A Topography of Climate Change Research - Response to Reviewers

Thanks to all reviewers for their detailed comments and feedback on our letter. We feel these have significantly improved our manuscript, particularly in terms of

- Improving, where possible given space restrictions of the letter format, the clarity of methods.
- Increasing the care with which we present our conclusions. This has been the result of comments from across reviewers. We now give a greater emphasis in our conclusion to the limits of the analysis we present, but also note where, notwithstanding those limits, the analysis can offer a complement both to our current understanding of the IPCC and to established IPCC processes in conducting future reports.

Below we have addressed each point in detail in red, referring to changes in the text where necessary.

Reviewer #1 (remarks to the author):

The manuscript "A Topography of Climate Change Research" presents an interesting study of the topics of climate change literature in comparison with IPCC report references. I only was able to get access to two PDF files, the manuscript and the methodological supplement. Other supplementary files are referred to (e.g., topic_comparison.xlsx, docs.csv, topics.csv, and doctopics.csv) in the PDFs, but I couldn't find these files in the submission system.

I am sorry the supplementary files were not available: these were also stored in an online repository: https://doi.org/10.6084/m9.figshare.9009665, which is pointed to in the methods section; we will make sure they are also available in the submissions system.

I wonder if the authors downloaded the around 400k papers from the Web of Science web version or if they have some in-house version of the Web of Science (WoS).

We downloaded the 400k papers from the Web of Science, and stored them in our own database. We have added the sentence "We downloaded the results of the query on March 19, 2019" to the first paragraph of the methods section.

Table SI.1 shows the proportion of references in IPCC reports of different working groups matched with WoS source papers. The proportions vary a lot (from 2% to 70%). The authors

should comment of this. I also recommend a sample check to see why so few references could be matched for some working groups and assessment cycles.

Thanks for highlighting this. We looked in more detail, and using a more sophisticated document embedding approach to identify similar documents, we were able to match some additional references. We have also amended the table, which was not quite clear. It showed the percentage of documents which could be matched to *any* document in our database (which originally was built to store documents from this query, but now also stores documents from other projects). It now shows only the proportion of documents which could be matched to documents from the query used for this paper. We have since checked for false negatives, by manually searching through a sample of documents that were not matched to documents in the query. For each reference, we searched on google as well as the web of science for the document it referred to, as well as to potential variants of the title. We reported if the document was in the web of science, and if so, if it was one of the documents in our query. Of those references which could not be matched 46% of documents were not in the web of science at all, 53% were in the web of science, but not in our query, and 1% were in our query, and mistakenly identified as not being so. A table of the documents sampled is included in the supplementary material, and this process is described in the relevant methods section.

Topics are defined by non-negative matrix factorisation which is an algorithm from multivariate statistics. The methodology is referred to as machine learning. I understand under machine learning that some training data set is provided to a machine learning system to learn associations. Afterwards, the trained machine learning system is used to assign topics to other publications outside the training data set. Therefore, I wonder if the term machine learning is correct in the applied methodology.

This understanding of machine learning refers specifically to *supervised* machine learning, where a set of labels are used to train a model to predict labels for unseen data. This is a subset of machine learning (although it is sometimes used interchangeably), which also includes *unsupervised* machine learning, under which topic modelling is commonly included. The task of topic modelling is to *learn* the set of topics that best represent the data, and non-negative matrix factorization is an approach to doing this. The abstract of the original NMF paper cited (Lee and Seung, 1999) includes the following sentence.

> Here we demonstrate an algorithm for non-negative matrix factorization that is able to learn parts of faces and semantic features of text

I hope this sufficiently clarifies, we do include the phrase: "Topic modelling is an unsupervised machine-learning technique"

The glossary in the supplemental PDF does not explain all abbreviations. I spotted GHGs, WHO, REDD, WG1, WG2, WG3, and AR1, ..., AR6. I wonder if AC for assessment cycle or AP

for assessment period would be better than AR, because the reports appear in single years but cover periods of years.

Thanks for pointing this out. The glossary was intended for those abbreviations which appear in table 1 but cannot be explained in the text. All other abbreviations in the text have been written first in full then with abbreviations in brackets.

In multiple instances the number of 400,000 publications (or more than that) is mentioned. However, the number of publications is below 400,000 in Figure 1.

I have edited the caption to make sure that it is clear that the labels in square brackets show the number of papers in each assessment period. I have also added the total number of papers published until the end of 2018, which is the dataset we use, and is 406,191. The reference to the number of papers in the topographic map has been changed to 378,000, reflecting the fact that only those papers for which an abstract was available are shown here.

On page 3, parts of Figure 2 are referred to as West and East instead on left and right. I find this unusual.

This has been amended to left and right.

Figures 2 and 3 should get meaningful axis labels.

Since the 2 dimensions are the result of a dimensionality reduction algorithm t-sne, which represents similarities and differences between documents across the 140 topics, the 2 axes can not be directly labelled. We have amended the caption to make sure this is clear, and also to include a link to further discussion of how this is done in the methods. Further, we include this sentence in the main text

The two dimensions represent a projection of the 140-dimensional topic scores of each document that seeks to preserve small distances between topically similar documents.

The values on the x axis of Figure 4c look strange. What does an assessment period occurrence of 4.3 mean?

I have amended the figure caption to make things clearer, and pointed to the methods where this is discussed in detail. The axis refers to the center of the topic's distribution across documents, so that a topic occuring only in the 5th assessment period would have a value of 5, while a topic occuring in equal numbers of documents across all assessment periods would have a value of 3.5. Since the number of documents is increasing, the overall distribution of documents across assessment periods is such that values for each topic vary between 4 and 5.

The section "Research representation in IPCC reports" stays at a purely statistical view. The authors should remind the readers of the aims of the different working groups of the different

assessment reports. The discussion should go beyond just comparing proportions of papers and proportions of references. IPCC reports probably references the most useful papers with regard to their aims. Maybe, the statistically found bias matches the aim of the working group. The manuscript would benefit significantly if the discussion would be extended in such a direction.

We have added substantially to the discussion (relevant additions underlined) to make sure this comes through clearly:

> Notwithstanding the over-representation of social science and under-representation of technical solutions in the IPCC with respect to the WoS, a perfectly proportional representation of the literature is of course not optimal. A recommendation that the IPCC cite more or less of any part of the literature is by no means the goal of such an analysis. The IPCC, as a community of scientific experts, is vastly better placed to decide what is relevant than any algorithm. As with many machine learning applications, we should be mindful of David Hume's is-ought problem. Machine learning can help us to more efficiently understand and describe the landscape of climate change literature, but cannot tell us how things should be. The results represent new knowledge about the interaction between the IPCC and the literature, which can have a variety of implications. If the IPCC needs to include more social science knowledge [12], our analysis suggests that this is a result of insufficient production or funding of social science research on climate, rather than IPCC bias. The under-representation of solutions-relevant topics (despite calls for solutions-oriented assessments), and the small proportion of social science research within these topics, suggests areas social science research may be usefully conducted, as well as opportunities for particularly fruitful interdisciplinary collaboration

Later we discuss explicitly the ways that human expertise and machine learning can take on complementary roles

> One advantage of topic modelling is that outcomes are not determined by any categorisation scheme imposed by the modeller, facilitating the discovery of "unsearched" for topics. Highlighting recent research on, for example, membranes, biochar or e-vehicles, could prompt discussion in the scoping process about their inclusion in chapter outlines. This mode of discovery can act as a complement to human expertise, which may be better at identifying under-researched niches, existing biases or knowledge requirements

We also caveat the headline claim made in the abstract:

> We update current knowledge on the IPCC, showing that, at least when compared to the baseline of the literature identified in the WoS, the social sciences are in fact over-represented in recent assessment reports, and that technical, solutions-relevant

The supplemental figures after page 10 should get labels and captions.

Full captions have been added.

In general, the manuscript is well written, but I spotted a few typos:
- Line 27: "... the ratio of studies cited in IPCC reports to the number or relevant studies ..." --> "... the ratio of studies cited in IPCC reports to the number of relevant studies ..."
Fixed

- Line 50: "Figure 2 shows a thematic or topopographic map ... ." --> "Figure 2 shows a thematic or topographic map ... ."
Fixed

- Caption of Figure 2: "Document positions are obtained by reducing the topic scores to two dimensions via t-SNE Documents are coloured by web of science discipline category." --> "Document positions are obtained by reducing the topic scores to two dimensions via t-SNE. Documents are coloured by OECD category."

Fixed

- Lines 149/150: "In the future, making use of more sources climate change knowledge ... ."

Fixed

Reviewer #2 (Remarks to the Author):

Dear authors,

I very much enjoyed reviewing this paper and believe that it represents a novel and significant contribution and recommend that it be published in the journal. I particularly enjoyed the detailed application of machine learning to address big literature and to improve evidence synthesis for decision making contexts. The application of machine learning to assess the representation of the available climate change literature and re-evaluate claims of bias based in IPCC is timely and significant.

However, I do have some comments and suggestions, please see below and in the attached PDFs and hope that thee are of interest and use.

First, I believe that the findings highlighted in the abstract should be moderated a little by stating that the work provides indication of bias, due to the limitations that you indicate (e.g. relevancy of work and lack of consideration of grey literature). Although of course this represents a valuable contribution to the debate.

We agree that this should be better caveated in the abstract, but purposefully avoid use of the word bias, speaking only of over and under-representation with respect to a baseline.

> We update current knowledge on the IPCC, showing that, at least when compared to the baseline of the literature identified in the WoS, the social sciences are in fact over-represented in recent assessment reports, and that technical, solutions-relevant knowledge - especially in the agricultural and engineering sciences - are under-represented

Points regarding the funding and production of more social science literature on climate change, not just greater efforts by the IPCC to include it, and that solution orientated topics may be under-represented are extremely important and could be highlighted in the abstract. It would also be nice to see a bit more of a discussion in the main text regarding interdisciplinarity and the need for this across topic areas, see comments on line 121 and 134 of the PDF.

We agree and add the following to the abstract

> The perceived lack of social science knowledge in solutions-relevant IPCC reports does not necessarily imply a bias towards the natural sciences. It rather suggests a need for more social science research with a focus on "technical" topics related to climate solutions

We also agree it would be good to talk in more detail about interdisciplinarity, although what we can say is limited by the fact that we have disciplinary information only at the document level, so we don't know about collaborative research by interdisciplinary teams. We therefore highlight this as an area for future research:

> A future topography could be improved by making use of more sources of climate change knowledge, extracting and classifying information from full texts, or exploring author networks and interdisciplinarity

Finally, in the conclusion we add the following underlined text to make the implication about interdisciplinarity explicit:

> The under-representation of solutions-relevant topics (despite calls for solutions-oriented assessments), and the small proportion of social science research within these topics, suggests areas for future highly relevant social science research, as well as opportunities for particularly fruitful interdisciplinary collaboration.

Some greater clarity regarding the methodology is needed, particularly for readers who may not be experts in the area. I think this would be improved by greater clarity on what was learnt from the scrapping of the IPCC reports, was this how topics were developed and learnt? How were the disciplines determined, through the machine learning, or author input? more clarity on t-SNE and the 2 dimensions should be provided as many readers will not be familiar with this.

Thanks for these points: we made additions to the first paragraph of the section: "Mapping out the landscape of climate change literature", and to the caption of the figure

> Figure 2 shows a thematic or topographic map of the 378,000 publications on climate change in our dataset with abstracts. Using non-negative matrix factorization [14], the 140 topics are machine-learned from the papers' abstracts (see methods for details). The topic scores of each document are reduced to the two dimensions shown through t-distributed stochastic neighbour embedding (t-SNE) [15]. 2 . The two dimensions represent a projection of the 140-dimensional topic scores of each document that seeks to preserve small distances between topically similar documents

We also provide greater clarity on the use of IPCC citations in the main text, adding the following to the first paragraph of "Research representation in IPCC reports", and making clear that the disciplinary identification is done by WoS (as in previous literature):

> By matching the documents in our dataset to a set of references scraped from all published IPCC reports, we assess the representation of a group of studies by comparing its share in IPCC citations with its share in the dataset of WoS studies on climate change (see methods).
> Figure 4.a shows that social science documents (as identified by WoS) were indeed under-represented in AR3, but by AR5 were the most over-represented discipline, with a share in the literature cited by IPCC reports 1.32 times higher than their share in our WoS dataset. Likewise, social \& economic geography, political science, and "Other social sciences" were better represented in AR5 than economics.

Just to reiterate, the topic model is trained on all documents (as described above), and we use the IPCC citations just to compare shares of documents that were cited by the IPCC with their shares in the wider literature.

I have also added some comments where I believe some minor clarifications or rephrasing is needed in the PDFs and I hope that this is useful.

1 - Rephrase? How we understand what?

This has been changed to "The massive expansion of scientific literature on climate change poses challenges for global environmental assessments and our understanding of how these assessments work"

7 - This should be slightly modified to explain that when compared to the published literature base the result indicate an over representation of social sciences etc. This is in reflection of the limitations that the authors recognises e.g. not taking into account the relevancy of topics to the IPCC and lack of investigation of grey literature

We have added the underlined clause to the abstract:

> We update current knowledge on the IPCC, showing that, at least when compared to the baseline of the literature identified in the WoS, the social sciences are in fact over-represented in recent assessment reports

11 - something here on the implications for interdisciplinarity would be very nice

As discussed in the point on interdisciplinarity above, we include an extra sentence in the abstract about the funding and production of social science knowledge, particularly in "technical" areas, but prefer to leave more detailed discussions of interdisciplinarity to highlighting it as a promising area for future research

14 - regarding climate change or across any topic?

We have added "on climate change"

18 - how was this information generated? Is this a new finding from the scrapping or does it need a reference?

We have added "(see methods)" to point to the explanation of how we generate this information.

23 - These abbreviations need writing in full

We have added the full names, or removed for space

25 - potential rephrase to be more focused on the challenge rather than a criticism, i.e. difficulties for established procedures in scientific assessments like those conducted in the IPCC in exploring the ever increasing literature base

We have changed "fail to address" to "struggle to address". Although I agree that the emphasis should be that this is a challenge rather than a criticism, we report an actual decline in the ratio of studies cited, which I think is fairly represented by "struggle to address".

27 - How was relevant decided? Some additional details required here, e.g. the number of relevant studies identified by....

We have changed "relevant studies" to "studies on climate change in the WoS", as of course not all of those studies may be relevant, but many other studies not in the WoS or not directly about climate change may be relevant. This is properly caveated in the reference given, and indeed is subject to the same limitations discussed here in the later part of the manuscript. However the caveats make this a rather conservative estimate.

28 - the exponential increase in the volume of literature means that the provision of ....

We have changed this as suggested

37 - i don't understand this sentence and think it needs a rephrase

We have changed

Finally, overwhelming amounts of literature may be a major reason why studies of scientific assessments [10, 11, 12] do not offer robust quantifications, for claims about the relationship between report citations and the underlying literature

To

Finally, overwhelming amounts of literature may be a major reason why studies of scientific assessments [10, 11, 12] do not offer robust quantification for their claims about the relationship between report citations and the underlying literature

47 - on?

Changed accordingly

50 - It would be good to see a few more details on how this was done

Given space limitations for the letter format, we would prefer to keep more detailed discussion to the methods section (which is referred to here), however we have added the sentence "The two dimensions represent a projection of the 140-dimensional topic scores of each document that seeks to preserve small distances between similar documents", and the figure caption has also been expanded.

52 - The papers in the IPCC reports? Were these used to learn the topics? If so this needs to be stated

This simply refers to all of the papers in the dataset, we therefore think it is sufficient to leave it as "the papers'", see response to general comment above

54 - each document found in WOS?

This again is each document in the dataset referred to at the top of the paragraph, which we think does not require additional description.

64 - Unclear what this means, communities of scientists/disciplines? I think some more details are needed to make this clear

We have changed to "scientific communities"; we simply meant that the map can be useful for the production of research (including secondary research) as well the production of climate change assessments

78 - topic maps or IPCC assessments?

The sentence was cut for space reasons: we discuss uses in the conclusion

82 - but depends on relevancy to IPCC questions and needs of decision makers rather than an overview of the entire literature base

Absolutely, in this paragraph we simply acknowledge that the baseline is imperfect and point to detailed discussion of the caveats in the discussion

> Where these claims were based on an analysis of IPCC citations [10], they assess this without measurable baseline. In view of the organisation's mandate to provide "comprehensive, objective, open and transparent" assessments of the available science

[7], our dataset of publications allows us - albeit imperfectly, as discussed in the concluding section - to study representation with a meaningful baseline. Further we provide an update to the last quantitative assessment of IPCC citations [10], which looked only at AR3. This baseline forms a starting point for informed discussion about how to represent the literature according to the IPCC's priorities.

87 - In comparison to the articles published and available in WOS in that timeframe

This is made clear by the explicit explanation in the previous paragraph:

By matching the documents in our dataset to a set of references scraped from all published IPCC reports, we can assess the representation of a group of studies by comparing the group's share in the subset of studies cited by the IPCC with its share in the dataset of WoS studies on climate change (see methods for further details

87 - rephrase to: social sciences other than economics

Changed accordingly

92 - higher than

Changed accordingly

99 - early this was abbreviated, please make consistent. perhaps 1st time written in full and then abbreviated thereafter
I think that was actually the first time, but we've gone through and double checked all use of abbreviations to make sure these are properly defined and consistently used

121 - Does this not highlight that within the research there are some communities/topics that do not combine technical solutions with assessments of policy or social implications, or do not do this well. Perhaps this is evidence of the need for more interdisciplinary work in some areas and the findings may indicate that this happens sequentially, i.e. after fields have become established and therefore there could be greater need for interdisciplinary teams from the beginning of research

Yes, see previous comments on interdisciplinarity for relevant edits

125 - Only compared to assessment of literature found through WOS. Again this does not consider the relevancy of the papers to the IPCC or questions being asked. It also does not

consider grey literature. I think the claim can only be that it indicates differences in representations between literature in IPCC reports and the published evidence base. More work would be need to investigate relevancy

This whole section is rephrased to stress that this imperfect baseline is simply a starting point:

Notwithstanding the over-representation of social science and under-representation of technical solutions in the IPCC with respect to the WoS, a perfectly proportional representation of the literature is of course not optimal. A recommendation that the IPCC cite more or less of any part of the literature is by no means the goal of such an analysis. The IPCC, as a community of scientific experts, is vastly better placed to decide what is relevant than any algorithm. As with many machine learning applications, we should be mindful of David Hume's is-ought problem. Machine learning can help us to more efficiently understand and describe the landscape of climate change literature, but cannot tell us how things should be

128 - the research community?
We have deleted "for the community", as to whom the implications are important does not actually need specifying here.

134 - And perhaps investigate why research in this area is not being incorporated

Yes, I think this is implied and for space reasons leave it as such

141 - and so could have application for a large range of disciplines and decision making contexts

As above

150 - more sources on climate change ?

This was a typo, fixed to "more source of climate change knowledge"

Many thanks once again for this interesting and valuable contribution.

Reviewer #3 (Remarks to the Author):

One of the major claims of this article is that it updates studies of IPCC, in particular discussions concerning the underrepresentation of the social sciences. You suggest that, contrary to the main thrust of this literature, the social sciences were overrepresented in the AR5 and technical,

solution-orientated knowledge (particularly in the agricultural and engineering sciences) are under-represented.

There are several elements of this argument, as it is currently presented in the article, that need to be clarified and perhaps nuanced.

1. First, you suggest that there is now a high demand for more solution-orientated knowledge within assessments, using the paper by Kowarsch et al. to support this. However, if we look at one of the main sources of data within this paper - the IPCC's own survey of views on its future work/reform - it indicates that 32 governments submitted comments and one third were supportive of a greater focus on solutions. This then is a sample of 16% of total IPCC member governments. Those familiar with the internal workings of the IPCC know that it tends to be the same governments responding to these surveys - with governments from the global North most frequent submitters. It is important from the outset that this demand to move towards technical, solution-orientated assessments, particularly at the government/policymaker level, is not overstressed (you repeat it throughout the article). In fact, this survey indicates this is a more widely held perspective within the scientific/expert
community rather than demand from governments.

I would question this somewhat. First, of the 32 government responses, 20 (63%) mention solutions (see
https://media.nature.com/original/nature-assets/nclimate/journal/v7/n6/extref/nclimate3307-s1.pdf). Second, the same sample of 16% of total IPCC member governments also represents more than 60% of global GHG emissions, and includes respondents with as diverse interests, needs and standpoints on climate change as the US, China, the EU, the Maldives, Mali, Thailand and Saudi Arabia. A full list of responding governments is available here:
https://www.ipcc.ch/apps/eventmanager/documents/11/280220141142-inf1_future_of_ipcc_govt_comments.pdf. While certainly not fully representative, I would argue that this is still the best evidence we have on government demands for the IPCC.

Still, we alter how we engage with this finding:

> We apply our topic map to understand how IPCC assessments represent the science and respond to policymakers' and consulted experts' demands for more solution-oriented knowledge

Although we stop short of fully caveating this past research for the reasons given above

2. Second, in your discussion of the literature on disciplinary bias you suggest that studies based on citation fail to assess the claim against a measurable benchmark, which your study

provides. However, the scholarship you cite is doing different things - some are looking at the disciplinary expertise of the authorship (Corbera at al), and some of the literature assessed (Bjurström and Polk). Your own data supports the bibliographic analysis on the AR3 (Bjurström and Polk), whereas the study of the AR5 was on authorship and you don't address that. Thus, i think your correction to the literature is overstated and your discussion could be more careful in its representation, discussion and suggested contribution to knowledge in this area, otherwise the important conclusion you draw on line 129-131 is lost, when perhaps it is really what your data adds.

The point I was trying to get across (Several studies have identified, made, or repeated claims of a disciplinary bias of IPCC assessments) was that Bjurström and Polk's work - the conclusions of which with regard to AR3 are as you point out supported by our analysis - have been repeatedly cited (56 times according to the web of science, with 44 of these citations by articles published after 2015), and cited by high profile articles published after the 4th and 5th assessment reports (hence the citation of Corbera et al., which was published in this journal in 2016). This gives the impression that a general bias against the Social Sciences is (still) an established feature of the IPCC assessment process. The fact that the analysis applied only to AR3, and the fact that any kind of baseline necessary to establish bias was missing, have been rather lost.

Corbera et al, in 2016, write, citing Bjurström and Polk:

> More recently, a variety of statistical techniques have been employed to quantify the disciplinary and geographic bias within the IPCC and environmental science research more broadly 16–19. The paucity of the social sciences has been of particular concern to those critical of the IPCC's technical and managerial framings of the issue and seeking to broaden social understandings of the climate change problem.

Further, although the Corbera study on AR5 was on report authorship, it is beset by exactly the same problem: implying bias on the evidence of unequal proportions, without a proper baseline.

We also make efforts to present this as an update rather than strictly a correction:

> Further we provide an update to the last quantitative assessment of IPCC citations [10], which looked only at AR3

Finally, when Bjurström and Polk are cited, the intensity of the conclusions often increases. Hulme and Mahony, in "What do we know about the IPCC" (2010), describe the findings as a "powerful bias to the natural sciences in the construction of 'IPCC knowledge' about climate change".

Especially given the title of Hulme and Mahony's article, but also in the context of the frequent and high-profile repetition of Bjurström and Polk's work, we think the framing of the results as a

update to widely held views in the literature is appropriate. However, based on this and other comments we have attempted to make clear that the baseline we use in the study is very much a baseline (and an imperfect one) rather than a benchmark or target, in a very much reworked first paragraph of our concluding section, which highlights the limitations (and advantages) of our methods, while making explicit that it is those conclusions you identify that is supported by our analysis.

Notwithstanding the over-representation of social science and under-representation of technical solutions in the IPCC with respect to the WoS, a perfectly proportional representation of the literature is of course not optimal. A recommendation that the IPCC cite more or less of any part of the literature is by no means the goal of such an analysis. The IPCC, as a community of scientific experts, is vastly better placed to decide what is relevant than any algorithm. As with many machine learning applications, we should be mindful of David Hume's is-ought problem. Machine learning can help us to more efficiently understand and describe the landscape of climate change literature, but cannot tell us how things should be. The results represent new knowledge about the interaction between the IPCC and the literature, which can have a variety of implications. If the IPCC needs to include more social science knowledge [12], our analysis suggests that this is a result of insufficient production or funding of social science research on climate, rather than IPCC bias. The under-representation of solutions-relevant topics (despite calls for solutions-oriented assessments), and the small proportion of social science research within these topics, suggests areas for future highly relevant social science research, as well as opportunities for particularly fruitful interdisciplinary collaboration

3. Third, in the paper you make an interesting proposition about how these machine-learning techniques could be useful at the scoping stage and you highlight the strengths and advantages of topographically mapping the literature. However, clearly it isn't a silver bullet - the IPCC scoping process is also about identifying and ensuring that demands and requirements of governments are met through the scoping and outlining of a report. Can we see this reflected in the discussion? At what point might it be used? What are its strengths as an analytical tool? What are its limitations? Where could it be misleading?

Briefly summarised: topographic mapping's advantage is that the topics are generated only from the documents themselves, without input by the modeller. This means that unexpected or unsearched for topics can be highlighted. On the other hand, topic modelling doesn't take account of knowledge requirements and can reproduce existing biases. I have added the following paragraph to highlight this:

As a guide for future assessments, the map could facilitate well informed decisions about the representation of different areas of climate literature, from the early scoping process,

Going further and more specifically into how machine learning can aid IPCC processes is probably beyond the scope of this study given space requirements but I think this rounds out the sketch of what contributions could be made. We have added the sentence "The possible contribution of, or drawbacks to data science methods for IPCC processes is an important area for future research" to highlight this as an area requiring more work.

Finally, just a couple of specifics:

1. You suggest that "202,000 papers were published in the Web 15 of Science (WoS)" since the last assessment, are these papers on climate change? Across all disciplines? Relevant to the IPCC assessment? It would also be good to see this more clearly specified when you say " A total of around 350,000 new publications" on …? In fact, your whole discussion in the opening paragraph is a bit vague, can you specify exactly what literature you are referring to? I know that you detail this in your methods section, but i think the opening paragraph needs to be clearer and more explicit about what you are referring to.

The language here has been clarified, with the addition of
202,000 papers <u>on climate change</u>
And
205,00 papers <u>identified in the same query</u>
An explicit link to the methods section is also given.

2. At the moment the colour schemes you are using in your maps of WGs (fig. 3) are not very clear - they seem to blur into each other and it is hard to make out the distinctions, could you try a different colour scheme with greater contrasts to ease with the visual interpretation e.g. black, grey and orange? I'm not sure which would be best, but maybe you could play around a bit with the colours.
Thanks for this comment, I have changed the color scheme to one slightly darker, and based each decision on 3-color qualitative color-blind safe palettes from http://colorbrewer2.org/#type=qualitative&scheme=Dark2&n=3. Other colour schemes with greater contrasts may make differences clearer, but risk making some groups appear larger or more important than others. I hope the new figure is an acceptable compromise.