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.

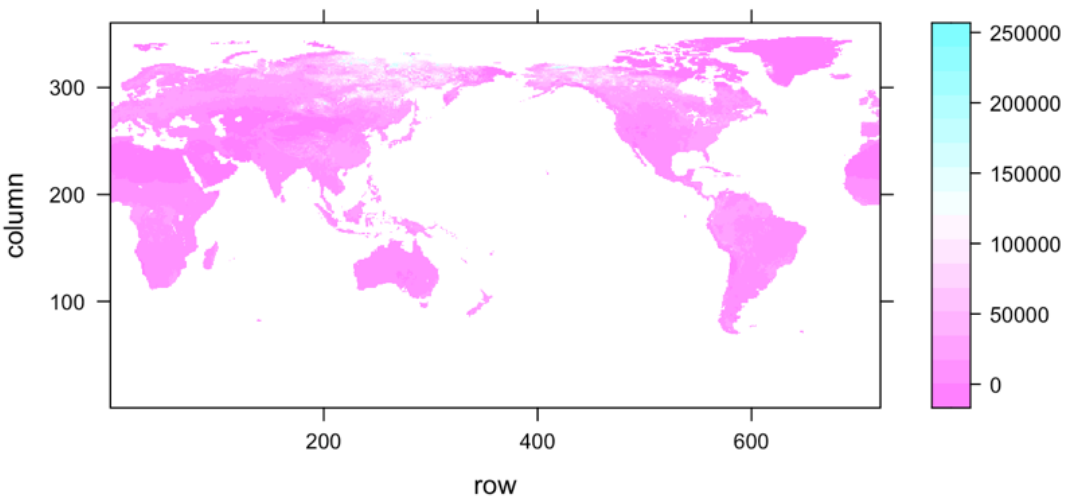
Model codes:

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?

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 For instance, in your ErosionMod the eroded phosphorus (pp2poc) is estimated using a fixed value from outdated reference (Meybeck (1982)) across all the soil types and biomes.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?

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?



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)?

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)).

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.

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?

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?

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.

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

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?

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

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?

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.

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?

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?

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)?

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

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 ELM-CNP has not improved much compared to the ELM-CN. This needs further explanation.

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

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

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 GOLUM-CNP (line 600). Did you test the model using the same leaf/wood/root C:P ratios from GOLUM-CNP to show this?

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

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.

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.

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.