Response to Anonymous Referee RC 1

We thank the referee for the careful reading and the useful comments and will adapt the manuscript accordingly. Below is a point by point reply with the referee's comments in bold font, our reply in italic font and the changes in manuscript in normal font.

1. Comment from referee:
   The authors assess temperature rises to 2100 only. For some scenarios the global mean temperature will continue to increase well after this date. Those cases should be acknowledged.

   Author's response:
   Such possibility can indeed occur and is implicitly acknowledged in Figures 1, 2 and 4. We do not treat these cases in much detail because a) the response function model is based on 140-year long simulations so extrapolations far into the future are more uncertain and b) such scenarios exceed our temperature targets and are therefore of limited interest in this study.

   Changes in Manuscript:
   In line 40, after "usually taken as the year 2100." we will add the following sentence: "The choice of a particular year is necessarily arbitrary and neglects the possibility of additional future warming."

2. Comment from referee:
   The authors note that 2K warming is commonly seen as a "safe threshold". It may be seen that way, but that is a value judgment subject to considerable uncertainty, and this should be acknowledged.

   Author's response:
   We agree with the referee.

   Changes in Manuscript:
   In line 32, after "the 2 K warming threshold commonly seen" we will add " – while gauging the considerable uncertainty – ".

3. Comment from referee:
   The assessment of delta T depends on the baseline period chosen. This point is addressed later in the report and is said to introduce a sensitivity to the PNR of up to 10 years. The new IPCC special report on warming of 1.5C and 2C indicates potentially large differences in delta T for different baseline choices. It would be nice to see the authors address this issue more explicitly to have confidence that their PNR sensitivity is as low as reported.

   Author's response:
   Within the scope of our model the effect of the baseline on the PNR is such that a lower baseline increases the currently realized warming. Therefore, a given temperature threshold is crossed at an earlier point in time.

   Changes in Manuscript:
   To clarify this point, the paragraph referring to the temperature baseline (from line 293, "This also") will be replaced by the following: "This also illustrates the importance of the temperature baseline relative to which ΔT is defined, as has been found previously (Schurer et al., 2017). Switching to a (lower) 18^{th} century baseline increases current levels of warming by 0.13 K (Schurer et al., 2017) and thereby brings forward the PNR. For example, for a maximum temperature threshold of 1.5K the PNR is brought forward from 2022 to 2016 in the MM scenario and from 2038 to 2033 for the EM scenario."

4. Comment from referee:
   The authors use the concept of "negative emissions" in their simulations, but don't say much about the feasibility of negative emissions. Some elaboration would be helpful for the reader.

   Author's response:
   It is not within the scope of this article to provide a detailed discussion of the question of feasibility of negative emissions, which is a research area in its own right. Scenarios such as the ones presented here and taken from Rogelj et al., (2016a) are usually based on cost-minimization in Integrated Assessment Models (IAMs), and are feasible within the constraints and choices enforced there.

   Changes to Manuscript:
   In line 223, at the end of the paragraph, we will add the sentence "For details on the scenarios refer to Rogelj et al., (2016a). With carbon budgets rapidly running out and the PNR approaching fast,

negative emissions accomplished by geoengineering may have to become an essential part of the policy mix. Such policies are cheap but may only be a temporary fix and lead to undesirable spillover effects on neighboring countries (e.g., Wagner and Weitzman, 2015). We abstract from these discussions here, since this is beyond the scope of the present paper".

We will add: "Wagner, G. and M.L. Weitzman (2015). *Climate Shock. The Economic Consequences of a Hotter Planet*, Princeton University Press, Princeton, New Jersey" to reference list.

5. **Comment from referee:**
The trajectory of warming from the present point to exceeding the specified temperature threshold will not be smooth as it will include multidecadal scale internal variability. That implies that the threshold will not be exceeded at a single point in time, but only in some average sense. The degree to which this is an issue depends on how well the CMIP5 runs represent multidecadal internal variability and how one treats temporal variability and overshoot in relation to the threshold. The authors could provide some discussion of this issue in relation to their analysis.

Author's response:
It is indeed the case that, due to internal variability, crossing the threshold takes place in some average sense. Commonly this is done by temporal averaging over 30 years. In our case, averaging is done in across the ensemble of simulations. Therefore, it is indeed possible to pinpoint the crossing of the threshold (at a chosen probability level) to a given year, as the large ensemble smooths out the variability (Figure 6). The model is not capable of accurately displaying modes of internal variability, nor is it designed to predict (in a one time-series sense) the crossing of the threshold.

Changes to Manuscript:

Before the final paragraph starting in line 345 ("We have shown the constraints..."), we will add the following paragraph:

"In this work a large ensemble of simulations was used in order to average over stochastic internal variability. This allows to pinpoint the point in time where a threshold is crossed at a chosen probability level. Such an ensemble is not possible for more realistic models, nor do GCMs agree on details of internal variability. Therefore, in practice, the crossing of a threshold will likely be determined with hindsight and using 30-year temporal means. This fact should lead us to be more cautious in choosing mitigation pathways."