RESPONSE TO REVIEWERS' COMMENTS

Dear Prof. Tomomichi Kato,

We would like to express our great appreciation to you and the reviewers for handling and reviewing our manuscript entitled "FORCCHN V2.0: An individual tree-based model for predicting multiscale forest carbon dynamics" (ID: EGUSPHERE-2022-49).

Those comments were very helpful for revising and improving our manuscript. We have studied the comments carefully and have made corrections which we hope meet with approval. A revised manuscript is being submitted for your consideration. Please see below for the point-to-point response. We have also highlighted the changes we have made in blue text (not including changes to the references) in our main manuscript (i.e. the track-changes file).

We are looking forward to your further decision.

Yours Sincerely,

Jing Fang, on behalf of all authors

Referee 1

In this manuscript, the authors developed an individual tree-based carbon model, FORCCHN2 by using the NSC pools to couple tree growth and phenology. It was tested that the model performed well in reducing uncertainty in predicting forest carbon fluxes. They described the framework in details and provide the source code of the model. The coding system is complete, and both research and development foundations were very solid. Besides, the model is very convenient to be called with other computer tools. I noticed that the results of application of the model in predicting carbon dynamics in the Northern Hemisphere was very ideal. I suggest accept the manuscript for publication after minor revisions. I have the following minor points need to be addressed by the authors.

Response: We sincerely thank the reviewer for your valuable feedback, and we think the reviewer's comments are very important to our work. The responses to the reviewer's concerns have been included in each comment. Our replies are given in blue text.

1. Section 2, for description of FORCCHN2, you just told phenology, you may mean phenophase, was controlled by heat and chilling. You'd better tell what model was used to decide phenophase by climatic variables. Beside, in phenology, we often say heat and chilling requirements.

Response: We agreed with the comments from the reviewer. Here, we revised the 'phenology' to 'phenophase' in lines 66 and 75. We also added the description of the climatic variables in the calculations of phenophase: 'The spring phenophase is decided by the effective temperature with Thermal Time model (Eqn 39-40), and the autumn phenophase is decided by the effective temperature and photoperiod with Cold Degree-Day model (Eqn 41-42).', in lines 76-79. According to the reviewer's comment, we changed the sentence to: 'The phenophase of spring and autumn in FORCCHN2 is controlled by heat and chilling requirements, respectively.', in lines 75-76.

2. Line 75-76: does phenology here mean spring phenology? As it is difficult for the calculation of chilling requirements for autumn phenology.

Response: In this work, the phenology meant the spring and autumn phenology. We thought the autumn phenology was more difficult to calculate than the spring phenology. We used the effective temperature and photoperiod with the Cold Degree-Day model to calculate the autumn phenology. The corresponding calculation and parametrization of this phenology had been tested in a newest paper, Fang et al. (2022). We had added the reference in this sentence.

Reference for the reviewer

Fang, J., J. A. Lutz,, H. H. Shugart,, Wang L., Liu F., and Yan X. 2022. Continental-scale parameterization and prediction of leaf phenology for the North American forests. Global Ecology and Biogeography, 00:1–13. https://doi.org/10.1111/geb.13533

Referee 2

This manuscript by Fang et al. describes the phenology and growth processes for the forest carbon model FORCCHN2, and they evaluate the model at 78 forest sites in the Northern Hemisphere. The paper presents a model that addresses carbon-cycle science questions relevant to the scope of EQU, and the model methods are presented in a reproducible manner. In addition to providing the source code and detailed description for the model, the authors have implemented a module to allow for a seamless integration of the model into a variety of software languages, allowing for user to run model predictions more conveniently with high efficiency.

Response: We thank the reviewer for the helpful comments. For the main concern of the reviewer, we make some explanations and revisions in both the responses and the manuscript. Detailed explanations on the concerns can be found in the following items as responses to each concern. Our replies are given in blue text.

General Comments

I feel that this manuscript warrants publication based on its reproducibility and presentation quality; however, I have some reservations regarding the significance and quality that I feel need to be addressed.

1) The substantial contribution to modelling science needs to be clarified, as the authors themselves state that the methods are compiled from previous versions with the only new methods being a software module rather than scientific concepts or ideas.

Response: Yes, we agreed that scientific concepts or ideas were the important parts. This paper was submitted to the section of 'Model description papers' in the journal of *Geoscientific Model Development* (*EGU* was the platform of preprints). For this section, the Aim and Scope were: 'Model description papers are comprehensive descriptions of numerical models which fall within the scope of

GMD. The papers should be detailed, complete, rigorous, and accessible to a wide community of geoscientists. In addition to complete models, this type of paper may also describe model components and modules, as well as frameworks and utility tools used to build practical modelling systems, such as coupling frameworks or other software toolboxes with a geoscientific application.', and 'The main paper should describe both the underlying scientific basis and purpose of the model and overview the numerical solutions employed. The scientific goal is reproducibility: ideally, the description should be sufficiently detailed to in principle allow for the re-implementation of the model by others, so all technical details which could substantially affect the numerical output should be described.' (see the manuscript types and the journal editorial of *GMD* in https://gmd.copernicus.org/). Therefore, we emphasized the reproducibility and implementation of the model in this paper. Besides, this was the first time to apply the FORCCHN2 model in the Northern Hemisphere and released our source code publicly. We hope that our model could be easily accessible to the wide community of scientists and thus they could use this model to evaluate the growth and carbon cycle of forests.

2) They state they apply the model on a global scale; however, they only evaluate the model at 78 Northern Hemisphere sites, which it misleading. While they present maps of outputs across the Northern Hemisphere, none of these large-scale outputs are evaluated against other satellite, modeled or derived products (i.e. fluxes from FluxCom or satellite-derived biomass). Additionally, several of the methods use the hard-wired date of January 1 for exchanges, which is likely not suitable for global use, particularly in the Southern Hemisphere.

Response: Because of the different phenology in the Northern and Southern Hemispheres, we applied this model in the Northern Hemisphere. The forests in the Northern Hemisphere were the important areas in forest studies. Therefore, we focused on the Northern Hemisphere forests and we revised the 'global scale' to 'hemispheric scale' in the whole of the manuscript. For the evaluation of large-scale outputs, we collected the carbon fluxes from the FluxCom dataset and the

aboveground biomass from the GLASS product (a satellite-derived product). We compared the predicted and observed results in Fig. S2 and Fig. S3. Some sentences had added to describe the comparison: 'As the comparisons, we use the aboveground biomass (AGB) from the GLASS product (a satellite-derived product, http://www.glass.umd.edu/Download.html) and the carbon fluxes from the FluxCom dataset (https://www.bgc-jena.mpg.de/geodb/projects/Data.php) to test our predictions (Fig. S2 and Fig. S3). Both predictions and GLASS observations present the tropical forests own the highest AGB and the boreal forests had the smallest AGB (Fig. S2). In terms of carbon fluxes (i.e. GPP, ER, and NEP), the resulting spatial pattern is consistent with the FluxCom dataset (Fig. S3). However, the GPP and ER derived from FORCCHN2 for some boreal forests are approximately 0.5 kg C m-2 year-1 smaller and for parts of eastern North America are approximately 0.5 kg C m-2 year-1 larger than those of FluxCom GPP and ER, respectively. Compared to the FluxCom NEP, the model overestimates NEP in some tropical forests and underestimates NEP in some boreal forests.', in lines 204-213.

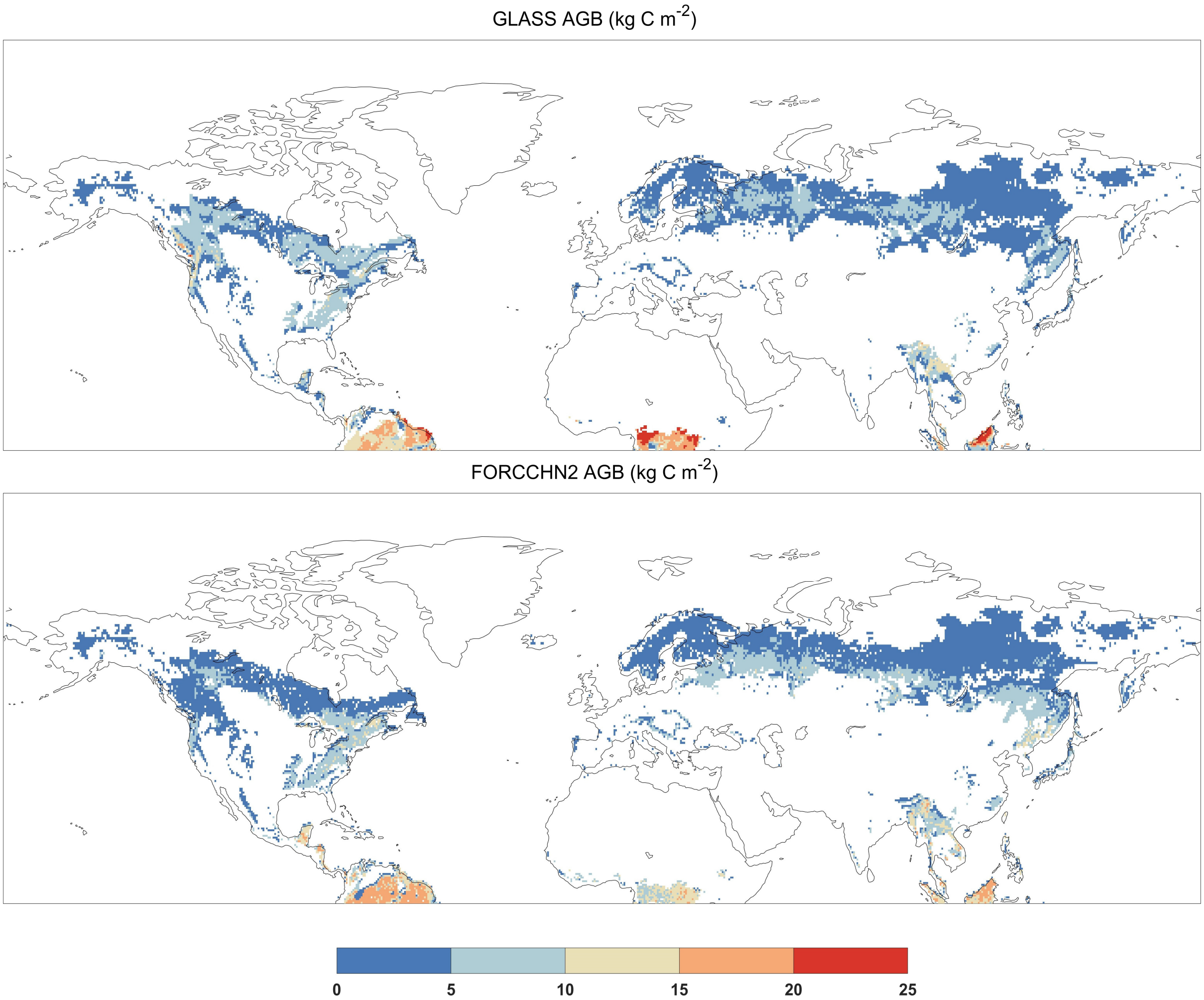

Fig. S2. The FORCCHN2-simulated and satellite-derived aboveground biomass (AGB) across the Northern Hemisphere. The satellite AGB are extracted from the GLASS product.

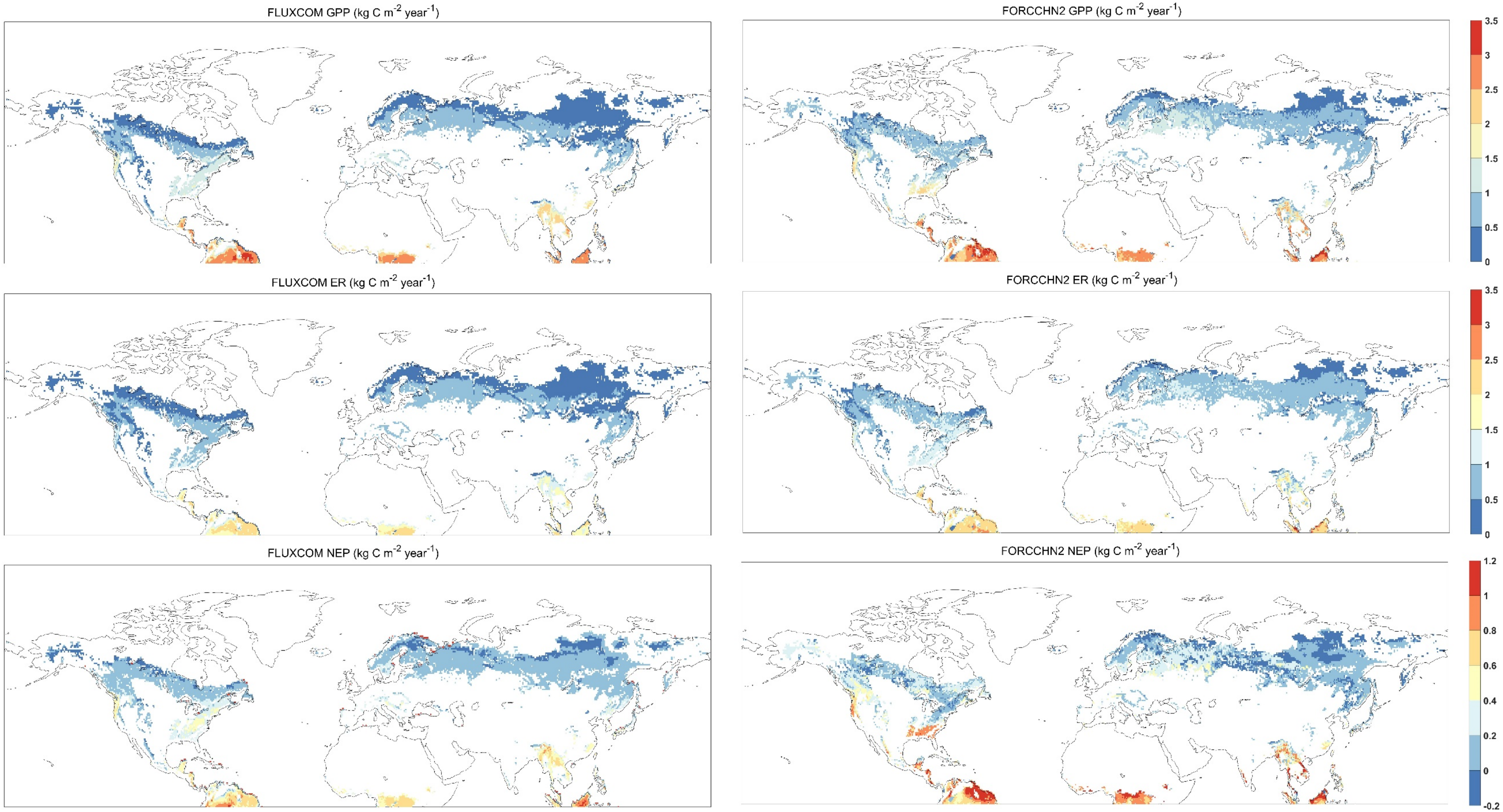

Fig. S3. The mean spatial distribution of FORCCHN2 and FLUXCOM gross primary productivity (GPP), ecosystem respiration (ER), and net ecosystem productivity (NEP) across the Northern Hemisphere during 1980–2013. The fluxes of FLUXCOM are extracted from the 'RS+METEO' dataset. The FLUXCOM NEP is equaled to the negative of net ecosystem CO2 exchange (NEE). The spatial resolution is 0.5°×0.5°.

Detailed Comments

1) I find calling the model a "tree-based" model to be misleading. First off, with the recent upsurge in literature on machine learning models, my first assumption was that this was a model using machine learning techniques from data, rather than a physical-based ecological terrestrial model. Second, individual trees are not actually being represented. While they do calculate tree height, diameter at breast height, and biomass as well as gap fraction between trees, these are all empirically calculated from LAI using similar methods to other terrestrial models. According to their documentation, it does not appear that each tree is allowed to grow separately, such as in a dynamical model, or that the model is even self-consistent between carbon uptake and growth, but instead is prescribed growth following plant functional type (PFT) equations. Additionally, the authors state that they use PFTs; however, their input data for forest type is a biome map rather than a PFT map. Given that input, it does not seem appropriate to call the model a "tree-based" model when it uses forest biomes such as mixed forest that incorporate not only a wide variety of species but also a variety of forest functional types (such as both deciduous and evergreen) to represent the forests being simulated.

Response: Thank you for your comments, the 'tree-based' may not be suitable for this paper. According to the description from Shugart et al. (2018), we revised the title to: 'An individual-based model for predicting multiscale forest carbon dynamics.'.

For all calculation processes, both this version and the previous version of our model were based on the individual trees. We had two input methods: one was to use the inventory data (i.e. DBH and tree species of every tree in one given plot) as the direct input, and we used this method when we had enough inventory data (this method could be found in Fang et al. 2021); another method was to use the satellite data (e.g. LAI) as the input, and the information of individual trees was evaluated by LAI. We realized that it was difficult

to obtain the data of individual trees in the Northern Hemisphere and thus used the satellite data in this work.

For the vegetation types, we followed the previous version to set them (see Ma et al. 2017). Each tree was given a PFT when we input the data. The PFT had included in Table S2. We can input the inventory data directly (each species belonged to one PFT), or we used the forest types from the satellite as the input and then used the random method (i.e. random function with Fortran language) to generate the possible PFT in the large scales. This initialization method of the individual tree had been successfully used in large-scale studies (Ma et al. 2017; Fang et al. 2020). Here, we added a sentence to describe it: 'The PFT of one tree is decided by tree species when using the inventory data or it is estimated by forest types and random function when using the satellite data.', in lines 97-98.

2) In the abstract (line 12) they state that the model can predict yearly phenology, but I'm not sure what that means. Phenology is the seasonal changes in vegetation, so do they mean that the model is capably of predicting inter-annual variability in phenology? This needs to be clarified.

Response: We thought this sentence would make a misunderstanding for the reviewer and the readers. The phenology meant the spring and autumn phenology dates. We changed the words 'yearly phenology' to 'spring and autumn phenological dates', in lines 12.

3) It is not clear to me how GPP is incorporated into the growth. The equations for leaf and fine roots growth in the supplemental material do not include any carbon from GPP, does that carbon get allocated out to the carbon pools to make the model self-consistent? If so, how is the carbon allocated? If not, then are the processes really separate and the carbon is not conserved in the model?

Response: In our model, the relationship between GPP and growth was indirect. We did not use the GPP as the basis of carbon allocation. The GPP first entered the NSC active pool and then the carbon was allocated by this pool (see Fig. 1 and Eqn 1). The allocated

carbon of growth was determined by the external environment and growth processes (see Eqn S30 and S31). Carbon was completely conserved in the model because this model was run with two carbon pools: one was the active NSC pool (daily) and another was the slow NSC pool (yearly). The slow pool was an NSC storage pool providing the necessary carbon for requirements when the contemporaneous active pool was insufficient.

4) Why are tree height and DBH only updated annually? Given that the model runs on a daily timestep, why can't all the pools and processes be updated daily? Additionally, where does the change in basal diameter and tree height increments come from and how are these then related to daily growth? Also, how well do these increments match up with annual GPP and what is the allocation of the GPP to these pools?

Response: The annual update of DBH was an assumption of our model. The assumption also was a common assumption in the dynamic models, such as the LPJ model (Lund-Potsdam-Jena model; Sitch et al., 2003), which calculated leaves growth and carbon flux at the daily scale and calculated wood growth at the annual scale. The model would be complex and the parameters would be too many if we calculated the wood growth at the daily scale (see Fang et al., 2020). To maintain sufficient efficiency and simple representation, we only updated the DBH by annual scale. However, we would keep developing this model if we had enough mechanisms and wood growth data on the daily scale. For the GPP, please see the response to Detail Comments 3.

5) Why does tree death only occur annually? If the NSC pools are updated daily, can't they become insufficient at any point in time? How realistic is it to have the tree mortality occur on 31st December and what are the impacts on the carbon cycle, especially given the claim that this model could be run globally and include forests that would be in the middle of their growing season? In section S1 line 179 states "photosynthate has been allocated to the growth of canopy height and basal diameter", how is this allocated? And how does this tie back to daily growth?

Response: Same as the LPJ model, the tree death also was an assumption of our model. Tree death was a complex process in the study of forests. For example, we set the tree would be dead when available carbon was below 0, but the specific carbon threshold of death had not been proved. We kept death and slow pool update at the yearly step was a common and simple assumption in the dynamic models. The growth of canopy height and basal diameter were the wood growth and they were calculated by Eqn S50. The wood growth was related to NSC pools at a yearly scale.

6) In equation S4, which of the (0.5, 1.0) is used in being subtracted from rh and what does that depend on?

Response: These were the same parameters followed as the previous version, FORCCHN V1.0 (see Ma et al. 2017). This model had been validated on the large scale.

7) In the S1 discussion on autotrophic respiration, it states in line 33 that "In Equation S1, t_resp represents..."; however, equation S1 is the GPP equation. Which equation is meant here and where is t_resp used?

Response: Thank you for checking. The t_resp was used in Eqn S6 and S7. We had revised it in the Supplement.

8) Why is the NSC updated only once per year, in Eq 2? Why not continually update this? What impacts does the sudden jump cause, particularly in respiration, and are these realistic? Additionally, what impacts are caused by the NSC active pool then being initialized to zero on the first day of the year? Wouldn't this cause a disruption for actively growing forests?

Response: Eqn 2 was the NSC slow pool. It was a carbon storage pool, which was updated on the yearly scale. The active pool was continually updated. The method of two pools had proved in a previous study, Richardson et al. (2012). The NSC active pool initialized to zero would not impact the forest because the NSC storage pool provided the necessary NSC for requirements when the contemporaneous active pool was insufficient (see Fig. 1).

9) In Table S5, where are the allocation parameters used? They are not given symbols and do not appear in the equations from what I could tell.

Response: These parameters are used as s_u in Eqn 16. We added the description and unit in Table S5.

10) In section 3, the inputs include soil and geography data, what are those and how do they come into play? In the equations in the supplemental material, I only saw LAI used as inputs, how is the soil initialized and what is the geography data used for?

Response: The soil data were used to calculate Eqn S14-S28. The geography data was used to calculate the intermediate processes, e.g. we used the latitude to acquire the day length. The soil was initialized with soil data and parameters (e.g. soil carbon pools were initialized by the soil total carbon and the allocated parameters).

11) Section 3 line 128 states "We can choose four output results", yet I only see 2 listed.

Response: Please see the description of the four outputs in lines 138-143.

12) Watch verb tenses throughout the entire manuscript and choose a consistent tense for the entire manuscript. It started as present tense, but then switches to past tense half-way through section 3.

Response: We had unified tense throughout the manuscript.

13) FLUXNET2015 is used for the site evaluation, which includes numerous options of ER and GPP. Is one of these pre-calculated options used? Is so, which one? If not, why not? And how was the selection of ER and GPP determined? The various methods each have advantages and disadvantages and can lead to substantial differences in the flux estimates. Additionally, why was the ER set to the night-time NEE? Are the daytime contributions added to this since you are using daily fluxes? I think that the ER being used does include

both the night-time and day-time contributions; however, lines 148-151 in section 4 are hard to follow between the three different statements for what ER is.

Response: Yes, the FLUXNET2015 had numerous options. In this work, we focused on introducing our model instead of evaluating the FLUXNET2015 dataset. We chose the Variable Ustar Threshold (VUT) Mean values from the dataset as the observations because the VUT had a relatively complete record. We extracted the flux data based on the mean value of the nighttime and daytime method. These methods had been described in detail by Pastorello et al. (2020). To avoid misunderstanding for reviewers and readers, we revised the description of FLUXNET2015 NEE and ER: 'The Variable Ustar Threshold (VUT) Mean values of FLUXNET2015 are used in this work. We extracted the flux data from the mean value of the nighttime and the daytime method. The nighttime method uses nighttime NEE data to parameterize a respiration-temperature model that is then applied to the whole dataset to estimate Ecosystem Respiration (ER). The vegetation GPP is then calculated as the difference between ER and NEE (Lasslop et al. 2010). The daytime method uses daytime and nighttime data to parameterize a model with one component based on a light-response curve and vapor pressure deficit for GPP, and a second component using a respiration-temperature relationship similar to the nighttime method (Pastorello et al. 2020).', in line 151-159.

14) Figure 2 is very tiny and hard to read. Additionally, it appears to be in alphabetical order and has labels for a) through e) that appear arbitrary? It would be more helpful to have these sorted by forest type and/or show the average fluxes per forest type.

Response: Limited by paper size, Figure 2 was hard to read. Here, we used the scatter plots to replace the previous figure (see new Fig. 2). The scatter plots were based on the total days of all studied sites. According to the reviewer's comment, we sorted the plots with forest types.

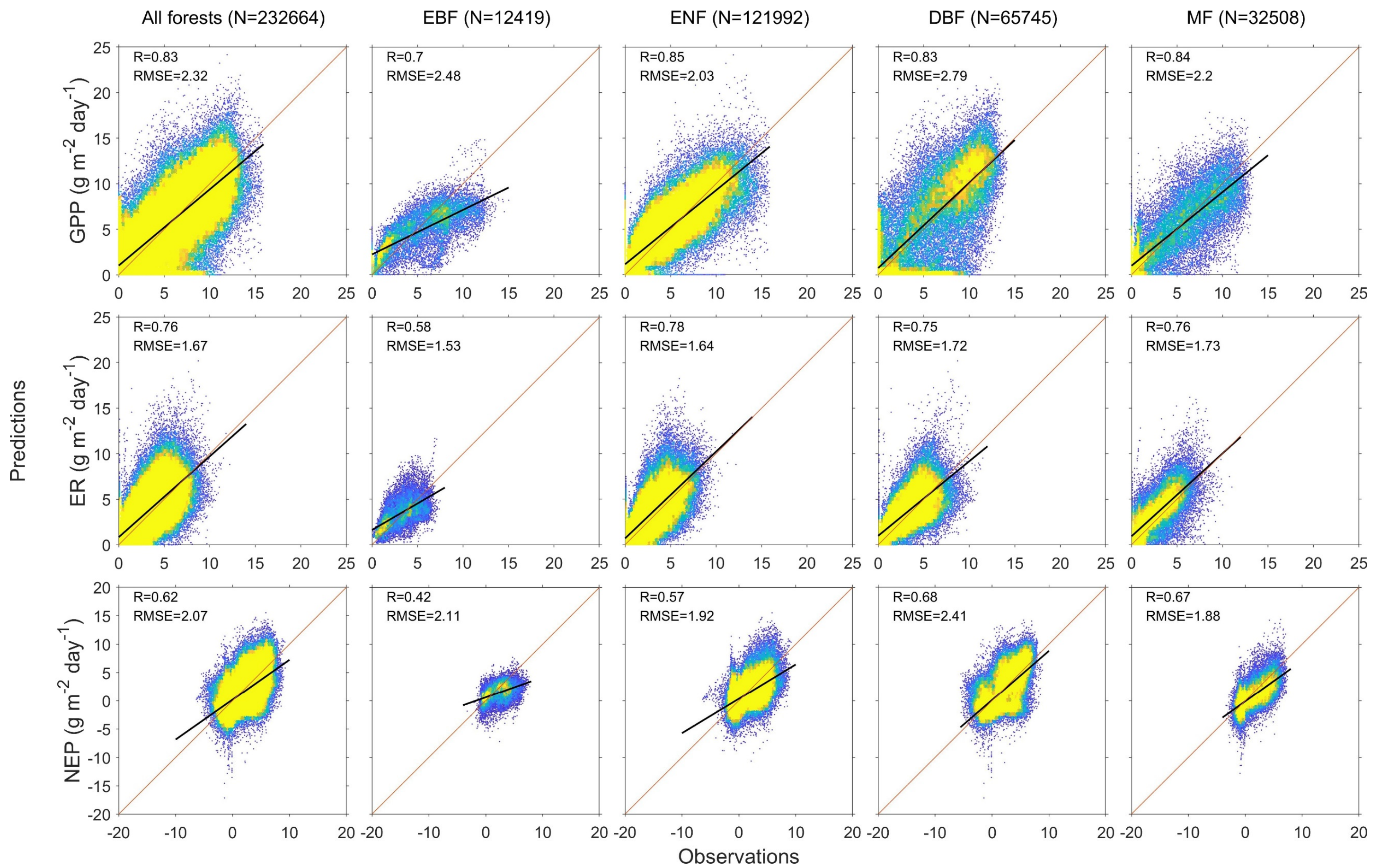

Fig.2. Heat plots showing the relationship between predictions and observations of daily gross primary productivity (GPP), ecosystem respiration (ER), and net ecosystem productivity (NEP) of the studied EC sites. N: the total days of all sites; R: correlation coefficient; RMSE: root mean square error (unit: g C m^{-2} day^{-1}). EBF: evergreen broadleaf forest; ENF: evergreen needleleaf forest; DBF: deciduous broadleaf forest; MF: mixed forest. Diagonal lines are 1:1 lines, indicating perfect agreement between predicted and observed fluxes. Black lines represent the linear regression. Colors indicate the percentage of pixels in each bin area (yellow is the densest).

15) For Figure 3, what is model efficiency? Also, the colors need to be labeled, which I believe are GPP (green), ER (blue), NEP (tan).

Response: Here, we added the calculation and description of E in Methods S4 (Eqn S60). The E value can range from −∞ to 1, and a value close to 1 indicates a perfect match between the simulated and observed data. We had added the index of E calculation in the manuscript. According to the reviewer's comment, we also added the color labels in the text of Fig. 3.

16) In section 4, line 160 states that the model has the best performance in capturing GPP dynamics, what evidence lead to this statement? Later in line 168 it is stated that "predicted ER performed lower than GPP", what is meant by "lower", do you mean that it does not perform as well or has higher errors?

Response: From the new Fig. 2, we could find the direct results of all predicted fluxes. The predicted GPP had the highest *R* (i.e. correlation coefficient) and thus it had the best performance. 'predicted ER performed lower than GPP' meant the median of *R* and *E* (i.e. model efficiency) from predicted ER was less than the median of *R* and *E* from predicted GPP. We added the corresponding description in lines 178-179.

17) Many forests in FLUXNET2015 are annual sinks of carbon due to stand age and regrowth. Since the model uses LAI as input, how well does it do at capturing this? And how well does the model then match or capture the annual growth of biomass and change in fluxes as forests mature?

Response: This was an interesting study of the carbon cycle. The model used the LAI as the initialization of vegetation information. The initialization meant we only need to input the LAI data once time (i.e. maximum LAI of 1980 in this study). The LAI of one time may not enough to explain this question. In this work, we focused on showing and introducing our model. The annual growth of biomass and change in fluxes as forests mature may be studied in our future works.

18) The input data used is described in section 5 lines 183-187, was this the data that was also used for the sites? Is so, this should be much earlier in the methods section. If not, what was the input data at each of the sites?

Response: We added the description of input data for sites: 'We also extract the climate data from the FLUXNET2015 dataset to drive the model. Soil data are taken from the Harmonized World Soil Database (HWSD) V1.2 (http://www.fao.org/soils-porta l/soil-survey/soil-mapsand-datab ases/).', in lines 160-162.

19) Figure 4 has unreadable font. Additionally, these large-scale carbon fluxes for the Northern Hemisphere should be evaluated in some way to claim that the results are reasonable, and in addition ideally the carbon pools should be evaluated as well. I believe further evaluation is necessary before the conclusions can state in line 210 that "FORCCHN2 was able to predict satisfactory carbon dynamics."

Response: We had remade the clearer figure than the previous figure in the manuscript (see new Fig. 4). For the evaluation, we had added the compared results of large-scale (the detailed response could be found in the second General Comments).

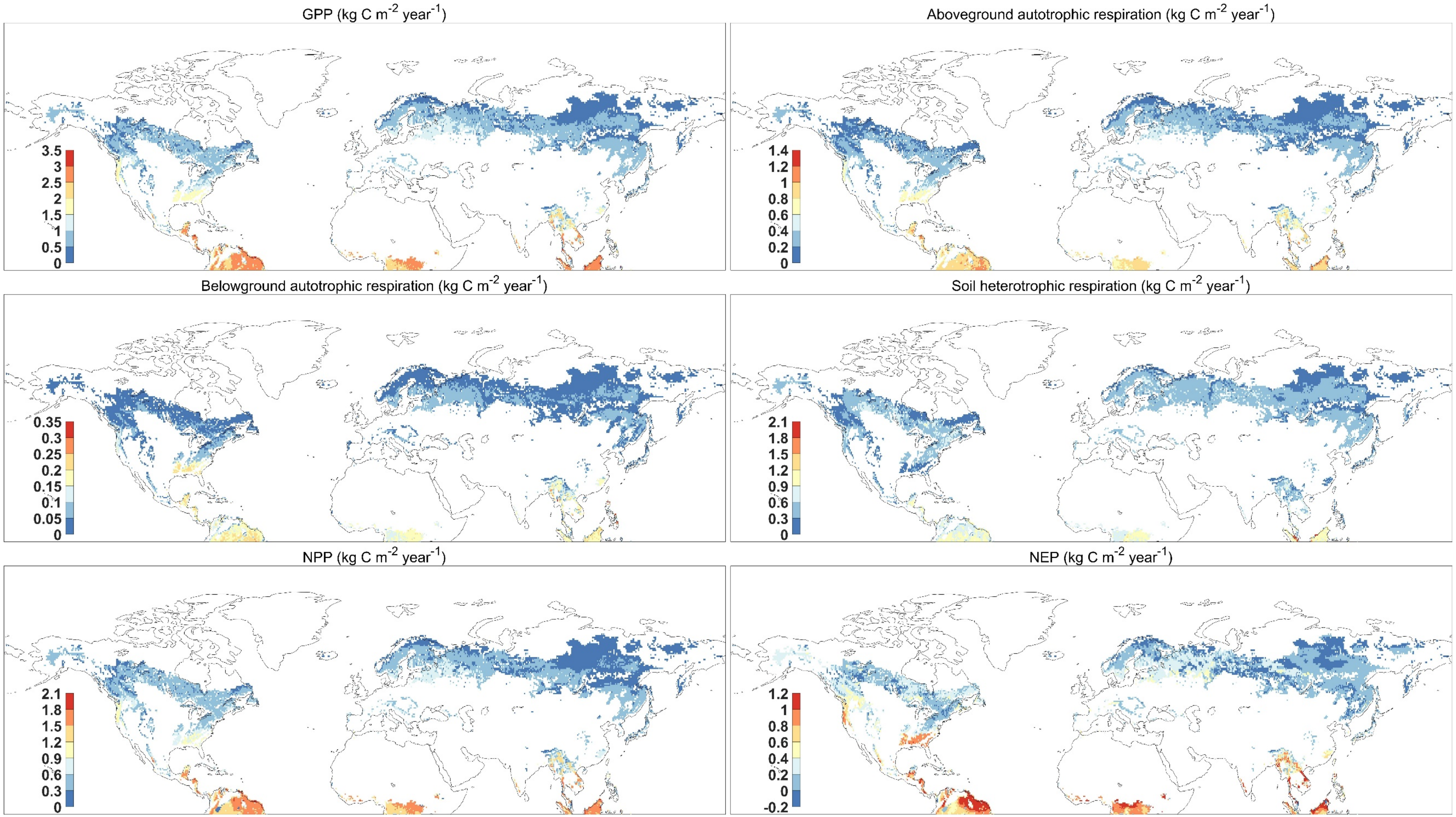

Fig. 4. The spatial distribution of mean GPP (Gross Primary Productivity), above- and belowground autotrophic respiration, soil heterotrophic respiration, NPP (Net Primary Productivity), and NEP (Net Ecosystem Productivity) predicted by the FORCCHN2 model for forest ecosystems of the Northern Hemisphere during 1980–2016. The spatial resolution is 0.5°×0.5°.

Technical Corrections

1) Section 2, line 71 should be "participating in the autotrophic respiration"

Response: According to the reviewer's comment, we had corrected this sentence.

2) In the supplemental material, all of the "Where" words do not start a new sentence and should be "where"

Response: We revised the words to 'where'.

3) In S1 line 106 should be "leaf growth is based on the assumption"

Response: We had added the word 'is'.

4) Section 2 line 81, remove first sentence of the paragraph or reword it as it doesn't make sense. (Major control equation of each individual tree.)

Response: We had removed this sentence.

5) Section 3 line 112 should be "adapt model runs to their"

Response: We had revised this sentence.

6) Section 3 line 118 should be "First, we installed and loaded"

Response: We had added the word 'we'.

Reference for the reviewer

Fang, J., J. A. Lutz, H. H. Shugart, Yan X., Xie W., and F. Liu. 2021. Improving intra- and inter-annual GPP predictions by using individual tree inventories and leaf growth dynamics. Journal of Applied Ecology 58:2315-2328.

Fang, J., J. A. Lutz, Wang L., H. H. Shugart, and Yan X. 2020b. Using climate-driven leaf phenology and growth to improve predictions of gross primary productivity in North American forests. Global Change Biology 26:6974-6988.

Ma, J., H. H. Shugart, X. Yan, C. Cao, S. Wu, and J. Fang. 2017. Evaluating carbon fluxes of global forest ecosystems by using an individual tree-based model FORCCHN. Science of The Total Environment 586:939-951.

Pastorello, G., C. Trotta, E. Canfora, H. Chu, D. Christianson, Y.-W. Cheah, ... and D. Papale. 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific data 7:225.

Richardson, A. D., Carbone, M. S., Keenan, T. F., Czimczik, C. I., Hollinger, D. Y., Murakami, P., ... & Xu, X. (2013). Seasonal dynamics and age of stemwood nonstructural carbohydrates in temperate forest trees. New Phytologist, 197:850-861.

Shugart, H. H., Wang, B., Fischer, R., Ma, J., Fang, J., Yan, X., ... & Armstrong, A. H. (2018). Gap models and their individual-based relatives in the assessment of the consequences of global change. Environmental Research Letters, 13(3), 033001.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., ... & Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology, 9:161-185.