AGU Advances

Authors’ Response to Peer Review Comments on

Observational constraints on the response of high-latitude northern forests to warming

Junjie Liu\textsuperscript{1,2}, Paul O. Wennberg\textsuperscript{2}, Nick Parazoo\textsuperscript{1}, Yi Yin\textsuperscript{2}, Christian Frankenberg\textsuperscript{2,1}

\textsuperscript{1} Jet Propulsion Laboratory, Caltech, United States
\textsuperscript{2} Caltech, United States

Authors’ Response to Peer Review Comments on Original Version of Manuscript (2020AV000228)

See next page.
The reviewers’ comments are in italics, and our responses are in blue.

Reviewer 1

Using spatial pattern of GPP inferred from solar-induced chlorophyll Fluorescence in combination with net ecosystem exchange (NEE) inferred from column CO2 observations made by OCO-2, the authors found three quarters of the spatial variations in GPP and in the fPAR by the high-latitude northern forest can be explained by growing season mean temperature (GSMT). The authors further substitute space for time, estimating the historical warming trend in GSMT leads to 20% increase in GPP and seasonal cycle of NEE by ~20%. Overall, the manuscript is well written and easy to follow. Should the authors address my concern in a revision, I could recommend publishing the manuscript.

We thank the reviewer for their constructive comments and suggestions.

In the Abstract, the CO2 seasonal cycle amplitude has almost doubled only in some aircraft sampling (700 mb) around 70°N, the Barrow data and 500 mb aircraft sampling show 50% or even less increment during the same period (Graven et al., 2013). Thus, it is misleading to simply claim CO2 seasonal cycle amplitude has almost doubled.

We agree and have rewritten the sentence as: “During the same time period, the CO2 seasonal cycle amplitude (SCA) has increased by ~50% or more.”

In the Plain Language Summary, the authors suggest that their results imply future warming would lead to further woody encroachment and forest transition towards deciduous trees. I do not find evidence from the analyses to support these claims, please revise.

In section 3.1, we have the following description: “The mean GS temperatures are 9.2°C, 11.5°C, and 13.6°C for shrubland, needleleaf forest, and deciduous boreal forest, respectively. Between the transitional zones, mixed forest types co-exist. The correlation is illustrative of the complex role that temperature likely plays on many aspects of the ecology of the HLNF. From this correlation, we might anticipate that warming would lead to succession of the dominant trees in a fashion consistent with the patterns observed spatially (e.g. evergreen forests will transition to deciduous forests as the GSMT reaches 13°C). Limited field measurement are consistent with this hypothesis (Kharuk et al., 2007; Bjorkman et al., 2018).”

We draw the conclusion of species succession from the relationship between species distribution and mean temperature. The conclusion is also consistent with paleo studies (e.g., Edwards et al., 2005).


The analyses were performed with a very coarse resolution (4° x 5°), which is hard for me to understand. The authors claimed that "transport model has smaller transport errors over high latitudes at 2° x 2.5° resolution than at 4° x 5° resolution." But why the CMS flux with 4° x 5° resolution, when 1° x 1° data can be available? The robustness of
the spatial analyses should be tested with varying spatial resolutions, at least for high resolution satellite data, to confirm the findings are not coincident with the spatial resolution selected.

The spatial resolution of top-down flux estimates of net ecosystem exchange (NEE) is limited by computation speed and spatiotemporal density of assimilated CO₂ observations. With the current top-down inversion framework, it needs ~one-month wall clock time to calculate three years of NEE. Increasing the resolution from 4° x 5° to 1°x1° would require ~20 times of current wall clock time.

More importantly, the CO₂ observations from current observing system is not dense enough to constrain fluxes at 1°x1° spatial resolution. Figure R1 shows the total number of monthly OCO-2 column CO₂ observations at each 1°x1° in June (a) and July (b) 2015. There is still vast space without observations in any single month at 1°x1° resolution. Any current available 1° x 1° flux products from top-down flux inversions strongly depend on the prior assumptions in the flux inversion.

Figure R1 The total number of monthly OCO-2 column CO₂ (X_CO2) observations at each 1°x1° in June (a) and July (b) 2015.

With high-resolution fPAR data and site-level data from FLUXNET 2015 dataset, we find that the conclusion does not depend on resolution. We added the following discussion in section 3.2 to address the dependency of the results on spatial resolution: “We use fPAR – T spatial relationship and its temporal prediction to test whether the conclusion depends on spatial resolution. The fPAR at 1° x 1° resolution has almost the same spatial sensitivity to temperature as at 4° x 5° resolution at seasonal time scale (0.050 vs. 0.054, 0.074 vs. 0.079, 0.056 vs. 0.055) (S16). As a result, the predicted fPAR trend is essentially identical to that at 4° x 5° resolution (0.021 ± 0.07%/year vs. 0.020 ± 0.07%/year) (Figure S17). This test indicates our method is robust at the spatial scale of hundred kilometers. An additional test with site-level data from FLUXNET dataset (Figure S14 and S15), which has spatial representativeness of 1-10 kilometers², further supports the robustness of the conclusion with respect to spatial resolution. Whether the method and conclusion are still valid at even smaller spatial scale needs further test with high-resolution data in the future.”
Figure S16 The fPAR – temperature spatial relationship during spring (left), summer (middle), and fall (right) based on fPAR over 2013-2016 at 1° x 1° resolution. Each point in the figure is a monthly value at a single grid cell within the respective season. The fPAR is from GIMMS 3g. The temperature is CRU 2-meter air temperature.

Figure S17 Same as Figure 3 in the main text, but based on data at 1° x 1° resolution instead of 4° x 5°.

The space-for-time substitution is in the core of the paper, while this substitution may reflect long-term response to climate change, it is also possible to be confounded with other co-varying factors, such as nitrogen deposition and moisture variability. The limitation of this approach
should be further discussed. If possible, it would be interesting to compared with tree-ring data, for example, to contest space-for-time approach.

To evaluate the space-for-time approach, we compared fPAR observations with the predicted fPAR in terms of trend and spatial anomalies in the paper (Figure 3). The consistency between observed and predicted fPAR supports the space-for-time approach. In addition, here we further compare observed and predicted GPP anomalies in 2015 and 2018 (Figure S13 a-d). The GPP observations in 2015 has been used in calculating the spatial GPP-T sensitivity used in predicting GPP, but GPP observations in 2018 were not been used in the fit. Though GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity, the observed and predicted GPP anomaly in 2018 shows similar spatial pattern and magnitude, which further support the space-for-time approach. Quantitatively, the R² is 0.93 between the monthly predicted and observed GPP between 2015-2017, and R² is 0.96 in 2018 (Figure S13 e).

We added following descriptions in section 3.2 in the main text: “Our analysis hinges upon the validity of space-for-time assumption. The comparison between fPAR observations and the corresponding hindcast provides compelling support for this assumption. In the Supplementary Material (SM.2), we further test the space-for-time assumption with GPP constrained with SIF and GPP from 11 sites of FLUXNET2015 dataset (https://fluxnet.org/data/fluxnet2015-dataset/). We compare observations and predictions in terms of seasonality and interannual variability (Figures S13 - S15). Both tests support the space-for-time assumption.”
Figure S13 Validation of space-for-time prediction of GPP. (a) Observed growing season GPP anomaly in 2015; (b) Predicted growing season GPP anomaly in 2015; (c) and (d) same as (a) and (b), but for 2018. (e) Relationship between observed and predicted monthly GPP over 50°N-75°N. The GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity used in the GPP prediction. (unit: gC/m²/day).

Performing similar analysis with tree-ring data seems to be an interesting idea, which could be done in a future analysis.

I still have conservation about calling the warming-induced increase in productivity as "thermal fertilization". The CO2 effect is considered "fertilization" because its impacts are non-negative, similar to fertilizer applications. However, the warming impacts, even in the northern high-latitudes, could also be negative due to higher vapor pressure deficit (e.g. Novick et al., 2016). This analogy is not quite suitable in my opinion.

We have changed the title to “Observational constraints on the response of high latitude northern forest to warming”
The reasoning regarding CO2 fertilization needs reconsideration. Even though the temperature relationship can well explain the interannual variations, the CO2 effects, as well as that from nutrient cycling, moisture condition may cancel out each other resulting in smaller residual magnitude. Similarly, the extrapolation to 2100 should also be drawn with cautions, as the 2-5 degree additional warming may goes beyond current norm and lead to non-linear temperature effects.

We replaced “would” with “may” in the following sentence: “Figure 2 implies that such a change may dramatically alter the distribution of tree species over the region:…”. Prior the last paragraph in section 4, we discussed other possible limiting factors on productivity with increasing temperature, such as summer drought and disturbance.

The response of GPP to temperature could be regulated by precipitation/moisture conditions (e.g. Wang et al., 2014; Reich et al., 2018). Should the authors find similar phenomenon? This is an interesting point. On large spatiotemporal scale we consider here, we did not find such phenomenon, but we do discuss potential water stress on productivity with increasing temperature in the future in Section 4.

Forkel et al. (2016) reported similar observed and simulated SCA change by one DGVM since 1980s. If considering the time period since the 1980s, will the authors still find their estimates of NEE increment much larger than most biogeochemical models? Forkel et al., 2016 used a DGVM that has been optimized by observations, which may be the reason that the observed and simulated SCA are similar. Bastos et al. (2019) analyzed SCA between 1980 and 2015 based on TRENDY v6, which still shows dominant CO2 effect over high latitude, much different from our conclusion.

Page 5: Sullivan -> Sullivan
Thanks, corrected.
Reviewer #2
Evaluations:

This paper estimates the effect of temperature on the net carbon balance of high northern latitude forests (>50 deg N) and finds that temperature effects contribute 56-72% to the observed increase in the seasonal cycle amplitude of atmospheric CO2 at high northern latitudes. The key to getting this estimate is their use of spatial variations between temperature and GPP (normalised by PAR), and the assumption that the derived spatial sensitivity can be used to model the temporal sensitivity and thus trends over the past three decades. The data underlying this sensitivity analysis is a (to me unclear mix of) sun-induced fluorescence data from OCO-2 and GPP estimates from the FLUXCOM empirical upscaling model. The contributions to seasonal CO2 amplitude changes are made by using the estimated increase in GPP and TER (total ecosystem respiration) in an atmospheric transport model and comparing simulated CO2 concentrations with observations derived from two aircraft campaigns.

This is a useful analysis, made possible by the combination of diverse types of data and the simple, yet elegant assumption of the space-for-time substitute. There are many choices and assumptions that need to be made to get from actual observations to the results presented here. Some are discussed, but many are not (see points under MAJOR below). Most importantly, no actual validation is presented and we don't know how robust and accurate the space-for-time substitute is for simulating GPP-temperature relationships.

The contextualization of results presented here is, in parts, not appropriate. It's argued here that "warming alone accounts for nearly all the observed fPAR trend is in agreement with prior observational studies that suggest CO2 fertilization is not a major contributor to carbon exchange in mature forests". However, their analysis (Fig. 4) suggests that the temperature trend explains only about half of the CO2 seasonal amplitude increase, still leaving substantial room for other effects, e.g., CO2 fertilisation.

The warming over the HLNF explains almost the total trend at the four surface sites that could be attributed to changes in HLNF. To clarify this point we added the following discussion in Section 3.3: “The surface sites may more directly reflect the changes of NEE over the HLNF, and frequent measurements make it more precise compared to aircraft observations. At the four surface sites, increase in the GSMT is calculated to contribute 56 - 72% of the observed trends in the CO2 SCA (Figure 4). The NEE over 50°N-75°N accounts for ~60% of the CO2 SCA at these surface sites. Assuming that a similar or even higher proportion (60-70%) of the trend in the observed CO2 SCA can be attributed to HLNF and the rest is due to changes in carbon cycle over low latitudes, then the increase in GSMT can explain almost the total observed CO2 SCA trend attributed to HLNF, which agrees with the dominant temperature impact on the fPAR trend discussed earlier. Since our focus is the temperature impact on the carbon exchange over the HLNF, the exact proportion of the CO2 SCA trend attributed to HLNF is out of scope of this study.”

The main result (important role of temperature for explaining the CO2 seasonal amplitude
change) itself is not a surprise, yet the estimate for effects on the CO2 seasonal amplitude presented here a useful addition to the literature. The contribution of the present paper is to translate a relatively simple estimate of temperature effects into CO2 seasonal amplitude changes. This should be better reflected in their presentation.

I'm proposing major revisions before this paper can be accepted for publication to address open points regarding validation, discussion of assumptions, and explanation of methodological choices, as listed below.

We appreciate very much the constructive comments. We respond the comments below in details.

**MAJOR**

- I don't like the term "thermal fertilization". In my understanding, 'fertilization' refers to increasing the availability of a substrate. Temperature is not a substrate. Warming rather acts on photosynthesis through the acceleration of biochemical rates and a relief of limiting effects under very low temperatures caused by photoprotection. I suggest to find a more appropriate title that clearly refers to temperature effects on the CO2 seasonal amplitude.

We have changed the title to: “Observational constraints on the responses of high-latitude northern forests to warming”.

- The results should be contextualised more appropriately, in particular with respect to novelty and consistence with earlier results. For example, the greening was attributed by Zhu et al. (2016; DOI: 10.1038/NCLIMATE3004) predominately to temperature at high northern latitudes. Keenan & Riley (2018; [https://doi.org/10.1038/s41558-018-0258-y](https://doi.org/10.1038/s41558-018-0258-y)) have attributed the greening at high northern latitudes to warming. I also don't understand the basis for the statement in the abstract saying "Most ... biogeochemical models ... generally suggest a dominant CO2 fertilization effect on changes on NEE in the HLNF (and thereby the CO2 SCA)." This is not true. Zhu et al. (2016) show an attribution of observed greening to individual drivers, suggesting that temperature is the dominant cause for high northern latitudes.

It is true that Zhu et al. (2016) showed dominant temperature effect on greening over high northern latitudes. But with the similar set of models, Piao et al., 2018 and Bastos et al., 2019 showed dominant CO2 fertilization effect on the CO2 SCA changes. To clarify this point, in section 4, we added: “While Zhu et al., (2016) show dominant climate impact on the greening trend (i.e., leaf area index) over high latitudes, later studies (Piao et al., 2018; Bastos et al. 2019) suggest a dominant CO2 fertilization effect on CO2 SCA changes with a similar set of TRENDY models.”

- Missing validation: This is, in my view, the weakest point of this analysis, but I think that the gap can be filled. Doesn't the interannual time scale provide information that can be used to validate the space-for-time assumption for GPP-temperature relationships. The paper uses spatial correlations to estimate multi-decadal changes. But couldn't the same relationship also
be used to estimate interannual variations in GPP? This may be limited by SiF data availability. Authors use OCO-2-SiF, but this is temporally and spatially sparse and covers only recent years. Alternatives may be GOME2-derived SiF or the new product by Duveiller et al., 2020. The same for fPAR: Fig. 3 shows a good correlation also at the interannual time scale. Couldn't this be presented as a validation of the space-for-time assumption? (Yes) For GPP, one could also use FLUXNET data, which may be limited in the number of sites and total number of years covered, but may still contain useful information for validating interannual GPP variations estimated from the spatial relationships.

Indeed, although we did not explicitly make this claim, Figure 3 is a validation of the space-for-time assumption. Here, we further test the space-for-time assumption with the interannual variations of SiF-constrained GPP (Figure S13) and GPP from 11 sites of FLUXNET 2015 (Figure S 14-15). Both tests support the space-for-time assumption. We added section “SM.2 Validation of space-for-time assumption with SiF-constrained GPP and GPP from FLUXNET2015 dataset” in the supplementary material.”

We added the following descriptions in SM.2: “Figure S13 a-d compare observed and predicted GPP anomalies in 2015 and 2018. The GPP observations in 2015 has been used in calculating the spatial GPP-T sensitivity used in predicting GPP, but GPP observations in 2018 have not been used. Though GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity, the observed and predicted GPP anomaly in 2018 shows similar spatial pattern and magnitude, which further support the space-for-time assumption. Quantitatively, the R^2 is 0.93 between the monthly predicted and observed GPP between 2015-2017, and R^2 is 0.96 in 2018 (Figure S13 e).

The test with FLUXNET observations further supports the space-for-time assumption, though with limited spatial representation. We select 11 sites from FLUXNET 2015 dataset (https://fluxnet.org/data/fluxnet2015-dataset/). These 11 sites are close or north of 50°N with at least eight years of data available. The northeast site among them is FI-Sod (67.4°N). Most of these sites locate in the west of Europe, limiting its spatial representation compared to SiF constrained GPP. Nonetheless, the spatial distribution of GPP is closely controlled by temperature, especially during spring and fall with R^2 equal to 0.5 and 0.8 respectively (Figure S14). The weaker relationship between GPP and temperature during summer is likely due to the narrow spatial representation and small T variability (8 degree Celsius vs. 15 degree in Figure S3) of these 11 sites. Substitute space-for-time, we predict monthly GPP at these 11 sites during growing season with the GPP-T spatial sensitivity shown in Figure S15. The prediction captures the seasonal cycle and interannual variability, though with large uncertainties in the summer months originating from smaller R^2 in the summer (Figure S15). The R^2 is 0.9 between the monthly GPP prediction and observations, and R^2 is 0.5 on interannual variability (bottom panel in Figure S14).”

We added the following in section 3.2 in the main text: “Our analysis hinges upon the validity of space-for-time assumption. The comparison between fPAR observations and the corresponding hindcast provides compelling support for this assumption. In the Supplementary Material (SM.2), we further test the space-for-time assumption with GPP constrained with SiF and GPP from 11 sites of FLUXNET2015 dataset (https://fluxnet.org/data/fluxnet2015-dataset/). We compare
observations and predictions in terms of seasonality and interannual variability (Figures S13 - S15). Both tests support the space-for-time assumption.”

Figure S13 Validation of space-for-time prediction of GPP. (a) Observed growing season GPP anomaly in 2015; (b) Predicted growing season GPP anomaly in 2015; (c) and (d) same as (a) and (b), but for 2018. (e) Relationship between observed (x-axis) and predicted monthly GPP (y-axis) over 50°N-75°N. The GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity used in the GPP prediction. (unit: gC/m²/day).
Figure S14 Seasonal GPP-temperature relationship based on 11 sites from FLUXNET 2015. From the top-to-bottom: spring, summer, and fall. Unit: gC/m²/day. These 11 sites and the corresponding latitudes are BE-Bra (51.3°N, mixed forest), BE-Vie (50.3°N, mixed forest), CA-Gro (48.2°N, mixed forest), CZ-BK1 (49.5°N, Evergreen Needleleaf Forest (ENF)), DE-Tha (51.0°N, ENF), DK-Sor (55.5°N, Deciduous Broadleaf forest (DBF)), FI-Hyy (61.8°N, ENF), FI-Sod (67.4°N, ENF), FR-Fon (48.5°N, DBF), NL-Loo (52.2°N, ENF), RU-Fyo (56.5°N, ENF)
Several key assumptions are not sufficiently discussed. For example, the space-for-time substitute assumes an immediate adjustment of vegetation structure. This is implausible and the analysis provides no insight into how big the related error is. The interannual time scale (also probably too short for vegetation structure to adapt) may yield relevant information. In the revision, we added the following discussion in section 4:

“Remarkably, the hindcast based on the spatial temperature sensitivity captures both the long-term trend and some fraction of short-term interannual variability (Figure 3 and Figure S13). We attribute this skill in part to the use of a single curve of the spatial data across multiple plant functional types and climate zones. The fit thus can capture possible species succession, which takes several decades, since the single curve assumes the same sensitivity to temperature among different forest types (Figure 2). At the same time, it also predicts with some skill the faster changes in GPP and other quantities associated with interannual temperature differences, because the single curve spans large range in temperature (Figure S3 and S6) and the temperature anomaly of any single yearly likely sits within the range. This will be especially true of springtime when the initiation of photosynthesis will be highly sensitive to small changes in temperature.”
Some methodological choices imply assumptions that may introduce systematic bias, but the paper provides no insight into how sensitive results are to these assumptions. In particular, I am left wondering whether an extension of the growing season can be captured by the methods applied. The temporal resolution of underlying data is monthly and sensitivities are derived for different seasons. Fig. S4 shows a smaller spring sensitivity than in summer months. If the start of the growing advances, a given month that was previously classified as 'spring' may change to 'summer' and a different sensitivity would apply, but the method inflexibly ascribes that month to 'spring'. Another point: The spatial resolution is 4 x 5 deg. Is this sufficient to accurately determine a spatial gradient? I may be getting this wrong, but then, this indicates that methods are not described in sufficient detail. Leading to the next point...

In the paper, we have the following discussion: “In predicting the GPP for 2009-2011 and 1958-1963 time periods, we assume that the temporal range of the three growing seasons is the same as in 2015-2017, which is a reasonable assumption since the temporal resolution of our study is month and the growing season length only increases by ~0.25 month for every ~1 degree increase of temperature (Figure S8).”

We added the following descriptions in section 3.2: “We have assumed that the temporal range of spring, summer, and fall in each grid cell is the same between 1960-2014 as in 2015-2017. It is a reasonable possibility that a given month may change from “spring” in the 1960s to “summer” in later time period, since our temporal resolution is only monthly. To test the sensitivity of the prediction to the separation of seasons, we move the ending month of spring at each grid cell one month later between 1960-1979. Figure S18 shows that the trend of predicted fPAR between 1960-2014 is somewhat higher (0.16 ± 0.03%/year), but within the uncertainty of the control prediction (0.14 ± 0.03%/year) where the same season definition is used throughout the time period. With the available of high spatiotemporal SIF observations from TROPOMI (launched in October 2017) (Köehler, et al., 2018), future studies can increase the temporal resolution and take into account the changes of growing season length in the prediction.”
We also tested the sensitivity of the method and conclusion to the spatial resolution we chose. We found that the conclusion is independent of spatial resolution on the order of hundred kilometers. We added the following descriptions in section 3.2:

“We use fPAR – T spatial relationship and its temporal prediction to test whether the conclusion depends on spatial resolution. The fPAR at 1° x 1° resolution has almost the same spatial sensitivity to temperature as at 4° x 5° resolution at seasonal time scale (0.050 vs. 0.054, 0.074 vs. 0.079, 0.056 vs. 0.055) (S16). As a result, the predicted fPAR trend is essentially identical to that at 4° x 5° resolution (0.021 ± 0.07%/year vs. 0.020 ± 0.07%/year) (Figure S17). This test indicates our method is robust at the spatial scale of hundred kilometers. An additional test with site-level data from FLUXNET dataset (Figure S14 and S15), which has spatial representativeness of 1-10 kilometers, further supports the robustness of the conclusion with respect to spatial resolution. Whether the method and conclusion are still valid at even smaller spatial scale needs further test with high-resolution data in the future.”

Figure S16 The fPAR – temperature spatial relationship during spring (left), summer (middle), and fall (right) based on fPAR over 2013-2016 at 1° x 1° resolution. The fPAR is from GIMMS 3g. The temperature is CRU 2-meter air temperature.
Figure S17 Same as Figure 3 in the main text, but based on data at 1° x 1° resolution instead of 4° x 5°.

- Methods are not clearly presented. A sometimes confusing mix of data sources is used (SIF, FLUXCOM GPP, fPAR, TRENDY [p. 9, top]) and it's not always clear why and how they are linked. For example, what is the rationale for analysing both GPP/PAR and fPAR? Such choices should be made explicit and motivated early on.

In the revision, we added a flowchart (Figure S1) to help illustrate the methods. In the text, we added: “This product is only used in defining seasons, since the product maintains essentially the same seasonality as original SIF observations during growing season, while filling the gaps of SIF observation during the late growing season and winter” in section 2.2.

In the revision, we discussed the relationship between GPP and fPAR in the introduction: “GPP is a function of fraction of photosynthetic active radiation (PAR) absorbed by plants (fPAR) (Farquhar et al., 1980), which in turn reflects forest structure (leaf area index, LAI) (Zhu et al., 2016).”
In section 3.1, we have the following discussion: “Why is GPP/PAR so sensitive to temperature in the HLNF? GPP/PAR is a product of amount of the fraction of PAR absorbed by plants (fPAR) and the light use efficiency (the efficiency of conversion of absorbed sunlight to assimilated carbon) (Farquhar et al., 1980).”

Another example for unclear methods: Is spatial sensitivity derived for monthly data pooled from all gridcells? Or aggregated over the growing season?

We clarify this point in the caption of Figure S2: “GPP/PAR vs. temperature (left column) and TER vs. temperature (right column) for spring (top), summer (middle), and fall (bottom) months. The color bars indicate latitudes. Each point represents values at a single grid cell of a specific month within the corresponding season. The uncertainty is 3σ of the exponential fitting.”

Is a constant growing season assumed per gridcell?

We clarify this point in Section 2.4: “In predicting the GPP for 2009-2011 and 1958-1963 time periods, we assume that the range of spring, summer, and fall per grid cell is the same as in 2015-2017, which is a reasonable assumption since the temporal resolution of our study is month and the growing season length only increases by ~0.25 month for every ~1 degree increase of temperature (Figure S8).”

In Fig. 4: It is unclear how data for the dashed line ("approximate observed SCA change from forest") was derived.

We clarify this in the capture of Figure 4: “Dashed black line: the approximate CO2 SCA change between HIPPO and IGY at 500 hPa attributed to the HLNF, which is the multiplication between observed CO2 SCA and the ratio between model simulated CO2 SCA forced by forest NEE over 50°N-75°N and the observed CO2 SCA (Figure S13).”

What was assumed regarding trends in GPP and TER below 50 deg N for the simulation of the CO2 seasonal amplitude?

To clarify the assumptions regarding GPP and TER below 50 °N, we rewrote the first paragraph of section 2.5: “We use GEOS-Chem transport model to simulate 3-hourly 10-member ensemble of CO2 concentrations during 1958-1963 (IGY), 2009-2011 (HIPPO) and from 1980 to 2011. The only surface boundary conditions are the predicted ensemble monthly NEE (i.e., TER-GPP) between 50°N and 75°N. Therefore, the differences in simulated CO2 are only due to the differences in the predicted monthly NEE. We assume the same diurnal biosphere fluxes during these two time periods, which are from CASA-GFED3.”

Also regarding the separation of GPP vs. fPAR changes: in Abstract, authors write: "... three quarters of the spatial variations in GPP and in the fPAR absorbed by the HLNF can be explained by the spatial variation in the growing season mean temperature". This would imply that there is no change in LUE (GPP = LUE * fPAR * PAR) [PAR should not change greatly unless there has been a trend in cloud coverage - has there?] Yet, at other points, LUE is
reported to have increased (p. 15: "... the sensitivity of GPP/PAR to GSMT results from the sensitivity of both forest structure (fPAR) and light use efficiency, with similar contributions (0.090 vs. 0.100) from each (Figure 2 b and c."). And in Discussions: "In particular, about 50% of the observed correlation between SIF and temperature is not explained by fPAR, but by the "light use efficiency". Both can't be true at the same time.

This is simply a confusion from our description. We have rewritten the abstract as: "... three quarters of the spatial variations in GPP can be explained by the spatial variation in the growing season mean temperature"

- The quality of the introduction is found wanting in several instances. For example:
  - Intro, p. 4: "Over the region northward of 50 degrees, precipitation substantially exceeds evaporation due to moisture convergence by the largescale eddy circulation, which results in a generally well-watered, thermally-limited ecosystem". The fact that these ecosystems are not water-limited is not only due to high precipitation, but also to low net radiation. Their ratio is relevant.
    We rewrote the sentence as: "Over the region northward of 50 degrees, precipitation substantially exceeds evaporation due to moisture convergence by the largescale eddy circulation and low net radiation, which results in a generally well-watered, thermally-limited ecosystem (Nemani et al., 2003)."
  
  - Intro, p. 4: "Plants generally grow faster as CO2 increases, which is called CO2 fertilization (Kimball et al., 1983; Long et al., 2004)" CO2 fertilization is used as a term to describe a positive response to CO2 of different processes, ranging from photosynthesis, growth (NPP), to the net C balance.
    We rewrote the sentence as: "Different biogeochemical processes ranging from photosynthesis, net primary production (NPP), to the net carbon balance respond positively to the CO2 increase, which is called CO2 fertilization (Kimball et al., 1983; Long et al., 2004)."

  - Intro, p. 4: "The optimal temperature for plants growth is higher than the mean annual temperature over high-latitude biomes (Huang et al., 2019)". Huang et al. quantify the temperature optima for photosynthesis, not growth.
    We rewrote the sentence: "The optimal temperature for photosynthesis is currently higher than the mean annual temperature over high-latitude biomes (Huang et al., 2019),"

  - Result conflicting a large body of work on VPD effects on GPP is left undiscussed ("As shown in Figure S10, GPP increases with the increase in vapor pressure deficit, opposite to the expected relationship for water-limited ecosystems")
    Figure S11 is used to support the argument that the high-latitude northern forest is not water limited in large spatial scale.

  In section 4, we discussed possible water limitation on the warming effect during late summer.
Reviewer #3 Evaluations:
Recommendation: Reject and encourage resubmission
Reviewer #3 (Formal Review for Authors (shown to authors)):

The manuscript "Thermal fertilization of the high latitude northern forests" by Liu and coauthors is a very interesting analysis that look at the effect of the sensitivity of GPP and TER to temperature and its effect on the amplitude of seasonal cycle of CO2. The authors use a combination of SIF and column CO2 observations from remote sensing, which is in my opinion the new interesting frontier for understanding global biogeochemical cycles. The authors used space for time substitution, by fitting relationships between GPP-T and TER-T in space and project in time. I find the article extremely interesting and congratulate with the authors for the idea and the analysis. The research question is timely and important. However, I have some comments that are related to the robustness of some assumptions, for instance in the space for time substitution approach, and the potential role of summer drought that is foreseen to increase in the higher latitudes. Please find my comments below.

Two small additional comments - please add the line numbers for the revision and please crosscheck the consistency between units and labels in the figure because often the same unit is reported in different ways (see comments below).

We thank the reviewer for their detailed comments. We have added line numbers in the revision, and made the units and labels consistent. Please see our responses below in blue.

Page 5 - I completely agree the increase in temperature will stimulate growth but another important question is to what extent the increase of temperature can be exploited considering the day length is not going to change and the forests are also light limited.
We added the following sentence in the Discussion: “Radiation may also limit the response of GPP to temperature increase in the future, especially in the fall (Zhang et al., (2020)).”

At the beginning of the introduction I would also introduce the expected effects of temperature on TER, considering the article deals also with TER.
We have added the following sentence in the introduction: “With the increase of GPP and temperature, total ecosystem respiration (TER) is also expected to increase (Davidson et al., 2006; Davidson and Janssens, 2006; Baldocchi et al., 2008).”

Page 5 - space for time would require more attention also with respect to the adaptation and also if tested for fAPAR GIMMS I would suspect GPP and TER is different..
Figure 3 is a validation of the space-for-time assumption. Here, we further test the space-for-time assumption with the interannual variations of SIF-constrained GPP (Figure S13) and GPP from 11 sites of FLUXNET 2015 (Figure S14 and Figure S15). Both tests support the space-for-time assumption. We added section SM.2 in the supplementary material.

“Figure S13 a-d compare observed and predicted GPP anomalies in 2015 and 2018. The GPP observations in 2015 has been used in calculating the spatial GPP-T sensitivity used in predicting GPP, but GPP observations in 2018 have not been used. Though GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity, the observed and predicted GPP
anomaly in 2018 shows similar spatial pattern and magnitude, which further support the space-for-time assumption. Quantitatively, the $R^2$ is 0.93 between the monthly predicted and observed GPP between 2015-2017, and $R^2$ is 0.96 in 2018 (Figure S13 e).

The test with FLUXNET observations further supports the space-for-time assumption, though with limited spatial representation. We select 11 sites from FLUXNET 2015 dataset (https://fluxnet.org/data/fluxnet2015-dataset/). These 11 sites are close or north of 50°N with at least eight years of data available. The northern most site among them is FI-Sod (67.4°N). Most of these sites locate in the west of Europe, limiting its spatial representation compared to SIF constrained GPP. Nonetheless, the spatial distribution of GPP is closely controlled by temperature, especially during spring and fall with $R^2$ equal to 0.5 and 0.8 respectively (Figure S14). The weaker relationship between GPP and temperature during summer is likely due to the narrow spatial representation and small T variability (8 degree Celsius vs. 15 degree in Figure S3) of these 11 sites. Substitute space-for-time, we predict monthly GPP at these 11 sites during growing season with the GPP-T spatial sensitivity shown in Figure S15. The prediction captures the seasonal cycle and interannual variability, though with large uncertainties in the summer months originating from smaller $R^2$ in the summer (Figure S15). The $R^2$ is 0.9 between the monthly GPP prediction and observations, and $R^2$ is 0.5 on interannual variability (bottom panel in Figure S15).

In the main text, we added:
“Our analysis hinges upon the validity of space-for-time assumption. The comparison between fPAR observations and the corresponding hindcast provides compelling support for this assumption. In the Supplementary Material (SM.2), we further test the space-for-time assumption with GPP constrained with SIF and GPP from 11 sites of FLUXNET2015 dataset (https://fluxnet.org/data/fluxnet2015-dataset/). We compare observations and predictions in terms of seasonality and interannual variability (Figures S13 - S15). Both tests support the space-for-time assumption.”
Figure S13 Validation of space-for-time prediction of GPP. (a) Observed growing season GPP anomaly in 2015; (b) Predicted growing season GPP anomaly in 2015; (c) and (d) same as (a) and (b), but for 2018. (e) Relationship between observed (x-axis) and predicted monthly GPP (y-axis) over 50°N-75°N. The GPP observations in 2018 have not been used in calculating the spatial GPP-T sensitivity used in the GPP prediction. (unit: gC/m²/day).
Figure S14 Seasonal GPP-temperature relationship based on 11 sites from FLUXNET 2015. From the top-to-bottom: spring, summer, and fall. Unit: gC/m²/day. These 11 sites and the corresponding latitudes are BE-Bra (51.3°N), BE-Vie (50.3°N), CA-Gro (48.2°N), CZ-BK1 (49.5°N), DE-Tha (51.0°N), DK-Sor (55.5°N), FI-Hyy (61.8°N), FI-Sod (67.4°N), FR-Fon (48.5°N), NL-Loo (52.2°N), RU-Fyo (56.5°N).
Controlling also for spatial variability of CO2 could be important - it is clear this is not as strong as the temporal but still is there a spatial gradient that can confound the relationship with T. A simple partial correlation analysis or variance partitioning can possibly help. Although in the northern latitude this problem could be minimized, at the fringe between more temperate and boreal areas the spatial variability in atmospheric CO2 concentration is not negligible and might confound the sensitivity to temperature. It could be this is tested already in the literature I am not aware of; in this case a citation would be enough. The authors refer to this at page 9 but looking at maps of atmospheric CO2 concentration seems to me that the differences are way larger than 10 ppm.

To address this comment, we evaluated the spatial relationship between seasonal mean surface CO2 from top-down flux inversion and normalized GPP in Figure R2. For all these three growing seasons, the CO2 spatial gradient is about 10 ppm, with maximum gradient ~15 ppm. However,
the relationship between seasonal mean surface CO\textsubscript{2} and normalized GPP on seasonal scale is weak, which implies that GPP spatial gradient is not due to CO\textsubscript{2} spatial variability.

Figure R2 GPP/PAR vs. surface CO\textsubscript{2} (left column) and TER vs. above ground biomass (AGB) (right column) for spring (top), summer (middle), and fall (bottom) months. The color bars indicate latitudes. The uncertainty is 3\sigma of the exponential fitting.

Page 6 last line - I would say that SIF is a linear predictor under high light conditions and add at seasonal temporal scale to be more rigorous. Under high light conditions but shorter time scale SIF alone might not be sufficient

We changed the sentence to: “It is sensitive to both plant structure and photosynthetic function, and is linearly correlated with GPP under high-light conditions at seasonal temporal scale (Frankenberg et al., 2011; Sun et al., 2018).”

Page 7 - the most updated and comprehensive article on FLUXCOM GPP is Jung et al., 2020

Thanks for this suggestion, we added the reference in the paper.

Fig S1 panel a - I suggest to include SIF in the y axis label and not only the unit
Fig S1 - I suggest to indicate clearly the meaning of the color bar (I guess is the latitude)
We made the suggested changes in the Figure. Now it is Figure S2.

Fig S1 Can one actually conclude that the T sensitivity is the same between SIF-T and GPP-T considering such a difference in the units? Can the authors test this by rescaling the Y axis to a common range, fitting the curve applying a bootstrap, and estimating the posterior distribution of the parameters (in particular for the coefficient of T). I guess will not change too much but I think is worth to try and to have a statistical confirmation about the statement. The discussion on the scaling factors GPP-SIF goes in this direction, but then I would say is even more important to either compute GPP from SIF and then calculate the sensitivity or try to make comparable panel a with b, c, d.
We added the second Yaxis representing the corresponding GPP in Figure S2 (original Figure S1). The right y-axis has the same range as the FLUXCOM GPP.

As suggested, we applied a bootstrap method to sample only 80% samples for each GPP-T sensitivity calculation. With 1000 times calculation, the mean sensitivity is 0.21 and standard deviation is 0.02, consistent with Figure S2 (original Figure S1).

Page 8 The dependence on TER to temperature can be related indeed to photosynthesis as the authors suggested (and this is quite clear from the curvature of TER-T relationship that increase from Spring to summer and decline in autumn). However, in particular for the spatial patterns (i.e. within the same season) there are other factors that can be more important and are related to standing biomass (and the latitudinal gradient is pretty clear in this sense) but also soil organic carbon and sensitivity of microbial activity to temperature. Photosynthesis is important but not the only driver of the relationship. I suggest to add few considerations in this direction and one article to cite is Davidson and Jannssen 2004 for instance (https://www.nature.com/articles/nature04514), or even better (https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2486.2005.01065.x).
The interaction aboveground biomass, soil organic carbon, and temperature can confound TER-T and I suggest trying to control for these controlling factors using aboveground biomass products and SOC from soil grid map. Otherwise the risk is that T sensitivity can be confounded and perhaps amplified.

Thanks for this suggestion. The available aboveground biomass and soil organic carbon data are not time dependent, while we examine the relationship between TER and temperature on seasonal time scale. Figure R2 shows that the spatial relationship between TER and aboveground biomass on seasonal time scale is negligible. In section 2.2, we added: “The temperature dependence of TER is not related to the amount of above ground biomass, which has a weak relationship with TER on seasonal timescale (not shown), with R² less than 0.07.”

We further added discussions on the impact of possible confounding factors on TER: “The spatial covariation between TER and temperature may include some confounding factors, such as the
available soil organic carbon and sensitivity of microbial activity to temperature (e.g., Davidson et al., 2006; Davidson and Janssens, 2006, Nottingham et al., 2019). Here, TER – temperature spatial sensitivity is a statistical relationship of the net impact of temperature on TER. A detailed process understanding on factor controls of TER still has large uncertainties and requires more research.”

Fig S2 - I suggest to include the label in the y axis and not only the unit
We added the label in the y axis as suggested.

Page 9 - when discussing the space for time substitution please cite also the recent discussion (on semi arid areas but some of the argument can be brought here). I like a lot the idea of the authors to look at the fAPAR in space and time, this is solid and give a lot of confidence on GPP, but still for TER I would be careful and I would try to analyze if there are other drivers of the relationship in space (that can have then an impact when used in time).

See our response above on the impact of possible confounding factors on TER.

Page 10 - another factor to include in the story is the potential summer drought due to increase in vegetation activity earlier in the season. This can both affect GPP and TER. This has been shown clearly for the high latitude ecosystem. I suggest the authors to add sentence that this can happen and discuss this point. See Buermann et al., 2018 Nature (https://www.nature.com/articles/s41586-018-0555-7).

In section 4, we added possible factors that could limit the response of plants to warming in the future: “The rapid increase in GPP with GSMT is certainly limited: the data shown in Figure 2 extends only to a GSMT of 16°C. It is expected that at some point, photosynthetic, water, and other nutrient limitations will limit further increase in GPP with temperature, especially during summer. Even today there is evidence that by mid/late summer such limitations come into play (e.g., Buermann et al., 2018), and browning trend has been observed over limited regions (Beck and Goetz, 2011) Radiation may also limit the response of GPP to temperature increase in the future, especially in the fall (Zhang et al., (2020)). In addition, the increase in abiotic (e.g. fire) and biotic (e.g. herbivory) disturbance with climate will also impact productivity (Sulla-Menashe et al., 2018)”

Figure 1 and throughout the manuscript - Here the units are in KgC/m2 while in other figure KgC m-2. I suggest to keep consistency. Sometimes there is no variable name in the y axis (only the unit) sometime both please
We have made the suggested changes in Figure 1.

Page 14 the dependence on LUE by PFT it is probably related to nitrogen content in leaves (for instance deciduous trees have larger N content per area than conifer). Perhaps something to add.
We revised the first paragraph in section 4 to reflect this point: “The estimated change in light use efficiency is correlated with biome (evergreen vs. shrub vs. deciduous trees, Figure 2c) and further work is needed to test whether all the differences in SIF of these biomes is indeed related to the differing photosynthetic efficiency of these tree types, or may (partially) reflect differences in
**N content**, or the optics of SIF within canopies differing in structure. Though the assumption that N fixation increases with warming to meet growth demands is qualitatively supported by literature (e.g., Schimel et al., 1997), more observations are needed to test such hypothesis.

*Figure S10 c - The spatial relationship GPP-VPD is driven by temperature. Also I would expect that variance of mean VPD in space is less than in time. My suggestion would be to remove this part. In my opinion is questionable that this relationship can stand in time.*

We agree that VPD is related to temperature, but it is a variable that has been broadly used to indicate the level of water stress on plants. Figure S11 (original Figure S10) is used to support the argument that the productivity is not water stressed on large spatial scale over the HLNF. We do not intend to imply that the relationship can stand in time.

*The section on the effects of the "thermal fertilization" on the amplitude of the seasonal cycle should be rediscussed in my opinion when the points above are analyzed.*
Authors’ Response to Peer Review Comments on First Revision of Manuscript (2020AV000228R)

See next page.
Dear Dr. Davidson,

We appreciate very much of your time and comments. We believe that have adequately addressed the reviewers and your comments. Please consider the paper titled: "Observational constraints on the response of high-latitude northern forests to warming" for publication in AGU-Advances.

All the best,

Junjie Liu and Paul Wennberg

09Oct, 2020

Dear Dr. Liu:

Thank you for submitting your manuscript entitled "Thermal fertilization of the high-latitude northern forests" [Paper #2020AV000228R] to AGU Advances. I have now received 2 reviews of your manuscript, which are included below and/or attached.

The revised manuscript has been reviewed by two of the original reviewers, who find that their concerns have been adequately address. I commend the authors for their thorough revision. Reviewer #3 lists a few minor edits needed with respect to figure legends and units. I found that the term "thermal fertilization" is still found on line 530-531 and in Figure 4. Because it is no longer defined, this could cause confusion for some readers, so it probably should be removed as it has been elsewhere.

We modified line 530-531 and Figure 4.

One of my original concerns has not been addressed regarding the implications of this work vis-a-vis changes in the net annual C sink. I worry that some readers may misinterpret the discussion of increasing growing season GPP and greenness to infer that the C sink has necessarily increased commensurately. I presume that the authors have chosen not to calculate an annual change in NEE from the data in the righthand panel of Fig S7 because that would have its own
set of uncertainties that are beyond the scope of the manuscript. Nevertheless, there should be a clear, brief disclaimer that a change in the annual net sink is not addressed and does not necessarily follow from the change in growing season GPP alone.

In the revision, we did calculate the net changes in NEE between 1960s and 2010s, which is $0.25 \pm 0.17$ GtC. The net NEE changes are uncertain, and we do not know how to adequately quantify that uncertainty. So in section 3.3, we added “The net carbon sink increased by $0.25 \pm 0.17$ GtC between 1960 and 2010. The estimate of net carbon sink has, however, large uncertainties that require further study to evaluate.”

Based on the review comments, your manuscript may be suitable for publication after minor revisions.

Reviewer #3 Evaluations:
Recommendation: Return to author for minor revisions
Significant: Yes, the paper is a significant contribution and worthy of prompt publication.
Supported: Please Select
Referencing: Yes
Quality: The organization of the manuscript and presentation of the data and results need some improvement.
Data: Yes
Accurate Key Points: Yes

We appreciate very much your time and thoughtful comments.
Reviewer #3 (Formal Review for Authors (shown to authors)):
I went through the answer of the authors. Indeed the authors did a good job.

My two main concerns were on the fundamental assumptions that certain relationship in space can hold in time - and can therefore be used to speculate on the effect of thermal fertilization (term that they removed and I agree) on the biospheric response. The authors did a good job in answering these questions and they proved in a solid way, in particular for the CO2 effect, that this is not the case and that the study is solid. Also I asked a number of editorial improvements and to add some citations that they introduced and discussed.

Despite scientifically the manuscript is mature the presentation is still not 100% satisfying. The figures can be still neater though, many Units are again missing or do not follow the scientific standards and the presentation could be better. This is of course a minor issue but I invite the authors to improve the presentation:
Figure 1 a, the unit of Temperature should not be celcius
We prefer to retain use of Celcius. While we acknowledge that Kelvin is the SI unit for temperature, Celsius is broadly used in scientific literature.

Figure 4 right axis - label is missing
We added label on the right y-axis.

Figure 2 unit missing on the
We added unit on x-axis of Figure 2.

Same in the supplements:
Fig S9, in my version the maps are overlaid with the name of the months
It is our intention to have the names of the month on each subplot, so it is easier for readers to link the figure with time.

Figure S16 Units of Temperature missing
We added the missing units of temperature on Figure S16.

Fig S19 label missing in the y axis (added)
We added label on y-axis as appropriate Figure S19.