Comment on bg-2022-114
Hongxing He (Referee)

In this study, Lin et al presented a double surface Langmuir adsorption isotherm in the QUINCY model and compared it to the traditional/simplified (single surface) Langmuir isotherm, that is mostly used in current TBMs. After model calibration, the authors argue the double isotherm shows a better representation of the inorganic P cycling. The improved P (ad)sorption model also suggests the current assumed P limitation in temperature forests was likely overestimated. Overall, the paper is well written, and the results are also sound. Given our current understanding of P limitation is still very limited, thus before the P models can be applied to make sound predictions, the model structure needs to be evaluated and discussed. Study like this paper thus contributes to improving the process description of P exchange in forest ecosystems and advances in C-P coupling in TBMs. I would thus recommend publishing this work at Biogeosciences. However, I have the following comments for the authors to consider during their revisions.

First, the authors did a literature survey to highlight uncertainties of the current parameterization of Langmuir isotherm in some TBMs. The parameter error in some previous TBMs was also noticed by myself when I develop a recent ecosystem CNP model. Thus I think it is important to highlight this for the community and a very good motivation for the current study. However, one would ask if this is just a parameterization issue or if it is a model structure issue (as the authors argued here)? Empirical data that fit different isotherm functions, including traditional Langmuir isotherm, generally show various but reasonably well-fit results (e.g. Brenner et al 2019, Lin et al. 2020, and much more). Thus, I would like to discuss this with the authors. First, which part of the improved model fits the measured data that could be attributed to the model structure, which part is from improved parameters? I have concerns about how the model comparison is made and how much conclusions can be drawn from such a comparison? In Line 138, the authors state separate calibrations were made for each site and each depth, what’s the influence of those separate calibrations for the comparison?
Second, as the authors argue the advantage of using double surface Langmuir, i.e. its higher buffer capacity. Then I would suggest a better separation of the influence of P release from other releases and uptakes? i.e. the feedback is of need. From the results, the main improvement is the ratio between Plab and exchangeable Pi (section 3.2). The P uptake across models seems rather similar (Fig. S2), i.e. for the P limited site LUE, the uptake PO4 for siLang, dbLang and 4pool model (Fig. 2Sf). The siLang shows a higher uptake in autumn but at an annual scale, the overall rates seem rather similar. The different approaches show a rather large influence on the C partitioning, (LAI, aboveground C, Fig. S2 bc). This is rather strange, what causes such large feedback on the aboveground plant properties, given the total P uptake seems rather similar? I also do not find evidence to support the statements of Line 203, i.e. the differing plant P uptake.

Third, the model performance of foliar P, Fig.4b shows a convergence of different models when P availability becomes smaller. In other words, in more P-limited conditions, the difference between the models becomes smaller, although all of them largely underestimated the measured P concentration. How come such large differences in the P-rich sites? Is this due to the calibration being mainly focused on the soil and thus less on the vegetation?

Some more specifics to consider:

Introduction

Line 36, missing references after “boreal forests are generally considered N limited”

Line 49-50, the argument is that organic P recycling is the major flux, while the geochemical P flux is small.

Line 50-55. In literature, several isotherms, or model functions, including double Langmuir, have been suggested to describe the phosphorus adsorption-desorption processes (i.e., McGechan and Lewis, 2002). I would also suggest not to use “a novel model concept, Line 54” as the authors propose in its current form. It has been in the literature for some time. I think the novelty is the implementation of the TBM models and evaluation of the implications? Besides, I am also lacking the field and experimental evidence to support the additional supplement of P from the adsorbed P pools. So, what is the, i.e. P isotopic data suggest, and do they support your hypothesis here? What are the mechanisms behind that? I would suggest adding those to the motivate current model development work.
Methods

Line 70, equ 1, the Langmuir isotherm, do the interaction with water considered? As the concentration also dependent on the water content at each time step?

Line 116 do you have leaf P/N concentration data over years? Or just sampled for one year?

Line 138 what is calibrated and what criteria were used for the calibration? Be specific here.

Results

Line 203 given the total P uptake by different approaches?

Line 246 the pool sizes differs also after the simulation, SOM top soil, the fluxes and the pools sizes. As also your sensitivity results show the importance of SOM pools for dbLang, Line 219, which indicates the potential feedback due to the biological mineralization. Also in your Line 243 on the plant and soil changes

So the different approaches show impacts on the fluxes and pool sizes. Why not show a complete P budget for each site with different fluxes simulated by various approaches? Also show the different pool sizes before the simulation and after the simulation, i.e. the pool size changes. This will give an overall picture of the ecosystems.

Some references mentioned:


Hongxing He

McGill, Montreal.