We wish to thank both referees for their positive and helpful comments. We answer each comment below.

Our comments are preceded by 'AU>>'.

----

1. OVERVIEW

The authors investigate the differences between two approaches used to represent sub-grid cell vegetation heterogeneity in dynamic global vegetation models (DGVM), using CLASS-CTEM. For pre-industrial equilibrium simulations and for transient simulations that do not consider land use change (LUC), the two approaches give similar global-scale results, despite substantial differences in various regions. However, when LUC is included, the two approaches give very different results for the atmosphere-to land global carbon flux over the 1959-2005 period. The three main strengths of the manuscript are the comparisons of CLASS-CTEM results with many other estimates (Table 2, Fig. 2, and Fig. 4), the in-depth explanations for a given grid cell in Russia, and the overall presentation quality (especially the Figures). The main weakness is the lack of sufficient discussion of what the outcomes mean for modellers that do not use CLASS-CTEM (see point 2.1). I consider that the manuscript fits well within the scope of Biogeosciences and suggest that it be accepted for publication, provided that the authors address the comments below.

2. SPECIFIC COMMENTS

2.1 The relevance of the study for modellers that do not use CLASS-CTEM is not obvious, for two different reasons. In order to improve the manuscript, I suggest (but do not request) that the authors address this shortcoming.

2.1.1 Besides CLASS-CTEM, the authors do not provide any example of a modern DGVM that resort to the mosaic approach as defined in the manuscript. As for the two examples of modern DGVM (besides CLASS-CTEM) using the composite approach as defined in the manuscript, I would rather argue that they resort to the mixed approach. In LPJ-DGVM, “the grid cell is treated as a mosaic divided into fractional coverages of PFTs”, for which “the physical environment is well mixed”, e.g., they share the same soil moisture (quotes are from Sitch et al., 2003). In CLM 4.0, all “the fluxes to and from the surface are defined at the PFT level, as are the vegetation state variables”, but the PFTs share a single soil column (quotes are from Oleson et al., 2010; see Section 1.1.2). The few other DGVM I know also resort to a mixed approach. How should these modellers interpret the results of the study? Should the mixed-approach DGVM be considered closer to the mosaic type (because PFT-specific fluxes are averaged over the grid cell) or the composite type (because the soil is the same for all the PFTs)?

AU>> This question about implications for other models is interesting, but presents us with a quandary about how to generalize to the diversity of models currently in use. While the reviewer feels that LPJ should be a mixed approach, the first author has experience using the LPJ model and believes that based on how the model is coded, it should actually fall into the composite category. This is a good example of how difficult it is to determine how the models are, in reality, structured from generalized descriptions in the literature. From this, it would be difficult for us to place other models in the spectrum. The reviewer is likely correct that CLM is more of a mixed approach and the Oleson et al. (2010) reference has been moved to that category.

It is likely that the composite and mosaic techniques present end-members for the range of models. Therefore it is perhaps best to think of the two versions of CLASS-CTEM (composite and mosaic) as upper and lower ranges of how the majority of models would be structured. <<AU

2.1.2 Regardless of point 2.1.1, how should the users of other models react to the study? To my knowledge, not many models (if any) besides CLASS-CTEM offer the possibility to choose between different (composite, mosaic, or mixed) approaches. If the study had concluded that one of the two approaches tested was clearly better, then users of other models could reflect upon the appropriateness of switching to this better approach. (Or maybe that the best approach is actually a mixed one, not tested in the study?) I understand that the authors do not have sufficient evidence to take a clear position, but additional discussion of the merits of at least the two approaches tested would improve the manuscript. Here are a few points to consider, at the authors will. (a) The two approaches are briefly discussed in the Introduction (end of page 16006); this text could be moved to the discussion and expanded, with references to support the claims that are made. (b) If each approach is better suited to specific vegetation types, does it mean that DGVM should change their approach from one grid cell to another? And maybe even through time in the same grid cell, if the vegetation type changes? If yes, based upon which criteria? (c) From an ecological perspective, do composite structural and physiological attributes make sense? Plants function as individual units... (d) See point 2.7 below.

AU>> Yes, the reviewer is likely correct in that many models do not offer the ability to change configuration between composite/mixed/mosaic. However, as we note in the paper, different landscapes are likely better represented by the composite or mosaic configurations. This is not something that we can expand upon with references, as it is a simple observation that landscapes with a homogeneous mixture of PFTs appear well represented by the composite configuration, while landscapes with distinct patches of non-overlapping PFTs appear to be well represented by the mosaic configuration. It is likely that for given a certain landscape, one configuration is better suited to simulated that landscape. However, on a global-scale which model configuration is more appropriate is not very clear. The merits of each configuration are really dependent upon the actual grid cell that is being simulated. It would be very interesting if it was feasible to switch configurations based upon the grid cell being simulated, using mosaic for some and composite for others or even evolving through time. This would likely be the most accurate way to simulate the terrestrial surface. However, at present, we don't have a parameterization that can accomplish this.

For the comment (point c above) about an ecological perspective, the structural and physiological attributes of the PFTs are aggregated only for the canopy and water balance in the composite configuration. This approximation is similar to assuming that the plants share a similar soil temperature and moisture. Error introduced by this approximation will depend on the PFTs being simulated and the landscape structure. These composite configurations are a common approximation (as in LPJ for example). Yes, ideally the interactions of each PFT with the atmosphere would be modelled explicitly, but all global-scale models must make compromises due to computational cost.<<AU

2.2 What is the fate of crops biomass? Normally crops should be harvested each year (otherwise, soil carbon could potentially build up to unrealistic high values). I assume that this is the case in CLASS-CTEM, right? If not, why? If yes, what happens with the crops biomass carbon (is it sent to the atmosphere immediately after the harvest)? These points should be addressed at the end of Section 2.1.

AU>>During harvest, the crop biomass is transferred to the litter pool. Crops are harvested every year. We have added in more information about the treatment of crops due to harvesting. <<AU

2.3 In Table 2, results from pre-industrial equilibrium simulations are compared with contemporary estimates, which is not very informative given all the changes (LUC, climate, CO2 concentration, etc.) that have occurred since 1861. Replace the “other estimates” by pre-industrial values; for example, some values are provided in Table 2 of Arora and Boer (2010).

AU>> We had previously not compared our results to other estimates of the pre-industrial period as comparing a model result to another model result does not necessarily result in greater certainty than comparing a pre-industrial model result to modern observations. Comparing a model to a model might just confirm that they agree, but in reality are both wrong. As well, the purpose of the comparison of the pre-industrial simulations to observations is to demonstrate that the observations of the differences between composite and mosaic are not attributable to the model having a generally unrealistic output. Regardless, we have added several pre-industrial modelling estimates alongside our model estimates in Table 2. <<AU

2.4 At first, the comparison of CLASS-CTEM results with the Houghton et al. (2012) estimate casts doubt on the capacity of CLASS-CTEM to simulate LUC impacts (end of page 16016 and Fig. 4b). Moving here some of the text from the Discussion (page 16020, lines 14+), or at least stating clearly that changes in pasture area were not simulated, could prove helpful. Urbanization (page 16020, line 18) is probably irrelevant, because global LUC datasets apparently do not account for it neither (Houghton et al.,2012; Section 5.2). I would however suggest discussing the role of soil carbon (see the end of Section 3.2.3 in Houghton et al., 2012): bookkeeping approaches apparently assume high soil carbon losses (likely included in the Houghton et al. (2012) data), whereas the results for CLASS-CTEM in Fig. 4b are for vegetation only.

AU>> CLASS-CTEM does indeed estimate LUC impacts on the lower-end of other models and the bookkeeping approach of Houghton et al. (2012). As suggested we have reminded the reader that our estimate does not include changes in pasture area, wood harvesting and logging, and shifting cultivation. We now note that Houghton also neglects urbanization. Our LUC estimates do not include degradation of the soil C pool directly as the reviewer notes. Although Fig 4b shows deforested biomass only, our soil C pools are influenced by LUC effects if those effects influence the soil moisture and temperature. As well, the rate of C inputs to the soil change under LUC. << AU

2.5 I appreciate the effort of the authors to better explain their results through the use of the H index. However, I have three issues with the use of the H index.

2.5.1 The authors ask us to compare Figs. 6 and 3 to see the association between the H index and the differences between the two approaches. It is obvious that desertic regions have low H and small absolute differences between the approaches. But the association between high H and high negative values (higher results for the composite approach) appears clearly to me only for south-eastern China and a few pixels in Mexico. For the rest of the world, I do not see much because the results are too variable over space... It is even worse for Fig. 3 vs. Fig. 5a, in particular for the "US Prairies" (which are not clearly identified) example. If the authors want to convince readers, I suggest (but do not request) that they compute a global-scale value for some association indicator between the two elements.

AU>>The heterogeneity index is not intended to be a prescriptive measure. We intended it to be used to highlight areas that could be expected to have greater differences between the composite and mosaic configurations due to the PFT spatial representation. It has, as a result, relatively low predictive power (e.g. $r = 0.148$ for GPP) as it does not integrate information about climate and soil conditions within the gridcells. The variability in space that the reviewer notes primaily reflects the impact of climate and soil conditions. We have added in some text to the MS to better explain what we believe to be the utility of the H index.<<AU

2.5.2 Either I do not understand the definition of the terms in Equation (5), either the claim that the H index takes "a value of 0 if an entire grid cell is occupied with only a single PFT" is inaccurate. Please look at the mathematical exercise in supplement and respond accordingly.

AU>> Yes, thanks to the reviewer for noting that and taking the time to check the H index. It was originally $1/N$, but should have been $1/(N-1)$. It is now fixed in equation 5 and the figure. <<AU

2.5.3 Regardless of point 2.5.2, I am not sure whether the H index is relevant for grid cells that are mostly covered by bare ground. What would be the value of the H index in a grid cell that is 90% bare, but has the remaining 10% equally divided among the 9 PTFs? I have a sense that the H index would be high. Yet having a high H index in such a case would be a poor indicator of possible important absolute differences in LUC impacts between the composite and mixed approaches, because there is very little vegetation in the first place. Please address this minor point in your response only, not in the manuscript itself.

AU>> In the reviewer's example above, the H index would indeed be high as the grid cell would have a high heterogeneity of land cover. This is correct behaviour according to how we defined the H index, which as mentioned previously, does not include information about climate. The LUC impacts are likely low in a grid cell that is already 90% bare, thus any differences between mosaic/composite would also be low. It is also hard to imagine how a gridcell that is 90% bare could have much land use since the land would likely be extremely arid, lack topsoil, or be very cold. The importance of LUC is quite dependent upon the amount of vegetation biomass present so LUC would also likely be unimportant on a grid cell such as this. <<AU

2.6 Once again, I appreciate the effort of the authors regarding the R_C index in order to better understand the results of Fig. 5b. But we now must look at three Figures (Figs. 5b, 6, and 7) simultaneously and, honestly, I almost do not manage to see anything out of it. Once again, I suggest (but do not request) that the authors provide a formal global-scale association indicator.

AU>> A formal global-scale association indicator would be very difficult to formulate. For the indicator to have any predictive power it would need to account for a great number of variables, such as PFT spatial heterogeneity, land use change intensity, climate and soil conditions. We have chosen to not attempt this. <<AU

2.7 Are the results from the mosaic approach credible under LUC? The first element of doubt is the following. Based on Fig. 4c, the impacts of LUC when only climate change is accounted for are 14.3 PgC (4.1 minus -10.2) for the composite approach and 7.6 PgC (0.0 minus -7.6) for the mosaic approach. When CO_2 is also accounted for, the impacts of LUC increase to 21.4 PgC for the composite approach, but *decrease* to 1.2 PgC for the mosaic approach (which is counter-intuitive, because CO2 fertilization should lead to more LUC-caused emissions). The second element of doubt is the following. For the specific grid cell analyzed, we see that the amount of cropland basically doubles from 30 to 60% between 1860 and 1940 (Fig. 8a). Yet during this time the amount of soil carbon slightly increases for the mosaic approach (Fig. 8e). This is counter-intuitive, because 1) conversion to cropland is believed to cause important losses of soil carbon (Houghton et al., 2012) and 2) the CLASS-CTEM parameter for soil respiration is much higher for crops than for natural vegetation (Table A1). I suggest (but do not request) that the authors consider discussing these observations, which are also related to points 2.8 and 2.9.

AU>> The reviewer is correct that, intuitively, CO2 fertilization should increase LUC emissions due to increased vegetation biomass. This is exactly the response we see in the composite approach for the Russian gridcell (and also globally as can be seen in the new figure described in response to the reviewer's point 2.9). The counter-intuitive response of the modelled terrestrial carbon budget to LUC in the mosaic approach was a surprise for us as well. In its normal operational mode in the Canadian ESM, CLASS+CTEM are used with the composite approach. The results shown here have given us an insight into the behaviour of the LUC parameterization. These simulations have highlighted the need to rethink how the model treats LUC so that irregardless of how the vegetation is spatially represented, the results are realistic. <<AU

2.8 Is the higher productivity of crops, compared to the natural vegetation they replace (page 16021, line 4 and page 16024, line 23), credible? This appears to contradict textbooks values, particularly for tropical and temperate forests (e.g., Tables 6.3 and 6.6 of Chapin et al., 2002). The authors need to discuss this point.

AU>> Table 6.4 in Chapin et al. (2002) (assumedly the reviewer intended this table, not 6.3) shows the aboveground productivity of crops to be more than double that of a temperate grassland (530 g/m2/yr vs. 250 g/m2/yr), while the below-ground productivity of the grasslands (500 g/m2/yr) is much higher than the crops (80 g/m2/yr). It is unclear how the values in Chapin et al (2002) are derived (as there is no information given) so difficult to parse how relevant these figures are to our simulation outputs. In CLASS-CTEM, the greater productivity of crops is primarily a reflection of their enhanced Vcmax values, the C3 crops in particular. Our Vcmax value for C3 crops is on the upper end of other models (Rogers, 2013) but actually below the mean value from a compendium of measurements as gathered by Kattge et al. (2009). <<AU

2.9 Could the lower LUC emissions under the mosaic approach result mostly from the build up of soil carbon following LUC, resulting itself from the cooler soil temperature of the new cropland (due to higher albedo?). This hypothesis is coherent with Fig. 8 and the related explanations (page 16022, lines 15+), but does it apply to the majority of LUC-affected grid cells or just to this single grid cell? In particular, what is the global impact of LUC on the soil carbon pool for each approach, both with and without CO2 fertilization? Please have a look at this hypothesis and respond accordingly.

AU>> Yes, this hypothesis partly explains the difference in mosaic vs. composite approaches' response to LUC. We use the gridcell in Russia as a case point for how this could occur. Stimulated by this comment we have added a new figure to the paper that shows the global evolution of the H_v and H_s pools through time. This figure clearly shows a large increase in the mosaic configuration soil + litter carbon pool with a slight decrease in the composite configuration over the 1959 - 2005 period. The H_v conversely shows a decrease for both the mosaic and composition configurations with a larger decrease for composite. Taken together these plots explain the large difference in Ẽ_LUC between the mosaic and composite configurations as seen in Fig. 4c and discussed in the text. Yes, it then does appear to be a relevant explanation for the majority of the gridcells. <<AU

3. TECHNICAL CORRECTIONS

3.1 Page 16008, line 20. Is the vegetation necessarily over snow? Or, in the case of crops and grasses, can the vegetation be buried by snow?

AU>> In CLASS, needleleaf and broadleaf trees are not buried by snow while crops and grasses can be buried depending on their LAI<<AU

3.2 Page 16010, line 10. For more information on NBP versus NEP, the authors could refer readers to Chapin et al. (2006).

AU>> Done.

3.3 Page 16011, lines 17+. To clarify, modify the text to: “As in McGuire et al. (2001) and Arora and Boer (2010), we diagnose $\{E_{LUC}\}$ according to [...]”.

AU>>Done.

3.4 Page 16011, last line. Explain how the “total 6 h precipitation amount was used to determine the number of wet half-hour timesteps”.

AU>>This is from Arora (1997). If the precipitation intensity averaged over the six hour period is > 0 mm/hr and < 0.024 mm/hr, all of that precipitation falls within one wet half hour. If the precipitation intensity (P_i) is >0.024 mm/hr & < 2.43 mm/hr then the number of wet half hours (wet) is found via:

$$wet = 2.6 \ln(6.93 * 6 * P_i)$$

if the precipitation intensity > 2.43 mm/hr, then wet = 12 half hours per 6 hour period. This description is more involved than is needed in the paper so we have just added the reference to the original work.

3.5 Page 16013, line 17. Although fire is not modelled explicitly, don’t these results indirectly include the impact of *some* biomass burning, i.e., the deforestation (permanent) fires that are responsible for a part of LUC? Please clarify accordingly.

AU>> Yes, some implicit burning does occur. This is now corrected.

3.6 Page 16016, lines 23+. The deforested biomass correspond to the change in H_V (or L) only, as explained in page 16010 around line 20, right? A reminder might prove useful.

AU>>H_L is correct and has been added.

3.7 Page 16016, line 28. The authors should briefly explain how they derived the 68.8 PgC value over the 1959-2005 period from Houghton et al. (2012). Unless I am mistaken, this value does not appear clearly in the Houghton et al. (2012) paper.

AU>> The 68.8 value is not in the paper, instead for increased accuracy we generated the 68.8 value directly from the data so that our time periods covered are the same. We have added in a short explanation of how the value was derived.

3.8 Page 16020, line 19. The reference should rather be Ramankutty et al. (2007). The authors should specify that the results of that study are for Amazonia only. If the authors rather refer to the range of results in Fig. 1 of Ramankutty et al. (2007), then the authors should cite the original studies.

AU>>Reference fixed. Estimates are global and from Houghton (2003).

3.9 Table A1. In the third column: replace “co-efficient” by “coefficient”; if the coefficient is really unitless write it explicitly, otherwise provide the units. In the fourth and fifth columns, the units are incomplete: there must be a time dimension (per year?) associated with these rates.

AU>>Yes, unitless, and both are per year. Done.

3.10 Fig. 4a). In the Figure itself, the reference should be Le Quéré et al. (2013).

AU>>Done

3.11 Fig 4c). In the Figure itself, put a minus sign in front of the two negative results.

AU>>Done

3.12 Fig. 4, in the legend. For (a), it should be ntilde{F}_Ln instead of ntilde{F}_L (as stated on page 16016, line 6). For (c), please specify that the results are for ntilde{F}_L.

AU>>Done for both.

3.13 Fig. 8a). In the Figure itself, the legend line for "Broadleaf evergreen" goes through the text. I would also recommend a different choice of colour, in order to clearly highlight C3 crops (put in red?).

AU>> Fixed the legend line but left the colours as is.

4. NEW REFERENCES
Chapin et al. (2006). Reconciling Carbon-cycle Concepts, Terminology, and Methods. Ecosystems 9, 1041-1050.
Chapin et al. (2002). Principles of Terrestrial Ecosystem Ecology. Springer.
Ramankutty et al. (2007). Challenges to estimating carbon emissions from tropical deforestation. Global Change Biology 13, 51-66.

AU>> References:

Arora, V.: Land surface modelling in general circulation models: A hydrological perspective, 155–192 pp., University of Melbourne, PhD Thesis, 1997.

Houghton, R. A.: Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, Tellus B Chem. Phys. Meteorol., 55(2), 378–390, 2003.

Kattge, J., Knorr, W., Raddatz, T. and Wirth, C.: Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models, Glob. Chang. Biol., 15(4), 976–991, 2009.

Rogers, A.: The use and misuse of Vc,max in Earth System Models, Photosynth. Res., 114(3), 2013.
<<AU

Anonymous Referee #2:

This work examines how the composite vs. mosaic approaches to address subgrid variability influence simulated fields using CLASS-CTEM. This is a story that needs to be told and the authors do this quite well. A worthy addition to Biogeosciences. I find this very well-written, coherent, and acceptable for publication largely as is, pending a few more minor technical issues outlined below. Note that I only list language/wording issues I found that R1 had not listed.

AU>>We thank the reviewer for the positive comments.

Pg 16009: "is influenced by the leaf phenological, light" You mean phenology here?

AU>> Yes it is referring to phenology but the full sentence was "The proportional allocation to each of these pools is influenced by the leaf phenological, light and root water status of the plant". So we are referring to the phenological status. It appears that both are correct (at least are used in the phenology literature). <<AU

Pg 16015: "the regions where composite simulates larger values" Missing word?

AU>>Fixed.

Section 3.1: Could you clarify which runs are used when you compare with reference datasets? Initially I thought you were comparing pre-industrial with Beer et al, which strikes me as the wrong thing to do... Later now, when I got to Table 2 I realized you are indeed comparing pre-industrial with current. I am troubled both by why this was done (why not use a transient run?) and that the mismatch is generally small. Have we spent so much scientific capital on understanding the effects of global environmental change wrt the carbon cycle only to find out that it’s moot point, that pre-industrial sims agree with current estimates of C cycling?

AU>> We have added comparison to preindustrial model estimates following the comment of referee #1. However, there is a great amount of overlap in modern and preindustrial estimates as can been seen in the revised table. As we expressed in reply to the first referee, the point of Table 2 is to demonstrate that our baseline model results are in-line with other estimates thus our model results, with regard to differences between composite and mosaic configurations, is not due to CLASS-CTEM producing results that are generally unrealistic. As well, to have used transient simulations we would have to include the influence of [CO2], climate and LUC. All of which have uncertainties associated with them that would reduce the utility of the comparison. Given the overlap in estimates between modern and preindustrial (revised Table 2), comparing preindustrial simulations to preindustrial model results and modern observations seems reasonable. <<AU

I am curious how grid cell size influences your results? This is likely beyond the scope here but it strikes me that the impact of composite vs. mosaic will vary as function of spatial resolution.

AU>>That is an interesting question that we do not know the answer to. It is beyond the scope of this manuscript, but would be interesting to look at in the future. <<AU

I would appreciate a compact treatment why composite vs. mosaic causes such a mismatch wrt LUC emissions. The Russian grid cell vignette, while useful, did not generalize sufficiently. That is, either a table with some simple heuristics or an a figure that shows the mismatch based on H and land cover changes etc. I think this would add value.

AU>> We created a new figure in response to this comment (and a similar comment from reviewer #1). This new figure shows the global evolution of the soil+litter and vegetation carbon pools. The principle difference between the mosaic and composite configurations response to LUC has been the gain of soil+litter carbon in the mosaic configuration with a loss in the composite configuration. This difference is the principle driver behind the large difference between the two configurations estimate of Ẽ_LUC. This figure and a full description has been added to the MS. <<AU