I recommend publication of this paper because it makes an important policy relevant contribution to the literature. It also builds nicely upon existing work in this field. The paper addresses the validity of integrated assessment models that have been, and may continue to be, used to inform climate change policy. This is an important topic for decision makers.

In particular, it builds on the work of van Vuuren et al. (2009) and Warren et al. (2010) which addresses the validity of the representation of climate change within the simple models. The significant advance that this paper makes upon the earlier works is that implications for the social cost of carbon are quantified. A significant dependence of the values of social cost of carbon upon the representation of climate change within the models is identified, with models being found to variously over or under estimate the values of the social cost of carbon as a result. This is found to be even more significant when the tails of the probability distribution for climate sensitivity are explored.

I recommend publication with revisions to address the following points.

A. Main points

Whilst the paper references the van Vuuren et al. (2009) and Warren et al. (2010) papers, it needs to do a better job of comparing the work done in this study with the work done in the two earlier studies. So, van Vuuren et al. and Warren et al. studied the representation of the following relationships in the simple models (i) relationship between emissions and concentrations (ii) between concentrations and radiative forcing (iii) between radiative forcing and temperature (iv) between emissions and radiative forcing and (v) between emissions and temperature. This study focuses on the relationship between radiative forcing and temperature, and on the relationship between these temperature changes and the calculated social cost of carbon.

It also needs to compare better the results of the previous publications with those from this study. The findings of van Vuuren et al. and Warren et al. are very similar to the findings in this study as far as the relationship between radiative forcing and temperature is concerned. Van Vuuren et al. considered transient responses and both short and long timescales. However, these calculations still need to be included in this paper because they form the basis of the SCC calculations which were not included in the earlier publications. Another important way in which this study builds on the earlier publications is that it includes the Monte Carlo analysis (p. 10-12) which shows that the discrepancies between the simple and complex models are more significant for the lower probability, higher consequence outcomes. This is a very important finding in terms of risk assessment. The paper needs to acknowledge that the earlier work did highlight the issues relating to temperature projection and the results in this study actually agree with these earlier studies, thus supporting them. It then needs to go on to say that the point of this new study is that it presents these two very important findings which build on the previous work: namely the SCC calculations and the Monte Carlo analysis.

The study also considers a scenario extending out to 2300, and is one of very few studies which attempt to do so. Whilst most economists would not consider such timescales, the scenario presented is a perfectly reasonable plausible future and is presented as such, and hence I consider it valid.

Thirdly it needs to examine the specific model versions studied in van Vuuren et al. and Warren et al. and explain whether the same model versions are used in this study, and whether these are the latest versions available.

B. Detailed points

P2 The statement 'their findings suggest that simplified IAMs produce results within the range of more complex models' is not quite right. For example van Vuuren et al. (p. 269) state that 'the response in DICE99 and FUND is slower than that of the comprehensive models included in the IPCC AR4 range'. The statement that 'the long term increase of all IAMs are within the sizeable uncertainty range ...' refers to the long term temperature increase out to 2200 or 2300. The introduction needs to quote that the Warren et al. highlight the potential implications of their findings for the values of SCC that might subsequently be derived, and that this study actually quantifies this. The introduction should also discuss that this paper differs from those two papers in comparing the results with a climate model developed by the authors.

P5. 'The parameters for the distribution are calibrated ...' make clear who calibrated this in what context.
P5. Use of 'slight': cf Fig 5, van Vuuren et al. Are the differences slight? They are perhaps small, and their existence does not invalidate the utility of this study.

P6. Define v. before it is used.

P7. 'these specific values are not crucial for the results presented in this paper'. Can you support this statement by either a logical argument or a statement that some sensitivity runs were carried out?

P10. More clarity needed on this page. Figure number missing. 'In the first experiment' presumably you mean 'the first experiment using the UDEB model'. Figure ?? shows the results from which models? Which parameters were varied? Keeping all this in the Figure caption makes it difficult to follow the procedure.

P10. 'Initially ...the temperature response ... project much slower warming ...' point out that this agrees with van Vuuren et al. (p. 268).

p.10. 'This is a significantly faster rate of warming'. The text around here needs rewording so that the reader can more easily follow the statements about faster and slower warming on different timescales.

p.10 ‘miss this characteristic’. I see that this is true but please explain more clearly to the reader by referring to the Figures, how it is shown there that characteristic is missed.

p.14 the comparison of UDEB with MAGICC, thus validating it against an accepted published code, should appear immediately after the description to UDEB.

P15. The text in the last paragraph needs clarifying with regard to the methodology used. It seems that you fixed the climate sensitivity in the models, but please reword.

P17. This is an extremely important finding. Make more of this in the abstract and conclusion.

p.17. Please explain to readers why when the climate sensitivity is higher, temperature shoots up and then falls, whereas in the median case with lower climate sensitivity it rises and stabilises.

p. 18 Table ??

p. 19. What did Stern use for n, as opposed to Nordhaus? Suppose n was lower? It might be interesting to see those results too in the table, if it does not involve a large amount of extra work.