Dear Reviewer,

Thank you for the constructive comments. We copied your comments below in blue font; our labeled responses are in ordinary black font. We have labeled our responses for ease in cross-referencing. Responses to your comments are labelled “RC1” and responses to a community reviewer are labeled “CC1”. We have not yet received responses from a second reviewer. We welcome further discussion during the discussion period.

General comments

This study integrated a representation of phosphorus-dependent relative allocation to root tissues into the ED2 model. The model was simulated at a site in a tropical dry forest in Costa Rica for +N, +P, and +NP fertilization treatments. Modelled results were compared to empirical observations. The model was then simulated over 30 years to examine the influence of the new process representation over a longer time scale. The comparison between a model and empirical observations of experimental manipulations of nutrient input is very useful for model development. However, it was unclear whether the process that was represented in this study (increasing allocation to fine roots with increasing soil P) is prevalent in systems outside of this site, what its underlying mechanisms are, and how it relates to other central processes (such as the relationships between allocation to fine roots and soil nitrogen or water). Furthermore, the statistics used to establish the results were unclear.

(Response RC1-1): Thank you for your review. We are happy that you found the comparison helpful and are very grateful for your constructive feedback. In our responses below, we have addressed the points of clarification that you raised. We also carried out a new statistical analysis, largely following your advice. We think that incorporation of these changes will result in a greatly improved manuscript.

Specific comments

1. The premise of this study needs to be better established. What is the mechanism underlying increasing allocation to fine roots with increasing soil P?

(Response RC1-2): This study investigated the consequences of three related premises: (a) plants increase fine root production with increasing soil P, (b) plants decrease fine root production with increasing soil P, and (c) plant fine root production is independent of soil P. Regarding (a), several studies have shown that roots proliferate in nutrient-rich patches (Pregitzer et al. 1993; Robinson 1994; Zhang and Forde 1998; Robinson et al. 1999; Fransen et al. 1999; Hodge et al. 1999; Jing et al. 2010; Li et al. 2012). This proliferation may be related to enhanced root morphological plasticity with increasing soil P, which increases nutrient uptake per unit construction cost (Fitter 1994; Eissenstat and Yanai 2002; Zhang et al. 2016). It may also be physiologically adaptive if it allows for faster uptake of nutrients from the soil (Jackson et al. 1990; Hodge et al. 2004). We discussed premise (b) in the original manuscript, lines 58-65. Regarding premise (c), it could be that fine root production depends more on water or N than on P. In this case, relative allocation to fine roots would not be directly sensitive to soil P. See our response RC1-5 below for more on this point. If the editor allows a revision, we would be happy to provide further discussion of these premises.
Does this response occur in other ecosystems or are the only observations from the Costa Rica site? The introduction gave several examples of how different ecosystems and different individuals within a given ecosystem respond differently to N and P fertilization. Are there any patterns that emerge across ecosystems? If this response is specific to a single or a small number of sites, why should it be represented in TBMs?

(Response RC1-3): Our observations are only from the Costa Rica site. Cunha et al. (2022) carried out a fertilization experiment on old-growth Amazonian sites and also found that relative and absolute allocation to fine roots increased under P fertilization. In another Amazonian fertilization experiment, Lugli et al. (2021) reported that absolute fine root production increased, but did not comment on relative allocation to fine roots. Given these studies, we think that this response is relevant to TBMs because it has been seen in a large proportion of tropical studies, even if the absolute number of studies with leaf/wood/root productivity measured separately is relatively small. In addition, the model simulations that we carried out can be useful for inspiring future experimental design and field work.

Previous global meta-analysis (Yuan and Chen 2012) showed that fine root production increased with P fertilization, but relative allocation to fine roots decreased. The (global) decrease in relative fine root production contrasts with the tropical forest studies of Waring et al. (2019) and Cunha et al. (2022). However, Yuan and Chen (2012) did note substantial variation across ecosystems. One potential confounding problem in making comparisons is that different experiments used different fertilizer amounts, affecting the fertilization response (Hou et al. 2020). Thus, we are reluctant to speculate too much on global patterns.

Have empirical studies indicated that this is important for larger C fluxes? This is somewhat touched on in the Discussion but its prevalence was not clear.

(Response RC1-4): We are not aware of its relevance for larger C fluxes on the basis of previous empirical studies. However, our expectation is that its relevance should be of interest (Line 69-73). Related to this point, we argue that some of our figures should stimulate hypotheses and prompt more field work in the future.

2. How do other factors interact to determine relative allocation to roots? Water and nitrogen should play important roles as well. Is it valid to only focus on P (especially given that the results suggest that it is increased water uptake that seemed to drive AGB)?

(Response RC1-5): We focused the relationship between soil P and relative allocation to fine roots on the basis of a previous fertilization experiment (Waring et al. 2019). That experiment showed that relative allocation to fine roots was impacted by P fertilization but not N fertilization. Furthermore, the experiment spanned years of relatively high and low rainfall, and thus allowed us to see how the actual fine root-to-leaf productivity ratio varied with rainfall. Specifically, rainfall in 2015 was 628 mm, rainfall in 2016 was 1754 mm, and rainfall in 2017 was 2050 mm. The observed leaf:fine root productivity ratio did not exhibit a clear relationship with rainfall. Averaged across treatments, this productivity ratio was 0.37 in 2015, 0.44 in 2016, and 0.24 in 2017. We also looked at the leaf:fine root productivity ratio by treatment. In all cases, this productivity ratio was intermediate in 2015, largest in 2016, and smallest in 2017.
Because the observations lacked a clear correlation with rainfall, we decided not to focus on the impacts of water on the leaf:fine root productivity ratio.

We acknowledge that our original manuscript did draw a correlation between simulated transpiration and simulated aboveground biomass (AGB) (Figs. 7a, 8). However, prompted by this comment as well as a comment from a community reviewer (response CC1-11), we carried out more detailed analysis of this point. Our current thinking is that the correlation between transpiration and AGB is causal. Instead, the higher transpiration in the “pos” schemes occurred because the “pos” schemes had larger leaf area. Transpiration per unit leaf area was similar in all schemes.

We therefore decided to look more deeply into why the “pos” schemes had relatively large AGB in Fig. 7a. Following a community comment (response CC1-9), we carried out additional simulations spanning a wider range of values for our allocation parameters $a$ and $b$ (these parameters are defined in Eq. 1 of the original manuscript). For each choice of $a$ and $b$, AGB at the end of 30 years is shown here (we only show the average of the plots without P fertilization, thus corresponding with the manuscript’s Fig. 7a):

This fuller analysis shows that AGB does not simply increase with $b$. To explain this pattern, we hypothesized that the highest AGB would occur when the relative acquisition of C:P was optimal. We therefore computed the ratio of nonstructural C to nonstructural P and plotted it against AGB:
The highest nonstructural C:P ratios corresponded most closely to leaf C:P (which was fixed at 600 g C (g P)^{-1}) and to the highest simulated AGB values. This result confirms the hypothesis.

Based on this result, it could be argued that our allocation rule for relative allocation to fine roots should be targeted to the nonstructural C:P ratio rather than soil P. We indeed think that such a scheme would work well in the unfertilized plots; however, it would not capture the P fertilization effect (increased relative fine root production with increased soil P).

Additionally, I would assume that the role of other plant mechanisms to increase P uptake would be important as well, such as phosphatase synthesis and arbuscular mycorrhizae. These are likely intricately linked to fine root biomass in real ecosystems. While these do not necessarily need to be examined or modelled, they should be at least recognized in the experimental setup and discussion.

(Response RC1-6) We agree that other plant mechanisms to increase P uptake could be important as well. In a revision, we would be happy to discuss additional plant mechanisms to increase P uptake, including phosphatase synthesis (Liu et al. 2015; Kong et al. 2016; Lugli et al. 2020) and arbuscular mycorrhizae (Hodge 2004; Comas et al. 2014; Eisenstat et al. 2015; Liu et al. 2015; Kong et al. 2016; Ma et al. 2018). For example, root phosphatases hydrolyze organic P-containing compounds and releasing inorganic P that is absorbable by roots; mycorrhizal associations are even more effective by enlarging the root absorbing surface per unit cost; both mechanisms provide additional P sources. Plants adaptively adjust their traits or metabolic processes in terms of effective P acquisition (Raven et al. 2018; Han et al. 2021; Aoyagi et al. 2022), and diverse P acquisition strategies are being evaluated from observations (Reichert et al. 2022).

Additionally, flexible stoichiometry could be important. How have other models approached these phenomena?
Current models approach stoichiometry differently. Some have fixed stoichiometries (e.g. JSBACH, Goll et al., 2012; CLM-CNP, Yang et al., 2014; JULES-CNP, Nakhavali et al., 2022). Some account for stoichiometric flexibility by prescribing ranges for each pool based on empirical studies (e.g. CASACNP, Wang et al., 2010; ORCHIDEE-CNP, Goll et al., 2017; QUINCY, Thum et al., 2019). ED2 has fixed stoichiometries in structural pools but non-fixed stoichiometries in non-structural pools (Line 131 in the original manuscript). Models adjusting relative allocation of new growth to fine roots mostly apply an idea that new growth is scaled by minimum of N and P stress scaling factor, rendering increased fine root production when P demand exceeds supply (for example, when the ratio of nonstructural C:P greatly exceeds the ratio present in plant tissues). By contrast, we parameterized the model only from the perspective of supply (soil P). We would be happy to provide more discussion on this point in a revision.

3. It was not made clear which PFT was being studied in these experiments. Were there multiple PFTs? Given that this is a dry tropical forest, do deciduousness and phenology play a role here? How could these results differ between tropical dry forests and tropical moist forests? Have similar experiments been conducted in tropical moist forests?

(Response RC1-8): The model included eight PFTs. Species are assigned to a PFT on the basis of three traits: wood density, specific leaf area, and legume/non-legume status. The binning of species into PFTs is discussed in Xu et al. (2016) and Medvigy et al. (2019). Deciduousness is not pre-assigned, but is rather an emergent model outcome (Xu et al. 2016).

In this dry tropical forest, intra-annual variability of precipitation does influence plant phenology: new leaves are produced in April and May and shed between January and March; stems did not grow during the dry season (Waring et al., 2019). Therefore, the fertilization was conducted only during wet season; measurements of leaf/wood/fine root production were conducted differently to accommodate seasonality (Waring et al., 2019). Our model simulations and the calculation of productivity were in accordance with this experiment.

Analogous experiments have been conducted in tropical moist forests (please see response RC1-3).

4. Using different statistical analyses for leaf, wood, and root due to patterns that emerged from the observation-based data may not be the best approach. It would be a more direct comparison to use the same statistical analyses for each tissue because the biases in the empirical observations may not be present in the model outputs. Figure 4 is a central figure but it is unclear whether it shows only the control treatment or an average across treatments. Regardless, this analysis should be conducted for each fertilization treatment independently given that the premise of the study is that fertilization treatment influences relative allocation. Furthermore, are the temporal trends important here given that the same amount of fertilizer was applied each year and the experiment was only 3 years long? Given that the primary focus is the difference between tissues rather than the difference between years, it may make more sense to aggregate across years for each tissue / treatment.
(Response RC1-9): We thank the reviewer for these suggestions. We re-analyzed these results largely following this advice. First, we averaged over years. The main drawback to this averaging is that we lose some statistical power because all t-tests would need to be done on only four replicates. However, when we carried out this analysis, we found that the impact on our results was small (see below). We also performed statistical tests by treatment. An advantage of following this approach is that we no longer had concerns about mixing samples that were drawn from different distributions. We therefore performed t-tests for leaf, wood, and fine root productivity. We no longer found it necessary to do the regression on fine root productivity. As a result, productivity of leaves, wood and fine roots were analyzed in exactly the same way.

First we looked at leaf productivity. We carried out t-tests to determine whether there were significant differences between simulations and observations. In response to a community comment (please see response CC1-9), we increased the number of parameterizations that we tested. If, for a particular treatment, the p value was less than 0.05 (the threshold which we took to indicate a significant different between the simulations and the observations), we indicated that in the table below. “C” means the control plots were significantly different, “N” means the +N plots were significantly different, “P” means the +P plots were significantly different, and “B” indicates that the +NP plots were significantly different. If an entry is empty, it means that no significant differences were found for any treatment. Here are our results for leaf productivity:

<table>
<thead>
<tr>
<th>LEAF PRODUCTIVITY</th>
<th>a=0</th>
<th>a=0.1</th>
<th>a=0.2</th>
<th>a=0.3</th>
<th>a=0.4</th>
<th>a=0.5</th>
<th>a=0.6</th>
<th>a=0.7</th>
<th>a=0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>b = -60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = -40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = -20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 60</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As in our original manuscript, almost all model parameterizations successfully predicted leaf productivity.

We then looked at wood productivity:

<table>
<thead>
<tr>
<th>WOOD PRODUCTIVITY</th>
<th>a=0</th>
<th>a=0.1</th>
<th>a=0.2</th>
<th>a=0.3</th>
<th>a=0.4</th>
<th>a=0.5</th>
<th>a=0.6</th>
<th>a=0.7</th>
<th>a=0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>b = -60</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = -40</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = -20</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 0</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 20</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Many parameterizations, especially with $b < 0$, $b = 0$, or relatively small $a$, did not predict wood productivity in the +P treatment. Parameterizations with $b > 0$ were mostly successful in all treatments. Again, this result is consistent with our original manuscript. However, here we have more information as compared to our original manuscript: we see that the model only ever had trouble with the +P treatment, and not the +NP treatment.

Finally, we looked at fine root productivity:

<table>
<thead>
<tr>
<th>FINE ROOT PRODUCTIVITY</th>
<th>a=0</th>
<th>a=0.1</th>
<th>a=0.2</th>
<th>a=0.3</th>
<th>a=0.4</th>
<th>a=0.5</th>
<th>a=0.6</th>
<th>a=0.7</th>
<th>a=0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>b = 40</td>
<td>C,N,P</td>
<td>C,N</td>
<td>N</td>
<td>C,B</td>
<td>C,P</td>
<td>C,P</td>
<td>C,N,P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b = 60</td>
<td>C,N</td>
<td>N</td>
<td>B</td>
<td>P,B</td>
<td>C,P,B</td>
<td>C,P</td>
<td>C,P</td>
<td>C,N,P</td>
<td></td>
</tr>
</tbody>
</table>

Only four parameterizations, all with $b = 20$ or $b = 40$, successfully predicted fine root productivity in all treatments. These four parameterizations also successfully predicted leaf and wood productivity in all treatments.

Again, the drawback of this approach is that it reduced the number of samples: each t-test was done with only four replicates. However, our new results are similar to what we reported in the original manuscript. As an additional test, we tried aggregating control and +N, and comparing that to the aggregation of +P and +NP. This procedure doubles the number of replicates. It gave very similar results (not reported here).

5. Is a 2 year spinup sufficient? Shouldn’t the spinup be run until an equilibrium is established?

(Response RC1-10): We did not want to spin up the model to equilibrium because the actual forest is only about 30 years old and is not in equilibrium. Rather, we used the observed stand structure and composition and observed soil nutrient status to initialize the model. Our two-year spin-up was done to initialize soil water. For further rationale, see Xu et al. (2016). We would be able to provide more justification of the spinup in a revision.
Technical correction

Line 81 “While models have rarely be validated on these time scales” I would argue that models are often evaluated over the past several decades (1960s to present).

Here we wanted to emphasize that nutrient fertilization experiments have usually lasted for a few years and hence the lack of long-term benchmarks for model validation. We can rephrase it to make it clearer.

Production units should be kg m\(^{-2}\) yr\(^{-1}\).

Thanks for pointing this out. We can correct the units in a revision.

Table 3 is challenging to interpret. Could this be transformed into a figure?

Yes, we can make such a figure. See the figures given in response RC1-5 for examples of the potential format.

Include other parameterizations in Figure 5 (additional panels).

As discussed above, this figure will be eliminated in the revision.

Include other treatments in Figure 8 (additional panels).

As discussed above, we plan on eliminating this figure.

Figure 6: Clarify if this is averaged across treatments or if this is the control treatment only.

This is the average across treatments. We can clarify it in the caption.

References


