Revision of "Global evaluation of ELMv1-CNP and the role of the phosphorus cycle in the historical terrestrial carbon balance" by Yang et al.

In this study, the authors evaluated the global application of the ELM-CNP model and used different data to evaluate model simulations. They compare the model performance against CN version as well as several models from CMIP6. Moreover, they compared their results against a data-driven model GOLUM-CNP. I am familiar with this model. Thus, it was interesting to see the global application of this model. While I appreciate the work, several points in the model codes, outputs, and manuscript need further clarification to make this work merit publication in GMD journal.

Response: We thank the reviewer for the time and efforts in reviewing our manuscript. We carefully addressed each comment as shown below and will revise our manuscript accordingly.

Model codes:
The simulation description states that the simulations were first spun up to bring C, N, and P pools to equilibrium. I believe this is not entirely correct. In your codes in PhosphorusStateUpdate3Mod.F90: you are ignoring the phosphorus pools during spinup but estimating only the fluxes. I see your comment in the codes that the rationale is not ending up with depleted pool size during the transient run, but how are you reaching a steady state from your spinup runs while ignoring the changes in pools?

Response: There are two versions of nutrient competition scheme in ELM, one is the relative demand scheme (RD) and the other one is the Equilibrium Chemistry Approximation (ECA). The default version of ELMv1 uses the RD nutrient scheme as described in this manuscript. The part of code that ignores the phosphorus pools during spinup only applies to ECA scheme (in the code: if ((nu_com .ne. 'RD') .and. ECA_Pconst_RGspin)). For the RD scheme, all the P pools are being updated during spinup.

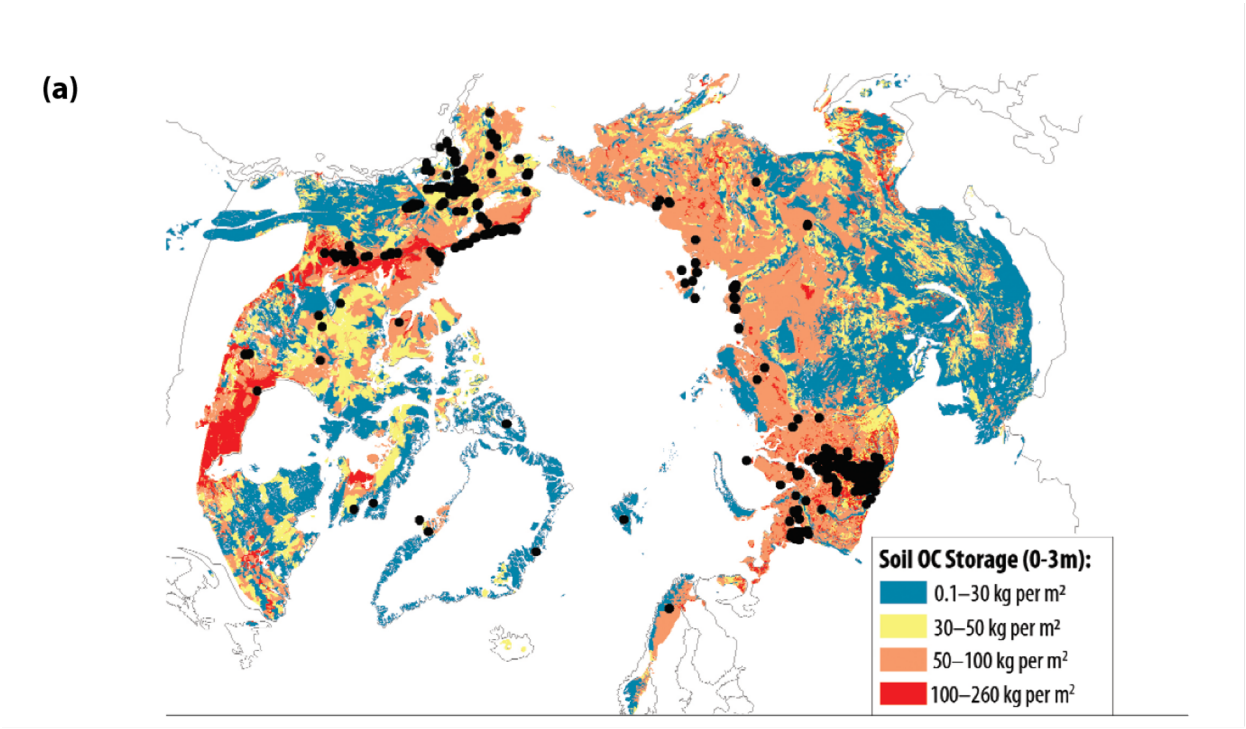
Also, in SoilLittDecompMod, you introduced a 'new' C:P decomposition. Is this rate fixed across all soil types/biomes, or is it changing (similar to the plant stoichiometries)?
Other parameters are also fixed for other processes. For instance, in your ErosionMod the eroded phosphorus (pp2poc) is estimated using a fixed value from outdated reference (Meybeck (1982)) across all the soil types and biomes. What is the rationale behind this? And couldn't you use the updated reference studies and different values per soil type?

Response: The "new" C:P decomposition rate is fixed for all soil types/biomes. Erosion module is not active in the default ELMv1 model, therefore it is outside of the scope of this study. We will clarify this in the revised manuscript.

Model outputs:
I processed one of your output files as an example. In your runs using the CNP model (ALL), there are some extreme values (for instance, in total soil organic matter carbon (TOTSOMC) +240 kg C m-2) (the following figure). What is the reason for such unrealistic values in

model outputs? Have you tried to detect these and understand the underlying issues? Were your final reported values excluding these extreme values?

Response: TOTSOMC is for total soil organic carbon for the whole soil column – up to 3.8m depth. There are a few grid cells in the arctic region that have very high soil carbon, as high as 240 kg C m-2. Total soil carbon to 3m depth in the permafrost region can be up to 260 kg C/m2 based on Schuur et al. (2015). Figure below is Figure 1a in Schuur et al. (2015) which shows the SOC in the interval of 0 to 3m depth of the northern circumpolar permafrost zone. The high soil carbon in these regions is mainly due to cold anoxic conditions. We did not exclude these high values in any of our analysis.



Model results:
Your model results show significantly underestimated leaching of P compared to Wang et al., 2018. Considering this underestimated leached fraction of soil P, your uptake is overestimated consequently (Table 2). Therefore, the available P for plant demand is overestimated as well. This could be the reason that despite the global application of P into the ELM model, still, you under/overestimate productivity similar to the pattern produced by the CN version of the model (figure 3) and overestimate the land sinking capacity (figure 2). This is in contrast with the objective of your study to quantify P limitation over land C sinking capacity. Moreover, in your code PhosphorusDynamicsMod (lines 334-439) you estimate the leaching only from sub-surface drainage flux. Do you have advection of soil P between soil layers? If not, why you do not estimate the leached fraction from each layer using the runoff/soil moisture (total water)?

Response: Our estimated leaching flux of P is lower than that of Wang et al., 2018, which could contribute to the underestimate of P uptake and overestimate of land carbon sink. We will add some statements in the discussion to discuss this uncertainty. Advection of soil mineral P and runoff was not included in this version of ELM, which could contribute to the lower P leaching. We are working on improving the representation of leaching fluxes through the ongoing efforts of coupling ELM with MOSART river biogeochemistry, which will be released in a future version of ELM.

Other comments:
Line 137: Prior to this paragraph, give a brief explanation on what are the P cycle interaction with C-N components (for instance P availability impact plant productivity (Vicca et al., 2012; Wang et al., 2010) or NPP (Aragão et al., 2009)).

Response: Thanks for the suggestion. We will add a brief explanation as suggested.

Line 206: I do not think this is correct. As explained in my comment on the model codes. Furthermore, it will be helpful to include the spinup results at the equilibrium in the supporting documents.

Response: As explained in response to the comments regarding model code, the part of code concerned was a special configuration (ECA) in ELMv1 and was not used in the default ELMv1 simulations as described in this study. The spinup results for all the P pools are included in the supplementary material (Fig. S1)

Line 219: Why by factor 10? Is there any reference for this value in accelerated spin-up for these pools? Did you test a range of factors to increase the turnover of this pool?

Response: The default mortality rate for the dead wood pools is 0.02/yr (a 50-year turnover time). We accelerate this mortality by a factor of 10, creating an effective turnover time of 5 years. The goal of accelerated spinup is to obtain steady state pool sizes quickly but as close as possible to those when the model is run with default turnover times. Koven et al. (2013) found that when accelerating decomposition pools in the soil too quickly, there were strong effects on the seasonal cycle that affected the steady state values - thus, there is a trade-off between faster acceleration and the disequilibrium between accelerated and non-accelerated steady states that requires a longer "final" spinup. The same effect occurs when accelerating vegetation mortality, and we found that accelerating to a 5-year turnover gave us a good balance between these factors.

Line 223: How did you deal with the Gelisol, Histosol, and Andisol which were not included in Yang et al 2013 but included in this study?

Response: For the grid cells that don't have values based on Yang et al.(2013), we applied the nearest neighbor extrapolation method to estimate the values.

Line 224: The rationale behind using a developed P map for initialization is not clear to me. I believe that the model should be able to reproduce the P-related dynamics from bare to aged soil without using the initialized map.

Response: Soil P transformations occur on geological time scales. It is not realistic to run a land surface model like ELM for millions of years. An approach that appropriately estimates P status for model initialization is more efficient than modeling P processes at these timescales to arrive at present day conditions. More details about the rationale of developing P maps for model initialization can be found in Yang et al. (2013). We will add a few statements in the text to emphasize this point.

Line 226: What is the period for this spin-up?

Response: As stated in the manuscript, we ran the normal spinup for 600 years.

Line 231: It is strange to see a very small variation in the labile pool. I understand that the occluded and parent material pools (due to very small rates used in the model) should not change much, but for the labile and adsorbed pools, it should not be the case. Is this because of the initialization of these pools using a global map?

Response: There are very small variations in the labile P because labile P is constantly interacting with other pools in the system. When the system reaches equilibrium, the inputs to and outputs from the labile P pool are balanced out so the pool size itself does not change much. Likewise, when the inputs to and outputs from adsorbed pools are balanced out during spinup, there are very little changes in the pool size itself.

Line 239: Name the environmental effects that you wanted to study, e.g. CO2/land use/climate impact or something else

Response: The environmental factors included in this study are CO2 forcing, land use and land cover change (LULCC), climate, and nitrogen deposition as summarized in Table 1. We will list the environmental factors in the revised manuscript.

Line 242: What was the rationale for bypassing the P limitation? Moreover, how did you prescribe enough P for each grid at each time step to exactly match the demand in the system?

Response: Here the idea is to run a simulation in which there is no P limitation on productivity and decomposition, in other words, the CN configuration. At each time step, we calculate the demand for P and the supply of P and supplement the difference between supply and demand.

Line 239-245: There is a repetition of the methodology here. If you have run with enough P that ignores the excess C as a result of P limitation, this is equivalent to the CN version run. My suggestion is to make these lines shorter and clearer.

Response: We will remove the repetition here as suggested.

Line 253: Firstly, this table can move to supporting document. Also, in most modeling papers in order to study different environmental factors' impact on the changes, there is one run with all the changes then the other factors attribute would be the run excluding it minus the run with all the changes. This way you keep the consistency between runs. What is the rationale behind your configuration with recycling all the other parameters except the study factor?

Response: We think it would be useful to have Table 1 in the main text. This table summarizes all the simulations in this study and the readers can reference back to the table when reading the results section.

Regarding the use of differencing to estimate single-factor effects in the model: There are two schools of thought on how to configure a multi-factor modeling experiment, and the Reviewer has described one while we prefer the other. The difference between them lies in how the interaction effects among the multiple factors are accounted for in the differencing of simulations. In our approach, we find the single factor effect by differencing a single factor experiment against a control with no factors varying. In that case there are no interaction effects among experimental factors mixed into the single-factor result. In the approach suggested by the Reviewer, where an all-factor experiment is differenced against an all-but-one-factor experiment, not only the single factor but also all of the two-way (or three-way, etc) interactions among factors are aliased into the single-factor result. Both are valid and useful as long as the differences between them are taken into consideration, but we prefer the simpler approach which avoids aliasing of the interaction effects, because it is easier to understand mechanisms related to differences. If needed, the total interaction effects are still quantifiable in the simpler approach as long as there is also a single all-factor run, as here.

2.3. ILAMB: These lines are too long and exhausting. You can summarize it in a few lines.
Line 277: which ones are CN/CNP models?

Response: We will shorten the description of ILAMB as suggested. ELMv1 CN/CNP models are not part of LS3MIP.

Line 284: The comparison against the steady-state model like GOLUM-CNP is not clear to me. If this is an intermodal comparison, firstly, it does not add any value to this study as the author stated as well as the uncertainty in equilibrium estimation by the GOLUM-CNP model (Wang et al., 2018). Secondly, if the intermodal comparison was the objective, why authors did not evaluate against a similar global process-based P-enabled model to the ELM-CNP such as ORCHIDEE (Sun et al., 2021)?

Response: For a nutrient-enabled model like ELM, it is important to also evaluate its performance on simulating nutrient pools and fluxes in addition to the evaluation of carbon pools and fluxes. The global land model benchmarking system ILAMB, however, does not include any observational dataset on nitrogen and phosphorus pools and fluxes. This is one of the major limitations of ILAMB and ongoing efforts are undertaken to address this limitation. GOLUM-CNP, unlike process-based models, provides estimates of present day nitrogen pools and fluxes by integrating observation-based estimates of C, N, and P pools and fluxes in terrestrial ecosystems into a diagnostic model framework. This observational-based dataset provided the next best thing for nutrient cycling evaluation. GOLUM-CNP has also been used in the evaluation of other land surface models (Sun et al., 2021).

Line 309: In IPSL-CM6A-LR, ORCHIDEE version 2 was used which did not include the P cycling. The comparison rationale is unclear.

Response: Results from several other land models from LS3MIP archive in CMIP6 are used to provide a context in terms of model performance. Although these models don't have an active P cycle, these model outputs are from LS3MIP offline simulations using the CMIP6 protocols, which is consistent with the simulations used in this study.

Line 316: Using your tool (https://compy-dtn.pnl.gov/yang954/_build/), as an example of selecting the tropic zones, your estimated RMSE score for C pools and fluxes using ELMCNP has not improved much compared to the ELM-CN. This needs further explanation.

Response: We agree that while ELM-CNP performs better overall in simulating carbon pools and fluxes, for some statistical metrics, ELM-CNP has not improved much compared to ELM-CN. It is challenging to show model improvements for all the metrics in ILAMB for a complex land surface model like ELM. It is important to note that ELMv1-CNP produces higher ILAMB scores for the integrated benchmarks such as global net ecosystem carbon balance and carbon dioxide concentration. These two integrated metrics are most critical to a land model in ESMs as they are most relevant to the coupling between land ecosystems and radiatively-forced climate change.

Line 334: Instead of an extra graph you could just report here the values for CN vs CNP.

Response: We prefer to keep the graph as it includes more information.

Line 345: Yet your error in estimated LAI is higher than other models in these regions (https://compy-dtn.pnl.gov/yang954/_build/)

Response: We agree there are models which have better LAI estimates in these regions. ELMv1-CNP simulated LAI in these regions are better than ELMv1-CN simulated LAI, which is the point we are trying to make here.

Line 360: As you state one of your biggest mismatches is in TRF, with overestimated P uptake (Figure S2) resulting in underestimated PUE (Figure 5). In the discussion, you state that this is mainly due to different plant stoichiometries between ELM-CNP and GOLUMCNP (line 600). Did you test the model using the same leaf/wood/root C:P ratios from GOLUM-CNP to show this?

Response: We hypothesized that the higher estimates of P uptake in ELMv1 compared to GOLUMCNP was mainly due to the lower wood C:P ratio used in ELMv1. We followed the reviewer's suggestion and ran an ELMv1 simulation using wood C:P ratio from GOLUM-CNP for tropical forests but found that ELMv1-CNP estimated P uptake in tropical forests is still higher than GOLUM-CNP. As pointed out by the reviewer in earlier comments, P leaching in ELMv1 might be underestimated and therefore P availability is over estimated, which could lead to higher P uptake in ELMv1. We will revise the discussions accordingly.

Line 441: How does ELM-CNP differ this much from (Yang et al., 2013), when you use its map for your initialization?

Response: We only initialized soil inorganic P pools using maps from Yang et al., 2013 (lines 222-223). Soil organic P pools are from ad-spinup and then allowed to interact dynamically with vegetation and the initialized inorganic pools until all the pools reach dynamic equilibrium state.

Line 448-453: I suggest rewriting this part and instead of using "In many parts of the world", you report the relative N/P uptake in major biome classes.

Response: Thanks for the suggestion. We will revise the manuscript to report the relative NvsP limitation for major biomes.

3.4. The effects of P limitation on the historical carbon cycle: Again, this whole paragraph is obscure. I suggest reporting the changes of P and C fluxes per major biomes, then the pools, and then. Reporting separately on environmental factors that impact these changes.
Additional note for figures: In some of the figures, units are missing. Please consider adding either on the plots or in the figure captions.

Response: One major component of this manuscript is to quantify the extent of P limitation on carbon pools and fluxes during historical time periods on the global scale. Therefore, in Fig 11, we show the ELMv1 estimated NPP, vegetation, biomass paired with the relative difference (in percentage) between CNP and CN. We will make sure the figure captions are detailed and easy to follow and units are provided for all the figures. Fig. 10 provides the estimates of cumulative global carbon fluxes due to each environmental factor and how those estimates are affected by including P cycling and therefore directly addresses one of the main questions in this manuscript. We feel like both figures are important to show

and decide to keep both. We will switch the orders of Fig. 10 and Fig. 11 to make it easier to follow. We will label units more clearly for each figure.

Reference:

Aragão, L. E. O. C., Malhi, Y., Metcalfe, D. B., Silva-Espejo, J. E., Jiménez, E., Navarrete, D., Almeida, S., Costa, A. C. L., Salinas, N., Phillips, O. L., Anderson, L. O. ., Baker, T. R., Goncalvez, P. H., Huamán-Ovalle, J., Mamani-Solórzano, M., Meir, P., Monteagudo, A., Peñuela, M. C., Prieto, A., Quesada, C. A., Rozas-Dávila, A., Rudas, A., Silva Junior, J. A., and Vásquez, R.: Above- and below-ground net primary productivity across ten Amazonian forests on contrasting soils, Biogeosciences Discuss., 6, 2441–2488, https://doi.org/10.5194/bgd-6-2441-2009, 2009.

Vicca, S., Luyssaert, S., Peñuelas, J., Campioli, M., Chapin, F. S., Ciais, P., Heinemeyer, A., Högberg, P., Kutsch, W. L., Law, B. E., Malhi, Y., Papale, D., Piao, S. L., Reichstein, M., Schulze, E. D., and Janssens, I. A.: Fertile forests produce biomass more efficiently, Ecol. Lett., 15, 520–526, https://doi.org/10.1111/j.1461-0248.2012.01775.x, 2012.

Wang, Y. P., Law, R. M., and Pak, B.: A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere, 7, 2261–2282, https://doi.org/10.5194/bg-7-2261-2010, 2010.

Yang, X., Post, W. M., Thornton, P. E., and Jain, A.: The distribution of soil phosphorus for global biogeochemical modeling, 10, 2525–2537, https://doi.org/10.5194/bg-10-2525-2013, 2013.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, Biogeosciences, 10, 7109–7131, https://doi.org/10.5194/bg-10-7109-2013, 2013.

Schuur, E., McGuire, A., Schädel, C. *et al.* Climate change and the permafrost carbon feedback. *Nature***520**, 171–179 (2015). https://doi.org/10.1038/nature14338

Sun, Y., Goll, D. S., Chang, J., Ciais, P., Guenet, B., Helfenstein, J., Huang, Y., Lauerwald, R., Maignan, F., Naipal, V., Wang, Y., Yang, H., and Zhang, H.: Global evaluation of the nutrient-enabled version of the land surface model ORCHIDEE-CNP v1.2 (r5986), Geosci. Model Dev., 14, 1987–2010, https://doi.org/10.5194/gmd-14-1987-2021, 2021