Response to reviews of “Earth system responses to cumulative carbon emissions”, Biogeosciences Discuss., 12, 9839-9877, doi:10.5194/bgd-12-9839-2015, 2015.

M. Steinacher and F. Joos, November 2015

We would like to thank all four reviewers for their very helpful and constructive comments and their careful assessment of our manuscript. We thank all four reviewers for their time and effort. This is much appreciated.

Please find below our detailed response (bold font) to the reviewers' comments (*italic*). We followed the advice by the reviewers, except for a few cases. A revised version of the manuscript with the modification highlighted (track-changed) is added to this response.

Referee #1

Overall the manuscript deals with a very relevant question, i.e. what is the sensitivity of the Earth system to a doubling of CO2 with respect to surface air temperature but in addition, and that is a new and enlightening aspect, also to a whole range of other target quantities of interest such as sea surface temperature, sea level rise, atlantic meriodinal overturning strength, surface ocean pH, surface aragonite saturation in the southern and tropical oceans as well as soil carbon stocks. Another new aspect is that the authors analyse these sensitivities based on a large ensemble of model simulations constrained by 26 different observational data sets, i.e. providing skill scores for each of the ensemble member. The manuscript is well written and the results are clearly presented. The manuscript only requires some minor clarifications listed below before it can be published.

Thank you. We appreciate that the MS is seen as well written and with results clearly presented.

P 9841, Ll 11-13: One could think of several other impact relevant parameters such as precipitation or extreme events (heat waves, droughts, floods). The exact choice is of course always subjective but maybe the authors could add an explanation why they have chosen their specific list of targets.

Extreme events are not well represented in our energy balance atmosphere model. We modified the sentence to read:
“To this end, we analyzed TCPRE for variables that we deemed impact relevant and also reasonably well represented in our model including surface air temperature, sea surface temperature, sea level, ocean acidity, carbon storage in soils, and ocean overturning.”

We also added a subsection (2.5) in the method section to further explain the motivation and the climate parameters selected for the analysis.

P 9842, L 24: It would be good to define 'metrics' here, in the following sentences the authors give example of these metrics but the term is never explicitly defined.

Sentence modified to include definition:
"Well-defined metrics that summarize the Earth system response to a given forcing by a single or few values are useful …"

P 9843, Ll 22-24: What are these 'recent studies', the authors should clarify this and provide some more details here, are they observational based, if yes, which observations, or model based. What is meant by 'low TCR'?

Sentence modified to read:
"There is an apparent discrepancy between the TCR estimated with the most recent set of Earth System Models (ESM) versus some recent studies that invoke observational constraints (Otto et al., 2013) and simplified models (Schwartz, 2012; Collins et al., 2013). These latter studies suggest the possibility of a TCR below 1°C, i.e. outside the very likely range given in the Fifth Assessment Report of IPCC (Collins et al., 2013)."

P 9844, L 1: It should be '21st' century.

Corrected.

P 9845, Ll 4-9: It would be good to provide some more details on the observational data sets: where do they come from, what are the temporal and spatial characteristics, what are the uncertainties?

We extended the text on the observational constraints significantly and added a new subsection (2.1.2) "Observational constraints and the computation of skill scores" in the method section. The extended text provides details about the origin, characteristics and uncertainties of the observational data sets.
In addition, the new figure A2 in the appendix gives an overview of the observational data sets.

P 9845, L 18: The cumulative skill should be the sum of all Sm over m. The authors should clarify this and add this in the manuscript.

We clarified this in the manuscript:
"3931 out of the 5000 ensemble members contribute less than a percent to the

cumulative skill $\sum_m s_m$ of all members *m* and are not used any further.”

P 9858, Ll 1-21: Does the observational constrain on the posterior distribution depend on the order of applying the different data groups, for instance would it make a difference if ‘heat’ is applied before ‘CO2’ compared to the other way around?

No, the posterior distribution (after all constraints have been applied) does not depend on the order of applying the individual constraints. Where we apply the constraints sequentially (Fig. 6b and 6d in the revised manuscript), the order obviously matters for the intermediary steps because the subsets of constraints will be different for different orders, but at the end when all constraints are included, they all end up with the same posterior distribution independent of the order.

We added a clarifying note to the text:
“Please note that the fully constrained posterior distribution does not depend on the order of applying the individual constraints.”

P 9861, Ll 15-20: What would be an appropriate time scale for SAT records to constrain TCR or ECS?

This is a question related to the ratio between signal (SAT change due to external forcing) and noise (SAT change due to internal, unforced variability). For the industrial period, internal variability (e.g., the recent slow down in the warming trend) does significantly affect trends on decadal to multi-decadal time scales and on the hemispheric and global scale. Also the determination of radiative forcing is generally affected by larger percentage uncertainties for short than for long periods. It is also a question of the spatial scale considered. On the regional or local scale (grid cell within a model) the influence of internal variability is generally larger than on large spatial scales. Number and quality of SAT data may degrade when going back in time making large-scale SAT reconstructions more challenging. The selection of century-long and hemispheric-scale SAT records appears appropriate in the context of our study.

Figures: The figures could be increased in size for better readability, esp figures 2 and 3.

We will ask the publisher to make the figures as large as possible when typesetting the final paper.

Referee #2 (A. J. Dolman)

This is a nice, timely paper addressing the issue of linearity between cumulative emissions and a range of climate and ocean variables. The authors also quantify the equilibrium climate sensitivity (ECS) and transient climate response response. For surface temperature this linearity is by now well established (if maybe not completely understood), for other variables not. The main contribution of the paper is that it sets out to investigate this is a structured way with an constraint model ensemble and a large set of emission scenarios. This allows the authors not only to address scenarios uncertainty but also model, response uncertainty. However, it takes several readings to get this message across to te reader, and the paper may benefit from some clear rephrasing of the goals and the steps to approach that. Perhaps a flow chart after figure 1 could do that. For instance section 3.1 is entitled climate response to an emission impulse, whereas really the objective is to see whether, using IRFs the response of any variable X is linearly related to the emission pathway. Keeping the reader fixed to that ultimate goal helps the readability. In general the paper is very densely written with little attention to helping the reader stay focussed on the key questions. Some rephrasing and additional lines would be easy to insert, but turn the paper into a real scientific paper rather that a report.

Thank you for these valuable comments. In response, we have expanded the abstract, introduction, the method, the result and discussion section to express and explain more clearly the goals of the manuscript and of individual subsections. We would like to refer the reviewer to the revised manuscript (in track change modus) supplied with this response for a complete overview of changes. A few revisions are given below to provide a few examples of our changes.

For example, the paragraph defining the goals of the paper has been expanded and moved towards the end of the introduction. The last paragraph of the introduction now explains the structure of the paper. The text reads:

"The goals of this study are (i) to establish the relation between cumulative CO_2 emissions and changes in illustrative, impact-relevant Earth System parameters, (ii) to quantify TCRE, TCR and ECS, and (iii) to establish the response of different Earth System parameters to an emission pulse, i.e. the Impulse Response Function. In analogy to TCRE, we introduce a new metric, the Transient Climate Parameter Response to cumulative CO_2 Emissions (TCPRE). $TCPRE^X$ is the change in a climate parameter, X, in response to cumulative carbon emissions of 1000 GtC. To this end, we analyzed TCPRE for variables that we deemed impact relevant and also reasonably well represented in our model including surface air temperature, sea surface temperature, sea level, ocean acidity, carbon storage in soils, and ocean

overturning. The link and the linearity between the responses in the different variables and cumulative CO_2 emissions is investigated in a structured way with an observation-constraint model ensemble and a large set of emissions scenarios. This allows us not only to address scenarios uncertainty but also model, response uncertainty. We quantify uncertainties related to specific greenhouse gas emission trajectories, i.e. scenario uncertainty, by analyzing responses to CO_2 emission pulses as well as to a set of 55 scenarios representing the evolution of carbon dioxide and other radiative agents. The response uncertainties for these scenarios are quantified with a ~1000-member model ensemble constrained by 26 observational data sets in a Bayesian, Monte-Carlo-type framework with an Earth System Model of Intermediate Complexity. The model features spatially-explicit representation of land use forcing, vegetation and carbon dynamics, as well as physically consistent surface-to-deep transport of heat and carbon by a 3-D, dynamic model ocean, thereby partly overcoming deficiencies identified for box-type models used in earlier probabilistic assessments (Shindell, 2014b, a). This allows us also to reassess the probability density distribution, including best estimates and confidence ranges, for the Equilibrium Climate Sensitivity (ECS), the transient climate response (TCR), and the Transient Climate Response to cumulative CO_2 Emissions (TCRE).

This paper is structured in the following way. In the methods section , the modeling [...]"

We have expanded figure 1 showing a flowchart to cover also 2xCO_2 and CO_2 emission pulse simulations.

The title of subsection 3.1 now includes the word linearity:
"Climate response to a CO_2 emission pulse: testing the linearity of the emission-response relationship"

And additional text at the beginning of this section reads:

"A main goal of this section on IRF is to discuss to which extent one may expect a close-to-linear relationship between cumulative CO_2 emissions and a climate parameter of interest. A linear relationship between emissions and parameter X has the advantage that the determination of $TCRPE^X$ does not depend on the choice of scenario nor the magnitude of CO_2 emissions. In addition, $TCRPE^X$ would in the case of linearity fully describes the parameter response to any CO_2 emissions. We start with a description of the model setup, followed by theoretical considerations, then discuss linearity in the context of CO_2-only scenario uncertainty by analyzing median model responses, then turn to the response uncertainty by relying on the full model ensemble and compare scenario and response uncertainty. Finally, we

briefly address scenario uncertainty due to non-CO_2 forcing.“

Other comments

Title. It may be worthwhile to insert the word linearity somewhere in the paper.

We changed the title to: “Transient Earth system responses to cumulative carbon dioxide emissions: linearities, uncertainties and probabilities in an observation-constrained model ensemble”

Abstract
L 3 is rather than are
Corrected.

L4 response rather than responses (and is rather than are)
Corrected.

L11 replace in steric by a steric
Corected.

P9841 L10-28. I would prefer to see this paragraph at the end of the introduction. **Done. Paragraph moved and expanded as explained above.**

P9846 l 1-3. How much cooling is effected by this assumption, and how does this effect the climate sensitive further down in the paper. This is an important issue, and while the authors appear to be aware of it, they skip over it a little too lightly.

The constant radiative forcing from aerosols assumed for the 23 AME scenarios after 2005 is -1.17 W/m2, whereas the average of the other 32 scenarios is -0.40 W/m2 by 2100 [range -1.19 – -0.10 W/m2] (Fig. A1f; Steinacher et al. 2013). This negative forcing has roughly the same order of magnitude as the positive forcing from non-CO_2 greenhouse gases after 2100 at the lower end of the scenario range, and therefore those effects tend to cancel out in the AME scenarios (Fig. A1b).

It is important to note, however, that this issue does not affect our estimates of TCR, ECS, and TCRE. Those metrics are calculated from CO_2-only simulations (1%/yr increase scenario) without aerosol and without non-CO_2 greenhouse gas forcing. The process of constraining the model ensemble is not affected either, because the model results are evaluated over the historical period where the observation-based aerosol forcing is prescribed rather than a scenario.

To emphasize this in the paper we added the forcing time series in the appendix

(Fig. A1) and extended the text:
"We therefore make the conservative assumption of constant aerosol emissions at the level of the year 2005 (-1.17 Wm^{-2}), which implies a significant cooling effect continued into the future in those 23 (out of 55) scenarios (Fig. A1f). Please note that this effect does not affect or estimates of TCR, ECS and TCRE, which are based on the CO_2-only simulations described below, nor the constraining of the model ensemble with observation-based data over the historical period."

P9846 l21 definitions
Corrected.

P9847 L21. I do not quite understand why the weighing with the model scores is performed here. I understood that you do that in the first steps, so is this really necessary? Or is is double weighting?
There is no double weighting. Equation 4 is the only step where the model score S_m is applied, when calculating the probability density functions. The steps described in the previous section concern only the calculation of the score S_m for each ensemble member. No weighting is performed in those first steps, the actual constraining of the ensemble is performed when calculating the PDFs (Eqs. 3-5).

We have extended the section 2.3 "Calculation of probability density functions" to explain the procedure in more detail (along with the general extension of the methods section; see responses to other reviewers).

P9853. L14. It may be better to use the verb used than ran in the case of the scenarios and refer to Steinacher et al (2013).

Changed to: "We used the model ensemble presented in Steinacher et al. (2013) for 55 multi-gas emission scenarios from the integrated assessment modeling community which range from very optimistic mitigation to high business-as-usual scenarios (Methods)."

P9853 L 28-29. I cannot see how the AME scenarios fit in 3c. Do you mean 3b? similarly on 9854 l 14-17.

Corrected. We had relabeled the figure without adjusting all the references in the text. Thank you for pointing this out.

P9861. L 17-20. It may be worth to have a look at van de Werf and Dolman, Earth Syst. Dynam. Discuss., 5, 529-544, 2014

Thank you. We refer to this interesting study in the introduction:
"van der Werf and Dolman (2014) applied a multiple regression approach using

historical temperature and radiative forcing data to find that recent temperature trends are influenced by natural modes of variability such as the Atlantic Multi-decadal Oscillation. They estimated TCR to be above 1°C using century-long records."
The study is also cited in the discussion section when we address the influence of internal climate variability on observation-derived estimates of TCR and ECS.

Referee #3

General Comments:

This study is heavily reliant of work presented in Steinacher et al. (2013) to the point where I had to reread that paper and its supplementary materials to understand what the authors had done in the present study. This degree of recursive reading is unnecessary for a journal like Biogeosciences which has no hard length limit. I recommend giving a full description of the perturbed parameters experiments in the methods section, instead of referring to Steinacher et al. (2013). I recommend adding a table containing the parameters that were perturbed, the prior ranges, and a description of how it was decided which parameters to perturb.

We have extended the method section significantly. In the new subsection 2.1.1 "Model parameter sampling" we provide an extensive description of the parameters and the selection process. Further, we added a table (Table A1) in the appendix with the parameters and their prior ranges.

From the supplementary information from Steinacher et al. (2013) it does not appear that the authors have perturbed any ocean carbonate chemistry or biology parameters. As TCRE arises partially from ocean carbonate chemistry modulating the airborne fraction of carbon (MacDougall and Friedlingstein 2015, Goodwin et al. 2015), this limitation should be very explicitly stated in the manuscript.

We included the following in the methods:
"The ocean carbonate chemistry and marine biology parameters are not perturbed in this study in order to save computational costs. The ocean chemistry is very well understood and the relevant parameters are already well constrained (Dickson 2002). The marine biology parameters are considered of secondary importance for this study, and when compared to the parameters affecting the physical transport and uptake of anthropogenic carbon (Joos et al., 1999; Plattner et al., 2001; Heinze et al., 2004; Gangstø et al., 2008; Kwon et al., 2009)."
And in the discussion section:
"Further, we note that no ocean carbonate chemistry or marine biology parameters were varied in this study."

The definition of TCRE used in this manuscript is confusing. TCRE is the constant of proportionality between cumulative emissions of CO2 and change in near surface air temperature (Gillett et al. 2013). TCRE does not include the effect of non-CO2 radiative forcing and certainly should not be used to describe changes in variables other than near surface temperature change, as is done here in equation 8. The response of peak temperature to cumulative emissions is the Cumulative Warming Commitment (Allen et al. 2009), which is the sum of TCRE and the Zero Emissions Commitment. I recommend that the generally accepted CO2 only definition of TCRE be used in this manuscript.

As discussed in the manuscript we did not find a consistent use of TCRE in the literature. We agree, however, that this is confusing and changed our notation throughout the manuscript. The general transient response in a variable X to cumulative CO_2 emissions is now denoted $TCPRE^X(t)$ and $TCPRE^X{}_{peak}(t)$ instead of TCRE(t,X) and $TCRE_{peak}(t,X)$. As suggested we use 'TCRE' according to the CO_2-only definition and we clarified this in section 2.2 "Definition of TCRE and $TCPRE^X$":

"TCRE is used in this study as defined by Gillett et al. (2013). Thus, TCRE is equivalent to $TCPRE^{\Delta SAT}$ derived from a simulation with prescribed 1% yr^{-1} atmospheric CO_2 increase and no other forcings."

The dataset of simulations used in this study present a good opportunity to discern the effect of non-CO2 forcing on the linear relationship between cumulative emissions and near surface temperature change. The authors have done some of that implicitly by calculating conventional TCRE (with the 1% experiment) and the proportionality between cumulative emissions of CO2 and temperature change with all forcing included. This could be made much more explicit.

Unfortunately this is not that easy. Nonlinearities in the response to cumulative emissions in the multi-gas simulations may arise from either the non-CO_2 forcing or the CO_2 emission pathway (scenario dependency). Those effects cannot be distinguished properly without performing additional simulations with the same CO_2 emission pathway but without the non-CO_2 forcing.

There is no description in this manuscript of how the model was spun-up. Were all 5000 model versions spun-up or was some other method used to save computing resources? The lack of proper spin-ups crippled the analysis of perturbed physics experiments conducted by the Hadley centre a decade ago (e.g. Collins et al. 2007). Therefore, it would benefit this study to have a full description of how this was done for the present experiments.

We added a new subsection "Spin-up procedure" (2.1.3) to the methods section that

describes how it was done and checked.

The final section of the results discussing the effect of datasets on the PDFs of TCR and ECS is very interesting and could be expanded for clarity.

Done. In particular the following text is added:

"Interestingly, also the ``$CO_2$'' subgroup narrows the probability distribution for TCR and ECS, although less than the SAT and ocean heat records. The ``CO_2'' subgroup includes data sets of the atmospheric CO_2 increase over the industrial period and observation-based estimates of the ocean and land carbon uptake for recent periods. Ocean carbon and heat uptake are governed by similar processes, namely the surface-to-deep transport of excess carbon and heat from the surface to the deep ocean. Apparently, model members that are not able to describe the ocean carbon uptake and the evolution in atmospheric CO_2 reasonably well, fail also to match observational records for SAT and ocean heat content. The PDF for the ``CO_2'' subgroup displays several maxima for ECS and similar for TCR. We are not in a position to provide a firm explanation for these maxima, but speculate that this result may be related to the limited number of members in our ensemble and that the multi-dimensional model parameter space is not completely sampled."

The discussion and conclusions section drifts into subjects not discussed in the introduction, such as the hiatus and the 2 K temperature change target. It may make more sense to split the discussion and conclusions section and include these subjects in a policy implication subsection. In its present form the conclusions from this study are not clearly articulated.

We split the section into a "Discussion" section with multiple subsections and a "Summary and Conclusions" section as requested, but did not introduce an additional policy implication subsection.

We revised the wording regarding the paragraph on internal climate variability (the hiatus) and its influence on estimates of TCR and ECS. It was not our intention to introduce new subjects, but to discuss our results for TCR and ECS. The following text was added:

"Another focus of this study is to provide observation-constrained estimates of the Transient Climate Response (TCR), the Equilibrium Climate Sensitivity (ECS), and the Transient Climate Response to cumulative CO_2 Emissions (TCRE) as determined from CO_2 -only scenarios. The recent slow-down in global surface air-temperature warming [...] On the other hand, our results do not confirm some recent studies (Otto et al., 2013; Schwartz, 2012; Collins et al., 2013) that suggest the possibility of a TCR below 1°C. Such low values for TCR are outside the very likely range given in

the Fifth Assessment Report of IPCC (Collins et al, 2013) and of this study.

The choice and record length of observational constraints may bias results for TCR and ECS. In particular, internal climate variability, e.g., associated with the Atlantic Multi-decadal Oscillation, may obscure the link between anthropogenic forcing and response (van der Werf and Dolman, 2014). [...]"

We also articulated the link between the 2 degree target and the subject of this paper, the transient climate parameter response to cumulative CO_2 emissions more clearly (section 4.2 "Climate targets, allowable emission and TCPRE^{X}").

The relationship between cumulative emission of CO2 and various Earth-system metrics was explored by Herrington and Zickfeld (2014). An intermediate complexity ESM was used by Olson et al. (2012) to estimate PDFs of TCR and ESC. These studies need to be cited and discussed within the context of the present study.

Done. These studies are now cited and discussed and we added additional citations to earlier probabilistic studies. E.g. "The 68% c.i. includes the scenario uncertainty range in TCRE (1.5 to 2.0°C) obtained by Herrington and Zickfeld (2014) with a single model setup and for a range of CO_2-only scenarios (with constant future non-CO_2 forcing)."

"This sensitivity is larger than found by Herrington and Zickfeld (2014), but simulated changes in AMOC are known to be model dependent."

"A tendency for the TCRE to decrease with increasing cumulative emissions is noted in earlier studies (Herrington and Zickfeld, 2014; Gillett et al., 2013; Matthews et al., 2009), while Krasting et al. (2014) find TCRE to be large for low and high emission rates and low for modern emission rates in idealized scenario in the GFDL model."

"However and in contrast to reduced-form, box-type, two-dimensional, linear response models, expert assumptions, or component models applied in many earlier probabilistic assessments (e.g. Wigley and Raper, 2001; Knutti et al., 2002, 2003, 2005; Schleussner et al., 2014; Bodman et al., 2013; Little et al., 2013; Harris et al., 2013; Holden et al., 2013; Bhat et al., 2012), the Bern3D-LPJ features a dynamic 3-dimensional ocean with physically consistent formulations for the transport of heat, carbon, and other biogeochemical tracers, similar to work by Holden et al. (2010) and Olson et al. (2012), and includes a state-of-the-art dynamic global vegetation model, peat carbon, and anthropogenic land use dynamics. The model is applied directly without using an emulator (Holden et al., 2010; Olson et al., 2012; Holden et al., 2015)."

Specific Comments:

Title: The title poorly described the study conducted. I would recommend something more like: "Estimating uncertainty in Earth system response to cumulative emissions utilizing a large perturbed physics ensemble."

We agree that the title was too general and we changed it to: "Transient Earth system responses to cumulative carbon dioxide emissions: linearities, uncertainties and probabilities in an observation-constrained model ensemble"
(See also answer to Referee #2.)

Page 9840 line 23: "damage from" not "damage by"
Corrected.

Page 9840 line 25: change to "from burning of fossil fuels"
Corrected.

Page 9841: The first paragraph is poorly written please re-write for clarity.

Done. The opening paragraphs are reformulated.

Page 9843 line 8: change to (e.g. Huber and Knutti . . .
Changed.

Page 9843 line 21: TCRE is the "Transient Response to Cumulative CO2 Emissions" not the "Transient Response to Cumulative carbon Emissions". Please change throughout the manuscript.
Changed.

Page 9843 line 16: Change "interesting" to "useful"
Changed.

Page 9843 line 21: Add citation to Gregory et al. (2009).

Done. Reference added.

Page 9843 line 23: Spell out "versus".
Changed.

Page 9843 line 26: Use "Shindell" instead of "He"
Changed.

Page 9844 first paragraph: This would be a good place to mention Olson et al. (2012).

Done. We added Olson et al., 2012 and some other relevant citations:

"Probabilistic assessments are widely used to determine probability distributions of model parameters, climate sensitivity and many other climate parameters (e.g. Wigley and Raper, 2001; Knutti et al., 2002, 2003; Murphy et al., 2004; Knutti et al., 2005; Schleussner et al., 2014; Bodman et al., 2013; Little et al., 2013; Harris et al., 2013; Holden et al., 2013; Bhat et al., 2012; Holden et al., 2010; Olson et al., 2012; Williamson et al., 2013). The Bern3D model was used in connection with an Ensemble Kalman Filter to constrain model parameters and regional air-sea carbon fluxes from observations (Gerber et al., 2009; Gerber and Joos, 2010, 2013)".

Page 9844 line 26: Table S1 seems critical to understanding this study. Please reproduce the table in this manuscript.

The table is now reproduced in the appendix (Table A1) and the method section has been extended in this respect (see also answer to general comment above).

Page 9846 second paragraph: Why were pulses done from 389 ppm instead of the pre-industrial climate?

Here, we follow Joos et al. (2013) and IPCC AR5. IPCC AR5 uses the pulse response for a background of 389 ppm to compute Global Warming Potentials. The response in atmospheric CO_2 to an emission pulse depends on the background conditions (Joos et al. 2013). In the context of future scenarios it makes sense to analyze the response to an emission pulse under present day conditions rather than preindustrial.

Page 9847 line 14: The peak temperature will not necessarily occur in 2300 CE. Is the temperature in 2300 being used as an approximation for the peak temperature? In many scenarios would peak temperature occur before 2300?

No. According to equation 2 the peak temperature response to cumulative CO_2 emissions is defined as the maximum temperature between 2000 and 2300 divided by the cumulative CO_2 emissions by the year 2300. The temperature peaks generally before 2300 CE, but this might not be true in all cases. We changed and extended the text to clarify:

"For the transient response analyses, $TCPRE^X(t)$ is computed for every year t in the range $2000 < t \leq 2300$ (i.e. 300 data points per simulation). In contrast, the peak

response is represented by only one data point per simulation. It is the value of X(t) at the time t_{max} , i.e. where the maximum absolute change between the years 2000 and 2300 occurs, divided by the cumulative emissions in the year 2300, E(t = 2300), and denoted $TCPRE^{X}_{peak}$ (t = 2300). Please note that the actual peak response might occur after 2300 CE, in which case $TCPRE^{X}_{peak}$ is only an approximation. Surface air temperature usually peaks before 2300 CE in the applied scenarios. Steric sea level rise, on the other hand, continues to increase after 2300 CE due to the large thermal inertia of the oceans."

Page 9847: The term "convex hull" needs to be explained here. I suspect that for most readers of BGS (including myself) without explanation "convex hull" will only incite dim memories of long-ago geometry classes. 12, C3842–C3848, 2015
The abbreviation PDF (probability distribution function) is never spelled out in the manuscript.

We extended this section for clarification and added a definition of the term "convex hull":
"Technically, we use the the convex hull which is the smallest region containing all points such that, for any pair of points within the region, the straight line segment that joins the pair of points is also within the region."

The abbreviation PDF is now spelled out in the revised section.

Equation 8: TCRE is strictly defined in terms of temperature. Adopting this formulation confuses the meaning of TCRE. Please remove this equation and find a better way of explaining what you have done.

We changed this and introduced the more general response $TCPRE^{X}$ (see answer to general comments above).

Page 9851 line 16: CE (Common Era or Christian Era) not AD (In the year of our Lord).

Changed.

Page 9852 line 23: The defining and most useful feature of TCRE is that it is nearly path-independent (Matthews et al. 2009) so this statement seems out of place.

Canged to the more general response $TCPRE^{X}$ (see also answer to general comments above):

"Consequently, the model uncertainty will dominate the uncertainty in $TCPRE^{X}$ and is larger than uncertainties arising from dependencies on the carbon emission pathway."

Page 9853 line 24 to 27: I think the authors are trying to state that the Zero Emissions commitment is nearly zero in this model. This could be stated much more clearly.

No. The SAT response to an emission pulse (Fig 2b), and thus the Zero Emissions commitment, is similar to other models (see e.g. Joos et al., 2013). The zero emission commitment depends on the scenario pathway and has not been investigated in this study.

We simplified the text to avoid confusion:
"We find a largely linear relationship between cumulative CO_2 emissions and both transient and peak warming (Fig. 3a and c) for the set of emission scenarios considered here. These linear relationships confirm the finding from the pulse experiment above, i.e. that the response in the global SAT change is largely independent from the pathway of CO_2 emissions in our model."

Page 9857 line 23 to 25: These questions are awkwardly worded.

Text is reworded to read:
"26 different observational data sets are applied to constrain carbon cycle and physical climate responses. This raises the question to which extent an individual data set or a group of data sets constrain the model responses and whether some data sets may unintentionally deteriorate estimates."

Page 9860 line 27: Citations should be added to discussion of the "hiatus".

Done. Several references are added:
"The recent slowdown in global surface air-temperature warming (Hartmann et al., 2013; Roberts et al., 2015; Nieves et al., 2015; Karl et al., 2015; Marotzke and Forster, 2015), termed hiatus, has provoked discussions whether climate models react too sensitive to radiative forcing."

Page 9861 final paragraph: This paragraph seems out of place with respect to the rest of the manuscript. This should at least be described in more detail earlier in the manuscript.

Done. This paragraph is reformulated to more clearly express the link to the material presented. It is not the concluding paragraph anymore as a "Summary and Conclusions" section is now included at the end of the manuscript.

Temperature changes expressed in Kelvin and Celsius are used interchangeably in the manuscript. Please pick one and standardise throughout.

Done.

Figure 1: Does not include the 1% or pulse experiments which are crucial for the present study. Maybe redraft to take these simulations into consideration.

Done. The figure has been revised as suggested.

References:

Allen, M. R., D. J. Frame, C. Huntingford, C. D. Jones, J. A. Lowe, M. Meinshausen, and N. Meinshausen, 2009: Warming caused by cumulative carbon emissions towards the trillionth tonne. Nature, 458 (7242), 1163–1166.

Collins, M., C. Brierley, M. MacVean, B. Booth, and G. Harris, 2007: The sensitivity of the rate of transient climate change to ocean physics perturbations. Journal of climate, 20 (10), 2315–2320.

Gillett, N. P., V. K. Arora, D. Matthews, and M. R. Allen, 2013: Constraining the ratio of global warming to cumulative CO2 emissions using cmip5 simulations. Journal of Climate, 26, 6844–6858.

Goodwin, P., R. G. Williams, and A. Ridgwell, 2015: Sensitivity of climate to cumulative carbon emissions due to compensation of ocean heat and carbon uptake. Nature Geoscience, 8 (1), 29–34.

Gregory, J. M., C. D. Jones, P. Cadule, and P. Friedlingstein, 2009: Quantifying carbon cycle feedbacks. Journal of Climate, 22 (19), 5232–5250.

Herrington, T. and K. Zickfeld, 2014: Path independence of climate and carbon cycle response over a broad range of cumulative carbon emissions. Earth System Dynamics, 5 (2), 409–422.

MacDougall, A. H. and P. Friedlingstein, 2015: The origin and limits of the near proportionality between climate warming and cumulative CO2 emissions. Journal of Climate, 28, 4217–4230, doi:10.1175/JCLI-D-14-00036.1.

Matthews, H. D., N. P. Gillett, P. A. Stott, and K. Zickfeld, 2009: The proportionality of global warming to cumulative carbon emissions. Nature, 459, 829–832, doi:10.1038/nature08047.

Olson, R., R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, and K. Keller, 2012: A climate sensitivity estimate using bayesian fusion of instrumental observations and an earth system model. Journal of Geophysical Research: Atmospheres, 117

Steinacher, M., F. Joos, and T. F. Stocker, 2013: Allowable carbon emissions lowered by multiple climate targets. Nature, 499 (7457), 197–201.

Referee #4

My general impression is that the paper addresses important research questions and delivers interesting information on the performance of a large model ensemble under various emission pathways. Yet it could benefit from putting more focus on the discussion of certain model aspects (e.g. the issue of linearity) rather than more describing simulation results.

Specific comments:

- Given the focus of this work, I am missing a more in-depth discussion of non-linearities in the model response. The authors discuss non-linearities mainly in the context of the impact of non-CO2 forcings and negative emissions. I consider it crucial to underline that the model used may not simulate non-linearities due to model limitations. Only a short comment in the Discussion section is made about potential model limitations. I think the paper would benefit from discussing (as far as feasible) non-linearities which are not captured by the model due to (needed and justified) simplifications in the model design (e.g. regarding the model description of carbon cycle feedbacks). Although to a lesser extent, the aspect of incomplete description of Earth System feedbacks is also an issue for state-of-the-art complex climate models (e.g. for permafrost-carbon feedbacks).

Done. The following text is added in the discussion:
"Results for changes in Atlantic Meridional Overturning (AMOC) are known to vary considerably among different models and our ensemble may not represent the full uncertainty in AMOC response. Important processes are not represented in Bern3D-LPJ. Most notably, the melting of ice sheets and glacier and its impacts on sea level and AMOC are not included. Consequently, only results for the steric component of sea level rise are reported and results for changes in AMOC should be considered with caution. Potential climatic "surprises" such as the massive release of methane from clathrates or permafrost are also not considered."

- The authors underline that they add new variables (surface ocean pH, calcium carbonate saturation states, and soil carbon) to the discussion of linearity in Earth System responses. Yet in the manuscript these climate variables are only shortly discussed when describing the simulated model simulations results (Figs. 2,5,6). A few implications of those results would be interesting to discuss. Also in this context: A description of what "surface aragonite saturation state" is and why it is of interest would be a helpful information to include into the manuscript.

Done. A new subsection "Selection of the climate parameters for analysis and computation of TCPRE" (2.5) is added at the end of the method sections. There, the

climate variables are discussed. Please see the track-change version of the manuscript for the related text.

- Section 2.4 and figure 4: To me it is not obvious that the issue of total vs. fossil-fuel only carbon emission is a key point for this study. I probably would put this discussion into a supplementary information section. If kept in the main manuscript, Figure 4 should be discussed in more detail (in the current manuscript there are only two short references made to Figure 4).

We moved the fossil-fuel only results to the appendix (former Section 2.4, Fig. 4 and Table 3).

- Additionally, in a supplement a figure could be shown illustrating the various emission scenarios used in this study (showing the temporal evolution of CO2 emissions).

We added this figure in the appendix (Fig. A1).

- Fig. 2c: An improved explanation would help better clarifying the meaning of Fig.2c for interpreting modelling results.

Done. Text in caption modified to read:
"(c) shows the mean age of past emissions over the historical period and for the four RCP scenarios (left axis), and the fraction of the emissions older than 30 years (right axis) versus calendar years. More than half of the emissions are older than ~30 years. The bulk of the emissions at any calendar year is thus in the age range (x-axis in the other panels) where the pulse response function varies within a limited range for surface air temperature (b), surface pH (g) steric sea level rise (d) and Atlantic meridional overturning (e). This implies an approximately linear relationship between cumulative emissions and responses in these variables."

- Forcing design: Is the assumption of constant CO2 concentration and radiative forcing after 2150 made in all scenarios? E.g. RCP8.5 stabilizes only after 2200. Then the labels in the figures should be modified accordingly (the now labelled "RCP8.5" scenario differs from the conventional RCP8.5 scenario) and the figure legend should hint to the stabilizing at 2150.

Strictly speaking, the RCPs are only defined up to 2100 CE. You are right, however, that our extension of the scenarios beyond 2100 is not identical to the "official" Extended Concentration Pathways (ECP). We have clarified this in the caption of Fig. 3:
"Please note that our extensions of the RCP scenarios beyond 2100 are not identical to the Extended Concentration Pathways (ECP, see Methods)."
In the method section we have added the following text:

“Please note that the extensions of the RCP scenarios beyond 2100 CE as used in the AR5 (Extended Concentration Pathways, ECP; Meinshausen et al., 2011) are not identical to the extensions applied here. Our extensions of RCP4.5 and RCP6 are similar to ECP4.5 and ECP6, but ECP8.5 differs significantly from our extension of RCP8.5, where atmospheric CO_2 is stabilized by 2150.”
Further, the newly added Fig. A1 shows the forcings and emissions used in this study and should clarify things.

- Constraining of model responses: I wonder to what extent uncertainty in aerosol forcing is captured in the performed model simulations. Uncertainty in aerosol forcing is a key factor when observationally constraining climate sensitivity and transient climate response. I understand that the ensemble describes various emission scenarios with differing pathways of aerosol emissions. But how do these emissions translate into a spread of direct and indirect radiative forcing in the model?

One of the 19 key model parameters that are sampled is a scaling factor ($RF_{aerosol,scale}$) and reflects the uncertainty in the aerosol forcing given by Forster et al. (2007). I.e. the aerosol emissions are translated to radiative forcing as described in Joos et al. (2001) and Strassmann et al. (2009), and then the resulting radiative forcing is scaled with the sampled parameter $RF_{aerosol,scale}$ in different model configurations to reflect the uncertainty.
The scaling factor $RF_{aerosol,scale}$ is now mentioned in the extended method section 2.1.1 “Model parameter sampling” and listed in the new Table A1. Further, we explain the translation of non-CO_2 emissions to radiative forcings in the method section:

“For these simulations, we prescribe atmospheric CO_2 and the non-CO_2 radiative forcing derived from the emission scenarios (Fig. A1) as described in Joos et al. (2001) and Strassmann et al. (2009).”

- When using the “CO2 Group” constraint, how is the multi-modal distribution of ECS (and of TCR) to be interpreted? I would have expected no effect on ECS and TCR as I understood that these are calculated based on simulations with prescribed CO2 concentrations (and therefore CO2 parameters should not affect the results)?

We have extended this section (3.3.1) and elaborate more on these questions:
“Interestingly, also the “CO_2 ” subgroup narrows the probability distribution for TCR and ECS, although less than the SAT and ocean heat records. The “CO_2 ” subgroup includes data sets of the atmospheric CO_2 increase over the industrial period and observation-based estimates of the ocean and land carbon uptake for recent periods. Ocean carbon and heat uptake are governed by similar processes, namely the surface-to-deep transport of excess carbon and heat from the surface to the deep ocean. Apparently, model members that are not able to describe the ocean

carbon uptake and the evolution in atmospheric CO_2 reasonably well, fail also to match observational records for SAT and ocean heat content. The PDF for the "CO_2" subgroup displays several maxima for ECS and similar for TCR. We are not in a position to provide a firm explanation for these maxima, but speculate that this result may be related to the limited number of members in our ensemble and that the multi-dimensional model parameter space is not completely sampled."

- Discussion / conclusion section: The hiatus discussion is not directly linked to the rest of the manuscript and seems like describing a new issue. Maybe placing into an extra subsection?

We revised the wording regarding the paragraph on internal climate variability (the hiatus) and its influence on estimates of TCR and ECS. It was not our intention to introduce new subjects, but to discuss our results for TCR and ECS. The following text was added:

"Another focus of this study is to provide observation-constrained estimates of the Transient Climate Response (TCR), the Equilibrium Climate Sensitivity (ECS), and the Transient Climate Response to cumulative CO_2 Emissions (TCRE) as determined from CO_2-only scenarios. The recent slow-down in global surface air-temperature warming (Hartmann et al., 2013; Roberts et al., 2015; Nieves et al., 2015; Karl et al., 2015; Marotzke and Forster, 2015), termed hiatus, has provoked discussions whether climate models react too sensitive to radiative forcing. [...] On the other hand, our results do not confirm some recent studies (Otto et al., 2013; Schwartz, 2012; Collins et al., 2013). that suggest the possibility of a TCR below 1°C. Such low values for TCR are outside the very likely range given in the Fifth Assessment Report of IPCC (Collins et al., 2013) and of this study.

The choice and record length of observational constraints may bias results for TCR and ECS. In particular, internal climate variability, e.g., associated with the Atlantic Multi-decadal Oscillation, may obscure the link between anthropogenic forcing and response (van der Werf and Dolman, 2014). [...]"

- The paper refers to "cumulative carbon emissions" but should refer to "cumulative CO2 emissions" which is the focus of this study.
Changed.

Minor comments/ suggestions:

- Abstract: use consistently 1 Digit estimates for indicated temperature changes
Changed.

- Page 9843: definition of TCR: refer to "atmospheric CO2 concentration". Although it is clear what is meant, I would check the manuscript where reported changes in "CO2"

should be rather described as changes in "atmospheric CO2 concentrations" to avoid misunderstanding.

Done. "CO_2" modified to "atmospheric CO_2" wherever appropriate.

- Page 9843: Definition of ECS: ECS itself does not depend on the rate of ocean heat uptake, while observationally constrained estimates of ECS do. . .

Text modified to read:
"TCR and ECS are metrics for the physical climate system and they do not depend on the carbon cycle response (e.g. Huber and Knutti, 2014; Kummer and Dessler, 2014). TCR and ECS depend both on multiple physical feedbacks such as the water vapor, the ice-albedo, or the cloud feedbacks. TCR depends also on the rate of ocean heat uptake. ECS itself does not depend on the rate of ocean heat uptake, while observationally constrained estimates of ECS do."

- Section 3.3.: What is the choice for the priors for TCR and ECS?

No explicit priors for TCR or ECS are chosen. The prior PDFs result from the unconstrained model ensemble, which is generated by varying 19 model parameters. One of those model parameters is the "nominal climate sensitivity", which translates to a feedback parameter in the model to produce approximately the specified ECS. This parameter is now explained in the extended method section 2.1.1 "Model parameter sampling" and listed in the new Table A1.

- Page 9847: A quick explanation of "convex hull" in the applied context would be helpful.

We rephrased this section (see also response to other reviewers) and added a definition of the term "convex hull":
"Technically, we use the the convex hull which is the smallest region containing all points such that, for any pair of points within the region, the straight line segment that joins the pair of points is also within the region."

- Page 9847, L.28: non-linearity for low-end scenarios hard to see in shaded area.

The reference to the corresponding panel was wrong (Fig. 3c instead of Fig. 3b). Figure 3b has no shaded areas. The reference is corrected in the revised manuscript.

- Page 9851, L.12: ". . .will be discussed later" Indicate in which section the discussion can be found.

Done. Changed to: "... that will be discussed later in section 3.1.4."

- Page 9854. L.18: "The median transient response. . ."
Changed.

- Page 9856. L3: It seems in the following peak responses are discussed. Then the legend of figure 5 should indicate this.

This was wrong and has been corrected:
"For AMOC, the response is somewhat stronger for low-emission than high-emission paths (Fig. 4). For 1000 GtC total emissions, we find a peak reduction in AMOC of −24 % (−35 to −15 %) (Table 2)."

- Page 9856. L 24: Reference should be given to Fig.6e/f. For the response of soil carbon a linear regression seems to only insufficiently describe soil carbon dynamics. . .

A reference to the figure is added. The sentence regarding the linearity of the response at the end of the paragraph has been deleted.

- Fig.2a,b: Here it could be helpful to shortly discuss/explain why (early time) SAT responses are declining with increasing peak emissions (from 100 to 5000 GtC) despite increasing CO2 concentration anomalies.

Text added in section 3.1.3:
"At first glance, it may be surprising that the responses in SAT, SSLR, AMOC, and pH do not much depend on the size of the emission pulse given the strong sensitivity of the atmospheric CO_2 response to the pulse size. For the physical variables, this is a consequence of near-cancellation of non-linearity in the carbon cycle and in the relationship between radiative forcing and atmospheric CO_2 (Joos et al., 2013). The long-term response in atmospheric CO_2 (Fig. 2a) increases with increasing emissions and the fraction remaining airborne is substantially larger for large than for small emission pulses. On the other hand, radiative forcing depends logarithmically on atmospheric CO_2 and the change in forcing per unit change in CO_2 is smaller at high than at low atmospheric CO_2 concentrations. As a consequence, the response in radiative forcing is rather insensitive to the magnitude of the emission pulse and so is the response in climate variables forced by CO_2 radiative forcing."

Spelling / Readability:

- Many of the figure legends (e.g. Fig.3,b&d) , axes description (e.g. ylabels in Fig.2) are hard to see.

We have increased the font size of the legends and axes labels where possible. Further, we hope that the figures will be scaled appropriately in the final paper so that the they are easy to read.

- Fig.5,6: subplot labelling a), b) etc. is missing

We have added the panel labels.

- Page 9846, L21: "define" instead of "defines"
Corrected.

- Use consistent notations, e.g. "confidence interval" or "c.i."
Changed.

- Page 9856. L21: "to an increases" -> "to increases". . .
Corrected.

- Page 9857. L15: "1500 yrs"
Changed.