18 September, 2018

Dr. Russell Monson

Editor-in-Chief, Oecologia

Dear Dr. Monson,

We would like to thank you and the reviewers for providing such helpful comments and suggestions. We greatly appreciate your time and expertise. We have reviewed the comments carefully and have revised the manuscript accordingly.

Please find below our responses to the reviewers’ concerns. Our response is given in blue Arial font. Excerpts from the manuscript are black Times New Roman with additions in red and deletions indicated by a strikethrough.

We believe that the revisions we have made have improved the paper and we hope that you will find the manuscript to be acceptable for publication. We look forward to hearing from you soon.

Sincerely,

Kara Gonzales and Ruth Yanai

Your manuscript has been reviewed for Oecologia and one of our editors has made a recommendation of Major Revision. I concur with this recommendation and believe that a major revision of this manuscript could produce an acceptable paper.

The Handling Editor and Reviewers comments appear below. As you will see, some concerns are significant and eventual acceptance is not a certainty.

\*In addition to addressing the reviewer's concerns, I request some editorial revisions: (1) Please do not use bold type in the axis titles or other labels on your figures. The font type should be unbolded Helvitica or Ariel.

The figures have been edited so that all text is an unbolded sans serif font.

(2) Many of the lettered labels on your figure clash with trend lines. This makes it hard to discern the letters. Please revise your figures so that the lettered labels within the figure panels have solid-fill white backgrounds or are otherwise set off from the lines that travel through them. At the moment it looks like a hash of letters and lines.

We understand the problem. For example, in Figure 7 there are two RM and a YB, of different treatments, for which the observations are nearly on top of one another. This is the original figure:



Clearly, these three observations cannot be separated by white without moving the points (which are currently accurate in x and y values) or blocking the ones they overlap.

Another possibility is to use open symbols:



While the points are distinguishable, we judged this to be an overall setback in legibility, as it requires the reader to match the symbols to species.

At last, we opted to keep the lettered species codes as the markers, but made the letters smaller and lighter weight rather than bold, as in the example below.



If you wish to revise your manuscript for further consideration, please address the points raised in the reviews and submit your revision by 18 Sep 2018 at<https://oeco.editorialmanager.com/>. In your response to reviewers, please indicate how you have addressed each individual reviewer's comments. Of course, if you disagree with specific comments raised in the reviews, feel free to respond accordingly. The original handling editor will be notified of your revision and will continue to handle your manuscript.

Please review our detailed author instructions for proper manuscript formatting: "Oecologia Author Instructions General" (Manuscript guidelines) and "Oecologia Author Instructions Figures" (Artwork instructions). These files are located at

<http://www.springer.com/cda/content/document/cda_downloaddocument/Oecologia_Author_Instructions_General.pdf?SGWID=0-0-45-1135637-p1081245> and

<http://www.springer.com/cda/content/document/cda_downloaddocument/Oecologia_Author_Instructions_Figures.pdf?SGWID=0-0-45-1135642-p1081245>

Please make sure to submit editable source files (e.g. Word).

We will return manuscripts which do not follow these guidelines.

Thank you for considering Oecologia as an outlet for your research results. We look forward to receiving your revised manuscript. If you choose not to undertake these revisions, then I wish you the best of luck in finding a more suitable outlet for your work.

Sincerely,

Russell K. Monson

Editor in Chief, Plant Physiological and Ecosystem Ecology

Oecologia

Comments to Authors from Handling Editor and Reviewers:

Handling Editor:

Three reviewers and I read this manuscript and there was significant interest in the topic and the results. Work like this is very rare, and valuable, and I think the paper will be of interest to readers of Oecologia. The reviewers (and I) were supportive of eventual publication, while recommending various levels of revision (most recommended major revisions). I concur with their assessment. I think there are three key things to address: replication, critical N:P ratios, and broader implications as outlined below. However, the reviewers also identified other areas for clarification and other types of improvement. A revision that takes these comments into account may be publishable in Oecologia.

A key concern from the reviewers is pseudoreplication. No reviewer conveyed that the lack of replication was a cause for certain rejection, but they did indicate that the sampling needed to be clarified and if indeed the study was not replicated (or was pseudoreplicated) then the issue needs to be addressed in the paper. Personally, I found the statistics section confusing on this point, but by the time I got to the results, it was pretty clear to me when the data were based on one stand and when on three. It seemed to me the authors were being very careful to let the reader know when things weren’t replicated at the stand level. Still, the reviewers and I see room for improvement in handling this important issue.

We have improved the clarity with which we describe the statistical tests, including the addition of a table (detailed below).

One key concern I have is distinct from those made by reviewers. This paper uses an N:P ratio of 16 as the cutoff for N vs P limitation. For example in the abstract: “In unmanipulated controls, foliar N:P ratios ranged from 17 - 24 and litter N:P ratios ranged from 18 - 42. These values are indicative of P limitation,”. My question is this: Is there any evidence from forest stands that 16 is the right number? To my knowledge, the use of N:P ratios above and below 16 as indicators of N vs P limitation has only been thoroughly tested in wetland systems. Never in forests. Why would we expect the ratio to be the same in forests? I wouldn’t. The Koersleman paper you cite only has wetland data. The Guswell paper is also based, almost exclusively on herbaceous plants and wetland plants, but it at lease mentions forests a few times. If you read the section on established vegetation N:P in Guswell you lose all confidence in the idea that the critical N:P of Koersleman’s wetlands is appropriate for NE forests. Guswell concludes, N : P ratios <10 and >20 often (not always) correspond to N- and P-limited biomass production, as shown by short-term fertilization experiments; however long-term effects of fertilization or effects on individual species can be different.” You are not alone; many people are using the 16 value, but it is not at all clear to me how our discipline went from the two papers you cite, to using the 16 value to make assertions about forest P limitation. I think you miss an opportunity here to help forest ecosystem ecologists understand what the critical N:P ratio is for forests. Instead of using 16 as a known critical value, can you use these data to calculate what the critical value might be for forests? Or for each of the species you studied?

Thank you for raising this important question. We now have data on foliar nutrient concentrations from multiple stands and species for several years post-treatment, and we hope to use these data to make a contribution to the literature on this topic. As for the present paper, we continued to use Gusewell’s thresholds of 10 and 20 for N and P limitation. We qualified the values from the Koerselman paper: “Ratios >16 and <14 indicate P and N limitation in wetlands (Koerselman and Meuleman 1996).”

We also toned down the wording referencing these ratios in the abstract: “These values ~~are indicative of~~ suggest P limitation, although this forest type has been assumed to be N-limited.” and in the results: “Foliar N:P ratios in the control plots suggest ~~indicate~~ that our study stands were P limited, averaging 26 for green leaves and 32 for leaf litter.”

I don’t think this sentence is supported by data from forests around the world: “Increases in N:P ratios 56 and shifts towards P limitation have been induced by N fertilization around the world (Menge 57 and Field 2007; Elser et al. 2009; Peñuelas et al. 2012)”. Menge is a model. Elser is specifically for lakes. Penuelas just lists a few case studies.

We removed Elser et al. 2009 and Penuelas et al. 2012, but kept Menge and Field 2007 because it reports results from a field experiment at Jasper Ridge. We also added two additional citations with results from Canada and Sweden:

Increases in N:P ratios and shifts towards P limitation have been induced by N fertilization and deposition in Ontario (Gradowski and Thomas 2006), Sweden (Hedwall et al. 2017), and California (Menge and Field 2007).

Line 359: I would delete “surprising”.

Done. The sentence is now: “The fact that there was no analogous reduction in foliar P concentration under N addition lends further support to our ~~surprising~~ finding of P limitation in this forest.”

Line 36-37. I think this conclusion to the abstract is rather weak. Perhaps the authors are trying not to reach too far based on one study, but if, as the into notes, “This is the first full-factorial test of N and P manipulation of nutrient resorption in a temperate forest.” Then couldn’t you link this site based work with broader implications for ecology? See reviewer #2 on this topic as well.

We ended up moving a version of the first sentence of the abstract to the end, which provides a broader conclusion for this work.

~~Temperate forests are generally thought to be N limited, but human activities, such as anthropogenic N deposition, can affect nutrient limitation in these forests.~~ Resorption, the process by which trees withdraw foliar nutrients prior to leaf abscission, is one of the most important nutrient conservation mechanisms in trees, and, along with foliar nutrient concentrations, can be used to infer nutrient limitation status. We collected green and senesced leaves of five species in early successional stands in the White Mountains of New Hampshire. In unmanipulated controls, foliar N:P ratios ranged from 17 - 24 and litter N:P ratios ranged from 18 - 42. These values suggest P limitation, although this forest type has been assumed to be N-limited. Additionally, N:P resorption ratios in control plots were <1, reflecting proportionately more conservation of P through resorption than N. Four years into a full-factorial NxP fertilization experiment, N and P additions had increased N and P concentrations in leaves; more importantly, P addition reduced N concentration, possibly indicating alleviation of growth limitation by P. Resorption of P was less proficient (indicated by the concentration of an element in leaf litter) with P addition, as expected. Resorption proficiency and efficiency (the proportion of leaf nutrients resorbed) of N increased with P addition, suggesting increased demand for N with alleviation of P limitation. Resorption of P was more proficient and efficient with N addition, consistent with exacerbated P limitation. Temperate forests on glaciated soils are generally thought to be N-limited, but long-term NxP manipulations in this biome have been lacking. Our results suggest that decades of anthropogenic N deposition may have tipped the balance to P limitation in these forests.

Reviewer #1: Review of the manuscript entitled "Nitrogen-phosphorus interactions in young northern hardwoods indicate P limitation: foliar concentrations and resorption in a factorial N and P addition experiment" submitted to Oecologia (OECO-D-18-00407). This paper deals with nutrient limitation of temperate forests and indicate that also forests under relatively low deposition of anthropogenic nitrogen can be P-limited. There is very little work done on P-limitation in temperate forests in comparison to for example tropical systems. The paper is very well written and can make a nice contribution to the literature on this topic after minor revisions.

Thank you!

My only larger remark is that the section "Data Analysis" (lines: 178-235) is hard to understand. It is difficult to follow which data was used for which analysis. I think this issue could be resolved by presenting the structure of the analyses in a table or figure.

Thank you for this suggestion. We added the following table. We agree that this is an improvement to the paper, even though this information is available in the text.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sampling Unit | C1 | C2 | C3 |
| Green leaves | Tree |  | Pre and Post |  |
| Litter (Resorption Proficiency) | Plot | Pre and Post | Pre and Post | Post |
| Resorption Efficiency | Plot |  | Pre and Post |  |

Due to this it also hard to judge if the analysis are appropriate or not, but there seems to be an issue with pseudo replication of the green leaves since they were harvested in one stand. This may be a smaller issue but needs to be addressed in the discussion and the statement on lines 199-200 corrected.

We had said: “Though there was no replication of treatments, the experimental design permitted analysis of the full factorial of N treatment, P treatment, and species.” We agree that this statement is confusing; the reason we can test for the effect of N and P addition is because there are two plots receiving each nutrient. We revised this sentence to say, “Though there was no replication of treatment plots across stands, the experimental design permitted analysis of the full factorial of N treatment, P treatment, and species since each treatment was applied to two plots and each species occurred in all four plots.”

Minor comments:

Line 94 and more places: give scientific species names when first mentioned.

Thank you -- all scientific names are now given at first mention in the introduction.

Line 226: which covariate?

This was an error in the use of the indefinite article! The sentence was “A covariate was included to control for variation among plots not due to treatment, improving our ability to detect treatment effects.” “ Now it says, “The covariates were included to control for variation among plots not due to treatment, improving our ability to detect treatment effects.” The previous paragraph introduces the covariates.

Line 228: which reasons?

We had said,“We report results both with and without the covariate for several reasons.” The reasons are given in the subsequent sentences. We corrected the sentence to: “We report results both with and without the corresponding covariate for the following reasons.”

Lines 410-413: this section is over-referenced since there are no controversial statements here. We realized that this sentence was somewhat repetitive and ultimately combined it with the preceding sentence. We also removed some of the references:

“At one ~~extreme~~ end of the spectrum, evergreens maximize nutrient use efficiency through low leaf nutrient concentrations, lower N resorption, ~~and~~ a longer leaf lifespan, and a lower potential growth rate compared to deciduous species (Aerts 1996; Killingbeck 1996; Vergutz et al. 2012). ~~Compared to deciduous species, evergreens have lower nutrient concentrations in foliage (Aerts 1996; Van Heerwaarden et al. 2003; Vergutz et al. 2012; Liu et al. 2014) and leaf litter (Killingbeck 1996; Van Heerwaarden et al. 2003; Kobe et al. 2005; Vergutz et al. 2012; Liu et al. 2014) and lower N resorption efficiency (Aerts 1996; Vergutz et al. 2012; Yuan and Chen 2015).”~~

Lines 422-424: an increase in foliar concentration may just as well be a sign of luxury uptake and no dilution from increased growth. Without growth data this is impossible to say.

Good point. We qualified this statement. “For example, pin cherry exhibited some of the greatest increases in foliar P concentrations in response to P additions (Fig. 3)~~, which~~ ; increased P uptake could help ~~it~~ pin cherry outcompete other species or persist longer if P is limiting (Fahey et al. 1998).”

Discussion: the title points out that this study was done in young forests. I really miss a discussion about how the patterns found here may change over time with succession both by time from disturbance and due to the differences between tree species found. How may then the results found here differ from for example older forest?

This is one of the things we hope to investigate with this long-term nutrient manipulation experiment! We are planning a paper reporting similar measurements in the older stands of the same study, collected the following year. For this paper, we chose to work in the young stands because they offered diversity in species and life history strategy, as described in the third to last paragraph of the introduction.

There is also some new literature on phosphorus resorption, phosphorus limitation of forest plant communities and the dynamics of phosphorus limitation under changing nitrogen deposition that could be cited.

Ji, H., Wen, J., Du, B. et al. Annals of Forest Science (2018) 75: 59.<https://doi.org/10.1007/s13595-018-0727-5>

Hedwall, PO., Bergh, J. & Brunet, J. Oecologia (2017) 185: 317.<https://doi.org/10.1007/s00442-017-3945-x>

Lin Yu, Giuliana Zanchi, Cecilia Akselsson, Håkan Wallander, Salim Belyazid, Modeling the forest phosphorus nutrition in a southwestern Swedish forest site, Ecological Modelling, Volume 369, 2018, Pages 88-100, ISSN 0304-3800,<https://doi.org/10.1016/j.ecolmodel.2017.12.018>.

Thank you for bringing these recent studies to our attention. We have added a citation to the first one in the discussion and to the second one in the introduction:

“Consistent with studies that have found nutrient resorption to be low when nutrient availability is high (Staaf 1982; Kobe et al. 2005; Hagen-Thorn et al. 2006; Vergutz et al. 2012; Yuan and Chen 2015; Ji et al. 2018), P resorption proficiency and efficiency were reduced in the P and NP treatment plots, due to higher P concentrations in both green leaves and litter (Fig. 1).”

“Increases in N:P ratios and shifts towards P limitation have been induced by N fertilization and deposition around the world (Menge and Field 2007; Elser et al. 2009; Peñuelas et al. 2012; Hedwall et al. 2017).”

Reviewer #2: Review for "Nitrogen-phosphorus interactions in young northern hardwoods indicate P limitation: foliar concentrations and resorption in a factorial N and P addition experiment" by Gonzalez and Yanai

General comments

This manuscript, by Gonzalez and Yanai, examines the foliar P and N concentrations and resorption dynamics in young northern hardwood forest stands that were subjected to a factorial N and P fertilization experiment. I found this manuscript to be generally well written and thoroughly interesting. The main message — that these forests appear to be limited by P rather than N — is important. I found the study design and interpretation of the results to generally be convincing. It also complements other recent work in this region.

Thank you.

While I was generally happy with the framework of the paper, there are some issues that could be addressed before publication is considered.

Much of the paper focused on how the evidence suggests P limitation and did a good job making that case. But this begs an obvious question that was not addressed clearly. Why did you find P limitation in hardwood forests previously thought to be N limited? Do you think that is a widespread (regional) phenomenon or something more localized to these stands, because of geology or other reasons? I'd like to see more, even if it is somewhat speculative, about the underpinnings of these results and how they fit in a broader picture of temperate forest ecosystem nutrient limitation.

In the introduction (second paragraph), a sentence that previously read: “Increases in N:P ratios and shifts towards P limitation have been induced by N fertilization around the world (Menge and Field 2007; Elser et al. 2009; Peñuelas et al. 2012).” now reads: “ Increases in N:P ratios and shifts towards P limitation have been induced by N fertilization and deposition around the world (Gradowski and Thomas 2006; Menge and Field 2007; Peñuelas et al. 2012; Hedwall et al. 2017).”

Additionally, in the first paragraph of the discussion, the sentence that previously read, “Forests in temperate regions are typically thought to be N-limited (McGroddy et al. 2004; Reich and Oleksyn 2004), but we found multiple indications of P limitation in our stands.” now reads: “Forests in temperate regions are typically thought to be N-limited (McGroddy et al. 2004; Reich and Oleksyn 2004). However, decades of anthropogenic N deposition in the northeastern United States might be expected to lead to altered biogeochemical cycling, and we found multiple indications of P limitation in our stands.”

Like many ecosystem ecologists, I admit I tend to think about resorption only generally as important to nutrient conservation and without devoting a lot of thought to details. From that perspective I liked this paper for its use of resoprtion dynamics to say something about the ecosystem nutrient limitation and how that is expressed across multiple dominant tree species. But I also found some details about the methodology to be lacking. Specifically, resorption efficiency was calculated as the difference between sunlit upper canopy leaves and litter which could have originated anywhere in the canopy. I know that foliar chemistry can vary quite a bit depending on height in the canopy. Is your method standard? I know it would be way harder to collect litter that was sunlit, but some mention of this is warranted and citations of other papers that have explored it more in detail if possible.

We clarified our sampling methods as follows: “Leaf litter was collected from two stands (C1 and C2) in 2009 and 2010, and from three stands (C1, C2, and C3) in 2014. Pre-treatment leaf litter was collected with multiple net traps distributed within each plot (See et al. 2015), while post-treatment litter was collected from the ground. ~~In both cases, litter was collected throughout the plot and composited before analysis, as the experimental unit was the plot.~~ In both cases, litter was collected throughout the plot and, since the experimental unit was the plot, composited before analysis. This sampling method results in a sample representative of the full canopy, while green leaves are sampled from the upper canopy. The variation in nutrient concentrations in foliage due to sampling position was only 12% in an earlier study in northern hardwoods (Yang et al. 2016). In all sampling years, litter was collected in early October following a rain-free period.”

I also have several specific comments listed below that I think should be addressed before publication of this manuscript is considered.

Specific comments

Title: The current title is unnecessarily long. I've heard that papers with shorter titles get more attention, so it seems like a good idea to go that route.

We considered a variety of shorter titles, such as:

“Factorial N by P addition experiment in young northern hardwoods indicates P limitation” 13 words

“Nitrogen and phosphorus foliar concentrations and resorption indicate P limitation in young northern hardwoods in a factorial N by P experiment” 21

“Nitrogen addition increases P conservation through resorption in young P-limited northern hardwood forests” 14

“Foliar concentrations and resorption in young northern hardwoods indicate P limitation” 11

We didn’t think any of them was better than the title we had selected (after long deliberation), which is: The title was: “Nitrogen-phosphorus interactions in young northern hardwoods indicate P limitation: foliar concentrations and resorption in a factorial N by P addition experiment.” 22.

We are aware of the correlation between short titles and citation rates, but we contend that this relationship may not be causal, but rather a consequence of good scientists (and good writers) choosing short titles for most of their papers. It is also possible that papers reporting specific studies, such as ours, require longer titles to describe them well, whereas a meta-analysis reporting global patterns in foliar nutrient resorption could have both a short title and a lot of citations. We would like to keep the long title.

Lines 20-24: Awkward transition.

The sentence originally read: Temperate forests are generally thought to be N-limited, but human activities can affect the biogeochemistry of these forests.

It now reads: “Temperate forests are generally thought to be N-limited, but human activities, such as anthropogenic N deposition, can affect nutrient limitation in these forests.” This flows better into the next sentence about nutrient conservation and limitation.

Line 20: "human activities"… are you getting at N deposition? Changing biogeochemistry is too broad.

As documented for the previous comment, this sentence now refers to “nutrient limitation” instead of “biogeochemistry,” and “human activities” now includes the example of “anthropogenic N deposition.”

Line 42: I suggest changing to "elevated" rather than increases since many areas are now experiencing decreasing atmospheric deposition.

The sentence is: “Increases in atmospheric nitrogen (N) deposition have profoundly affected the biogeochemistry of forests through changes in limitation status, pH, and net nitrification.”

We have chosen to leave the word ‘increases’ because the sentence is referring to decades of increases above historic levels and the ongoing legacy of this increase, rather than a recent trend of decrease.

Line 43: profoundly affected…in what way?

We added the text in red: “Increases in atmospheric nitrogen (N) deposition (Galloway et al. 1995; Galloway 1998), have profoundly affected the biogeochemistry of forests through changes in pH, net nitrification, and nutrient limitation status.”

Line 58: I suggest adding a reference for this sentence.

“Limitation status affects nutrient uptake and conservation in plants (Güsewell 2004).”

Line 63: may confer a competitive advantage-true if nutrient availability is low

The sentence has been improved: “Nutrient resorption lessens plant dependence on external nutrient supplies (Aerts and Chapin 1999), and thus may confer a competitive advantage under nutrient limitation (Fahey et al. 1998).”

Lines 94-97: Awkward sentence. I'd start something like "For example, pin cherry is a fast-growing, short-lived species in northern hardwood forests that has high nutrient concentrations…"

Yes, the subject was too far from the verb: “Pin cherry, for example, a fast-growing, short-lived species in the northern hardwood forest type, has high nutrient concentrations.” It now reads, “Pin cherry (*Prunus pensylvanica* L.f.), for example, is a fast-growing, short-lived species in the northern hardwood forest type, with high nutrient concentrations.”

Lines 96-97: For this comparison cite something to indicate beech has low nutrient concentrations.

“American beech (*Fagus grandifolia* Ehrh.), a shade-tolerant, long-lived species in the same forest type, might be expected to exhibit more conservative strategies (Tripler et al. 2002).”

Line 101: unneeded parentheses before Fisk

Removed.

Line 117-119: Why include this hypothesis (regarding this effect of P addition on N resorption) if you don't have a reason to expect it?

We stated previously in the introduction that a consequence of co-limitation might be the possibility of the allocation of a non-limiting nutrient towards the acquisition or conservation of a limiting nutrient; thus, even though the mechanisms were not as clear as for the converse, it seemed plausible that P addition might have an effect on N resorption. The lack of a mechanism is also addressed in the Discussion. It’s interesting and we hope future research will address this!

Line 120-121: Just for caution's sake, I'd recommend adding "To our knowledge…"

Done: “To our knowledge, this is the first full-factorial test of N and P manipulation of nutrient resorption in a temperate forest.”

Line 129: hyphen in well-drained soils

The US Soil Taxonomy does not place a hyphen there. “Well” is an adverb and it has every right to modify an adjective. Compound modifiers should be hyphenated.

Line 133: mean temperatures are highly dependent on elevation

The weather station from which we obtained temperature data was at a lower elevation than our study sites, so we deleted the sentence about temperature, having already characterized the climate. Fortunately, the elevation range among the three stands is small. “All three stands are located at similar elevation, 340-590 m above sea level.”

Lines 139-142: Can you cite a paper with more detail on the fertilization method?

Yes, we cite Fisk et al. (2014), which provides more information on the fertilization methods. “Four 50x50 m treatment plots were established in each stand as part of a larger study on multiple element limitation and have been fertilized annually since 2011 in a full factorial design (Fisk et al. 2014).”

Line 144: Change to "Five tree species.."

Done.

Line 150: do you mean at least half of the sampled trees?

This sentence has been improved, sorry for the confusion. “Trees >10 cm dbh in our plots are tagged; about half of the individuals sampled in 2010 had dbh > 10 cm and thus were tagged. These trees were also chosen for sampling in 2014~~, so at least half the sampled trees were sampled both pre- and post-treatment.~~”

Line 152: not clear where 60 cases came from. Stands x plots x species?

“Leaves were shot from 2-4 trees (3 trees in > 90% of 60 cases (trees x species x plot) in 2014) of each of the five species in each of the four plots.”

Line 159: This answered a question I had earlier. I'd suggest moving it into the preceding paragraph.

The reviewer is referring to the last sentence in the section quoted below. The sentence does not belong in the previous paragraph, which is specific to leaf litter collection in the field. The selection of 10 leaves (green or senesced) occurred in the laboratory after collecting them in the field. We added a few words to make this explicit.

“Leaf litter was collected from two stands (C1 and C2) in 2009 and 2010, and from three stands (C1, C2, and C3) in 2014. Pre-treatment leaf litter was collected with net traps hung in each plot, while post-treatment litter was collected from the ground. In all sampling years, litter was collected in early October following a rain-free period.

 “In the laboratory, we selected at least 10 leaves per tree (green leaves) or per plot (litter) ~~were selected~~ for analysis, avoiding those that showed evidence of disease or damage from buckshot or herbivory.”

Line 167: change to "at 470". - add muffle furnace or whatever is appropriate

Done, it was a muffle furnace.

Line 171: add "analysis" after ICP-OES; change ten to 10

Done.

Line 172: this is an example (of several) where your writing for the methods switches between passive and active voice. I suggest being consistent with one or the other.

The sentence is: “During ICP-OES, a blank was run after every ten samples and an in-house quality control after every five; we re-calibrated the machine if >5% drift was observed in the in-house standards.” We appreciate that others may disagree, but thought that the way we wrote the methods (including sometimes switching between active and passive voice) gave the section a good flow and made it interesting and dynamic to read.

Line 321: This section includes interpretation of the N:P as indicative of P limitation. I'd save this interpretation for the discussion section where you can cite your references

We worked hard to write the results in a way that would allow readers to appreciate their meaning. One way to do this is to pose interesting questions in the Introduction and relate the results to those questions. There are other ways to organize papers; in geology, the interesting questions are not raised until the Discussion (and readers skip the Results section). We err on the side of relating our results to the questions posed in the Results, though interpretation better developed, with reference to other studies, in the Discussion.

Line 351: "in our forest type" is the Goswami et al 2018 study the same plots or the same forest type? This gets to my question about how widespread is P limitation?

Goswami’s study involves 13 stands, of which we used 3 (none of them mature). So we now say: “These indicators are consistent with a recent finding of increased diameter growth of trees in response to P but not N addition in mature stands in our study system ~~forest type~~ (Goswami et al. 2018).”

Line 367-371: This is repetition of the text above.

We removed some of the preceding text: “Nitrogen resorption, however, was incomplete in both control and N plots, suggesting that N is less in demand ~~and that N conservation through resorption is not a priority~~. The relative completeness of resorption may reflect the relative costs of acquiring nutrients via uptake from external sources versus recycling from internal sources.”

Line 399: Extra parentheses around Townsend et al.

Removed extra parentheses.

Figures: I'd like to see the figures easier to interpret. Some are great (Fig 1), while others are messy (Fig 2 and Fig 4 especially). If you can't find another way to depict them that is easier to read, some information could goes in a table.

We agree that some of the figures have a lot of information and are not immediately straightforward. We did quite a bit of searching on various ways to graph ANCOVA results, and the most common approach, by far, is to present them in the graph style that we used. We were unable to come up with a better way to present the same information in a concise manner while also reflecting the statistical test. We also did not want to put the information in a table because it would, in our opinion, be even harder to see any patterns in the data. Therefore, the figures themselves are the same, but we did adjust the font style and size of the markers to reduce overlap in points.

Reviewer #3: This is a fairly straight forward study reporting potential forest P limitation in a region expected of widespread N limitation. The results of this study are novel because they provide empirical evidence that forests in Northeastern United States may actually be limited of P (or co-limited) rather than the assumed N limitation. Overall, the authors report that under native ambient conditions, the NP ratio of tree leaves are well in the zone of P limitation, and fertilizing with P reduces the NP ratio, and likely P limitation.

My primary concern is over the unbalanced sample design, issues with replication, and some assumptions in site comparability. While I don't doubt the results and I do find them compelling and reasonable, I find the approach of the analysis troubling. More justification is needing (Data Analysis section) on how the statistical approach to the unbalanced and low/no replication is appropriate to produce robust scientific results. In other words, are the results scientific evidence or anecdotal? Perhaps provide evidence that the sites have similar soil N and P concentrations? If the sites are very similar, then it could provide more evidence to support your approach.

This was not well described in our paper. We clarified the issue of replication of the treatments in the factorial design, as described above. In addition, the ANCOVA with pre-treatment data increases confidence that the results were due to treatment effects, and not random variation among plots.

Specific Comments:

Page, Line

52 Mistake/typo, NP ratios above 16-20 indicate P limitation, not N limitation.

Thanks! Corrected.

124 Please provide some basic soil properties like: total CNP, pH, WHC, texture, etc so the results from this study could be compared to other studies.

Good point. We added references to previously published soils information from the study (Vadeboncoeur et al. 2012, 2014).

Vadeboncoeur M.A., S.P. Hamburg, J.D. Blum, M.J. Pennino, R.D. Yanai, and C.E. Johnson. 2012. The quantitative soil pit method for measuring belowground carbon and nitrogen stocks. *Soil Sci. Soc. Am. J.* 76(6):2241–2255. DOI:10.2136/sssaj2012.0111

144 Were trees samples even if they were very close to the plot edge? In other words, did you only sample within the core of the plot to minimize trees growing also in non-treated soil?

The following sentence was added to the methods: “Treatment plots measure 0.25 ha (50 m x 50 m) and all measurements are made in the inner 30 m x 30 m, allowing for a 10-m buffer to avoid edge effects.”

147-154 This whole paragraph is very confusing due to poor flow and needs to be reworded to improve clarity. I suggest keeping to simple sentences. Overall, I still don't clearly understand exactly which samples were harvested beside all the green leaves were just from one stand - but what is the replication (# of plots) within a stand? Based on 139, each stand only has one replication. As such, I'm very sorry to say this appears the results are anecdotal evidence.

Each stand has the full factorial treatment layout. “Four 50x50 m treatment plots were established in each stand as part of a larger study on multiple element limitation and have been fertilized annually since 2011 in a full factorial design. One plot in each stand has been treated with N alone (30 kg N ha-1y-1as NH4NO3), P alone (10 kg P ha-1y-1as NaH2PO4), N and P together (same application rates), or control.” For green leaves, only one stand was sampled. We say more clearly now why this allows for statistical analysis: “Though there was no replication of treatment plots across stands, the experimental design permitted analysis of the full factorial of N treatment, P treatment, and species since each treatment was applied to two plots and each species occurred in all four plots.”

For litter, three stands were sampled. This is complicated, indeed, and should be easier to follow with the addition of the table suggested by Reviewer 1 (shown above).

156 What was the size of the nets and how many nets per plot? As written, it seems only one net per plot per treatment, per stand. So, in 2009 & 2010, n=2 and in 2014 n=3.

We added more information to clarify the statistical design and indicate that the sample is representative of the whole plot, which is important. “Pre-treatment leaf litter was collected with multiple net traps distributed within each plot (See et al. 2015) while post-treatment litter was collected from the ground. In both cases, litter was collected throughout the plot and composited before analysis, as the experimental unit was the plot.” Analyzing the litter separately for each plot would not add statistical power because treatments were applied at the plot level.

164 Please report replication for the N analysis.

Our lab does not run samples in duplicate or triplicate, but relies on frequent analyses of the reference material. We report these results: “For the standard reference material (NIST 1515), recovery of N was within 5% of the certified value for 22 of 24 samples, and within 10% for all samples.”

168 Did you just average the duplicate sample; it is not clear in the manuscript.

Yes, duplicates were averaged, and the following sentence was added: “Replicate samples were averaged for subsequent analysis.”

255 Can you provide the degree of change as a percentage?

The sentence includes the following changes: “green leaf N concentrations were ~~significantly~~ 12% higher with N addition (p = 0.003) and 9% lower with P addition (p = 0.02), with no significant interaction (p = 0.58).”

277 Just to be clear, this was just performed on one (C2) site, correct?

Yes, we have data for efficiency in one stand, as noted in Line 282 and made more clear in the new Table 2.

294 How can you calculate resorption for three stands when green leaves were only harvested in one stand? You need to provide more/better support that this approach is valid and sound. This seems inconsistent from line 207.

The sentence at Line 294 and the following sentence clarify that proficiency (mentioned in the first sentence) is compared in three stands but that resorption efficiency (the subsequent sentence) is only analyzed in one stand. This is also reinforced by the addition of Table 2.

362 What does biochemically complete actually mean? What is the 0.7% and 0.05% referring too? Concentration in litter? That seems to be too prescriptive considering trees have very different nutrient management strategies. Regardless, this concept needs better justification.

We removed the word “biochemically” and clarified that this refers to leaf litter: “In hardwoods, resorption may be considered ~~biochemically~~ complete ~~in hardwoods~~ below concentrations of 0.7% N and 0.05% P in leaf litter (Killingbeck 1996).

365 Explain the logic of what this provide further evidence of P limitation.

The sentence was: “By this definition, P resorption was complete in the control plots, but incomplete in plots with P added, which provides further evidence of P limitation.”

We hope that this sentence will follow better given the change to the preceding sentence. We already established that we are using the degree of P resorption as indicative of the degree of P limitation. Adding P results in less resorption thus less P limitation.

366 Do you have soil nitrification rates for this experiment? If so, increases in nitrification due to the treatments could support the idea of nitrogen supply excessing demand.

The referenced sentence is: “Nitrogen resorption, however, was incomplete in both control and N plots, suggesting that N is less in demand...”

We did observe greater resin-available nitrate in the plots with N and NP added (Fisk et al. 2014). It doesn’t seem interesting to report that we added enough N to exceed demand; this is a reflection of the choice of fertilizer addition rate. It’s more interesting to report that N in the untreated controls was less in demand.

386 Very interesting, so what does this mean? Please discuss this in more detail and try to fit it with broader theory.

We moved one sentence and a paragraph break to make clear that we are discussing the surprising observation. It was clumsy of us to follow that statement (“It was more surprising…)” with a paragraph break and a sentence that did not belong in the topic position. Here are the two affected paragraphs.

“The MELNHE experiment allowed us to test not only for single-element resorption responses but also for multiple-element interactions. Functional links between N and P mean that the concentration of one nutrient influences concentrations of the other, within species-specific ranges of N:P ratios (Mohren et al. 1986). Phosphorus resorption in our study sites prior to fertilization was higher in stands with high soil N (See et al. 2015). We found similar results after experimentally manipulating soil N availability: P resorption was most proficient, efficient, and complete in the plots fertilized with N. This was driven by both higher green leaf P and lower litter P when N was added.

→ “It was more surprising that we also found evidence for the converse: N resorption was more proficient and efficient in plots treated with P, resulting from lower litter N concentrations. ~~Functional links between N and P mean that the concentration of one nutrient influences concentrations of the other, within species-specific ranges of N:P ratios (Mohren et al. 1986).~~ The mechanisms for N to facilitate P resorption are clear, because N is required to build the enzymes required to resorb P. It is less clear by what mechanisms available P could be used to improve N resorption, but the value of resorbed N is greater when P limitation is relieved. Theoretically, plants should distribute effort towards maintaining stoichiometric balance, by increasing acquisition and conservation of the most limiting nutrient (Chapin et al. 2002; Harpole et al. 2011; Rastetter et al. 2013).”

Our close study of this paper during revision resulted in a number of other minor editorial improvements to the paper, which are documented with tracked changes in Word. We hope you like it!