Reply to Reviewer 2

May 15, 2019

First, a more interesting question about SRM is, to me, not the precise level at which we should be using it now but whether we should start using it at all. The current model has no force that would lead a policymaker to delay introducing SRM. I can think of two relevant forces. First, there may be a fixed cost to beginning SRM, whether political, economic, or ecological. Second, the authors assume (I think) that the risk of SRM failure is constant over time. A more realistic model would have a small chance of new research finding a fatal flaw in SRM as long as it is not used and a much larger chance of actual consequences revealing a fatal flaw once it is used. I am suggesting that the probability of failure should be low as long as SRM is not used and then be higher once it is used. A policymaker might then choose to delay beginning SRM until it is really needed, since SRM may not last for long once it is begun. I strongly encourage the authors to explore these (or other) ways of making the paper about when to begin using SRM.

While in the standard scenarios (section 3.2) we indeed assumed constant failure probabilities for simplicity, the failure probability is time dependent in the "realistic storyline" (Section 3.3), although it is still exogenous, i.e. independent of whether SRM is used yet.

Unfortunately, in the dynamic programming procedure, very high additional computational costs are incurred for adding another state variable, for example for tracking the history of whether and when SRM has started. This makes it very cumbersome to include costs depending on how long ago SRM started (e.g. an initial cost to be payed when starting SRM).

As a compromise, we will run a set of simulations with various (high) costs in case of SRM failure, similar to the termination shock damage in section 3.3. The aim is to investigate whether the fear of very large risk associated with SRM failure might lead to delay (of strong decrease) of deployment. In particular, it could be interesting to see whether SRM is (nearly) suppressed until the climate tipping threshold of 2K is reached.

Second, the interesting aspect of SRM value is, to me, its insurance value. The authors analyze a dynamic programming problem, so they have the machinery to answer that kind of question, but they fail to get there. Instead, they analyze perhaps the least interesting form of "tipping" that they could have.

Their tipping point permanently reduces economic output, but the interesting aspect of SRM in a world with tipping points is the potential for SRM to manage a tipping process that is underway. Had the authors chosen to model tipping points in the way that Lemoine and Traeger (2014) did (where a tipping point is a sudden change in a parameter of the physical system, which manifests itself in temperature and economic output only over time and in a fashion that depends on subsequent policy), SRM could play a key role by allowing the policymaker to intervene after triggering a tipping point so as to mitigate the consequences of that tipping point. Alternately, had the authors studied uncertainty about future warming, SRM could have been used to control the consequences of ending up in a high-warming world. Finally, if the model allowed for the more interesting type of tipping point (or for warming uncertainty) along with a reason to delay beginning SRM, then we may see the policymaker delay SRM until a bad state of the world warrants using it. Learning whether/when that plan is optimal would be interesting..

https://are.berkeley.edu/~traeger/pdf/Lemoine%20Traeger_ Watch%20your%20Step_ AcceptedAEJPolic

This is indeed a very interesting suggestion, thanks. We now run an additional simulation (based on the Abatement+SRM scenario) with a different type of tipping point: With a certain likelihood, crossing a threshold of 1.5K leads to an additional positive radiative forcing (thought of as being the consequence of methane release form thawing permafrost).

The preliminary result from this simulation is, however, a bit boring: the policy maker uses a little more SRM than in the standard case to stay away from the 1.5K threshold, so the methane tipping point is only reached in case of SRM failure (in which case the policy maker responds by increasing abatement).

As mentioned in the previous comment, it is hard to make the policy maker delay SRM through a cost or failure probability that depends on history. However, we will investigate whether the combination of high failure costs or higher SRM-related damages (i.e. higher ψ_S) with methane-related tipping can bring about the "insuring behaviour" against a bad state of the world which you mention.

SMALLER COMMENTS

The authors should spend more space clearly laying out the methodological, calibrational, and conceptual contrasts with Heutel et al. They also should elaborate on the differences in results and speculate as to their origin,

Main Methodological differences between our study and Heutel 2018:

- Their optimisation scheme is a 4-day look-ahead scheme, which is not suitable for long-term optimisation.

- They use a different technique to take into account uncertainty of "how harmful unmitigated CO2 is" and how damaging SRM is, namely by assuming the policy maker to be uncertain about climate sensitivity and SRM damage (without learning). In contrast, we use tipping and SRM failure. (Incorporating parameter uncertainty would be very computationally demanding using dynamic programming.)
- Their climate-related damage function is split in a different fashion, ignoring residual climate change (they do include direct effects from CO2, though).
- SRM implementation costs and damages are both implemented in an implausible fashion, namely as being proportional to the *fraction* of CO2-induced radiative forcing removed by SRM. I.e. it does not matter whether the CO2 forcing is strong or weak; the damage induced for compensating X % of this forcing remains the same. In particular, the fact that a high CO2 content requires absurdly high SO2 emission rates (making a full compensation of global warming undesirable) is not included; the maximal damage from SRM is 3% of GDP (full compensation of CO2-induced forcing); we have no such limit.

The effects of the first item onto the results are hard to predict - their optimisation scheme *might* yield unreliable results, but we cannot tell a priori if this will have a strong impact and to which direction.

The second item: Uncertainty in the "harmfulness of CO2" (climate sensitivity in Heutel; tipping in our study) has a larger policy impact in Heutel, at least for the Abatement+SRM-like scenarios in absence of SRM failure, because in our study, Abatement+SRM keeps below the tipping threshold.

The third item would - at least if the SRM-related damage and implementation costs were equal - probably lead to more SRM in Heutel, because SRM always diminishes climate-related damage (namely, the 80% associated with temperature). In our results, even if SRM caused no damage or cost on its own, it would be optimal to compromise between temperature-related and rainfall-related (residual) damage. However, this might be a minor effect.

The last item causes the most obvious differences. In Heutel's case, the higher the CO2 concentrations (and thus the more global warming), the higher the fraction of greenhouse gas forcing that on wants to balance by SRM, as one deletes more global warming damage for the same SRM-induced cost. So Heutel should be more (less) inclined to rely on SRM for high (low) CO2 concentrations, compared to our model.

It seems that in their deterministic results (Heutel et al., 2018; fig. 2), their peak SRM happens to be somewhat similar to ours: In their year 2140, the highest atmospheric CO2 content is about 1800GtC, leading to a forcing of about 6W/m^2, of which they compensate about 50%. In our model, compensating 3W/m^2 of radiative forcing requires 27Mt(S)/yr; similar to our peak abatement of 35Mt(S)/yr (which we reach later than 2160, namely around 2220; see our fig. 2b). However, around 2020, Heutel et al. compensate 10% of a radiative forcing of 1.6W/m^2, which is equivalent to about 1.5Mt(S)/yr in our model. In our results, the injection rate around 2020 is almost 10W/m^2 (compensating 1.5W/m^2 of radiative forcing). So basically, especially during the first 100 years or so, while CO2 concentration is not yet so high, Heutel et al. use considerably less SRM, because they overestimate its cost (i.e. ignore that for low CO2 concentrations, there is relatively little radiative forcing that needs compensation). This might explain why they get higher temperatures than we (see their fig. 2d and ours).

Although they do not explore it, if Heutel et al. had used some high-emission scenario like SRM-only, they would unrealistically find that they can always eliminate 80% of their climate damage (the temperature contribution) for a max. cost of around 3%GDP (max. SRM damage), which would imply that SRM is a very good emergency technology (in case that for some reason abatement is not working or in case of unexpected CO2 release from permafrost).

We will include an abbreviated version of this comparison (focusing on the fourth methodological difference) into the discussion section of our paper.

-Technical notes: 1) The authors describe how they mitigate the problem being non-smooth, but it seems far more straightforward and more realistic to simply assume that the chance of tipping is smooth. Assuming no risk of tipping below some temperature is arbitrary.

2) The authors impose an upper bound on injection rates. This is a curious choice. I would prefer the authors to model the source of the constraint. As it is, either the constraint binds or it does not. If it binds, then it is important, the authors should highlight it in their results discussion, and the authors should consider eliminating it. If it does not bind, then it is unimportant and the authors should consider eliminating it (or at worst justify it in terms of numerical convenience but say it is irrelevant in practice)

1) Any choice of ill-constrained parameters is inherently arbitrary. Ours was inspired by the Paris agreement of staying "well below" 2K warming, because additional warming would be "too dangerous". We could of course have picked a different threshold, such as 1.5K or 0K, or a quadratic increase of tipping probability with temperature. Note that although the 2K (or maybe 1.5K) threshold may be hard to justify on physical grounds, it does match the current political approach of trying to avoid crossing a certain temperature threshold.

We may use a lower threshold when introducing the "methane tipping point" (see second major remarks).

2) Yes, indeed, a threshold was needed for numerical reasons. It has no practical consequences except towards the end of the SRM-only run (which is a somewhat unrealistic scenario anyway). We will clarify this in the manuscript.

Sensitivity checks: I want to see more sensitivity analysis with respect to the fairly arbitrary damage parameters, and especially to the ψS that I'm not *sure has been explored. This should be the core of the results. I am far more interested in learning about what we need to believe to get near-term SRM to be high or low than in learning about the spread of future policy.*

We already have a sensitivity run with ψ_S doubled; this leads to a slightly faster abatement (6 years) and lower peak SRM (ca 30% reduction), but no qualitative change in policy. We will also include a simulation with ψ_S halved.

-I wasn't happy with the discussion of precipitation damages in 2.1.3. There are plenty of ways precipitation can matter, independently of temperature. Think of crops. One could have just as easily said that many temperature channels are mediated by precipitation. Plus geoengineering changes precipitation in ways that are not merely determined by temperature. For instance, might geoengineering change patterns of rainfall (such as monsoons)? Perhaps these kinds of effects could be captured by ψS, about which very little is said other than that its level is arbitrary.

I do not fully understand this remark. We did assume that precipitation can matter independently of temperature, that's why we split the damage function into a precipitation and a temperature component (which are differently affected by SRM).

It was mentioned at the end of 2.1.1 that global mean precipitation changes serve as proxy for the residual climate change - similar to the way that Nordhaus uses temperature change as an indicator for "climate change", including changes which are not directly caused by temperature (e.g. rainfall changes). Nordhaus needed only one climate indicator, because he could assume that all climate change somehow scales with temperature, while we must differentiate between climate change that can / can not be mitigated by SRM.

An interesting aspect about precipitation change is that SRM and CO2 have opposing effects and can balance each other. Therefore even if SRM caused no damages or costs on its own, it might be beneficial to compromise between reducing warming and precipitation changes. This interplay would be lost if precipitation damages were simply included into ψ_S.

We will add a remark for clarification.

-The objective in (10) must be wrong. I think the policymaker must be maximizing expected welfare, not just welfare. And in light of that comment, the motivation for the final paragraph in 2.2 doesn't make a lot of sense. A risk-averse policymaker (as modeled here) has already chosen policy in light of the range of possible cases. It is not internally consistent to postulate caring about percentiles of the subsequent policy performance because of risk aversion.

On eq. 10: indeed, in the stochastic case one must optimise *expected* welfare. We will correct this.

Concerning the motivation of the final paragraph of 2.2: I do not fully agree. Even if one chooses expected welfare as objective for the optimisation, it is interesting to at least look at other criteria such as the percentiles used here (for example, a policy maker might still want to know the probability of "something going quite wrong" under a certain policy). In particular for the realistic storyline scenario, in which SRM is only available with certain probability, it is also insightful to consider not only the mean but include more measures (to verify that the increase in expected welfare compared to Abatement-only comes from those simulations where SRM became available).

We will clarify in the text that the additional measures are not used as optimisation objectves but for additional information only.

-In s 3.2, the authors report that abatement decreases after triggering a tipping point. This is purely an artifact of the type of tipping point modeled here. It would probably not arise if abatement had any role in controlling the consequences of a tipping point, as in Lemoine and Traeger 2014.

Indeed, our preliminary results from the "methane" tipping point (see major remark 2) suggest that with this kind of tipping point, abatement goes up, not down, after hitting the tipping point. We will clarify this, also in the context of the new tipping point.

-Why is the scc unaffected by the possibility of tipping in the abatement+SRM world?

There might be a small effect which however is so small that it vanished when rounding to whole dollars. The reason why this effect is small is that tipping is quite unlikely in this scenario, especially at the early (less discounted) time steps, because unless SRM fails, the temperature is kept below the tipping threshold.

-It seems to me that the final sentence on page 16 would be stronger with a comparison between maximized welfare in the "realistic" world and welfare in the same world if the policymaker used the policy from the abatement-only world. How much could today's policymaker gain by anticipating the future possibility of SRM?

That very last sentence on p16 was meant not so much to stress the potential gain, but to warn against slackening abatement while SRM is uncertain. We will stress your point in earlier in the paragraph and reformulate the last sentence to make this more clear. In particular, we will clarify that for the "realistic" scenario, there is a welfare gain by almost 200% in those cases where SRM becomes available (compared to the "abatement-only" case).

-The discounting result in 3.4 was one of the more interesting results of the paper: a higher discount rate favors SRM because it causes damages, may fail, and doesn't curb future CO2. I would suggest highlighting it more and perhaps doing more with it.

We will highlight the result by mentioning it in the conclusions (to further stress that abatement is indispensable, especially if one “cares much about the future”, while SRM is more a short-term measure).