

We thank both referees for their careful consideration of and comments on the manuscript. Specific responses to each point raised are below, in red text.

Review of Koven et al., "The effect of vertically-resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4"

This paper sets out to document the development and testing of improved soil C and N biogeochemistry in the CLM land surface model. It is a very clearly written manuscript and is an excellent example of developing, implementing, testing and evaluating different options for a model formulation. If only all -CN soil model development was performed so well!

As well as a very clear documentation of the developments, and reasons for them, in CLM4, the manuscript examines some important scientific questions. In particular the study finds a need to introduce a parametrisation of reduced SOM turnover with depth, and discusses why this might be required. It is clear that such a parametrisation improves the fit of model simulations to data (with use of isotopic data being particularly important) but it is not yet known why. Far from being a failing, this is a good example of how detailed model testing and evaluation can highlight potentially important missing processes - future development may now address this issue and attempt to solve the reasons behind it.

I recommend this paper be published with minor revisions. A few specific comments and suggestions follow.

Chris Jones

We thank the referee for this assessment.

- One motivation for trying the CENTURY scheme for C-cascade is to address the total amount of carbon stored, and you find that it improves (increases) this. But could you not simply adjust the turnover times in the original CLM scheme? If a soil-C model gives a factor 2 wrong storage for about right NPP inputs, then it can't just be the cascade scheme that's wrong - there must be a problem with the mean lifetime. There must be lots of tunable scope in such parameters as tau, especially for global application - is it really necessary to adjust the cascading of carbon rather than the lifetime of each pool? If you simply tune the turnover times in the CLM scheme you could easily get the right global amount - so is there a reason not to? e.g. does this break the site level evaluation?

We agree that a more simple model parameter change could have led to

increases in the total C storage. However, as you say, there is enormous tunable scope in the soil C cascade, and therefore the problem would be underdetermined if we were to try to tune each of the decomposition parameters against a relatively small set of integrating observations such as total SOM C and its  $^{14}\text{C}$  age. Instead, we chose to explore the effects of using an entirely different soil C cascade as a way to explore the aggregate differences between the two approaches. Given that the Century approach has been widely used and tested in the community, it was a good reference technique from which to start, and then go on to explore the depth effects,  $^{14}\text{C}$  ages, etc. Interestingly, as can be seen in the differences between figs. 5d and 5e, and between figs. 6a and 6b, the replacement of the decomposition cascade itself has a relatively minor effect on the total carbon stocks, but a much larger effect on the radiocarbon ages; so the taus alone are not responsible for the C stock biases present in the base version of CLM. Further, we do not argue here that the Century approach is unique in satisfying the dual constraints of radiocarbon and total carbon stocks, but that it does provide a closer starting point than the base CLM approach, in particular for satisfying the  $^{14}\text{C}$  age profile constraints, and this is why we adopt it here.

- Having shown some initial maps of carbon content globally you don't discuss global quantities or use these to evaluate the schemes behaviour in its global, 20th century simulations. If you, say, adjust turnover times by  $\pm 20\%$  you would immediately adjust your stores by the same fraction and probably the magnitude of your response to changes. The site level evaluation of amount and profile are extremely important, but getting a good match to global amounts and distribution is also important - I felt you could have made more quantitative use of your observational (or at least data-based) spatial fields.

The revised manuscript will include estimates on the global integrated C change for the transient 20<sup>th</sup> century simulations. For the absolute changes, we will include discussion of the total integrals for each of the simulations as well.

- can you describe your experiment set-up in more detail? e.g. presumably it is land- surface only driven by obs meteorology and  $\text{CO}_2$ . What about N deposition? what about land-use change? any other land management? I think you also need to discuss the inputs from the land-surface model INTO the soil components. At no stage do you show how well your litter inputs to the soil compare with obs - which are probably pretty sparse, but you can assume they are close to NPP on annual timescales and we do have some global fields/estimate of NPP. Likewise are you simulating or prescribing the vegetation PFT cover? or what about soil physical properties? does CLM do a realistic job of soil T and moisture? if the hydrology or soil physics is wrong then your rate modifiers will be and the soil BGC won't match the obs, even for a "perfect" -CN model. I think you should acknowledge that the evaluation is complicated by not

always being able to constrain where in the chain of processes errors creep in - it may not always be the soil BGC

The model is driven offline by a reanalysis meteorology (Qian et al, 2006). N deposition is time-varying based on the Lamarque et al. (2005) simulations, and land-use for the 20<sup>th</sup> century transient runs is based on Hurtt et al. (2006) land-use datasets. More description of these details will be included in the revised manuscript.

For the more comprehensive set of benchmarks, these are being shown in a separate paper (Lawrence et al., *in prep*), documenting the behavior of the CLM4.5 model for a large suite of biogeophysical and biogeochemical benchmarks. This will include temperature, moisture, litter, NPP benchmarks as you suggest.

- p.7207, line 16. You say all the rate modifiers are between 0-1. Is this true for temperature? how do you keep a Q10 between 0-1?

You are correct here; the Q10 can exceed 1 above the reference temperature; we will correct this in the revised manuscript.

- eqn. 8. can you define "As"? Presumably related (but is it exactly the same as) atm C14 from figure 3?

$A_s$  in equation 8 is the  $^{14}\text{C}/\text{C}$  ratio in a given sample. So this is not exactly the same as the  $\Delta^{14}\text{C}$  shown in figure 3, which has different units. Equations 8 and 9 allow the expression of the raw isotopic ratios as  $\Delta$  notation. This will be defined in the revised paper.

- is it worth adding an Appendix, or supplementary info, with a more complete set of technical details of the model? such as how many vertical levels you use, do you have a list of new prognostic variables? It is not clear always in the process discussion whether some of the N-related variables in particular are state variables, or diagnostic. Any new ancillary data that are required - such as soil texture, depth etc? Plus maybe a list of data used in the evaluation - other modelling groups might like to be able to access these site level data for example

A formal technical note has been prepared with a much more exhaustive set of technical details for the CLM4.5, in Oleson et al., 2013. We will add some relevant details, as you suggest, here. The base version of the model uses 15 soil levels, but only 10 of these are hydrologically and biogeochemically active, with a maximum depth of 5m. For the vertical soil model, no new ancillary data are required beyond what is described here; soil texture is already used for soil hydrologic calculations, and soil depth is globally uniform.

- can you check a few of the entries in the tables? e.g.: - p.7216, line 25 (should refer to table 1?) - says the "A" factors collapse the turnover times onto that of the fastest pool. But "soil 3" has  $A=5$ , but there is more than a factor of 5 between soil 3 and soil 1  $\tau$ ? - some of the table 2 numbers don't match figure 2 - e.g.  $r_j$  for L2-S2 -  $T_{ji}$  for CWD sums to 2?

We will correct the differences you point out; e.g. the CWD should be the same for the 2 cascades in table 2 and sum to 1, not 2. Also, the text should be revised to say that the idea of the spinup is to collapse the slow pools onto an approximately 1 year  $\tau$ , such that they are sufficiently responsive to the changing inputs but still allow the annual cycles of N mineralization to be approximately what they would be in the full model.

- p.7224, lines 10-20. Can you describe how you spin-up your permafrost? in reality much of the carbon here accumulated in warmer climates which subsequently froze, so it's not possible to spin-up this pool under constant climate. Did you initialise from obs?

We follow the method of Koven et al., 2009 for this, which attempts to spin-up the near-surface permafrost (to a depth of less than 3m) rather than initialize from the observations. The idea of this approach is that, as you suggest, in reality much of the carbon in these soils has accumulated in prior climates, but that there is some continuity between active-layer carbon and carbon in the upper permafrost through transport of material by cryoturbation and other soil processes. To include deeper (>3m) carbon in syngenetic permafrost such as Yedoma, loess, or deltaic sediments, we would need to initialize the C from maps, but we do not include these pools in these simulations.

- p. 7225, line 12. Now your soil-CN model has the capability to affect vegetation productivity, which is not the case in -C only models, can you show then how vegetation carbon/storage/productivity etc are affected by the addition and development of soil-N processes? Should this not also form part of the evaluation? The paper is about the role of this on carbon dynamics, but you only really show soil-C results.

Fig. 13 disaggregates the 20<sup>th</sup> century C changes into soil litter and vegetation changes and shows them as functions of latitude. This shows the basics of what you are suggesting. The revised CN coupling lifts some of the N limitation throughout the model, and this allows vegetation to grow more in response to the 20<sup>th</sup> century changes.

Review of “the effect of vertically-resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4” by Koven et al.

This is a clearly written paper. It documents the new soil biogeochemical models as implemented in one of leading global land surface models, CLM. The authors clearly explained what the new models can do, and cannot do. This work is a significant contribution. I recommend it for publication after some minor revisions.

We thank the referee for this assessment.

P3. On L13-15. It is not always true that single-layer approach has an implicit assumption about the depth (30 cm) of soil C the model represents. In theory, the single-layer model simulates all C in the rooting zone. Many single-layer models underestimates soil C because they overestimate the soil C turnover rates, as limitations from oxygen and other factors are not accounted for.

We agree; the actual assumptions in a single-layer soil biogeochemical model are that all C inputs are included in the box, but the actual vertical structure of such a model is left undefined. What is typically defined is the vertical domain of the physical properties (e.g. temperature, moisture) used to calculate environmental control on turnover, and secondly, a depth interval is usually chosen for comparison of the model against observational estimates. However, such a comparison is somewhat arbitrary as the bottom boundary of single-level models is not well defined.

P9. Eqn (7). In the field, it is difficult to distinguish between the effect of  $r_z$  and other effects, such as the effect of oxygen supply. If the effect of oxygen has been accounted for by  $r_o$ , why the effect of microsite anoxia is also included in  $r_z$ . Does this account for the effect of oxygen twice? Should  $r_o$  be dependent on the anaerobic fraction (see your eqn 12)? Much of the work by Arah and his colleague is related to soil aggregates.

A central point of this paper is that we make no a priori assumptions here about what is responsible for the  $r_z$  term that we include in the model; thus oxygen may be one such factor. The explicit oxygen term we use for  $r_o$  includes some, but not all, possible oxygen effects. Comparing the  $r_o$  term to the Arah and Vinten term we use for determining the oxic/anoxic fraction for the nitrification and denitrification losses, the latter also includes uptake rates itself, and this is why it has minima both in shallow tropical soils (where uptake of oxygen is high) and in deeper soils globally (where oxygen concentrations are lower). Given that the reduction of soil decomposition rates are seen globally, even where oxygen uptake is small at depth, suggests that the Arah and Vinten approach will not appropriately limit the decomposition with depth. Other oxygen-limitation

mechanisms, such as exclusion of oxygen by larger aggregates at depth than are used in the Arah and Vinten approach, are possible, but we do not try to explicitly model them here. The larger point is that some process, beyond what we do model, is needed to capture the reduction in decomposition rates at depth.

P13. Eqn (12). In theory,  $\bar{A}_c$  increases with a decrease in soil temperature. Is that dependence significant? particularly for soil at high latitudes?

We have not performed sensitivity tests of this, but since the relative change in oxygen concentration, as a function of temperature, is proportional to the relative change in absolute temperature, we expect this to be small (e.g. a 3% change for a 10 degree change at typical conditions). Also, the exponent of this term, which is less than one, will further reduce the sensitivity to this term.

P13. Eqn (13). As explained by Houlton et al. (2008, Nature) and Wang and Houlton (2009, GRL), equation (13) would predict higher N fixation rate in evergreen tropical forest than in tropical savannah, which is contradictory to observations. Does BNF in eqn (13) include both symbiotic and asymbiotic N fixation? A better way to capture the seasonal variation of BNF is to use the variation of canopy LAI and mean temperature. This may be future work for CLM, but needs to be discussed. The dependence of N fixation on temperature has been provided by Houlton et al. (2008). Wang et al. (2007, GBC) developed the first model of N fixation, and Wang and Houlton (2009) provided the first estimates of global BNF.

It is true that equation 13 would, e.g. predict higher N fixation in tropical forests than savannahs, contrary to observations. As discussed here, this is future work and some discussion of the shortcomings of the current CLM BNF parameterization is already included here. We will include more discussion of these shortcomings in the revised version.

P14, L 15. See my comments above. Houlton et al. (2008), Wang et al. (2007, GBC) and Wang and Houlton (2009) should be cited here, as they provided the first explanation of global N fixation variation.

Will include this in revised version.

P14, L25-26. It remains controversial whether plants prefer  $\text{NH}_4$  to  $\text{NO}_3$ . This should be noted here.

Agreed, this is an open question, and we will include this in revised version.

P17. L1-3. The recent work on accelerated spinup by Xia et al. (GMD 2012) should be cited here.

Will include this in revised version. However we note that this paper is not directly relevant to the effort described here since we do not use an iterative equilibrium-calculating approach as is described in the suggested manuscript.

P17, section 2.6. It should be stated more explicitly whether variation of vegetation type in each land cell from land use and fire is accounted for, because this has implications on your simulated global carbon budget (Figure 12).

These variations are included, and more discussion of this will be in the revised manuscript.

P25, L18. Also cite the work by Aerts and Chapin (2000, Adv. Ecol. Res).

Will include this in revised version.

P25, L26. Galloway et al. (2004)'s estimate BNF of 120 Tg N yr<sup>-1</sup> does not include asymbiotic N fixation. More recent estimate of N export to river has been provided by Dumont et al (2005, GBC), and more recent estimate of denitrification loss by Bai, Houlton and Wang (2012, bio- geoscience) using 15N data.

Will include discussion of this in revised version.

Figure 11. Your estimate of denitrification loss is much higher than the estimate by Bai et al. (2012). This needs to be discussed.

We believe this is related to the too-low dissolved N losses, which require that denitrification losses are high to compensate. More discussion of this point will be in the revised manuscript.

Figure 12. If land use change is included, your simulated global land carbon change by all three versions of CLM agree with the other estimates, given the uncertainty of CO<sub>2</sub> emission from land sue change. If land use change is not included, you should simulate a net land carbon sink about 80 150 Pg C from 1955 to 2005. Please clarify.

Land use is included in these estimates. The revised manuscript will include the estimated change from the global carbon project (Le Quere et al., 2013) budget.