Response to Referee #1

We are grateful to the reviewer for their time and energy in providing helpful comments and guidance that have improved the manuscript. In this document, we describe how we have addressed the reviewer's comments. Referee comments are shown in italics and author responses are shown in regular text.

This paper describes a new model for terrestrial biosphere processes that can be run in offline mode and coupled to an atmospheric chemistry-climate model. Aside from the model description, the authors show substantial validation of the model (YIBs) in three different configurations. The model performs well compared to the observations presented and to other more established terrestrial biosphere models. Compared to other models it also has a fairly mature treatment of ozone damage and BVOC emissions. This represents a substantial contribution to the earth system science community, as evidenced by the fact that YIBs has already been used in previously published work. The paper is well structured and well written, and the abstract is clear and concise. It is appropriate for GMD and I recommend it for publication in this journal. However I do have recommendations as outlined below.

→ Thank you for the positive evaluation of this study and support of the YIBs model.

1. Model spin up: Why are different periods used for each of the experiments? In the supplement, it is stated that 80 years was long enough to get a net land sink of 2 PgC yr^{-1}. Was this initial condition only used for the global offline simulations? Why was only 60 years used for the online simulations?

For the site simulations, 30 years is a very short spin-up period, and it's likely that the respiration fluxes are still a function of the initial soil carbon. A longer spin-up would affect both the annual total and seasonal cycles of ecosystem respiration, and would therefore alter the NEE results. Since it is unknown to what level the soil carbon is in equilibrium at each site, I recommend removing the discussion of site simulated NEE. It could warrant a whole paper on its own, but as it stands this does not substantially add to the paper.

→ We confirm and verify that all 3 experimental configurations (site level, YIBs-offline and YIBs-online) use derived soil carbon pools from a 60+80=140 year spin–up as the initial conditions. The 140-year soil carbon pool spin-up period is performed with YIBs-offline for recycling year 1980 forcings (year 2000 for YIBs-online initial soil carbon pools). Site-level and global on-line simulations involve additional 30-year spin-ups.

YIBs requires minimum hourly resolution meteorological input data, which is only available from 1980. Hence, we do not adopt preindustrial (e.g. 1750 or 1850) conditions as the base period for the spin-up.

We completely re-wrote the Supplementary Information to give a much more transparent and lucid description of the soil carbon spin-up process:

“Soil carbon pool spin-up process using YIBs-offline

Studies investigating terrestrial carbon fluxes usually initialize models for the preindustrial period when human perturbations are negligible and soil carbon pools are considered to be at an equilibrium state (Huntzinger et al., 2013; Sitch et al., 2015). YIBs requires (minimum) hourly meteorological input data fields, which are only available from 1980. Therefore, to spin-up YIBs, we adapt the approach of Wutzler and Reichstein (2007), who applied a transient correction to the soil carbon pools so that the simulated stocks match observations.

We apply YIBs-offline to spin-up the soil carbon pools that provide the initial conditions for all 3 experimental configurations (site-level, YIBs-offline and YIBs-online). The spin-up process proceeds in 2 stages. In the first stage, we set a uniform initial height H_0 for each PFT (Table 1) and run YIBs-offline for 60 years using fixed CO_2 concentrations and meteorological forcings for the year 1980 to allow vegetation carbon to equilibrate. By the end of the first stage spin-up run, the year-to-year variations of global average tree height, LAI, GPP, and NPP are all within ±0.05%. In the second stage, the derived equilibrium tree heights from the first stage and the global soil carbon content at the top 30 centimeters (Batjes, 2009) provide the initial conditions for an 80-year run (again using fixed CO_2 concentrations and meteorological forcings for the year 1980) until the transient NEE is equal to -2 Pg C a^{-1}, a value supported by observations and multi-model ensembles (Piao et al., 2013).

Following this 2 stage 60+80=140 year total spin-up, the derived soil carbon pools represent the state for the year 1980. For YIBs site level simulations: the derived spun-up soil carbon pools for 1980 are used as initial conditions. For YIBs-offline simulations: the derived spun-up soil carbon pools and equilibrium tree heights for 1980 are used as initial conditions for the 1980-2011 transient period. For the on-line simulations with NASA ModelE2-YIBs: the entire 140-year process described above is repeated with fixed WFDEI meteorology and atmospheric CO_2 conditions for the year 2000 values. The derived spun-up soil carbon pools and equilibrium tree heights for 2000 are used as initial conditions for the present-day climatological coupled global carbon-chemistry-climate simulation.”

We further clarified the spin-up procedure and extensions for each configuration in the main manuscript text:

4.1 Site-level simulations (Lines 557-562):

“The soil carbon pool initial conditions at each site are provided by the 140-year spin up procedure using YIBs-offline (Supplement). An additional 30-year spin up is conducted for each site-level simulation using the initial height H_0 for corresponding PFT (Table 1) and the fixed meteorology and CO_2 conditions at the first year of observations. Then, the simulation is continued with year-to-year forcings at the specific site for the rest of measurement period.”

4.2 Global off-line simulation (YIBs-offline) (Lines 572-573):

"The soil carbon pool and tree height initial conditions are provided by the 140-year spin up procedure using YIBs-offline (Supplement)."

4.3 Global on-line simulation in NASA ModelE2-YIBs (Lines 591-595):

"In NASA ModelE2-YIBs, initial conditions for soil carbon pools and tree heights are provided by the 140-year spin-up process described in the Supplement using YIBs-offline but for year 2000 (not 1980) fixed WFDEI meteorology and atmospheric CO_2 conditions. The NASA ModelE2-YIBs global carbon-chemistry-climate model is run for an additional 30 model years."

Since the site-level simulations use soil carbon pools from 140-year total spin-up as initial conditions for 1980, and then an additional 30 years spin-up for each individual site, we would like to retain the discussion of site-level simulated NEE. Our main reason is that NEE is the directly measured quantity whereas GPP is derived from the directly measured NEE. However, we do certainly agree with the referee that NEE is sufficiently complex, sensitive and critical to warrant an entire study in its own right.

2. GPP: Please state where the GPP data comes from: was it downloaded from a website? Processed by the authors from NEE? Also, there is uncertainty associated with the flux tower GPP, as it is calculated from the measured NEE, which is itself uncertain (e.g. Papale et al. 2006, or see biome-dependent uncertainty estimates in Luyssaert et al. 2007, both attached). It would be useful to know the ability of the model in light of the uncertainty. For example, in Figure 4: a relative bias of 50% in GPP would be very high if the uncertainty in the GPP is around 20%. It would be useful to include uncertainty in the Flux Tower analysis – for example, the standard deviation of measured GPP could be used as a proxy for uncertainty in the flux, or general guidelines for biome-level uncertainty in Luyssaert et al. 2006.

→ In the section 2.1 (Lines 132-135), we state the origins of the GPP data:

"To validate the YIBs model, we use eddy covariance measurements from 145 flux tower sites (Fig. 1), which are collected by the North American Carbon Program (Schaefer et al., 2012) (K. Schaefer, personal communication) and downloaded from the FLUXNET (http://fluxnet.ornl.gov) network.

The error bars in Figs. S2 and S3 show that derived FLUXNET GPP uncertainties in terms of standard deviation of the interannual variations are small for most sites without crop rotation. We agree that intraspecific uncertainties could be large. As a result, we check the intraspecific variations in GPP and NEE and present results in Table S2 (we put in SI because the paper is already lengthy). We add the following statement (Lines 644-649)

"YIBs model performs simulations at the PFT level while measurements show large uncertainties in the carbon fluxes among biomes/species within the same PFT (Luyssaert et al., 2007). The simulated intraspecific variations (in the form of standard deviation) are smaller than the measured/derived values for most PFTs (Table S2), likely because of the application of fixed photosynthetic parameters for each PFT (Table 1)."

3. Judging from Table 1, the tundra PFT is most like a shrub but with reduced productivity. Does phenology apply to the tundra PFT?

→ Yes. We clarified in the revised text: (Lines 325-327):

"The shrub phenology applies for shrubland in tropical and subtropical areas, as well as tundra at the subarctic regions, though the phenology of the latter is usually dependent on temperature alone because the climatological soil temperature is <12 °C."

4. Model description: A few questions about the model formulation:
- In Equation 19, (1-λ)NPP is allocated toward the different vegetation carbon pools. What happens to the rest of the carbon assimilated through NPP (λNPP)?
- Nitrogen in wood: From equation 23c, it looks like all wood respires. Is it accurate to assume that all wood respires? In TRIFFID there is an additional parameter for calculating N in stems that approximates the respiring stemwood based on height (e.g. Friend et al. 2003).
- Is this model available for people to use? Is there a website?

→ (1) The rest of carbon is not used for carbon allocation in the current version, but is included in the litterfall for soil respiration. We adopted the carbon allocation scheme from TRIFFID, which also simulates dynamic changes in vegetation fraction. The portion of NPP (λNPP) is used for the spreading but is turned off in the current YIBs version. In future, we expect to extend the YIBs model to a fully dynamic vegetation model with additional treatment of fractional changes. We clarify (Lines 433-434):

"In the current model version, we turn off the fractional changes by omitting λNPP in the carbon allocation but feeding it as input for the soil respiration."

(2) We rechecked and found that the original function 23c was not presented correctly as reported in Cox (2001) and in the YIBs code. Yes, both TRIFFID and YIBs calculate N of stem based on tree height and LAI. We revised equation 23c correctly: $N_w = n_{wl} \cdot n_0 \cdot \eta \cdot H \cdot LAI$. We show the values of η in Table 1.

(3) We have added a new section "Code availability" (Lines 799-804):

"The YIBs model (version 1.0) site-level source code is available at https://github.com/YIBS01/YIBS_site. The source codes for the global off-line and global on-line versions of the YIBs model (version 1.0) are available through collaboration. Please submit request to X. Yue (xu.yue@yale.edu) and N. Unger (nadine.unger@yale.edu). Auxiliary forcing data and related input files must be obtained independently."

5. As explained in Section 5.1, there is a high correlation between modelled and observed GPP at the ENF sites. Is this definitely due to the frost hardening? Was the correlation evaluated with and without frost hardening? There are many other temperature dependencies in the photosynthesis equations so it seems possible that other factors are affecting the GPP in winter.

→ We performed additional sensitivity tests that turned off the frost hardening for ENF. Simulated GPP without frost hardening on average correlates with observations by 0.81, very slightly lower than the value of 0.84 with frost hardening over the 54 ENF sites. In winter (December-January-February), the average modeling bias is -0.20 g C m^{-2} day^{-1} (-15%) with frost hardening and 0.50 g C m^{-2} day^{-1} (38%) without. We do agree that other factors may also contribute to the modeling biases. We revise the statement (Lines 619-620):

"The correlation is also high for ENF sites even though phenology is set to a constant value of 1.0."

6. Table 3 and related text in Section 5.2: This section of text is difficult to follow, mainly because the text explains carbon fluxes for large regions (ie: All tropics) in terms of % of the global total, while the table shows actual fluxes for smaller regions – this makes it cumbersome to cross-reference the table. One suggestion is to add columns for All Tropics and All Temperate regions.

Also, is 46% meant to refer to both the NPP and Rh?

The final sentence about the NEE differences between the tropical and temperate biomes needs some further evidence or explanation: If the warmer biomes have a higher dark respiration, this implies that they also have higher Vcmax and possibly GPP, which would contribute to a more negative NEE in the tropics (unless tropical GPP is limited by radiation, Je, which in some regions is likely the case – in this case higher Vcmax would result in increased Rdc but not higher GPP). Was dark respiration output by the model?

→ We have added one column for tropics as suggested. In the original manuscript, we were summing up values in the Amazon, Central Africa and Indonesia as the "tropics" definition. We revise the tropics definition to encompass the latitude belt (23°S-23°N). Values for temperate regions are very close to (global - tropics) and are not included. Using this revised "tropics" definition, both NPP and Rh account for 57% of the global total.

Yes. We agree that our explanation for NEE differences was not clear. In the revised text, we have deleted the original explanation to avoid confusion. NEE is the net difference between 2 large and similar quantities: Rh and NPP, making NEE highly sensitive to even small uncertainties in Rh and NPP and therefore to many propagating factors (e.g. temperature, PAR differences). For example, in our calculation, tropical Rh accounts for 57.3% of global total and NPP accounts for 56.6%.

7. Figure 6: It would help to have a legend showing the color-coding for the seasons.

→ We have added color indicators at the bottom of Fig. 6.

8. Conclusions and discussion: I disagree with the statement that "The vegetation parameters, Vcmax25, m, and b are not well calibrated for the tropical forest rainforest biome due to the limited availability of tropical site measurement data (Fig. 1)." While it is true that relative to other biomes, there is a lack of data for tropical forest biomes, this is not what is shown in Figure 1. This figure shows there is only 1 flux tower used in this study (not the authors fault, there is a lack of flux towers in the Tropics especially ones with enough data to calculate GPP from the NEE). However, there is a fair amount of data which could be used to calculate photosynthesis parameters in the Tropics, see for example Figure 2 in Kattge et al. 2011.

→ We agree with the referee, and the original statement as written was incorrect. We change the sentence to emphasize the point we intended to make: "The vegetation parameters, $V_{cmax\,25}$, *m*, and *b* (Table 1), are fixed at the PFT level, which may induce uncertainties in the simulation of carbon fluxes due to intraspecific variations (Kattge et al., 2011)". (Lines 772-774)

9. Why was the MTE from Jung et al. 2009 used, instead of the MTE described in the 2011 paper? The 2009 product was made to reproduce a model (LPJmL), using fPAR simulated in LPJmL.

For clarity, we add the statement (Lines 155-156):

"This product was made to reproduce a model (LPJmL) using the fraction of absorbed PAR simulated in LPJmL."

We do not consider this choice a shortcoming because (i) most of the satellite products used in the validation necessarily involve a model at some stage in the processing; and (ii) Figure 7 shows that the two Jung et al. MTE GPP products used for large-scale GPP evaluation differ by only ~15 Pg a^{-1}, much smaller than the spread in the model intercomparison.

10. Technical comments: Abstract: An opinion: I'm not sure if the word "inextricably" is the best choice for beginning a paper attempting to explain these connections. Perhaps "intricately" is a better word?

→ We agree and changed to "intricately" following the referee's suggestion.

11. Dark respiration is referred to as Rd in equation 5 and Rdc in Equation 22.

→ The symbol R_{dc} has been changed to R_d in the revised text.

12. Page 3164, Line 20: remove "vs. higher"

→ Corrected.

Reference:

Cox, P. M.: Description of the "TRIFFID" Dynamic Global Vegetation Model, Hadley Centre technical note 24, 2001.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001-2013, doi:10.5194/bg-6-2001-2009, 2009.

Kattge, J. and co-authors: TRY - a global database of plant traits, Global Change Biol, 17, 2905-2935, doi:10.1111/J.1365-2486.2011.02451.X, 2011.

Luyssaert, S. and co-authors: CO2 balance of boreal, temperate, and tropical forests derived from a global database, Global Change Biol, 13, 2509-2537, doi:Doi 10.1111/J.1365-2486.2007.01439.X, 2007.