While I think this was a well conceived, designed, and conducted study, I believe that the authors may have pushed their results a bit farther than warranted. Particularly given the scope of the laboratory study. The simulation results were used to compare to reach-scale studies despite the simulation ignoring a multitude of other P uptake and respiration mechanisms common in streams and likely affected by temperature in complex ways. I don't think the simulation study is a problem, but I think it should be discussed for what it is: a simple simulation to develop testable hypotheses. It is not representative of real world expectations and shouldn't be considered as such. There were also some issues with the methods and particularly the lack of detail on statistical analyses performed. I think this manuscript has the potential to be a valuable contributor to the field and poses some interesting next steps that must be considered as we continue to push further into coupled C-N-P cycles and expectations with warmer climates.

We acknowledge that our simulation models only encompass a limited proportion of the total uptake that occurs in a forest stream and that we don't include abiotic sorption, uptake by primary producers, or uptake by microbes on fine particles, wood, or buried sediments. In revising this manuscript, we would like to modify the language we use to describe the results of our simulations to more specifically describe what they represent. We would like to frame the results of our simulations as describing "long-term cumulative uptake" and not "Ecosystem uptake". Additionally, in revising the manuscript we plan to be more careful in how we state the implications of the findings from our experiments and simulations.

Specific comments:

Line 16: I don't know what a 0.48 and 1.02 eV value means for temperature dependence. Is this standard unit/metric used to compare temperature dependence of various processes? Not sure if this is the best choice for the abstract.

Quantifying the effect of temperature on rates using the activation energy from the Boltzmann-Arrhenius equation is the convention in the field of metabolic ecology (e.g., Brown et al 2004, and other citations in the manuscript), and eV is a standard unit of kinetic energy.

Line 17: for ranges (0.12 to 0.48, or 11 to 212) I encourage authors to use "to" instead of a hyphen because a hyphen could be misconstrued to represent a negative sign.

We can make this change when revising the manuscript.

Line 34: Should this be increases in productivity rather than increases in growth? The Rasmussen citation quantified stream metabolic activity (GPP, ER) not growth.

The reviewer is correct. This should be increases in productivity.

Line 40: If the authors are using U in the nutrient spiraling sense, isn't U directly correlated with nutrient availability? It's in the calculation, isn't it?

We agree that there is good information about the effect of concentration on rates of nutrient uptake. In revising this manuscript, we would like to revise this topic sentence to clarify that we

mean that the joint effects of temperature and nutrient concentration on uptake are unknown. We plan to revise this sentence to read: "Mechanisms explaining the joint effects of temperature and nutrients on mass-specific rates of nutrient uptake (*U*) remain poorly resolved."

Lines 60 – 61: I mean, maybe? But the Michaelis-Menten kinetics (Vmax, Ks, etc) might not kick in until super high concentrations, though.

We plan to make this statement less definitive by revising this passage to read "Consequently, the proportion of dissolved nutrients taken up by the microbial community may decline with increasing nutrient concentration (O'Brien et al., 2007)."

Lines 76-78: How much of this negative effect of temperature could be due to canopies opening up in cooler months leading to more light and subsequently more autotrophic nutrient uptake? Even forested headwater streams have open canopies sometime and that short window of autotrophic activity could offset heterotrophic decreases maybe?

We thank the reviewer for pointing this out. This same point comes up again later in comments on the discussion. We address this comment there.

Lines 103 – 104: When were bags deployed initially? There are two collection dates but only one incubation date. Not a big deal but if you are going to report collection date, report deployment date, too.

The bags were deployed on 17 November 2020. We can add this date to a revised version of the manuscript.

Line 105: Were fragments a consistent size? Or was there a targeted size? Why cut the leaves into smaller fragments? How much leaf litter material (mass) was added to each bottle? How much water (1L?)

The leaf fragments were cut to a target size of 1.5 cm x 1.5 cm squares. The leaves were cut to allow measurements in smaller volumes of water. We do not have an estimate of the mass of leaves in each 1-L bottle. We added approximately 1 L of water to each bottle. We can add this information to the methods in a revised version of this manuscript.

Line 113: Were blanks measured initially and after the end of the incubation the same way? I worry about displacement/replacement of water due to the initial DO measurement given how large DO probes can be compared to a scint vial.

The blanks and samples were measured the same way before and after the experiment. Roughly 15% of the water in the vial was displaced during the measurement of the initial DO concentration, and it was immediately replaced with streamwater from the same source that was used to fill the vial. We can add this information to the methods when revising the manuscript.

Line 113: Was there at least an attempt to add a similar amount of leaf material to each vial?

Yes, we attempted to add a similar amount of leaf material to each vial.

Line 117: I suggest writing this out as an equation. Were the incubations done in the dark? I don't see anywhere suggesting that. If incubations were not done in the dark then this approach yields NEP, not respiration.

We can add an equation to a revised version of the manuscript. The incubations were done in the dark. We will add this information to a revised version of the manuscript.

Line 123: nominal pore size?

The nominal pore size was 1.0 micron. We can include this information in a revised version of the manuscript.

Line 126: U is traditionally reported in units of mass per area per time (e.g., mg P / m2 / h). I think the approach the authors have taken here to estimate uptake as mass of nutrient per mass of leaf per time is fine but I think that something other than just "U" should be used here. Also, I'm not the biggest fan of the calculation for U as I would greatly prefer an initial and final sample collected from the same sample container. The authors are assuming that all incubation vials started with the same conditions. I don't know how I feel about that assumption. Were individual tubes amended with P? Or was a reservoir amended with P and then added to the tubes? Also, the drastic differences in incubation time is strongly suggesting an assumption of linear P uptake which I don't know can be expected to hold true across different concentrations.

For the sake of clarity, we would like to use U, which we define clearly at first usage. For each temperature treatment a reservoir of streamwater was amended with P, shaken to mix the solution, and then this nutrient-amended water was added to the vials that were incubated. Our analysis does assume that all vials within a given temperature treatment had the same initial concentration. The three blanks that were measured from each temperature treatment generally had measurement of P concentration that were within 1-2 $\mu g\ L^{-1}$ of one another, which gives us confidence that the reservoir was well mixed, and that the initial conditions were similar among vials. The approach we used allowed us to forgo removing water to measure P concentrations at the beginning of the experiment. When designing this experiment, we were concerned that taking an initial sample from each vial would change water volumes and could have a large effect on rates given the small volumes of water we used. We used shorter incubation times for the warmer incubations in this experiment because we expected uptake and respiration to proceed more quickly. Our hope was that similar masses of P would be taken up across the different temperature treatments to limit the importance of kinetic effects.

Line 135: Where was this categorical block effect included? What are the different experimental batches? I don't understand this statement at all.

We ran the experiment twice to increase our sample size and so wanted to account for any differences between the first and second batch of the experiment in our statistical models. There may have been differences in the microbial communities, the streamwater we collected, or other factors between the two batches of the experiment we ran, and we wanted to account for these in

our statistical models. The block effect was included in the statistical model by adding a categorical variable that indicated which batch each measurement was from. In practice, this meant that the model estimated a different *y*-intercept for the data from each batch, and a single slope. We plan to modify this sentence to make our statistical design clearer: "The two dates the experiment was run on may have had differing biological or environmental conditions, so we included a categorical effect of date in our statistical models to account for any differences."

Lines 136 – 138: What? I do not understand this statement at all. There was a model that compared respiration and Usrp to each other and a categorical variable that indicated if the model was for respiration or Usrp? What kind of a model? How was it evaluated? What was the dependent variable? The dependent variables? More detail and description is needed here.

The model compared the slopes of the relationship between temperature and each rate (respiration and uptake), which are the activation energies. The model is essentially an ANCOVA, but with temperature transformed so that slopes represent activation energies. Because the question we are testing is whether the slopes are different, we are interested in evaluating the significance of the interaction term in this model. We plan to modify this sentence to: "We used an ANCOVA-type linear model to evaluate whether the response of respiration and Usrp to temperature were different. We fit a linear model that described the natural log-transformed rates of respiration or uptake rates as a function of the standardized Boltzmann temperature. The model included an interaction between temperature and the type of rate (i.e., respiration or uptake). A significant interaction term in this model indicates that the slopes of the relationships between temperature and these rates differ."

Line 138: A significant interaction term between what?

The interaction term in the model is between temperature and a categorical variable that describes whether the rate is respiration or uptake (see revised text above).

Line 143: Why was the centered inverse Boltzmann temperature used as the predictor variable? Why not just temperature?

By regressing natural log-transformed rates of biological processes against the inverse Boltzmann temperature the slopes of these relationships can be interpreted biologically as an activation energy. We plan to add more description of this to a revised version of the manuscript.

Line 149: Were leaves weighed at the end of the experiment tin the same manner? How were initial SRP concentrations achieved?

These methods were the same as those we described above for the first experiment. The leaves were weighed and the P concentrations were achieved by adding a concentrated solution of phosphorus to streamwater. We will add this detail to a revised version of the text.

Lines 150 – 154: So 6 temperatures * 8 SRP concentrations * 3 bottles per treatment = 144 individual incubations. Is that accurate?

Yes, that is accurate.

Lines 145 – 173: So basically Usrp was regressed against temperature for each initial SRP concentration and then the slope of those regressions were compared across initial SRP concentration? Was a regression or correlation or something done here? There don't seem to be any stats, it reads like the authors plotted these out and visualized them but that's not a real satisfying analysis in my opinion. The same general though holds for the M-M analysis, too.

The regressions that we did in the section were not just visual - we fit regression models to the data. We can include the parameters and model fits in an appendix of a revised version of the manuscript.

Line 199: It seems like there should to be an analysis section. Or more detail needs to be given for the analyses in the individual sections (as was described in some of my previous comments). How were the simulation models assessed/evaluated?

Because there are three distinct units (two experiments and the simulations), we prefer to include the relevant analysis within each section, instead of describing all the analyses in one unified section. We evaluated the outcomes of the simulations by simply comparing the estimated effects. We did not include statistics for this section. We can clarify this in the revised text of the manuscript.

Line 203: canonical is an odd word choice here.

We can drop this word from a revised version of the manuscript, although this word is often used in the metabolic theory literature to describe this important value of the activation energy of cellular respiration.

Line 226: These are interesting results. It's definitely a very simplified model, but I think that is acknowledged and it points towards interesting (and testable) mechanisms changing P dynamics with future warming expectations. Obviously there are many other things to consider (e.g., changes in animal behaviors altering the decomposition of leaves, shifts in phenology matching shifts in climate, changes in N dynamics and broader stoichiometric questions…) but still an interesting exercise.

Thank you, we agree that our approach is simplified in terms of parts of the ecosystem that we model. We appreciate that other studies could add additional components such as animal behavior, phenology, and broader stoichiometric issues.

Line 256: Again, I wonder how much temperature is correlated with canopy cover in some of these whole system nutrient spiraling studies. I also think it's difficult to compare a scint vial's worth of Usrp to a full stream nutrient release. The authors have quantified the effect of temperature on leaf respiration and leaf-based Usrp. They did not measure anything about other components of the ecosystem that could/would change with temperature (e.g., sediment uptake dynamics, hyporheic processes, autotrophic uptake (which would increase with decreasing temps due to autocorrelation with canopy cover). While I think the simulation study was a valuable

exercise, I don't think that these results can really be extrapolated and compared to reach-scale results/studies.

We plan to add acknowledgement of the potential role for autotrophic nutrient uptake (and other processes) in modifying the trends we observed to this section of the manuscript. While we don't contend that the rates we measure are equivalent to rates measured at the whole-stream scale, we do think it is useful to compare the uptake measurements we made in the laboratory with uptake measurements made at the stream-reach scale.

Lines 262 – 263: This statement is unfounded. I disagree that the current study separated the contribution of physiological and biomass-mediated effects of temperature on ecosystem-level nutrient uptake. As mentioned in the previous comment, the study separated the contribution of these temperature-based mechanisms to affect leaf litter respiration and Usrp. Even in the most detritally-driven ecosystems (of which, Coweeta stream are definitely up near the top), there are still a multitude of other autotrohic and heterotrophic compartments contributing to ecosystem-scale respiration and nutrient uptake. This section should be either deleted or modified to be more accurate for what the study actually did do (which is still a valiant effort!).

We appreciate this point and would like to change how we refer to the results of the simulation throughout the paper. As we mentioned above, we would like to change "ecosystem-level nutrient uptake" to "long-term cumulative uptake". In that narrower context we believe that we do separate these direct and indirect effects.

Line 265: But this ignores potential increases in sediment-based Usrp. Or autotrophic. Or hyporheric. Maybe the leaf litter breakdown is fueling more labile DOM to reach interstitial spaces where sorbed P can be broken down.

We agree that our model does not include these factors. In revising this manuscript, we plan to define the results of our simulations more narrowly, which should address these concerns.

Line 295: These are great caveats to include. I don't know how the authors can make the bold claims such as in lines 262 – 263 and then simultaneously acknowledge all of these issues.

In revising this manuscript, we would like to describe the results of our simulations using language that is more specific, as we have described above.

Line 305:Not sure how this study addresses/provides insight into the effect of observational scale on temperature sensitivity.

The comparison we are looking at drawing here is between the measured instantaneous effects and the modeled long-term effects. In revising this manuscript, we will change this from "observational scale" to "time scale".

Line 305: The majority (entirety?) of the discussion focuses on the simulation experiment, which is the weakest part of this paper in my opinion. I think more general discussion of temperature dependence of biogeochemical processes and how that can affect things more broadly would

worth including initially, and then a toned down version of the focus on the simulation model. E.g. 'The results of our lab incubations would theoretically imply xyz. Our simulation studies confirm some of these expectations but revealed somewhat contradictory patterns due to abc.'

We note that this reviewer has the perception that the discussion focused too much on the simulation model in this version of the manuscript. We will try to seek a better balance when revising the manuscript.

Figure 1: I know the stats and fits are included in table 1, but I still think they'd be good to include on the figures, personally. I recommend including the equations for each line as well as stats (r-2, p-value).

When revising the manuscript, we can add this information to the figure caption.

Figure 2a: Are each of these significant regressions in panel a? Doesn't seem like it, which would entail no relationship between SRP uptake and temperature at certain initial values. A table (supplemental?) with slopes, r-squared, etc supporting these model fits would be useful. I'm not sure what in table 1 is showing model fit for these lines. Maybe a more details statistical analysis section in the methods would help clarify things a bit.

In a revised version of the manuscript, we can add a table of the model fits in equation 2a and 3a to the appendix, and revise the figure captions.

Figure 2b: It almost looks like this is a hump-shaped relationship maxing out at mid-concentration. Inhibitory effect of high P? Any reason this particular form of curve was used?

If there was an inhibitory effect of high P, then we would expect lower rates of P uptake at all temperatures in the highest P concentration. We don't really see this in Figure 2a, where we visualize the rates directly. In Figure 2b we see that the response of P uptake to temperature does peak at intermediate concentrations, but we don't think that inhibition at high P can explain this. We used this fit because saturating functions are often used to describe the effect of increasing concentration on ecological processes.

Figure 3: Now by the third figure, I'm really having difficulty connecting individual panels from figures 1 – 3 to stats and model fits from table 1. Make it easy for me (and the other readers) by putting this info on the individual panels.

When revising the manuscript, we can add this information to the figure caption.