Response Letter

Title: Soil nutrient competitive traits of plants, microbes, and mineral surfaces explain nutrient acquisition in tropical experimental manipulations

General Response:

We would like to thank the three anonymous referees for their constructive comments. In this revision, we added the Gelman-Rubin criterion to quantitatively describe parameter convergence. Since some parameter posteriors do not follow a Gaussian distribution, we added a quantile-based uncertainty reduction criterion, which does not require any assumption of the parameter distribution. As suggested by reviewers, we also added more discussion about model uncertainty, model parameter nonconvergence, and impacts of soil heterogeneity on model predictions.

The response letter is organized by (1) reviewers' major comments in blue; (2) authors' response in black. The corresponding changes in the manuscript are in red.

Reviewer 1

I would like to congratulate the authors for a well-written model description paper on a very interesting topic. The paper describes the model very satisfactory and it is well calibrated. The only complaint about the paper is that I don't feel the model is tested to the extent that is required to see if the model 1. makes a difference, 2. can capture the critical plant-soil nutrient competition correctly to enhance the capability to model plant productivity properly.

Response:

We appreciate the reviewer's comments. The ultimate goal of this study is to capture critical plant-soil nutrient competition correctly in the context of large-scale prognostic model (like CLM). However, the version of the model we are discussing in this paper is solely a diagnostic module of soil biogeochemistry. Therefore, the model lacks the capability to show the improvement in plant productivity.

We acknowledge that it is important to extend the work presented here to explore impacts to plant productivity under different nutrient competition treatments. As mentioned in the Discussion, we are focusing ongoing work on this issue. We are incorporating the N-COM model into CLM and ALM to investigate the potential improvements in terms of e.g., gross primary production.

Reviewer 2

1. In lines 350-351 of the revised manuscript, you state you could "ensure thorough convergence" of your three MCMCs. To prove convergence you are referring to the trace plots in figure A1, but they rather prove that a lot of your parameters have not converged and the chains are generally not well mixed. In my first review I also recommended to use the Gelman convergence criterion. With your additional two chains you can now calculate the Gelman convergence criterion for every parameter.

Response:

We appreciated the suggestion of using Gelman-Rubin criterion for convergence checking. Based on three MCMC chains, we calculated Gelman-Rubin convergence criterion and added a new figure (Figure A2) to quantitatively illustrate the convergence for each parameter. We found that 11 (out of 20) parameters were well converged (Gelman-Rubin criterion <= 1.1).

One interesting point is that small posterior uncertainty (or large uncertainty reduction) does not necessarily mean a small Gelmen-Rubin criterion. This is because the Gelman-Rubin criterion focuses on multiple chains and compares within-chain variance (W) with between-chain variance (B). When the value of B is comparable to or smaller than W, multiple chains are thought to converge, while posterior uncertainty (or uncertainty reduction) depends on one chain only.

2. Contrary to the statement in the main manuscript, you acknowledged in the reply to the reviewers that some parameters do not show convergence: "We noticed that the some of the parameters were not well constrained, which was reflected in their relatively large posterior uncertainty and small uncertainty reduction. We argued that the nonconvergence resulted from data paucity rather than a short MCMC chain." Here, you are confusing convergence of MCMCs with how well a parameter is constrained by the data. Paucity of data should not hamper convergence. If a parameter is not well constrained by data, the posterior should follow the prior. However, the MCMC algorithm should be efficient enough to be able to sample parameter as prescribed by the prior.

Response:

We agree that our original statements on convergence of MCMCs are confusing, because it is difficult to know whether nonconvergence resulted from data paucity. However, we argue that data paucity could be one of the reasons why the MCMC chains did not converge. We added more discussion of this topic in the revised manuscript:

"Convergence of model parameters is reported with the Gelman-Rubin criterion (Figure A2). Using this criterion, eleven (out of twenty) parameters are found to converge (Gelman-Rubin <= 1.1). One reason for the lack of convergence of the remaining parameters is likely data paucity and resulting equifinality. In particular, starting from different prior values, MCMC calibrations may result in different converged posteriors but predict similar dynamics. In this regard, high frequency measurements may improve model calibration [*Tang and Zhuang*, 2008]."

3. As a way forward, I could recommend to either try adaptive MCMC algorithms or to just use the best parameter set obtained with the current setting. For the second option, you would acknowledge that you did not manage to retrieve a proper posterior sample. Rather you would state that the calibration was explorative to study competiveness effects between plants and microbes and how the model fares for a specific site.

Response:

Thanks for the recommendation. We believe the second option better fits the objective of this study. We focus here on developing and testing a new competition theory and model development, rather than developing model calibrations for robust prediction. The main idea is to test whether or not the new competition hypothesis allows credible prediction of nutrient competition at tropical forest sites.

Even though some parameters are not well constrained, we are still able to investigate scientific questions using the new model. We acknowledge in the manuscript that the calibration is not perfect. Clearly, more work is needed for parameters tuning, uncertainty analysis, and reduction.

To describe this limitation forthrightly, we have added the following text to the Results and discussion section: "We acknowledge that, for large-scale model application, more work on parameter tuning and uncertainty analysis is needed. However, even with

these caveats, the model predictability is reasonably good when applied to the tropical forest fertilization experiments described in Section 3.4."

4. How is this equation solved?

$$\frac{d[POx]}{dt} = F_{others} + \frac{VMAX_P^{surf} \cdot KM_P^{surf}}{(KM_P^{surf} + [PO_x])^2} \cdot \frac{d[POx]}{dt}$$

Response:

This equation is solved using the Euler method:

First move $\frac{d[POx]}{dt}$ term on the right hand side, to the left, we have:

$$\frac{d[POx]}{dt}(1 - \frac{VMAX_P^{surf} \cdot KM_P^{surf}}{(KM_P^{surf} + [PO_x])^2}) = F_{others}$$

Then divide the two sides by $(1 - \frac{VMAX_P^{surf} \cdot KM_P^{surf}}{(KM_P^{surf} + [PO_x])^2})$, we have:

$$\frac{d[POx]}{dt} = \frac{F_{others}}{(1 - \frac{VMAX_P^{surf} \cdot KM_P^{surf}}{(KM_P^{surf} + [PO_x])^2})}$$

5. In line 286-287, you state that "Mineral surface "effective enzyme" abundance (E_P^{surf}) is approximated by $VMAX_P^{surf}$ – [SP]." I would appreciate if you could motivate this equation with one or two sentences.

Response:

Mineral surface adsorption is treated as a POx molecule captured at an unoccupied site on the mineral surface. The capability of mineral surface adsorption is proportional to the total number of unoccupied sites. The sum of occupied sites and unoccupied sites is the maximal adsorption capacity ($VMAX_p^{surf}$), and the number of occupied sites is equal to adsorbed P ([SP]). Therefore, the number of unoccupied sites is ($VMAX_p^{surf}$ – [SP]).

Reviewer 3

1. From the sensitivity analysis (Fig. 3), I infer that consistently across all temperature/moisture scenarios, the predictions are sensitive to the same parameters. How can the conclusion be drawn from this, that uptake is more regulated by internal kinetics instead of temperature or moisture? To my understanding, one cannot draw this conclusion.

Response:

Sorry for the confusion, we have modified our statement about the sensitivity analysis to be: "The sensitivity analysis indicates that the model is highly sensitive to kinetics parameters (*e.g.*, KM). Furthermore, the model is consistently sensitive to the same parameters across all temperature and moisture conditions."

2. Contrary to the statements in the paper (L. 351), Fig. A1 clearly shows that the chains starting from different position did not converge to the (same) limiting distribution. They seem to sample different (local?) minima (e.g. plot of Km^plant_NO3). Hence all conclusions on the sampled parameters are all very vague.

Response:

Thanks for identifying this issue. As described above in response to a similar criticism from Reviewer 2, we employed the Gelman-Rubin criterion (Figure A2). We found that 9 parameters did not converge well, of which KM^{plant}_{NO3} had the poorest convergence. This result demonstrates that different starting points (priors) result in different posteriors. We believe that this Equafinlity issue [*Tang and Zhuang*, 2008] is difficult to avoid, particularly when calibration data are limited. In the revised manuscript, we added a paragraph to discuss this issue. We have added to the Results and discussion section to address this issue. Please also see our responses to reviewer 2, questions 2 and 3.

3. Moreover, in Fig. 4, the posterior sample is at the edge of the prior distribution. This indicates an inconsistency between the model- data- and prior. E.g. the model and data suggest a much higher plant uptake rate of P (k^plant_P) than previous knowledge. This means, either the previous knowledge or the model cannot be reconciled with the data.

Response:

We agree that k^plant_P posterior sample is at the edge of the prior distribution. However, we argue that this result does not necessarily mean the model cannot be reconciled with the data. The solution phosphorus seasonal dynamics (Figure 4) indicated that there was a strong sink for soil phosphate from June to December. Plant P uptake could be a large sink. In this sense, the posterior model (high P uptake rate, k^plant_P) is consistent with the data.

4. Contrary to the text (L. 438), Fig. A1 suggest that also plant uptake of NO3 (K^plant_NO3) seems important, as the data significantly changes the prior.

Response:

We agree that in Figure A1 $k^{plant}{}_{NO3}$ is significantly different from its prior. It seems important. However, we did not have much confidence in this result. First, the posterior uncertainty of $k^{plant}{}_{NO3}$ is large (Figure A3). In addition, uncertainty reduction associated with this parameter is small (Table 4). In other words, the difference between prior and posterior parameters may result from random effects of MCMC sampling, but not inferred by the calibration data.

5. The calculation of gross nutrient mineralization and immobilization fluxes based on current stoichiometry of soil organic matter (SOM) pools becomes clearer in the revised version. Potential fluxes are calculated so that current C/N/P ratio is not changing. However, the adjustments due to competition with multiple consumers alter those potential fluxes and the stoichiometry over time. The rational and consequences of these choices need to be elaborated.

Response:

The SOM C/N/P ratio is fixed when we calculate the potential fluxes. We also adjust the potential flux due to competition to ensure fixed stoichiometry over time by imposing a limitation factor on potential decomposition fluxes:

Actual_C_flux = potential_C_flux * min(Nlim, Plim)

Actual_N_flux = potential_N_flux * min(Nlim, Plim)

Actual_P_flux = potential_P_flux * min(Nlim, Plim)

Nlim and Plim are soil decomposition nutrient limitation factors due to multiple consumer competition. We added several sentences in the Appendix to elaborate this point.

6. The paper now includes some information on the cost-function used in the calibration. However, the handling of different data streams is not sufficiently described by the word "including time series" L.343. I guess the different time series were just concatenated to one big series and enter the cost function without weights. To evaluate the results it is important is to state the number of records per data stream in Table 4 and the average variance of an observation (here corresponding to the magnitude of the observations). From Fig. 4 I get that there are more observations in the gas-data (CO2 and N2O) streams. Hence I expect the calibration to choose parameters so that these streams are better matched than the streams with sparse observations.

Reponses:

Thanks for the suggestion. In table 3, we added the number of records per data stream but we did not include the average variances of each observation, because they are not reported in their original papers. We agree that parameters associated with processes that are observed more frequently may be better constrained.

7. Eq. (30) compares a parameter of log-normal distribution in the prior, i.e. the standard deviation of a log-transformed value, to a parameter of the normal distribution in the posterior. I cannot interpret this. Fig. 2 is more helpful. For a similar quantitative uncertainty reduction measure, I suggest to base it on inter-quantile ranges, e.g. the 10-90% range instead of sigma. For the prior it can be calculated from the distribution, for the posterior estimated from the sample. This is simpler, straightforward and easier to interpret.

Response:

Thanks for the suggestion. We used quintile rage 25%-75% to calculate the uncertainty reduction.

$$UR_Q = (1 - \frac{Q^{75}_{posterior} - Q^{25}_{posterior}}{Q^{75}_{prior} - Q^{25}_{prior}}) \cdot 100\%$$

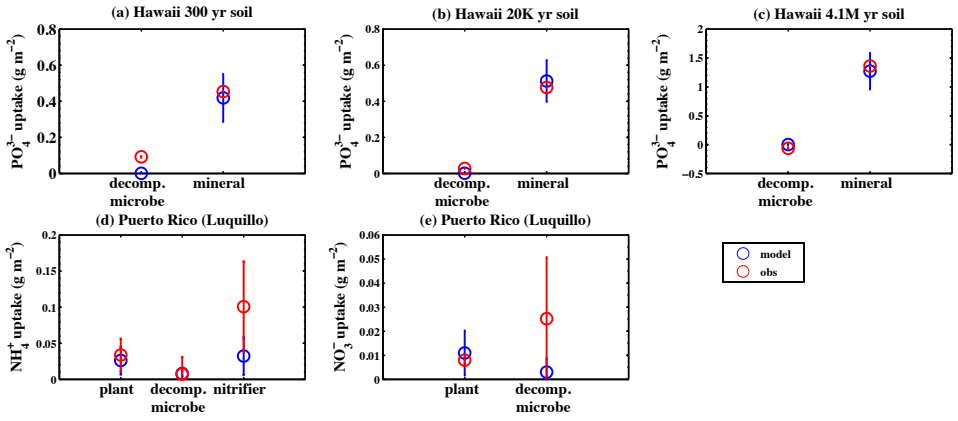
The corresponding changes are added in Table 4.

Specific comments

L421: "prognostic prediction could be uncertain": Better do it instead of guessing. I.e sampling from the parameter space and generate several predictions. Then plot the uncertainty of the model predictions.

Response:

We agree with the reviewer on the importance of prognostic uncertainty, which is an important component of model prediction. To address this comment, we randomly sampled 100 parameter sets from their posteriors, and calculated the uncertainty of our model prediction. Error bars are added in Figure 5 to illustrate the prediction uncertainties.



L439: The new paragraph seems quite unmotivated. What are "those fractions"

Response:

The new paragraph introduces the calculation of ECA, which characterizes the nutrient limitation of plant potential nutrient uptake rate. We agree that "those fractions" is ambiguous. The sentence is rewritten to be "We re-organize the right hand sides of Eqns. 14 – 22 to be the product of potential nutrient uptake rate and an ECA limitation term"

L450: To me the analogue of root spatial occupation is not intuitive. If diffusivity limits a process a high concentration of Enzymes at one site e.g. hotspots like rhizosphere or litter layer will not increase uptake when the substrate access is limited by diffusion.

Response:

From a microsite perspective, nutrients diffuse from surrounding soil towards roots at the microsite. In this case, you are right that high concentration of enzymes will not increase uptake, because the nutrient diffusion rate limits the nutrient concentration at the root surface. However, from the whole soil perspective, soil is highly heterogeneous. If the nutrient (e.g., PO_4^{3-}) diffusion rate is slow, plants may grow more roots to explore a larger volume of soil. In this case, root occupancy becomes an important regulator of nutrient uptake. To address this reviewer comment, we have clarified in the revised manuscript that the discussion is focused at the whole-soil scale.

L461 Thanks for now acknowledging soil heterogeneity. I still would expect some discussion on the consequences of ignoring this heterogeneity in the model. The ratios of enzymes between competing groups will be very different in different in rhizosphere and bulk soil. Will this lead to an underestimation of overall plant competitiveness in the model? What other predictions will be strongly affected?

Reponses:

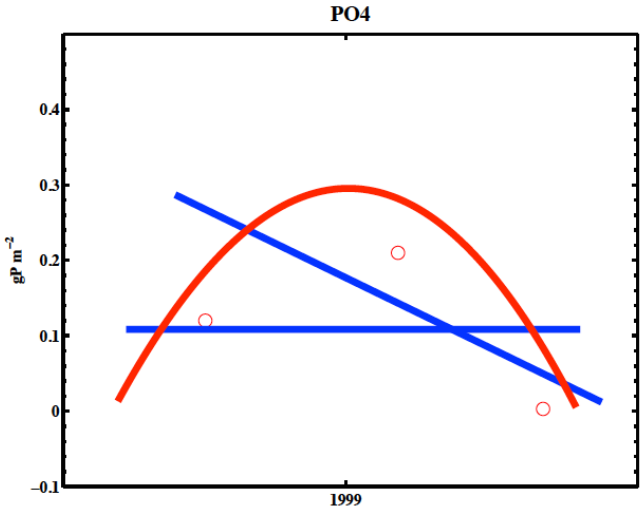
Thanks for the suggestion. We have added more discussion about the potential impacts of ignoring soil heterogeneity in large-scale models: "Although beyond the scope of the current study, the consequences of ignoring the rhizosphere versus bulk soil heterogeneity warrants further investigation. Large-scale models aim to quantify ecosystem level dynamics, although they are usually driven by parameters inferred from *in situ* field observations. In the absence of a model that explicitly represents this spatial heterogeneity, it is difficult to quantify the impacts of using inferred rhizosphere decomposer affinities on model predictions of the whole soil [*Schimel et al.*, 1989].

L493: The statement of more data leading to better constrained posteriors need to be backed up. I still think it will not have a big effect (by looking at Fig4, where the model error is larger than the observation error). By generating more artificial observations (but

as far away from the model results as in Fig 4) and repeating the model calibration, one could check if posterior is narrower.

Response:

Thanks for bringing this issue up. To clarify, we meant to imply that more data across the whole year, which would better represent seasonal dynamics, would lead to better-constrained posteriors. For example, only three measurements of soil free phosphate were made during 1999. Many detailed dynamics are therefore missing. It is difficult to tell whether the real POx dynamics were linear (blue) or polynomial (red) (see following figure). In this case, more POx observations in January, February, August, and September would be extremely helpful to constrain POx associated parameters. However, in this case, artificial observations would not be helpful, since the real dynamics are unknown.



L496: Thanks for the additional discussion. This demonstrates how we can gain insight from the modeling exercise.

Response:

Thanks for the positive comment.

L516: More frequent measurement of the already richer streams will probably be of little help (See my comment on more data (L493) and on multiple data streams. I suggest additional measurements of the sparse streams.

Response:

Sorry for the inappropriate statement. We have replaced “higher frequency CO2/N2O sampling would significantly benefit the model uncertainty reduction” with “higher frequency sampling of sparse measurements (*e.g.*, free phosphate) would significantly benefit the model uncertainty reduction”

Fig. 2: Thanks for the figure. This greatly helps to understand the calibration.

Response:

Thanks for the positive comment.

Fig. 5: Uncertainty of model prediction is missing. See comment L421.

Response:

We have added the model prediction uncertainty in Figure 5.

Fig A2. Are multimodal distributions close to Gaussian?

Response:

No. Besides fitting the posterior parameters to a Gaussian distribution, we also use Qantile (25% ~ 75%) to calculate the uncertainty reduction (see table 4 in manuscript). The latter approach does not require any assumption of the parameter distribution.

Appendix B2: Missing description of VMAX. I Suppose it is k + [E_tot] typo: simplify after B1

Response:

Description of VMAX is given in (B8).

When stating the objectives p4063, (2) seem to be a means of achieving (1), rather than an objective.

Response:

Thanks for pointing this out. We agree that the objective (2) is confusing and have therefore removed it in the revised manuscript.

P4066L18, the “respectively” is ambiguous.

Response:

Sorry for the confusion, we have removed “respectively”.

References

Schimel, J. P., Jackson, L. E. & Firestone, M. K. (1989). Spatial and temporal effects on plant-microbial competition for inorganic nitrogen in a California annual grassland. Soil Biology and Biochemistry 21, 1059-1066.

Tang, J., & Zhuang, Q. (2008). Equifinality in parameterization of process□based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics. Journal of Geophysical Research: Biogeosciences (2005–2012), 113(G4).