticity in supply and demand. Some of these models focus on specific sectors (partial equilibrium), while others focus on overall macro-economic impacts (general equilibrium). Another important distinction is the focus on optimisation versus simulation. The optimization models MERGE_ETL, MESSAGE and REMIND include foresight of future supply and demand to reach an optimal least cost solution. The simulation

Table 1 Integrated assessment models used in this study

Model	Model category	Solution dynamics	Policy costs
IMAGE	Partial equilibrium	Recursive dynamic	Area under MAC curve
MERGE_ETL	General equilibrium	Intertemporal optimization	GDP loss
MESSAGE	General equilibrium	Intertemporal optimization	Consumption loss
REMIND	General equilibrium	Intertemporal optimization	Consumption loss

model included in this study (IMAGE) has no foresight, but is still able to derive least-cost climate policies in a recursive dynamical way by means of the sub-model FAIR-SiMCaP (Den Elzen et al. 2007). The Supplementary Material provides additional information about the climate modules used by the models and shows these perform well in emulating climate mechanisms.

All models have been used to generate cost-optimal trajectories towards the same global radiative forcing (RF) targets for the year 2100. The models aimed to meet two RF targets: 2.8 and 3.7 W/m^2 increase compared to pre-industrial levels. The former is a stringent climate goal associated with a 2 degree temperature rise, while the second leads to an approximate 2.5 degree increase, which allows for more flexibility in emission reduction strategies.[1] In combination with each target, 12 scenarios were prescribed, which differed in the use of a climate metric and the overall climate target (see Table 2). The employed metrics are: the 100 year GWP based on the SAR, the 100 and 20 year GWPs based on AR4, the 100 year GTP and the time-varying GTP. In addition, MERGE_ETL made use of an additional time-varying metric, MERGE_RF (t). This metric aims for a least-cost solution to reach the RF target, based on the Global Cost Potential (GCP) approach by Manne and Richels (2001) (see Table 2). The behaviour of this metric is very comparable to GTP(t). Models were allowed to simulate an overshoot in RF (and temperature) in the years before 2100. Furthermore, scenarios were based on the assumption of full globally integrated carbon markets, implying equal marginal CO_2 and non-CO_2 GHG abatement costs across all regions. Full technology availability was assumed. The model projections were compared in terms of climate policy costs; CO_2, CH_4 and N_2O emission profiles; carbon price profiles; global temperature change in 2100 (compared to the pre-industrial level); and maximum global temperature change before 2100.

4 Results

4.1 Emission profiles

We first look at the impacts of the different metrics on emissions (Fig. 1). In the discussion of results, we focus on the 3.7 W/m^2 scenarios. The results for the 2.8 W/m^2 are very similar, and are therefore only briefly summarised while detailed results are found in the Supplementary Material.

For all models, the use of different metrics results in clear differences in reduction strategies for methane. Typically, higher methane emission reductions correspond with higher metrics values for methane, as this increases the relative value of the gas in reduction strategies. In the case of MERGE_ETL and REMIND differences between scenarios at the end of the century are only small. The reason is that unlike in the other models, maximum reductions are effectively reached in all scenarios (except GTP-100). The marginal abatement cost (MAC) curves, used in the model to calculate emission reductions at various price levels, limit the maximum reduction potential (see Supplementary Material for analysis of the methane MAC

---

[0] For the radiative forcing target the so-called "AN3A" metric was used to generate results comparable to the widely used Representative Concentration Pathways (RCPs) (van Vuuren et al. 2011). This metric includes all anthropogenic forcing agents except direct forcing from land albedo changes, mineral dust and nitrate aerosols. This means that the total radiative forcing target is +/− 0.2 W/m^2 higher.

Table 2 Climate metric scenarios

Climate metric	Target	CH4 (CO2 = 1)	N2O (CO2 = 1)	Information
GWP 100 (SAR)	2.8 W/m^2 or 3.7 W/m^2	21	310	GWP metric used in 1st commitment period of the Kyoto Protocol
GWP 100 (AR4)		25	298	GWP metric used in 2nd commitment period of the Kyoto Protocol
GWP 20 (AR4)		72	289	
GTP 100		0.4	265	Numerical values based on (Shine et al. 2005) *
GTP (t)		0.8 (2010) - 102 (2100)	280 (2010) - 216 (2100)	Numerical values based on (Shine et al. 2005) *
MERGE_RF (t)		RF target dependent	RF target dependent	Approach based on (Manne and Richels 2001)**

*In this study, the simple approach from Shine et al. 2005 to calculate GTP100 and GTP(t) was used. Note that the resulting values for methane are markedly lower (by up to a factor of 10 for GTP100) than the values for GTP presented in the recent IPCC assessment (Myhre et al. 2013). The implications of this approach to calculating GTP are discussed where relevant in the main text; the specific metric values generally do not alter the main conclusions of this study, but they have been shown to influence calculated mitigation costs for some models in specific circumstances

**The Global Cost Potential (GCP) proposed by Manne and Richels, used as a basis for the "MERGE_RF" metric, also included economic considerations (as represented in the MERGE model) and was originally used in combination with a temperature change target. Substitution in MERGE_ETL is derived from radiative forcing expressions from the IPCC Third Assessment Report (IPCC 2001)(Table 6.2)

curves). At a certain high amount of emission reductions, higher prices hardly influence reduction rates. For REMIND, this can be seen in Fig. 1 where all scenarios (except GTP-100) show almost equal methane emissions in 2060 (close to 243 Mt) until 2100 (close to 210 Mt). This increase in reduction potential over time can be attributed to technological learning and is based on Lucas et al. 2007.

In the GTP-100 scenario, models consistently show relatively high methane emissions compared to other metrics. This is a result of the low metric value. The difference of GTP-100 to GWP-100 methane emissions is somewhat greater in REMIND and IMAGE than in the other models. This can be attributed to the higher methane abatement potential considered in these two models. With GTP(t), methane emission reduction also starts late because of flexible valuation, yet it generally leads to the lowest emissions in the long run compared to other scenarios. Here, IMAGE forms an exception. In the model, GTP(t) does not lead to minimum methane emissions before the climate target is reached, despite high metric values later in the century. This is caused by the assumed limitation of year-to-year changes in methane emission reduction levels, as described by Van den Berg et al. (2015). Annual changes in methane reduction rates are assumed to be limited to 2.5 % to 5 % of the baseline emissions, depending on the source. This inertia effect implies that in extreme delay scenarios, it is not possible to achieve similar reduction rates by the end of the century. Therefore, reduction measures have to be taken early enough to fully exploit the potential in later years. This assumption is also relevant because IMAGE is a simulation model, and does not have perfect foresight in its annual investment decisions (that are thus purely guided by the current value of the metric and gas-specific reduction curves).

In general, CO_2 emission profiles are relatively similar for the different metrics in the different models. Scenario differences in CO_2 emissions are mainly the result of compensating

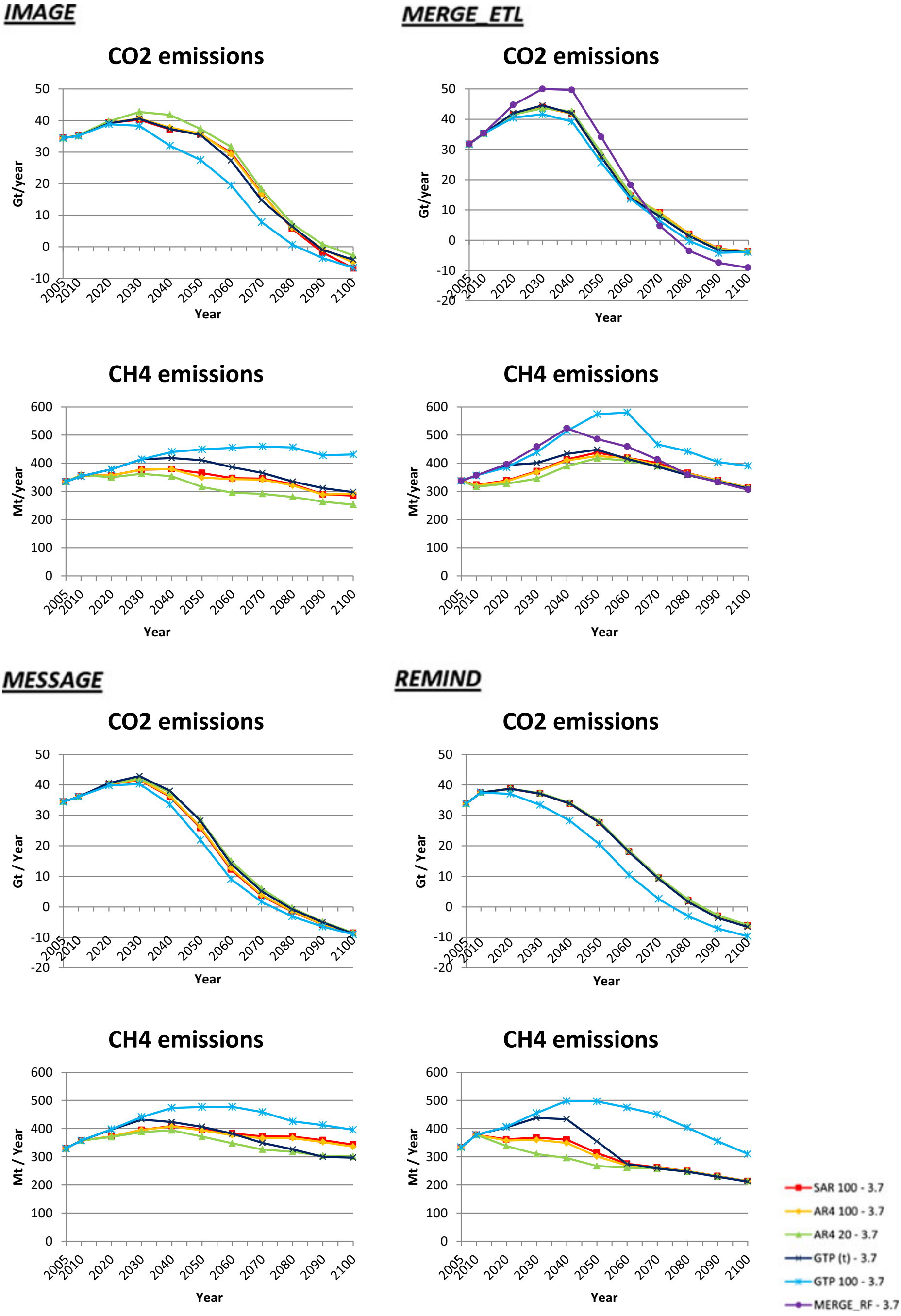

Fig. 1 Scenario results: Emissions

for higher or lower methane emissions. Only the MERGE_RF metric leads to high CO_2 *and* methane emissions early in the century. With this metric, MERGE strongly favours late century mitigation of GHG emissions, including a much larger deployment of bioenergy in combination with carbon capture and storage (BECCS) to ensure negative carbon emissions.

The choice of a metric has a very small effect on the N_2O emission profile (shown in the Supplementary Material). Differences between scenarios are never larger than 5 % (except GTP-100 at 2.8 W/m^2 in IMAGE due to a very high carbon price) and are smaller than 0.5 % for most years and all models.

4.2 Policy costs

Figure 2 shows the carbon price and policy cost profiles in all models for the 3.7 W/m^2 scenarios. The different methods used in the models to calculate policy costs imply that absolute values cannot be compared among them. Therefore, we concentrate on analysis of the relative differences in costs across scenarios. In Table 3, this is shown for all scenarios (including those with a 2.8 W/m^2 target), with the total integrated discounted policy costs in 2100 expressed in relative difference to SAR GWP-100 (discount rate = 5 %).

Overall, cost differences as a result of the use of different GWP metrics are small at the global scale. Particularly, substitution across different values for GWP-100 has hardly any impacts: The AR4 GWP seems to lead to a slightly more cost-efficient climate policy than the SAR GWP, considering the similar result for both forcing targets. Models disagree on the effect of changing the GWP time horizon from 100 to 20 years, but again, the overall effect on policy costs is relatively modest. Only IMAGE in the 3.7 W/m^2 projects significantly lower costs for the GWP 20 (−5.7 %), due to an early start in methane abatement leading to a higher reduction potential (in contrast to GTP (t), further explanation below).

The only metric that consistently leads to higher policy costs is the 100 year GTP. All models agree that using this metric would lead to much more expensive climate policy. Because of very low valuation in the metric, mitigation of methane reduces considerably and as compensation, CO_2 emissions are reduced much more at a far higher carbon price (see Fig. 2). This clearly shows the advantage of a multi-gas strategy. For IMAGE in the 2.8 W/m^2 scenario, this effect almost leads to a doubling of policy costs. The reason is that in this extreme case, the target can only be met with a very early start of CO_2 emission reduction at a much less discounted (and thus higher) carbon price. Note that the numerical value for methane in this metric is uncertain and very low in this study. If a higher value is used, e.g. 4 in Myhre et al. (2013), the policy costs are expected to be lower.

The GTP(t) is closely related to the Global Cost Potential (Tol et al. 2012), and therefore can be expected to be close-to-optimal in terms of the costs to reach a prescribed climate target (Shine et al. 2007). This was confirmed by the earlier modelling study by Reisinger et al. (2013), which made use of MESSAGE, but not found by the other models participating in this study Fig. 3 shows the methane abatement and policy cost profiles in all the models for the GTP(t) 2.8 W/m^2 scenario. Using the GTP(t) metric leads to a shift of methane abatement closer to the target year, which could reduce the mitigation costs by avoiding too early reductions of methane. MESSAGE, and to a lesser degree MERGE_ETL do indeed show lower costs due to optimal timing of emission reductions (up to 4.7 % in the 2.8 W/m^2 scenario). The main reason that MESSAGE shows the lowest costs for GTP(t) is the assumed increase in methane reduction potential at higher carbon prices towards the time horizon. This is shown in a detailed analysis of the methane marginal abatement cost (MAC) curve analysis in the Supplementary Material. This is in line with the earlier findings from Reisinger et al. (2013).

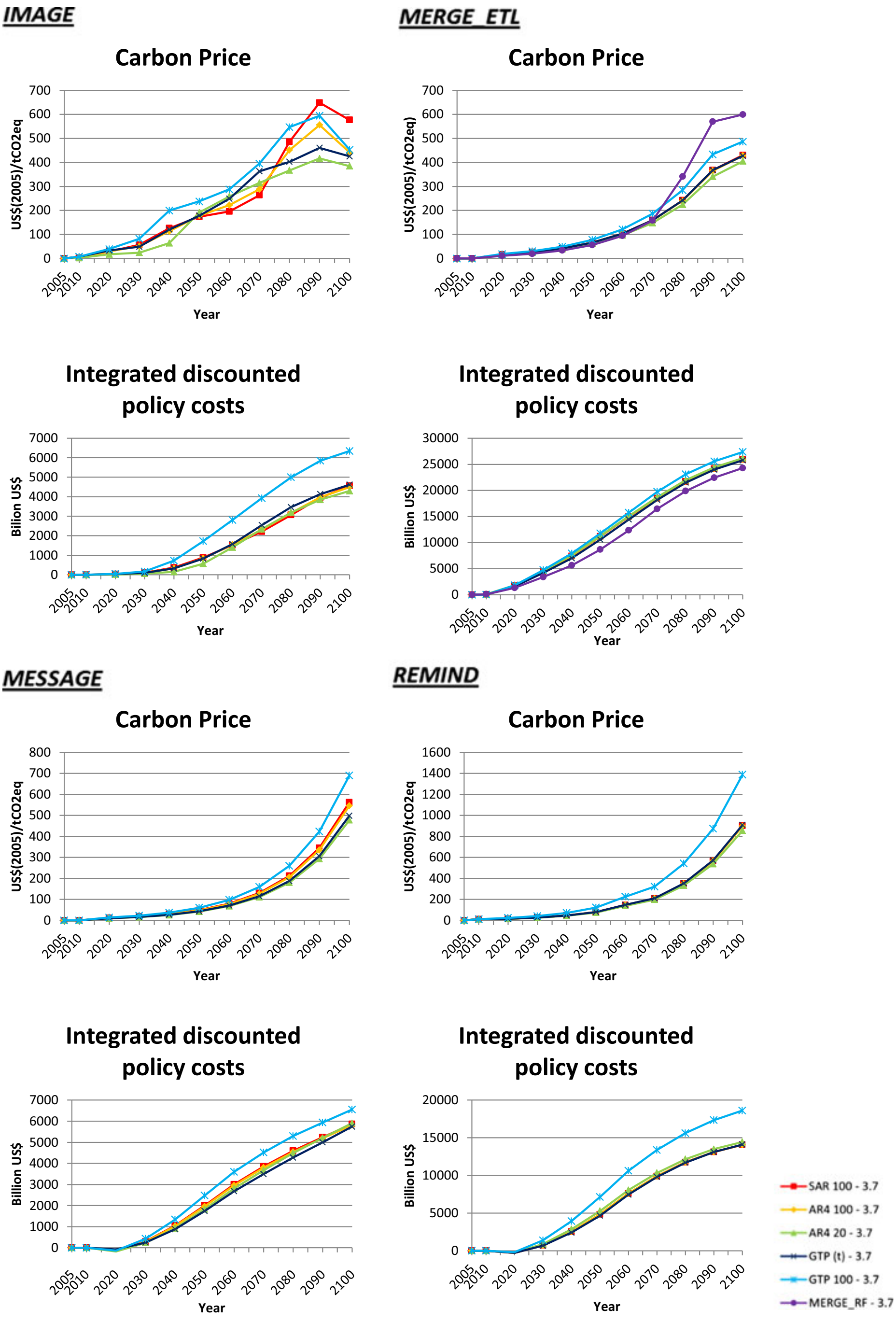

Fig. 2 Scenario results: Carbon price and policy costs (note: different scales on y-axis)

However, there are several factors that might counteract the cost advantage of GTP(t). One reason is that the metric is aimed at a temperature target instead of a radiative forcing target,

Table 3 Integrated discounted policy cost in 2100, for all models and all scenarios, relative to SAR-100 (= 100 %)

RF target	Metric/Scenario	IMAGE	MERGE_ETL	MESSAGE	REMIND	Average
3.7 W/m2	GWP 100 (AR4)	−1.4 %	0.2 %	−0.8 %	0.0 %	−0.5 %
	GWP 20 (AR4)	−5.7 %	1.0 %	0.8 %	2.6 %	−0.3 %
	GTP 100	39.2 %	5.6 %	11.8 %	32.2 %	22.2 %
	GTP (t)	1.4 %	−0.6 %	−2.0 %	0.2 %	−0.3 %
	MERGE_RF (t)		−6.3 %			−6.3 %
2.8 W/m2	GWP 100 (AR4)	−1.2 %	0.3 %	−1.9 %	0.1 %	−0.7 %
	GWP 20 (AR4)	−2.2 %	2.0 %	0.9 %	1.4 %	0.5 %
	GTP 100	89.6 %	7.3 %	13.7 %	20.6 %	32.8 %
	GTP (t)	5.3 %	−0.1 %	−4.7 %	1.2 %	0.4 %
	MERGE_RF (t)		−1.3 %			−1.3 %

although that effect is small given the large correlation between these parameters. The different scenarios that focused on the same forcing target using different metrics lead to very similar temperature levels (see section 4.3).

Another factor is the lack of methane reduction potential (as explained in the MAC analysis in the Supplementary Material). REMIND shows costs that are almost equal to GWP 100 for GTP(t) (in the 3.7 W/m^2 scenario) up to slightly higher policy costs (in the 2.8 W/m^2 scenario). The main reason is that mitigation is limited as maximum reductions are effectively reached, leading to similar CH_4 and CO_2 trajectories (see section 4.1). The slightly higher cost for GTP(t) is caused by stronger CO_2 mitigation to compensate for higher CH_4 emissions at the beginning of the century. It is important to note that the values used in the GTP metric to ensure optimal substitution depend on climate model assumptions such as the CO_2 background concentration. Here, we used the numerical values of GTP(t) as provided in Van den Berg et al. (2015) based on the original study (Shine et al. 2005). The results would have looked different if the climate modules native to the integrated assessment models had been used to derive the GTP(t) values. In fact, in a single model study in which climate metrics were compared with REMIND, GTP(t) did lead to slightly lower policy costs, due to optimal tuning to the model (Strefler et al. 2014).

For IMAGE, GTP(t) clearly leads to higher costs, particularly in the 2.8 W/m^2 scenario. This is a result of an assumed inertia effect in the upscaling of methane abatement measures, which prevents a fast increase in emission reductions (as explained in the 4.1 section). Without this effect GTP(t) is shown to lead to least cost trajectories in IMAGE (van den Berg et al.

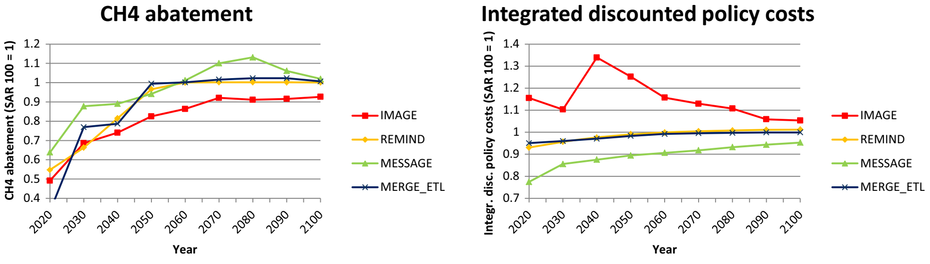

Fig. 3 Methane abatement and integrated discounted policy cost profiles for the GTP (t) scenario, relative to SAR-100 (= 1), for all models (RF target =2.8 w/m^2)

2015). The rapid change in methane emission related policies might therefore be another reason why GTP(t) is less cost-efficient in practice. For the same reason, the 20 year GWP is the most cost-efficient metric in IMAGE. In that scenario, methane emission reductions are maximized because of an early start in implementing abatement measures.

The MERGE_RF metric, only used in the MERGE model, leads to considerably lower policy costs, especially in the 3.7 W/m^2 scenario. This metric is, however, optimally tuned to the MERGE model and will not be optimal in another model. In addition, the potential limitations associated with GTP(t) will also apply to this metric.

4.3 Temperature

The effect of the choice of a certain metric on temperature change, compared to pre-industrial temperatures, is very small. However, GTP(t), GTP-100 and the MERGE_RF metric do lead to slightly higher than average maximum transient temperatures and temperature change rates. In Fig. 4, this is shown with a policy cost/temperature change plot, based on the maximum temperature until the year 2100 (upper panel) and the maximum temperature change *rate* until 2100. The scenario results have been normalized by the values for SAR-100 and are given for all models and both forcing targets.

When only considering the maximum temperature change (upper panel), the highest temperature change levels result from GTP-100 in the 2.8 W/m^2 scenario in REMIND (6 %) and from MERGE_RF in the 3.7 scenario (4.2 %). This roughly corresponds to a 0.1 °C higher maximum temperature than SAR-100. In all models, except IMAGE, GTP(t) also leads to higher

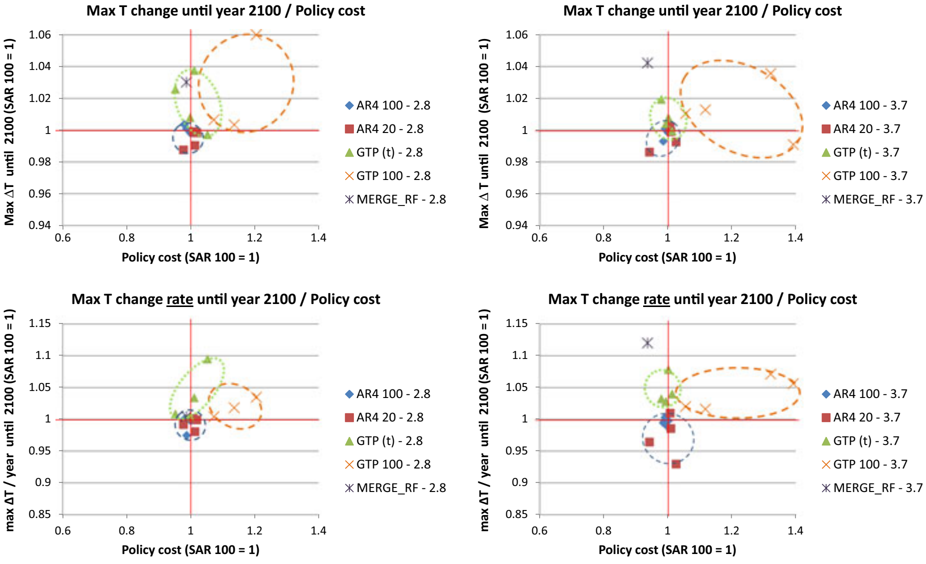

Fig. 4 Policy cost/Temperature change plot. Shown for both forcing targets and all models. Normalized by SAR-100 (= 1), indicated by the red lines. GWP metrics are encircled by blue lines, GTP (t) by green lines and GTP-100 by orange lines. **Upper panel:** Integrated discounted policy cost/Maximum temperature change until the year 2100. **Lower panel:** Integrated discounted policy cost/Maximum temperature change *rate* until the year 2100 (note the different y-axis scale). The GTP-100 2.8 W/m^2 scenario is not shown for IMAGE on this scale (with a policy cost of 1.9 times that of SAR_100, a 5 % lower max T and a 5 % higher max T rate).

maximum temperatures. For GTP-100, GTP(t) and MERGE_RF, the slightly higher temperature is the result of late methane mitigation. Only in the case of MERGE_RF this is further aggravated by late mitigation of CO_2 emissions. One exception to this effect is the GTP-100 result from IMAGE, particularly in the 2.8 W/m^2 scenario. There, due to very early CO_2 mitigation, the maximum temperature change before 2100 is actually 5 % lower, but at a policy cost of 1.9 times that of SAR-100. For that reason, the metric falls outside the scale of the figure and using it for reducing temperature change can be considered highly unrealistic. All GWP metrics (blue circles in Fig. 4), as well as the GTP(t) metric are within 1.5 % or 0.04 °C of the SAR-100 result. Within that range, the slightly lower temperatures in the AR4-20 scenarios are seen across the models. This can be explained because of an early methane mitigation resulting in early radiative forcing reduction and a decrease in temperature.

A roughly similar result emerges when considering the maximum temperature *rate* until 2100 (lower panel). GTP(t), GTP-100 and MERGE_RF generally lead to higher temperature change rates due to a lack of methane mitigation in the first decades (up to 12 % higher for MERGE_RF in the 3.7 W/m^2). The maximum ΔT/year is approximately 0.03 °C/year, so the expected difference in temperature change rate with these metrics is small in absolute terms, not exceeding 0.0036 °C/year and generally leading to less than half that value. GWP-20 can potentially lead to a slightly lower maximum temperature change rate, due to higher short-term methane abatement.

The small differences between metrics imply that the choice of a metric hardly needs to be motivated by the effect it has on the maximum temperature change. This also holds for temperature differences in the target year (see Supplementary Material). Related to this, it can be argued that a metric based on radiative forcing is quite suitable for policy involving temperature change targets, given the proportionality between radiative forcing and temperature levels in the target year.

5 Discussion and conclusion

By using a multi-model analysis, this study aims to understand how different climate metrics influence climate change mitigation strategies and costs. The multi-model approach avoids differences in results due to different experimental set-ups. The models in this study base their projections on a large body of work and aim to include most factors that are relevant to climate change mitigation. The conclusions of this study relate to policy relevant parameters such as emission pathways, policy costs and temperature changes. As such, some conclusions have been found in earlier, single-model studies, and are found to be robust despite the differences across the models. In this paper, we show how model assumptions on emission reduction potentials, inertia and foresight can differently affect the resulting effects of various metrics.

In the past, models have reported different findings with respect to the impact of metric choice on mitigation strategies. Despite model differences, they all find that the different consequences of metrics are primarily caused by diversity in methane emission abatement strategies. Metrics have a clear impact on methane emission reduction levels. Only when models reach their maximum abatement potential, different metrics lead to similar emission levels. This also implies that different model assumptions on methane abatement potential lead to different projected outcomes for the same metric (e.g. as described below for GTP (t)). N_2O emissions reductions are hardly influenced by the metric choice (usually less than 0.5 % difference between scenarios). Differences in CO_2 emissions across scenarios are relatively small and tend to compensate for higher or lower methane emissions.

The time varying GTP(t) can lead to cost optimality under perfect world conditions, but could lose this advantage when implemented. Differences in the projected cost-effectiveness of the metric trace back to model assumptions on methane abatement. Two models showed the time-dependent GTP metric as defined in this study to lead to slightly lower costs than other metrics for achieving a long-term climate target, in line with previous work. An analysis of the models' marginal abatement cost curves for methane showed that the possibility of additional methane reductions at higher mitigation costs contribute to the cost-effectiveness of the metric. However, other model outcomes indicated that the advantage of avoiding too early methane reductions might be counteracted by technical limitations in combination with imperfect foresight. This inertia effect implies that methane reductions have to start long before the target year, making the metric ineffective and costly. Another reason for increased policy costs might be a deviation from the optimal time-variant trajectory of CH_4-to-CO_2 exchange ratios. In all models, we found the cost difference between the time-varying GTP(t) and 100-year GWP to be relatively small (<5 %).

Models consistently show that most GWP metrics that are considered for policy making lead to very similar global mitigation costs. The reason is that all GWP metrics allow for sufficient non-CO_2 emission reduction. Especially substitution between different values for GWP-100 from different Assessment Reports (AR) does not lead to important changes in overall global cost levels (with an average difference in policy costs of −0.7 % to −0.5 % between the second and fourth AR). The same is true for changes in the metrics with different time-horizons (with an average difference of −0.3 % to 0.5 % between a 100 and 20 year horizon).

The 100 year GTP with a low valuation of methane emissions leads to high policy costs in all model projections. Compared to GWP 100 it led to an increase of 6 % to 40 %, and in a single case 90 %. This high cost increase can partly be explained by the very low methane valuation used in this study and implies that constant time horizon metrics need to valuate methane mitigation sufficiently in order to be cost-efficient.

Models agree that GTP(t) and GTP-100 would lead to a small increase in the maximum temperature *rate of change*. However, the effect of the use of different metrics on maximum temperature change is very limited. Although the induced temperature profile could be a relevant strategic consideration in climate policy, the choice of a metric does not have to play a large role. Given the proportionality between radiative forcing and temperature levels in the target year, it can be argued that a metric based on radiative forcing would not lead to ineffective policy aimed at reaching temperature targets.

From a global perspective, and in the long term, the 100-year GWP metric seems to lead to relatively attractive outcomes in terms of mitigation costs and climate outcomes, and no reason is found to replace it as the most common metric used in climate policy. However, there are possible considerations that could lead to alternatives. For policy making, the choice of timing of methane reductions impacts short-term co-benefits with respect to air pollution, costs and temperature. As such, the choice of metric can be used to influence policy decisions in the short term. Alternatively, it is also possible to consider separate abatement strategies for long-lived and short-lived greenhouse gases so that independent choices can be made with respect to the different advantages and disadvantages of reducing short-lived GHG emissions. This would lose, however, the advantage of common framework for short and long-lived Kyoto gases.

Acknowledgments The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2010 under grant agreement n°265139 (AMPERE) and under grant agreement n° 282846 (LIMITS)

References

Den Elzen MGJ et al. (2007) Multi-gas emission envelopes to meet greenhouse gas concentration targets: costs versus certainty of limiting temperature increase. Glob Environ Chang 17:260–280

Deuber et al. (2013) Physico-economic evaluation of climate metrics: a conceptual framework. Environ Sci Pol 29:37–45

Ekholm et al. (2013) Robustness of climate metrics under climate policy ambiguity. Environ Sci Pol 31:44–52

Fuglestvedt et al. (2003) Metrics of climate change: assessing radiative forcing and emission indices. Clim Chang 58:267–331

IPCC (1990) Climate change: the intergovernmental panel on climate change scientific assessment. Cambridge University Press, Cambridge

IPCC (2001) TAR WG1, climate change 2001: the scientific basis, contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

IPCC (2009) Meeting report of the expert meeting on the science of alternative metrics. Plattner G-K, Stocker TF, Midgley P et al (eds). IPCC WGI Technical Support Unit, Bern, Switzerland, pp75

Johansson (2011) Economics- and physical-based metrics for comparing greenhouse gases. Clim Chang 110: 123–141

Johansson et al. (2006) The cost of using global warming potentials: analysing the trade off between CO2, CH4 and N2O. Clim Chang 77:291–309

Kandlikar M (1995) The relative role of trace gas emissions in greenhouse abatement policies. Energy Policy 23: 879–883

Kriegler D. et al. (2014) Making or breaking climate targets: the AMPERE study on staged accession scenarios for climate policy. Technological Forecasting and Social Change.

Lucas D et al. (2007) Long-term reduction potential of non-CO2 greenhouse gases. Environ Sci Pol 10:85–103

Manne, Richels (2001) An alternative approach to establishing trade-offs among greenhouse gases. Nature 410: 675–677

Myhre et al. (2013) Chapter 8: anthropogenic and natural radiative forcing. In: IPCC (2013) WGI climate change: the scientific basis. Cambridge University Press, Cambridge, UK, and New York, USA

O'Neill BC (2000) The jury is still out on global warming potentials. Clim Chang 44:427–443

O'Neill BC (2003) Economics, natural science and the costs of global warming potentials. An editorial comment. Clim Chang 58:251–260

Reisinger et al. (2013) Implications of alternative metrics for global mitigation costs and greenhouse gas emissions from agriculture. Clim Chang 117:677–690

Shine KP (2009) The global warming potential – the need for an interdisciplinary re-trial. An editorial comment. Clim Chang 96:467–472

Shine et al. (2005) Alternatives to the global warming potential for comparing climate impacts of emission of greenhouse gases. Clim Chang 68:281–302

Shine et al. (2007) Comparing the climate effect of emissions of short- and long-lived climate agents. Philos Trans R Soc Lond Ser A Math Phys Eng Sci 365:1903–1914

Smith et al. (2013) Sensitivity of multi-gas climate policy to emission metrics. Clim Chang 117:663–675

Strefler et al. (2014) Economic impacts of alternative greenhouse gas emission metrics: a model-based assessment. Clim Chang 125:319–331

Tol RSJ (1999) The marginal costs of greenhouse gas emissions. Energy J 20:61–81

Tol et al. (2012) A unifying framework for metrics for aggregating the climate effect of different emissions. Environ Res Lett 7:044006

UNFCCC (2011) Greenhouse gases, sectors and source categories, common metrics to calculate the carbon dioxide equivalence of anthropogenic emissions by sources and removals by sinks, and other methodological issues. United Nations Framework Convention on Climate Change, Durban

van den Berg et al. (2015) Impact of the choice of emission metric on greenhouse gas abatement and costs. Environ Res Lett 10:024001

van Vuuren et al. (2006a) Long-term multi-gas scenarios to stabilise radiative forcing - exploring costs and benefits within an integrated assessment framework. Energy J 27:201–233

van Vuuren et al. (2006b) Multi-gas scenarios to stabilize radiative forcing. Energy Econ 28:102–120

van Vuuren D et al. (2011) The representative concentration pathways: an overview. Clim Chang 109:5–31

Weyant et al. (2006) An overview of EMF-21: multigas mitigation and climate change. Energy J