

Review ESSD-2018-120, carbon budget

Very impressive compilation and synthesis effort! A very good presentation and a good match to ESSD; one hopes for broad use and strong impact of these annual carbon budgets.

Not surprisingly given the amount of material processed and the number and range of contributors, the manuscript needs a few improvements. I hope that the authors regard my close reading and consequently rather-long list of comments, questions and proposed corrections as helping to make a good product better.

Check capitalisation! You use Tropics and tropics (often within the same paragraph!). You use “Marine Boundary Layer” and “marine boundary layer” on the same page (Page 19). To make your synthesis more effective you also need to impose consistent spelling and punctuation. Make these changes now to save time at the proofreading stage? Data set vs datasets - please choose and make consistent.

Reader encounters reference to Supplement several times, but reviewer finds no supplement to evaluate! Three supplementary figures actually appear in an Appendix, albeit referenced and labelled as S1, S2 and S3. Meanwhile, Tables A1 to A6 carry an ‘Appendix’ label but do not appear with Figures S1, S2 and S3 in an Appendix. I might understand why these particular tables carry an Appendix label and why these three figures should reside in an Appendix: to include technical details not of interest to general readers but relevant to carbon cycle colleagues? If true (or for whatever other valid reason) we need some revision, reorganisation and re-labelling here! If you want a proper Appendix, label and put all tables and figure in it as appropriate.

Please adopt and implement an informative flux labelling convention. For many in our earth systems community, fluxes upward carry positive sign, downward a negative sign. Not everyone follows this convention (no surprise) and I recognise that you specify fluxes in Gt per year rather than watts per metre squared (unfortunately those who measure $W/m^2 s^{-1}$ generally do not follow the positive up negative down guidelines) but you definitely have fluxes (emissions) going gravitationally upward and fluxes (sinks, land and ocean) going gravitationally downward. Therefore when you specify, e.g. on Page 35 (and in many many other places in the manuscript): “total atmosphere-surface CO₂ flux from process models for 2008-2017 is 3.7 ± 1.2 GtC yr⁻¹” you either need to carefully specify in every case which direction you mean - in this case you would write ‘atmosphere-to-surface’ rather than “atmosphere-surface” - or a reader needs to recognise from the sign (positive or negative) which direction the fluxes flow. Figure 8 highlights this need but you face the problem throughout the manuscript that readers do not know from your text or numbers in which direction the fluxes you cite actually operate. In the new Figure 2 you show, for example, ocean-to-atmosphere natural cycle flux of 90 up and atmosphere-to-ocean return flux of 90 down (and land with atmosphere likewise, 120 up matched by 120 down) with human impact superimposed but in neutral units. You should thus describe an ‘ocean bidirectional exchange with atmosphere of 90 GtC y⁻¹’ or you need to show 90 up as positive and 90 down as negative. The challenge becomes acute in Figure 8. Globally, you show combined land and ocean downward fluxes (sinks!) in positive numbers (e.g. currently roughly +5 GtC y⁻¹) but you also show that in the 1960s and in occasional years since net land flux has gone negative (upward, land as a source!), as low as (or as much as) -1.5 to perhaps -2.0 GtC y⁻¹. Globally, ocean fluxes stay downward/positive values/sink, but when you sort by latitude, the tropical ocean (which, disconcertingly, has its own ordinate range different to any of the other 11 panels in this figure) shows a clear ocean source but now designated by negative fluxes! For purposes of this figure you have in fact adopted a convention - positive downward and negative upward - but you have not declared it and you do not follow it consistently in the text. I agree that you can extract a lot of useful information from the flux products but you owe readers much more clarity on which direction you mean and when. You can implement this clarification in text probably more easily than in numbers and equations, but we need a good response on this issue.

You have already done something similar for the budget imbalance, specifying it as positive some years and negative in others, because as the mathematical residual of equation 1 it will necessarily have small positive values some years but small negative values in other years. Note

also that when you specify rates of change in % (e.g. change in growth rates of CO₂ emissions as on Page 34), you carefully and correctly specify either increase or decrease, using language to indicate the direction of change, or - as in line 24 on Page 37 - you carefully use negative percentages. We need similar careful language for flux descriptions!

After multiple readings, I still find myself confused by the Discussion. Here you have a chance to 'comment' on strengths and weaknesses and to recommend (or at least list) possible improvements. As I read it, the discussion omits or takes a too-modest approach to positive features and strengths. This reader finds the discussion almost apologetic in places, when in fact the authors could expand and expound on this statement (from Page 33 line 8): "the near-zero mean and trend in the budget imbalance is an indirect evidence of a coherent community understanding of the emissions and their partitioning on those time scales". What would make the evidence more direct? Do the authors feel that longer (paleo?) time scales (e.g. ice core data) have something useful to lend to this accounting? The discussion also fails to follow-up on specific issues and uncertainties from the text, e.g. the so-called 'loss' of land sink capability in Section 2.7.4, the various factors that go into coming-year projections discussed in Section 3.4, the important outcomes and clear cautions about cumulative emissions accounting in Section 3.5, etc. You likewise have an opportunity (responsibility?) here to raise highly-relevant but - to this point - not-yet-discussed issues. E.g., how would we know if either the land or ocean sink changed significantly? What measurements continued over what time period, and what model confirmation, would we need to quantify and certify such a change? Such a discussion would add substantial relevance to your results and contentions on variability. Similarly, how quickly and with what precision could we - using these data and models and this budget approach - detect and confirm an intentional drop in emissions? How much induced carbon capture sink could we detect, and how soon? Of course an external record for these topics often exists in scientific literature but readers across a wide spectrum - the broad stakeholders listed in the Conclusion - and particularly a research community interested in but not directly contributing to this budget would very much benefit from explanatory and enlightened discussion on these issues here! This annual carbon budget effort also constitutes a unique, very necessary, very positive yet challenging mixing of model products with observational products; many readers will not encounter such a breadth of model intercomparisons with observations elsewhere. This particular version of the budget has attempted a positive step forward by introducing model metrics - twice (unfortunately) in the Discussion text and (not very helpfully) in Figures S1 and (better) in S2 - but in both cases with a very tentative presentation. If metrics seem important, and better exposed and evaluated from the point of view of a carbon budget than from some future technical CMIP6 overview (or, god forbid, an IPCC report), shouldn't the authors spend more time in the discussion on what the metrics indicate (not much, evidently) or on what better metrics they might apply or imagine. In various specific sections below I have made so bold as to suggest alternate language, occasionally even for a full sentence or two. I looked closely at the Discussion with similar modifications in mind but feel that the actual text of the discussion belongs entirely and exclusively to the authors. I understand political pressures from various constituencies and also the fact that each annual budget runs right to the edge of release schedules, but I strongly urge the authors to reconsider this Discussion in light of the overall positive effort and some of the clear messages that emerge from that effort.

Good update and changes to Figure 2.

Apparently the word processing software allows line breaks at hyphens, including at the superscript hyphen yr⁻¹. This results in unfortunate line breaks in many locations in the text leaving an orphan superscript 1 at the start of the following line. Authors and proofreaders will need to watch carefully for these errors in the final product.

I don't know what ESSD or Copernicus recommend or require, but most manuscript reference lists today seem to use doi in place of pages and dates? Perhaps a change for a future version?

Page and line specific comments:

Page 8 line 2: "estimates of EFF globally and nationally CO₂ emissions" Something awkward here? Because you have defined global and national CO₂ emissions as EFF in the prior sentence, here you only need to say 'global and national EFF'?

Page 8 line 8: “Official UNFCCC national inventory reports for 1990-2016 for the 42 Annex I countries in the UNFCCC (UNFCCC, 2018), as we assess these to be the most accurate estimates because they are compiled by experts within countries that have access to detailed energy data, and they are periodically reviewed.” Awkward language here, allows some confusion. I think you mean (roughly) national inventory reports from UNFCCC. We assess these national reports to provide the most accurate estimates because they are compiled by experts within countries who have access to detailed national energy data and because they receive periodic review by UNFCCC.

Page 8 line 16: Again awkward language allows subject-predicate confusion. I think you mean ‘provide more details for each dataset and describe additional modifications required to make the dataset consistent and usable.’

Page 12, lines 8 to 13: Check punctuation here, I find commas or semicolons missing. Also “and (3)” followed by “and (4)”. Authors will know best how they want to separate and punctuate this list.

Page 13 lines 5 and 7: You designate the two (oil and natural gas) PPAC references here in the text as 2018a and 2018b but in the reference list you show them both as PPAC 2018. Change something somewhere?

Page 13 line 18: “which is much less strongly seasonal because of strong weather variations” I had to think about this for a moment. I think you mean that shorter-term cold or warm weather periods dominate coal usage patterns due to their impact on residential and commercial heating or general energy use, as opposed to a smoother winter/high summer/low pattern of use? But, if correct, the same pattern or factors would apply to US and China as well? Already included in CCIA projections for China or EIA projections for US, but not in projections for Europe? If I understand the statement correctly, I find it strange that you only invoke it for Europe. If short-term weather dominates, and varies widely, for Europe then your extrapolation from the prior year for missing winter months of the current year seems unreliable?

Page 14 line 27: “Additionally, they represents permanent degradation of forests by lower vegetation and soil carbon stocks for secondary as compared to the primary forests and forest management such as wood harvest.” Awkward. They (meaning bookkeeping models) ‘represent long term degradation of primary forest as lowered standing vegetation and soil carbon stocks in secondary forests and also include forest management practices such as wood harvests’ - correct?

Page 18 lines 4 to 9, confusing and somewhat redundant. You have a good test here but it gets lost in the language. With the DGVMs you run two scenarios, one of change (described as the “first” runs in line 4) and one with fixed invariant pre-industrial land use cover (described as the “second” in line 6). But in line 7 you say “allowing the models to estimate, by difference with the first simulation”. Confusion arises over which models - first, second or all - you refer to at the start of that sentence and therefore which differences you actually calculate. At line 8 you use the word “prescribed” when in fact in both the change and invariant simulations you ‘prescribe’ land use forcing. From the clear statement on line 9 a reader finally understands both the sequence and the purpose. I also don’t understand why in line 6 you write “as further described below” in a location and with punctuation that clearly implies that the phrase applies specifically to the second set of invariant simulations when in fact below (where ‘below’ as I understand it could refer both to Section 2.2.3 on uncertainty in ELUC or more likely to Sections 2.5.1 and 2.5.2 on using DGVMs to diagnose SLAND) you provide additional details about primarily the first active-change DGVM simulations.

I suggest a gentle rewrite along these lines:

Two sets of simulations were performed with the DGVMs. A first set included historical representations of changing land cover distributions, climate, atmospheric CO₂ concentrations, and N deposition. A second set adopted time-invariant preindustrial land cover distribution. Because dynamic evolution of vegetation biomass and soil carbon pools occurs in the first

simulations but not in the second, ELUC is diagnosed in each model as the difference between these two simulations. As discussed in more detail in Section 2.5.2, we only retain model outputs with positive ELUC during the 1990s (Table A1). Using differences between the two DGVM simulations to diagnose ELUC allows the DGVMs to account for the loss of additional sink capacity (around 0.3 GtC yr⁻¹; see Section 2.7.3) while bookkeeping models do not.'

Page 18 line 23: Eliminating the first two words of this sentence would correct the singular/plural problem and clarify the intent. E.g., instead of 'We assess an uncertainty of xxx reflects' a reader would see 'An uncertainty of xxx reflects our best value judgement ...'.

Page 19 lines 4,5: Confusion here. Earlier you said you used GFED and GFED 4s for fire occurrence and emissions, correctly referencing van der Werf et al. 2017. But, as referenced in van der Werf et al. 2017, GFED 4s derives from MODIS MCD64A1 c6 burned area determinations. Here, however, we find a different MODIS product, MCD14ML, accompanied by both the Giglio and van der Werf citations. Please check! Different MODIS product or typo? There is also a right parenthesis ')' missing here somewhere.

Page 19, line 6: "fires season" or fire seasons? Following in line 7, most burning, at least as detected by satellite, occurs in tropical Africa. Deforestation and associated emissions do or do not show the same geographic pattern as burned areas? Why or why not?

Page 19, Section 2.3.1. The legend to Figure 1 quotes Scripps as 1958 to 1979 followed by NOAA from 1980 to 2018. Those dates differ slightly from what we read here? Figure 1 legend also refers to an essential overlap intercalibration period of more than 20 years, not mentioned here?

Page 19 line 29: "Marine Boundary Layer" here but "marine boundary layer" earlier in line 16. Chose one appropriate capitalisation scheme and use it consistently?

Page 20 lines 2,3: Replace existing with: 'The second and third uncertainties, summed in quadrature, add up to 0.085 on average'?

Page 20 lines 6 to 12: Confusing, redundant and disorderly. Small changes could give a substantially better accounting:

'... assess the total atmospheric CO₂ burden. An excess measured at any station will not exactly track changes in total atmospheric burden, with offsets in magnitude and phasing owing to finite rates of vertical mixing and stratosphere-troposphere exchange. For example, excess CO₂ from tropical emissions may arrive at distant stations in the network after a delay of months or more and the signals will continue to evolve as the excess mixes throughout the troposphere and stratosphere. This effect must be very small on ...'

Page 20 line 20: for years prior to 1980?

Page 21 line 17: 'first halves of years show' or 'first half of the year shows'?

Page 22 line 1: statistical "spread" or geographic "spread" or both? Following to line 2, do we need a citation to support "... methods that are deemed most reliable for the assessment ..."? Following further from line 2 into line 3 "IPCC did not revise its assessment in 2013" - what does this mean? No further attention to the issue? I think you mean that IPCC confirmed their confidence in these ocean sink estimates as recently as 2013?

Page 22 line 7 "This estimate" Which estimate? The sentence follows an extensive list of observational constraints, but you expect a reader to remember the earlier IPCC estimate? Give the reader a little more guidance here?

Page 22 lines 18,19: Confusing. Suggest instead 'Several other ocean sink products based on observations are also available but they continue to show large unresolved discrepancies with observed variability.'

Page 23 line 9: ‘... in the source of the atmospheric forcing data, spin up strategies, and in the resolution of the oceanic physical processes ...’?

Page 23 lines 12,13 ‘changes in river organic carbon (discussed in Section 2.7.3)’?

Page 23 line 26 ‘... variability in ocean biogeochemistry models.’?

Page 25 line 12: 16 DGVM models (also 16 listed in Table 3 on page 63) but Figure S2 shows only 15?

Page 27 line 6: instead of rivers, suggest ‘river’s’ or ‘riverine’ or ‘fluvial’

Page 27 line 18: need a change in wording. Instead of “is missing in the combination of approached used here to estimate both” suggest ‘is missing in the combination of approaches’ or more simply ‘is missing in the approaches used here ...’.

Page 29 line 16: transferred in the open ocean or transferred to the open ocean?

Page 30 line 2: readers will perhaps know CMIP but likely will not know TRENDY for DGVM. Include some acronym definitions or references here?

Page 30 lines 7 to 9. 0.4 GtC per year over decade 2005 to 2014 gives 4 GtC per recent decade. Extrapolated over nearly 15 decades (1870 to 2017) does not give 20 GtC accumulated as stated here. Instead, closer to 60 GtC? The 20 GtC comes not from simple mathematical extrapolation of the most recent rate but from the time history of the OSCAR model ensemble? Including some scaling of the rate in past decades? Please check and explain. As written could allow or encourage errors by this reader and others?

Page 31 line 21: Estimates of total land fluxes (SLAND – ELUC) from the DGVMs is consistent or are consistent?

Page 32 line 3: 193%?

Page 36 line 15: were, not where

Page 39 line 19: total LUC emissions? total fire emissions? Help us out here, we can’t tell what you mean.

Page 40 line 17: would further exacerbate or further exacerbates

Page 41 line 9: You have shown repeatedly that mean B_{IM} remains small without a trend; at one point earlier you discussed the generally positive implications of a small imbalance, e.g. that all elements from all sources result in fundamental agreement on basic terms (Section 3.1.4 on Page 33). In the previous sentence you point out occasional higher B_{IM} values and substantial unexplained variability. This represents a valid but still somewhat mixed message. Therefore, in this particular sentence, I suggest perhaps the word ‘persistent’ to replace the word “large”: ‘A persistent budget imbalance ...’ instead of “Such large budget imbalance ...”.

Page 41 lines 10, 11: This statement seems more negative or cautious than the assessment back in Section 3.1.4.

Page 41 lines 23 to 25: I understand what you mean here but the language tends toward opaque jargon. Could you write instead (shorter, clearer): ‘... is not new and highlights the difficulty to quantify complex processes (CO₂ fertilisation, nitrogen deposition, climate change and variability, land management, etc.) that collectively determine the net land CO₂ flux.’?

Page 41 line 27: ... uncertainties that have been a demonstrated documented

Page 41 line 31 to Page 42 line 1: Not a proper sentence. Write instead something like: 'When assessing SLAND using DGVMs, uncertainties relate mostly to limitations in understanding and representation of fundamental processes as evidenced by ...'?

Page 42 line 2: multiple studies

Page 42 line 10: 'quality of the energy statistics and of the emissions factors remain the largest source of uncertainties for'

Page 42 line 11: again, 'documented' instead of "demonstrated"?

Page 42 line 13: "atmosphere introduces additional errors that have not yet been" If they haven't been quantified, we can't know if they represent significant additional errors?

Page 42 lines 17 to 20: awkwardly written paragraph with very mixed messages. For the first half, I think you want to say something like 'changes and improvements to the budget might either increase or decrease accuracy or variability and in any case might represent process that act neither globally nor simultaneously'? I can not agree with the next sentence as written "It is also possible that further yet unknown processes are not taken into account." Surely this statement, written so blandly and open-ended, invites confusion and criticism; it serves to undermine all that you have presented? Thinking in budget terms, we have almost absolute confidence in atmospheric CO₂ concentrations - what better examined, reproduced and reproducible, tested, intercalibrated, verified, re-verified dataset do we have about our planet? From the remaining budget terms we can then estimate best and worst cases. Best cases, realised in many past years, show a budget imbalance of essentially zero. Once or twice we might experience the ignorant good luck that emission errors exactly compensate sink errors - thus giving a false budget imbalance close to zero - but as Figure 3 clearly shows for perhaps as many as 50 or 60 or more years of the past 120, we have zero budget imbalance. You say this several times: mean budget imbalance essentially zero with no trends. Either we have confidence in that record or we don't, but we find nothing there to suggest fundamental missing components of the carbon cycle? Taking now the worst case, based on the most recent decade, we could envision a budget imbalance as large as 3 GtC y⁻¹ (working from Table 6 and assuming we err on the high side of all emissions and on the low side of all sinks). Setting aside for the moment that only in the late 1940s before Moana Loa CO₂ records start, we might imagine one or two years where budget imbalance approached 3 GtC while in every other year of a nearly 120-year record we never measured an imbalance anywhere close to 3 GtC, what "unknown processes ... not taken into account" can we imagine that would force a 30% error in either emissions or sinks? I know the sink side better than the emission side but I can not envision an omission or error that large in emissions or sinks! Some widespread but hidden emission source? Inversions and isotope analyses would already have identified such a source? Processes in soil or ocean, out of sight of satellites and difficult to reach by current sampling methods, that consume Gt of carbon per year? Dozens of researchers over at least four decades will have tried to attack current accounting and propose or discover new sinks, but (apparently) our understanding of basic carbon cycle processes remains valid. Do we have continuing errors - certainly! Does unexplained variability persist, to a degree that clouds our predictive skill - absolutely! But do we miss a fundamental term in the carbon cycle of today's planet? No, and everything GCP and others have published and reported for more than a decade reinforces that point. I suggested at the top several interesting topics for discussion; I contend that this paragraph and particularly this sentence need serious and substantial revision.

Page 42 paragraph beginning at line 21: I applaud this inclusion but this paragraph repeats almost exactly what you say again on Page 43 in lines 17 to 20, and you say it better (shorter, clearer) there.

Page 43 lines 2 to 4: Awkward, imprecise. Suggest something such as: 'Merging these terms has limited usefulness, however, as it involves mixing direct and indirect processes and bringing in errors from other components; source, sink and uncertainty signals become more difficult to interpret.'

Page 43 line 8: “different boundary limits among models.” What does this mean? Geographic boundaries, e.g. at coastlines between land and ocean? Process boundaries, e.g. how various model treat (or do not treat) soil carbon respiration or deforestation/reforestation? Component boundaries among atmosphere, land and ocean? Specificity and clarification necessary.

Page 43, lines 9 and 10: Does this sentence refer to land fluxes only or to carbon fluxes (land and ocean) generally?

Page 43, line 11: Annual estimates, which continue to generate more uncertainty than longer-term means of trends, may improve with improvements in data quality and timeliness.

Page 43 line 17: If you had truly large discrepancies you wouldn't construct long-term much less annual budgets. You should eliminate the word 'large'?

Page 43 lines 17 to 20: Good sentence about why you introduced model metrics. Keep this and eliminate the highly similar paragraph on page 42?

Page 44 line 13: Extraneous characters at the end of that line.

Page 44 line 29: that requires a careful compilation and synthesis of observations, statistical reports and model products.

Page 45 line 8: change of which sink? Ocean, land, both? Also, you should refer to ocean heat content rather than ocean temperatures?

Page 45 line 9: better = improved, more frequent, more precise?

Page 45 line 12: If your budget is correct, the sinks do more than affect, they determine?

Page 45 line 15: add some urgency here? 'Building this scientific understanding to meet urgent social needs depends on more frequent, robust and transparent ...'?

Page 45 line 23, 24; here; and F Joos ... Likewise line 25: measurements; and V. Kitidis ...

Page 46 line 3: profile data; and Japan Met....

Page 46 line 5: evaluations

Page 47 line 22: something wrong or missing with the citation here? Using doi would fix these problems?

Page 48 line 5: “submitted” used twice

Page 49 line 3: spelling and formatting errors in Collier et al. reference

Page 50 line 50: formatting error in the Hooijer reference

Page 51 line 25: Hurtt et al. in prep has an update by now?

Page 51 line 43: spelling and formatting errors in Joetzjer et al. reference”

Page 52 line 37: capitalisation errors

Page 54 line 38: “in review” used twice

Page 55 line 10: formatting error

Page 55 line 28: “2018” used three times

Page 56 line 23: you do not need “2018” twice here

Page 56 line 27: spelling / formatting error

Page 57 line 3: formatting error

page 57 line 13: formatting / spelling error

Page 57 line 15: formatting / spelling error

Page 58 line 7: formatting / spelling error

Page 58 line 50: formatting / spelling error

Page 59 line 7: van der Werf paper published in 2017, please update the reference

Page 70 line 2: should this read “of at least $\pm 0.3 \text{ GtC y}^{-1}$ ”? (As used for G_{ATM} later in this table)

Page 71 line 8: do we need definition of V_{cmax} here?

Page 77: Surprised to not see USA National Science Foundation listed in Table A5 because you several times mention NCAR and CCSM (and also HIPPO, flown on an NSF aircraft) in text, other tables or references?

Page 89 Figure S1: What should we learn from this Figure? To avoid the IPSL model? 0% IAV variability would indicate perfect fidelity with SOCAT fluxes? This figure as well as the text in Section 3.1.3 show or describe difference in percentage ranges, e.g. 30 to 50% compared to 50 to 70%. The authors offer no explanation (e.g. why should tropics differ from higher latitudes), no indication of ranges or uncertainties, and no conclusions. As a reader nothing useful leaps out at me here? Will a GOBM owner/operator extract something useful? Absent some clues from the authors, one wonders.

Page 90 Figure S2: Again, not much explanation from the Figure nor from the text. Familiar model by parameter matrix, but what should a reader conclude. Does the figure belong better to the DGVM owner/operators? Can / should the authors provide some general clues? The figure legend basically lists the validation data sets and the difference between ‘absolute’ and ‘relative scores’ but not much about what we should learn or expect. That no model offers reasonable skill at LAI or NEE but that most of them can replicate run-off? The text and figure fail to convince me of relevance to global carbon budgets? Do DGVM need to score high on all these parameters to better predict carbon sinks?

Page 91 line 7: land or ocean data, rather than “land of ocean”? Again, what should we conclude from this figure or from the associated text? That all four inversion techniques vary widely (and wildly) among themselves and in relation to aircraft data over most of the globe except perhaps in the northern half of the northern hemisphere? Has CGADIP done quality control or error analysis on any of the aircraft data profiles? Why do we see no error bars? What combination of profiles and sampling rates allows one to extract more than 200k data points (legend of Figure S3) from 40-some aircraft campaigns (Table A6)? Trust the authors to know their product and various means of validation but one really stretches to find relevance or utility in this figure and this data?