Response to reviewer's comments 2 February 2018

We thank the reviewers for their comments on our manuscript, which has greatly helped to clarify the information presented. Please see the point-by-point response below.

The reviewer's comments are **in blue**, our reply **in black**, and the new text integrated in the manuscript **in red**.

Corinne Le Quéré, on behalf of the author team.

Anonymous Referee #2

This is a more or less a standard update of the now-accepted GCP data products describing the global carbon cycle and its evolution over time. As such, there's not much to review about most of it other than clarity and completeness as this is a product with proven value and importance. I have enormous respect and admiration for this effort and the community uses this product with confidence in the care with which it is prepared.

All this of course comes before a "but" and the but in this case in GCP's decision to change the terrestrial flux, after estimating land use, to using process models instead of treating it as the residual, and then incorporating an actual separate error term, similar to the innovation in Schimel, Stephens and Fisher (2015). This is a bold move, as the land has been treated as the residual for decades and the terrestrial ecology community has long been concerned about the lack of incorporation of any knowledge of ecosystems, their response to climate and CO2, and other disturbances into the standard definition of the carbon cycle.

Thank you for your comment, we sympathise with your position. Our decision to move forward this year with a separation of the land sink from the residual budget was not taken lightly. It resulted from a community consultation at the International CO_2 meeting in August in Interlaken, and from multiple follow up discussions. The biggest decision factor has been the recent evidence based on observed oceanic constraints that the ocean models used in our carbon budget may be substantially underestimating the decadal and semi-decadal variability in the ocean sink. Therefore although the ocean models may be better constrained from physical processes than the land models, there are still issues to be resolved regarding their variability. Therefore the past assumption that most of the errors in the carbon budget originates from errors in the land sink is clearly wrong. We think the present approach is therefore the best available option from the perspective of the full carbon budget.

We note that we provide all the data and anyone who wishes to include the residual as part of the land sink can still do so. We hope that the identification of issues with individual components, such as those identified by the reviewer here, will trigger improvements in the models. The new Table 10 included in our manuscript is an initial attempt to prioritise the research needs. The uncertainties also become more visible and transparent for all the component terms by replacing the residual sink with the imbalance approach.

We have added a paragraph of explanation for our change at the end of Section 2 (Methods). The new paragraph reads: "The use of DGVMs to assess S_{LAND} (3 above) with the introduction of the B_{IM} (4 above) is a substantial difference from previous Global Carbon Budget publications. This change was introduced after a community discussion held at the 10th International CO_2 conference in 2017, in recognition of two arguments brought forward by the community. First, recent evidence based on observed oceanic constraints suggest that the ocean models used in our global carbon budget may be underestimating the decadal and semi-decadal variability in the ocean sink (Landschutzer et al. 2015; DeVries et al 2017). Second, the growing need to verify reported emissions with Earth System observations requires that we progress rapidly towards the resolution of remaining inconsistencies in the Global Carbon Budget (Peters et al. 2017). Furthermore, reviewers of Le Quéré et al. (2016) requested that this new edition of the Global Carbon Budget focuses on what we don't know, rather than on what we know. We introduce this change in anticipation that it will trigger new ideas in the way we think about the Global Carbon Budget, produce new, more stringent constraints on each of its components, and result in more evident and transparent attribution of uncertainties. "

However desirable this innovation is, the approach used raises some questions. The models used represent the state of the art, but this same ensemble is well-known to diverge dramatically when run into the future, showing that the past is matched with very different sensitivities and feedback strengths. Thus, averaging these different realizations of our knowledge is, on the face of it, a reasonable thing to do, independent of any evidence that the ensemble is biased.

The models also diverge substantially on an annual to decadal basis, and this is reflected in the uncertainty presented (±0.5 to ±1.0 for 1-sigma uncertainty) and shown in Figure 6. Also, the reviewer is not factually correct when saying these models are well known to diverge dramatically in the future. As far as we know, the ensemble of TRENDY models used here have not been used for any future simulations.

However, there is considerable evidence the ensemble is biased. Papers by Cox, Hoffman, Schimel and especially Lovenduski and Bonan suggest that the mean of the ensemble does not match other themselves highly uncertain estimates of climate or CO2 sensitivity, and there's very little evidence that the central tendency of the ensemble is closer to the truth that some other part of the distribution. Lovenduski and Bonan even argue that the uncertainty of terrestrial models, a state of the art set, can't even be meaningfully reduced with available data. They show that an ensemble, with an extreme weighting against observations so that only a very few models contribute still produces very divergent future carbon budgets. This implies that even the models that best fit data do so for very different reasons, and do not represent a process consensus.

Large model divergence in future scenarios is also widespread for ocean models, particularly for the Southern Ocean (e.g. Nevison et al. GRL 2015). The work of Cox (Nature 2013), Hoffman (JGR 2014) and Lovenduski and Bonan (ERL, 2017) all make use of CMIP5 coupled Earth System Models which have different climate representations that amplify the differences. Our DGVMs are forced by observed climate and therefore the inter-model differences are significantly lower, as only attributable to the vegetation models.

In spite of these inter-model differences, there are multiple constraints on the ensemble of 15 DGVMs presented here that support that their use for the assessment of the land sink is close to as good as the use of the carbon budget residual. These are:

- The multi-model-mean (MMM) of 3.0±0.8 GtC/yr for the last decade is well within the uncertainty from the residual of the other terms (3.6±1.0 GtC/yr);
- The MMM roughly doubled between the 1960s and the 2007-2016 period, as for the budget residual;
- The MMM interannual variability is as large as and consistent with (in direction) the budget residual.

These evidences indicate that there is no clear bias in the MMM as suggested by the reviewer. This was already highlighted in recent publications (e.g. Piao et al GBC, 2013), as well as in IPCC AR5 (Figure 6.16).

Clearly there are weaknesses of the models including those highlighted by this reviewer. However, the budget imbalance has reached a level where it is becoming more productive to scrutinise by itself than to include it with the land sink because substantial variability in the other terms could account for part of it. This (the separation of the residual from the land sink) is also a necessary step to move toward an eventual verification of reported emissions using Earth System observations as highlighted above (see Peters et al. NCC 2017).

The paper, fatally in my estimation, fails to provide a critical and unbiased assessment of the evidence or lack thereof for the underlying credibility of these models. Unlike ocean models, where O2 and the decadal carbon inventories, as well as a fairly strong theory in the Revelle factor, provide some sense of confidence about the basic processes in the ocean carbon cycle and its integral outcome over decades, there is not yet a similar set of observational constraints for the land. The level of observational constraint required as described in the paper is so minimal as to be almost meaningless in terms of establishing model credibility and based on model benchmarking papers allows models trivially tuned to participate.

We do not share this reviewer's assessment that so much more is known and so much more data is available about the ocean carbon sink than about the land carbon sink, as evidenced by the difficulties of ocean models to reproduce the Southern Ocean CO_2 variability, most likely associated with the poor representation of sub-grid scale processes.

Atmospheric O_2 is also a strong constrain on the terrestrial carbon sink, as is the seasonal cycle of atmospheric CO_2 and its interannual variability, which have been demonstrated to originate from the terrestrial carbon cycle. Further constraints include satellite LAI and NDVI, flux tower data and manipulative experiments such as FACE. There is also a lot of empirical evidence to constrain the growth and loss rates of PFTs on land.

Therefore we find that assuming all uncertainties in the carbon balance originates from the land is flawed, and that separating this uncertainty at the earliest opportunity (this year) is a positive development that will support improved constraints in the near future. Understanding the carbon imbalance could be the greatest source of improvement that our community could see in the coming decade.

To assist in this, we have introduced Table 10 which lists the major sources of known uncertainty, which we hope the community can focus on resolving. We have also triggered sub-projects separate from this global carbon budget that aim to better understand and to improve both the land and the ocean models that underpin our analysis, and to further scrutinise the uncertainty in the atmospheric growth rate (see reply to Reviewer 1). We think that terrestrial models can greatly benefit from this step.

As a result, instead of providing a well-understood if unhappy quantity, the land as residual accumulating all the budget uncertainty, the budget now provides a quantity whose credibility is not only questionable, there is no actual basis for giving it any credence on time scales affected by feedbacks. I would not use this quantity in combination with other data, and I would not consider modeled interannual variability to be of any proven value yet. I understand the desire to incorporate knowledge about the land in parity with land use, the oceans and the atmosphere, as well as emissions, but absent a data constraint, simply using a model mean here is not the right approach.

We provide all the data here and make it easy for users to make their own assessment of the land sink if they want to use a different strategy. The readers could still estimate the land residual sink by difference between emissions and atmospheric increase plus ocean sink. Our new approach only adds new information (our best estimate from bottom up models of the land sink).

We have added the following paragraph in the discussion to highlight different choices the user can make: "Although we have presented six components of the Global Carbon Budget individually, different aggregation of terms are possible. In particular S_{LAND}, E_{LUC} and B_{IM} could be aggregated into land fluxes and total uncertainty, as traditionally done, which would result in generally lower uncertainty compared to each term individually (see Table 6). This information is limited in usefulness however, as it mixes direct and indirect processes and bring in errors from other components and hence the signal becomes difficult to interpret. However providing a realistic assessment of uncertainties for S_{LAND} and E_{LUC} is also difficult. Here we have used the model spread as a measure of uncertainty, which may be on the one hand underestimated because it includes only partly uncertainty in the underlying observations, and on the other hand overestimated as it includes artificial spread from different boundary limits among models. Therefore further work is needed not only to better quantify the fluxes but also to better describe and quantify the uncertainty and reduce them where possible. "

Speaking (or writing) as a member of the carbon community, I would not use this product or support its use in assessments and policy. I think this is -in all frankness-an understandable but terrible decision and feel it undermines the credibility of the GCP data product, established over many years. I can't recommend this paper for publication prior to a far more thorough analysis of the value of the ensemble mean, which I think, if done honestly, won't support the mean sensitivity as being the most likely given data constraints. Again, I understand the desire to use terrestrial biogeochemical insight, and the limitations of data based products such as FLUXCOM, but this is the wrong decision at the wrong time for such a vitally important product.

We think our decision is well supported and timely when considering the recent developments emerging from the ocean carbon cycle and the need to work towards the verification of reported emissions. This change is also supported by the other two reviewers. We have developed plans to improve our assessment of both the ocean and land CO_2 sink in the coming year, particularly through the better use of observations to evaluate models and inform our analysis. However a more detailed analysis of the value of the ensemble mean beyond the global measures presented here will take some time to implement and unfortunately this cannot be submitted within the timeframe of the response to this review.