Reply to Anonymous Referee #2 comments on "Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation" by J.-F. Exbrayat et al.

In the following, we provide answers (in blue) to discussion points raised by referees (in black).

General comments

Exbrayat and co-authors should be commended for tackling this broadly interesting topic with a novel experimental approach. At best this manuscript can provide a thoughtful examination about how model assumptions regarding nutrient limitation and environmental scalars effect the sign and strength of land C dynamics in a global model. At worst the findings could be seen as different model configurations provide different results. In revising this manuscript care should be given to make the paper more of the former, and less of the later.

In my estimation the take home messages from these results are quite tractable. Global models need to consider nutrient limitation to get global C response to elevated CO2 correct (sensu Hungate et al. Science 2003; Figs 2 & 3; section 4.1). Ultimately, the fate of additional NPP over the historic period and in future scenarios depends on Rh, and assumptions made about the temperature and water sensitivity of organic matter decomposition in soils (Davidson and Janssens Nature 2006, Friedlingstein et al. J. Clim. 2006; Figs 4 & 5). In the end, a C-only model stores too much carbon, especially in soils at high latitudes using temperature functions that slow organic matter decomposition and have longer turnover times (Fig 6). Finally, results over the historical period (and in future projections) depend on the amount of CO2 fertilization (i.e., nutrient limitation) and initial conditions, which are determined by environmental scalars (Fig 13). If this is the general message, it's muddied at present by an over-abundance of results. If it's not the core message, then the central story is obscured by an over-abundance of results.

We thank the reviewer for their insightful comments and suggestions to improve the presentation of the manuscript. Therefore, we simplified Fig. 3, moved previous Fig. 7 and 9 to the supplement (Fig. S1 and S2) in which we also added a figure that helps understanding the post-1960 step change in land C uptake. We also moved former Figure 12 earlier in the text and described equilibrated carbon pools in the results part. We provide answers below to specific comments and hope that we successfully accommodated the reviewer's comments.

There don't seem to be any nutrient x climate interactions, where different temperature functions potentially alter productivity by changing nutrient mineralization rates. Is this a fair assessment?

Our study focuses on the sensitivity of microbial decomposition to environmental factors and its implication from the spin-up procedure. We do not investigate the nutrient modules in detail as this falls beyond the scope of this manuscript. We however agree with the referee that carbon and nutrient cycles are tightly coupled and recognise that it would be interesting to explore that space in the future.

Results and Discussion seem to be very convoluted, with several interesting results and analyses introduced in the discussion (e.g. Figs. 10 & 13). This approach is distracting, and the two sections should be merged, or care taken to separate them.

We have reshaped the results and discussion part. Total soil carbon at equilibrium is now addressed in part 3.1 ll. 190-196 and numbers are included in Table 3:

> Equilibrated total soil carbon is presented in Table 3 for each model version. Globally, for the same combination of f_T and f_W, soil carbon in CN mode equilibrates at a level between 72 and 86% of the C-only mode while differences between CN and CNP are negligible. However, regardless of the biogeochemical mode adopted, differences in f_T and f_W introduce a 4.5 times difference between the version that simulates the largest soil carbon pool (SOILN f_W with PnET f_T) and the version that simulates the smallest (the original CASA version) in response to the same steady boundary conditions.

Maps of regional soil carbon at equilibrium are now presented as early as Figure 4 and described in the part 3.2 ll 252-260:

> The soil C density at equilibration for all CN simulations is shown in Figure 4. Large differences are observed in pool sizes as a function of f_W and f_T (similar patterns exist in C-only and CNP simulations). For example, the K1995 and PnET f_T both equilibrate at much higher carbon density than CASA functions in the mid- to high-latitudes in the northern hemisphere. PnET also has a higher soil C density in warmer regions. Differences implied by f_W are more localised and do not seem to depend on a latitudinal temperature gradient. SOILN equilibrates at a higher level of soil C in dry regions of south-west Australia, southern Africa and the western edge of South America while the two other f_W provide relatively similar results when use with the same f_T.

Finally, we dedicated the whole of section 4.1 of the discussion to the reasoning around explaining differences in equilibrated soil carbon as a function of differences in fT and fW ll. 337-369:

> Spinning up a model means integrating the model with steady boundary conditions until the trend in carbon pool is negligible, or $R_h \approx NPP$. In our simulations, all model versions were brought to equilibrium until C pools achieved a steady-state. This is a standard procedure (e.g. Wang et al. 2010; Xia et al. 2012) that would most likely have been used in all CMIP5 simulations that incorporated carbon. According to Equation (1), in ESMs, the amount of decomposition, and therefore R_h, is controlled by a time-invariant reference k parameter, the $f_W(\theta_s) \times f_T(T_s)$ product and the amount C_s of carbon available in soil for decomposition. Model equilibration consists of achieving the carbon pool size needed to simulate R_h at a level that compensate for NPP while integrating the model under steady boundary conditions,. Given that NPP is similar between our simulations (Supplementary Fig. S3, S4 and S5), it is only our modifications to f_W and f_T that have led to total soil C in our CN model to range from 765 Pg C to 3495 Pg C. This approximates the six-fold range found in CMIP5 models (Todd-Brown et al., 2013).
>
> As shown in Figure 4, differences in total soil carbon at equilibrium are due to large regional differences, especially at high latitudes. This reflects the relative position of these functions for cold temperatures (Figure 1): the CASA f_T is

> systematically above the two other f_T for soil temperatures below 10°C. It therefore requires less substrate to simulate R_h at a level that compensates for the same NPP than K1995 and PnET. Conversely, PnET causes the model to equilibrate at a higher soil C density in warmer regions as it is well below the two other functions for soil temperature corresponding to Africa and South America. As there are dry and wet regions at any latitude, differences implied by f_W are more localised. However, the same relationship between the relative positions of the curves can be seen. The most noticeable feature is that SOILN, a very limiting f_W in dry conditions, equilibrates at a higher level than the two other f_W in south-west Australia, southern Africa and the western edge of South America where it requires more substrate to achieve the same R_h to compensate NPP.
>
> We adopted a colour scale similar to figure 3 of Todd-Brown et al. (2013) to present soil C in different CMIP5 ESMs. The regional differences implied by the different f_W and f_T map particularly well onto the diversity shown by the CMIP5 models. Of course, CMIP5 models also vary in the number of pools they employ and are likely to use different values of *k* as shown by Todd-Brown et al. (2013) with their reduced complexity models. We do not explore this in detail here but we suspect that these similarities between our simulations and CMIP5 results nevertheless strongly indicate that the formulation of the time and space invariant f_W and f_T is a key source of uncertainty in these models.

Throughout the manuscript would be improved by using consistent language for terms used synonymously to aid in clarity. For example uncertainty, range, and standard deviation seem to be used interchangeably throughout the text. Additional examples are described below.

Following the referee's advice, we have thoroughly revised the text to avoid redundancies in terms used to describe the same notions. For example, we now refer to cumulative NEA to describe cumulative net carbon uptake since 1850, and annual NEA to describe annual values (instead of NEE, or land sink).

Finally figure captions should be more descriptive so they stand alone, and agree with text in the main body of the manuscript.

We have revised figure captions and harmonised terms used to describe the same phenomenon (see previous comment). For example, the legend of Figure 2 now refers to NEA instead of land sink:

> Figure 2. Global cumulative human carbon emissions and modelled net ecosystem accumulation (NEA) throughout the historical period using CASA-CNP in C, CN and CNP modes. In each panel, the shaded area represents the range of NEA simulated by the 9 combinations of f_W and f_T in the indicated nutrient mode while thin lines represent individual simulations identified by f_W. The insert in the right panel shows the relationship between cumulative human emissions and atmospheric CO_2 (in ppmv) over the same period as the model is driven by this latter.

Specific comments

It seems like more information could be included / displayed on Fig 2. If the greatest variation in NEA is driven by ft (Figs 4 & 5) may there be some value is displaying models results with different ft parameterizations with different colors lines in Figure 2?

We agree with this statement and we now use a different style for each fT. The description of Figure 2 has also been update in the results section ll. 198-201.

> Each panel in Figure 2 shows results for all 9 combinations of f_W and f_T for a given C-only (Figure 2a), CN (Figure 2b) or CNP (Figure 2c) mode using thin lines and their style represent which f_T was used in each model as indicated.

and ll 220-222

> It is interesting to note that in each mode there is a great interaction between fT and fW as shown by the position of model versions using the same temperature function alternatively at the higher or lower end of the simulated range.

Fig 3: In a complex paper with lots of multi panel display items can this figure be simplified? It seems like the point of this figure is that the C only model generates a large terrestrial C sink with high variation, compared to previous estimates. Is this accurate? I'm not sure much value is added by showing multiple time periods. Since the temporal results warrant little discussion (section 4.1), can just one temporal period be displayed and discussed? Also, here NEA is used synonymously with average land sink, are they the same? If so, can just one term be used to clarify text? Finally, if the authors feel strongly that a 10-panel figure is warranted, why aren't Sitch et al (2008) results shown in each panel?

We agree with the reviewer and we replaced Figure 3 by a two-panel figure showing the average NEA and its variability between 1959-2005 only. We also replaced "land sink" by NEA to harmonise the notations as suggested by the referee.

Post card maps (Figs 4-9 & 11-12) are of limited usefulness, especially when may of them make the same point (i.e., Ft has a strong effect on global C results). Relevant regional results should be highlighted- and interpreted.

f_T may have a stronger effect on global C results, but different f_W involve more regional differences, a feature of interest that requires maps to be illustrated. We however recognise that some Figures are partly redundant (e.g. Figure 7 and 9) and so have moved them to the supplementary material as Figure S1 and S2.

We have described regional results in detail in part 3.2 while part 4.2 discusses them. We hope that edits made will satisfy the referee.

Fig 6-9: What time frame we looking at here? Positive values show: : : what? Are stipples showing significant differences (calculated somehow?) It seems like the color bar alone shows the sign of change. More broadly can Figs 7 and 9 be removed? It seems like Fig 7 is redundant to Fig 6, and Fig 9 repeats patterns shown in Figure 8. Qualitatively, it seems like most of the difference in NEA are driven by differences in soil C- largely at high latitudes, and it's not surprising that these patterns are magnified at high latitudes where K1995 and PnET temperature scalars are much lower than CASA's temperature function.

These figures correspond to the differences in cumulative NEA since the beginning of the transient simulations (1850), i.e. the net amount of carbon taken up by the land surface. As indicated in the Figure captions, stipples represent areas where the sign of NEA changes in C-only or CNP mode as compared to the reference CN mode. In other terms, adding or removing nutrient limitation in these regions transforms sinks into sources and conversely. This is described in the text ll. 299-303:

> This is most obvious in the mid- and high-latitudes of the northern hemisphere where the difference between C-only and CN simulations in NEA exceeds 5 kg C m^{-2} (Figure 7) which is large enough to change the sign of NEA and change these regions from net sources in CN simulations to net sinks in C-only simulations (stipples on Figure 7), especially with the K1995 f_T.

We agree that figures 7 and 9 are partly redundant and we moved them to supplementary information.

Figure 10 and discussion on page 10243: How does the data in this graph related to NEA? It seems like there's a lot of information that's synthesized in this figure, but I can't really put my head around what it means. I'm also confused what's really being shown, for example the black line is the standard deviation of the mean (text) or the mean (caption). What is the signal and noise and how are they calculated. Qualitatively it looks like there's relatively little signal (trend), and a great deal of noise (variation) in all figures, but that there's less signal and more noise in the CNP simulations, correct? As far as I can tell, this is driven by less signal, not more noise from the choice of environmental response functions- but I may be mistaken? Finally, what causes the step function ca. 1960 in the C-only model?

NEA and NEE are essentially the same flux in this paper and we have harmonised the notation on Figure 9. We also changed the sign so that positive values now correspond to an accumulation, or a sink, which is consistent with the notation used throughout the rest of the manuscript.

We have clarified the description of this figure in the text ll. 421-425:

> We define the "signal" as the temporal variability in NEA/NPP, calculated as the standard deviation of the annual mean NEA/NPP (the annual mean is represented in black in Figure 9). The "noise" is calculated as the intra-annual variability between combinations of f_W and f_T. It corresponds to the standard deviation of the distance between all models and the mean NEA/NPP for all years (full ranges with maximum distances are in grey in Figure 9).

The referee is right in attributing the differences in signal-to-noise ratio to a weaker signal in CN and CNP simulations. We acknowledge this in the discussion ll.427-432:

> This indicates that the uncertainty due to f_W and f_T relative to the variability in NEA increases when NP limitations are added. From Figure 10, it is clear that this

> results from the lack of response to post-1960 increase in atmospheric CO_2. Differences in the formulation of these response functions contribute to a relatively larger part of the simulated ranges, making their correct definition increasingly important as the representation of the terrestrial carbon uptake improves.

We also added a supplementary Figure that helps to attribute the step around 1960 to a sudden increase in the growth rate of atmospheric CO_2. This has been added to the main text as ll. 410-414:

> As a result, CN and CNP modes (Figures 9b, 9c) do not exhibit the post-1960 step change in NEA that corresponds to a greater carbon sink in the C-only model (Figure 9a) in response to the sudden increase in the growth rate of atmospheric CO_2 concentrations (supplementary Fig. S6).

Fig 11 and discussion on pages 10244-10245. In trying to explain the mechanistic rationale for changes in C storage pointing to a bunch of maps is not very useful or quantitative. Could the authors show results from regression or correlation analyses showing how changes in NPP and or environmental scalars (fw and ft) drive changes in soil C pools and / or the strength of the terrestrial C sink?

We thank the referee for this comment. As we already moved two figures to the supplementary material, we decided to keep the maps in the now Fig. 10 to have a better representation of spatial changes of the environmental control on respiration. However, we have performed linear regressions that quantify what we describe in the maps. We now provide this measure of correlation (actually R^2) between this change in environmental control and the change in soil carbon within each version ll. 460-462:

> This does not correlate with the change in soil C (Figure 6) and, based on linear regressions (not shown) the change in the environmental control on R_h never explains more than 12% of the variability of the change in soil carbon.

In general I think conclusions should be revised. The conclusion that ft (or fw) is of greater importance with nutrient limitation isn't clear to me. How was this determined (I think it relates to Fig. 10)? Can the authors provide more interpretation about the mechanisms involved here?

We have modified the conclusion to better reflect the main point of our paper that fT and fW control initial conditions, hence the response of the land balance to global warming, and that N and P remove the high response to quickly increasing atmospheric CO_2 concentrations ll. 513-535:

> We have used 27 combinations of f_T, f_W and nutrient limitations in an Earth System Model to explore how the land carbon balance responds to changing atmospheric CO_2 over the period 1850-2005. Various formulations of f_T and f_W generate a range of equilibrated soil carbon stores very similar to the six-fold range of global soil C achieved by CMIP5 models regardless whether nutrient limitation is implemented. That is, the range in soil carbon in CMIP5 is likely the result of equilibration methods.
>
> Implementing N and P limitations on plant productivity in the CASA-CNP ecosystem model better constrains the simulation of the historical response of the terrestrial C cycle irrespective of the f_T or f_W used because of the lack of response to post-1960 rapid increase in atmospheric CO_2. However, in these simulations the

> initial carbon pool size is the main driver of the response of soil carbon to global warming. As the magnitude of the available substrate controls the sensitivity of R_h to changes in temperature and moisture, larger pools are more likely to deplete under global warming. Due to the size of soil carbon pools even small changes in forcing can lead to large C losses and drive the whole land C balance response to warming. Therefore, the wide range of responses in CMIP5 in terms of soil carbon may well be an artefact of the initialisation procedure used.
>
> Based on our experiments, we recommend representing at least CN interactions in Earth System Models in order to capture the correct magnitude of historical land-atmosphere carbon fluxes and the response of the system to increasing atmospheric CO_2. The other clear implication of our results is that a more concerted effort in how microbial decomposition processes are represented in Earth System Models is required. We need to address how equilibrium should be defined or constrained to match some estimates, how nutrients should be represented and how we develop these efforts with limited global data bases of soil carbon.

Technical corrections

P 10232 L 14-17. It seems worth noting that only 3 of the CMIP-5 models include N limitation (and two of them use the same land model CLM). While this has changed, few of the CMIP5 models represent nutrient limitation.

We modified the following sentence to emphasise that two models with N use the same land model ll. 73-79:

> However, despite the recognition of the importance of interactions between these biogeochemical cycles, interactions between terrestrial C and N cycles are represented in just three of the ESMs used in the Coupled Model Intercomparison Project, Phase 5 (CMIP5; Taylor et al., 2012), among which two models use the Community Land Model as their terrestrial component: CCSM4 and NorESM (Todd-Brown et al., 2013). Meanwhile, the terrestrial P cycle is omitted in all CMIP5 simulations (Todd-Brown et al., 2013).

P 10233 L 8-9 fw is used synonymously with SMFR (and M in figures), as are ft and STRF (and T in Figures). One abbreviation for the same functions seems adequate, and would aid in understanding. As they are used in eq. 1, Please use fw and ft throughout (or write out temperature function and soil moisture function.

We followed both reviewers advice by replacing the terms STRF and SMRF by the more readable f_T and f_W throughout the text. We also replaced the labelling in Figures.

P 10236, L 17: consider replacing “the near future” with “future analyses” since no future projections are presented in the current work

We agree with the reviewer and replaced words accordingly.

Caption for Fig 2 does not match text in section 3.1, with terms used interchangeably (e.g. NEA and net ecosystem productivity). Are thin black line individual model runs with different combinations of fw and ft? If so, does the shaded area represent range of results of 9 combinations of fw and ft?

We have clarified the legend of Fig 2:

> Figure 2. Global cumulative human carbon emissions and modelled net ecosystem accumulation (NEA) throughout the historical period using CASA-CNP in C, CN and CNP modes. In each panel, the shaded area represents the range of NEA simulated by the 9 combinations of f_W and f_T in the indicated nutrient mode while thin lines represent individual simulations identified by f_W. The insert in the right panel shows the relationship between cumulative human emissions and atmospheric CO_2 (in ppmv) over the same period as the model is driven by this latter.

P 10237, L 1: consider rewording "there are very major changes"

We deleted "very major changes" and the sentence now reads ll. 204-206:

> However, there are major differences between the results from the C, CN and CNP modes and between the various moisture and temperature functions.

P 10237, L 21: remove "In terms of the mean terrestrial sink". To many clauses in this sentence make it difficult to understand.

This was done.

P 10238, L 27: remove "average land carbon (vegetation + litter + soil), or"; NEA has already been defined.

We rephrased the sentence to ll. 263-264:

> Figure 5 shows NEA between 1996 - 2005 and 1850 - 1859 for each f_W and f_T in CN simulations.

Fig 4: (and corresponding results) here NEA is convoluted with "land carbon", which I believe are the same thing? If so, please use a consistent term throughout.

The referee is right in that the change in land carbon is equivalent to the cumulative NEA since 1850. We rephrased the legend accordingly:

> Figure 5. Cumulative NEA during historical simulations as represented by the difference in the average land carbon between 1996 - 2005 and 1850 - 1859 in CN simulations. Each sub-panel corresponds to a single combination of response functions as indicated.

P 10242, L 14-27: While I agree with the sentiment of this discussion, nothing in this analysis makes me think the added complexity of adding P limitation to the model structure is warranted- while CN and CNP results overlap (Fig 2 & 3) if the later was penalized for its added complexity it seems like a "worse" model configuration.

CN and CNP models are indeed very similar for historical runs, but are likely to provide similar simulations for different reasons as they exhibit some regional differences (Fig. 8). Therefore, it is not sure whether this common behaviour will be held in the future when stronger forcing will be imposed to the system. We acknowledge this in the discussion part ll. 432-437:

> It is however worth noting that CN and CNP simulations are quite comparable under historical forcing despite the added complexity of P limitation. However, as indicated by maps in Fig. 8, CN and CNP simulations seem to provide different regional responses that sum up to a similar global signal. This characteristic may involve heterogeneous responses of CN and CNP modes to future climate change, something that remains to be explored beyond the work presented here.

P 10244 L 10-11 where are results showing that NPP is similar between common nutrient simulations (C-only, CN, and CNP)?

We added Figures of the average NPP simulated by each model version in each nutrient limitation mode in the supplementary material (Figure S4 to S6). We now acknowledge these figures in the main text ll. 444-447:

> Since we prescribed atmospheric CO_2, NPP is very similar for all our C-only model simulations (Supplementary Fig. S3). NPP is also very similar within all the N-limited simulations (Supplementary Fig. S4), and the NP limited simulations (Supplementary Fig. S5).

Fig 12. Are these 1850 or 2006 results? It seems odd to introduce these results at this stage of the manuscript.

Soil carbon density is presented for 1850. We followed the advice of the reviewer and this figure has now become Fig. 4 and is called much earlier in the text. We updated the legend to make clear that we present values as of 1850.

P 10247, L 1-11: while I completely agree this discussion seems outside the scope of the data being presented here.

We agree that this part is not directly related to our experiments as we do not propose a new enzyme-based model. However, our results *add* to these concerns on the adequacy of first-order parameterization of microbial decomposition. To make it clearer, we removed the last part of the sentence that explicitly evoked the physiological control on decomposition and it now reads ll. 498-500:

> This adds to recently stated concerns that the current parameterization of decomposition is not representative of our understanding of this process (Allison et al., 2010, Schmidt et al., 2011, Todd-Brown et al., 2012).

P 10247, L 21-27: This text should be removed or revised. I would not highlight the unforeseen importance of environmental scalars on calculating equilibrium soil C pools and sensitivity to climate change (see Davidson and Janssens Nature 2006, Xia et al Geosci. Model Dev 2012 and Xia et al. Glob Ch. Bio 2013, and references therin- alsoTodd-Brown et al. 2013- which should be cited (L 27)).

We have changed the conclusion (see previous replies).