Anonymous Referee #2

Though I would like to be encouraging of work in the general direction of confronting ecosystem process models with emerging data, unfortunately this effort does not provide a good example. There are many problems with this work including many detailed below, but the biggest problem is that the findings, interpretations, and conclusions are not at all supported by the work that has been done (see V. below).

Reply: thank you for your critical but precious comments. Substantial changes are made, including several numerical experiment results according to comments from you and reviewer 1. Over half of the original figures are revised, updated and/or newly provided. We hope these revisions can meet your expectations. We also apologize for the late resubmission because the additional numerical experiments did take time.

I. Model Calibration Is Not Described: The paper suggests that it performs a model calibration but there is no information on this. A set of model (IBIS) parameters are apparently calibrated with flux tower data on GPP and ET from select sites and with plot-level aboveground biomass data. However, there is no description of the model calibration, and no parameter uncertainty or parameter correlation (equifinality) analysis

Reply: The parameters (Table in MS) were manually calibrated using a try and error method until the model simulation results of GPP and AGB matched the observations. In brief, we use the observed GPP and AGB to constrain model parameters until R^2 for most sites arrived maxima.

Trial and error, or trial by error, is a fundamental method of solving problems. It is characterized by repeated, varied attempts which are continued until success, or until the agent stops trying (Source: https://en.wikipedia.org/wiki/Trial_and_error). Though this manual calibration method is time consuming, it is commonly used among the modelers (ecologists and hydrologists etc.). Some automation calibration methods are proposed; however, these methods also have the problem of low efficiency such as the Monte Carlo method (See Xiao et al., 2014).

Four Fluxnet sites, representing different woody PFTs, were randomly selected to test the AGB uncertainties due to τ_w (Fig. 1, i.e. Fig. 7 in revised MS). Five hundred τ_w values were randomly chosen between the default and calibrated values using the Monte Carlo method. The simulated AGB is shown to be sensitive to τ_w for all sites, resulting in a large variation in τ_w by the year of 2010. All the sites show an increasing trend during the test runs, except for the tropical deciduous site (Au-How). Variations in AGB are around 50 Mg ha^{-1} by 2010 for the two temperate PFTs (Us-Me2 and US-Ha1), indicating large uncertainties caused by τ_w. This further reveals the necessity to accurate estimate τ_w for model simulation.

These explanations are provided in the revised MS now.

Xiao, J.F., Davis, K.J., Urban, N.M., Keller, K. (2014) Uncertainty in model parameters and regional carbon fluxes: A model-data fusion approach. Agricultural and Forest Meteorology, 189-190, 175-186.

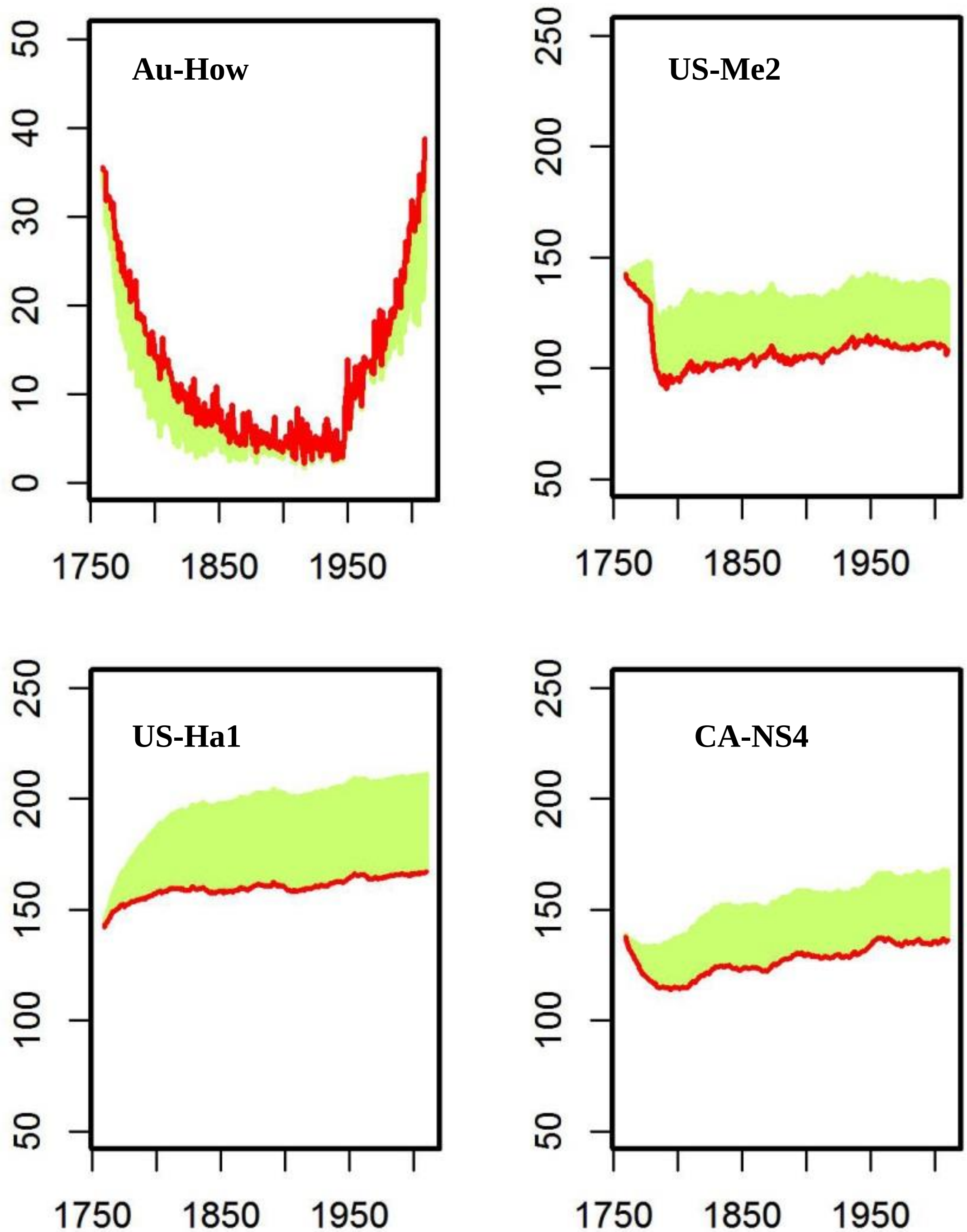

Fig. 1 Simulated temporal trends of AGB during 1759–2010 for different Plant Function Types (PFTs) by IBIS. The green lines show the 500 test runs using the random τ_w data ranging between the default and calibrated values; the red line shows the result of the calibrated τ_w. All the test sites were randomly selected from Fluxnet.

II. Data Sources Are Not Disclosed: The paper does not cite its data source(s) for the plot-level aboveground biomass dataset that it apparently used for calibration (though maybe just for evaluation?). It is suggested that most of the data come from China, though the Figure S1 shows a broad global distribution. Individual citations for all data sources must be provided, and the methods must explain the methods of data collection for each of those sources. It's inadequate to simply say "from the literature" and show a map of locations.

Reply: additional supplementary materials including the information on forest plot are provided. Please note that collection of the forest plot data is time consuming and we are conducting several studies using this data. Before our papers are published, we do not want to provide the detailed information for all the forest plots.

III. Model Set Up Incompletely Described: There is no description of the modeling procedure. Was a spin-up performed to bring carbon pools to some equilibrium state? How were PFTs assigned to grid cells, and/or does the model simulate PFT distributions that match with the other datasets? There is a risk of the wrong PFTs being simulated, for example where land use has substantially altered the PFT from a model estimated dominant PFT (e.g. if deforestation removed trees with grasses, crops, or savanna instead).

Reply: thank you for your comments. There is a spin-up period for our model simulation. We spun-up the model for 190 years (1759-1948) and then conducted transient simulations starting from 1948. Climate data in 1901 were used for the years before 1901. This information is included in the revised manuscript.

The model allows for the coexistence of different PFTs in a single grid cell. However, a dynamic vegetation mechanism is used to simulate annual changes in vegetation structure through PFT competition for light, water, and other nutrient resource pools (Kucharik et al., 2006). The competition among PFTs is driven by differences among carbon balances resulting from phenology, leaf form, and photosynthetic pathways (Foley et al., 1996; Kucharik et al., 2000). This information is included in the manuscript.

It is difficult to compare the model simulated PFT with observation, considering the spatial resolution and different PFT or land use classification. Here we show you the model simulated PFT in 2004 with that from MODIS data (UMD classification scheme). Because MODIS (actually also for other global data) does not include the climate zone information (boreal, temperate and tropical), we roughly compared the forest coverage between IBIS simulated and MODIS observations (Fig. 2). The model can generally map the different PFT distribution; however, as you said, wrong PFT is also observed but mainly for understory, such as shrublands in northeastern Asia.

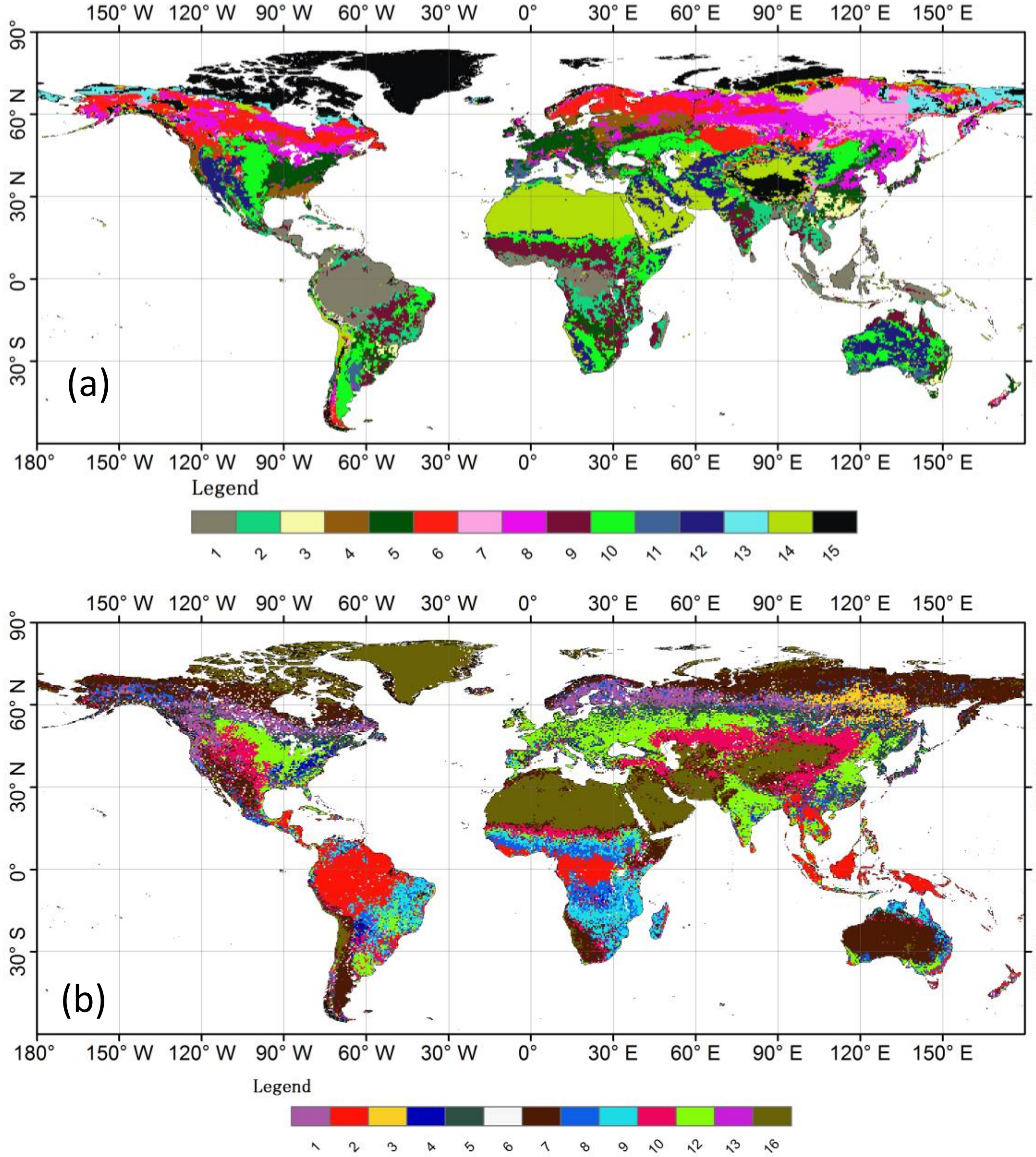

Fig. 2 Comparison of (a) IBIS simulated PFT and (b) MODIS land use. The plant functional type (PFT) numbers defined in IBIS are as follows: 1, tropical broadleaf evergreen trees; 2, tropical broadleaf drought-deciduous trees; 3, warm–temperate broadleaf evergreen trees; 4, temperate conifer evergreen trees; 5, temperate broadleaf cold-deciduous trees; 6, boreal conifer evergreen trees; 7, boreal broadleaf cold-deciduous trees; 8, boreal conifer cold-deciduous trees; 9, evergreen shrubs; 10: cold-deciduous shrubs; 11, warm (C4) grasses;12, cool (C3) grasses; 13-15, non-vegetated areas. MODIS UMD land use: 1, Evergreen Needleleaf forest;2, Evergreen Broadleaf forest;3, Deciduous Needleleaf forest;4, Deciduous Broadleaf forest;5, Mixed forest;6, Closed shrublands;7, Open shrublands;8, Woody savannas;9, Savannas;10, Grasslands;11, Permanent wetlands;12, Croplands;13-16 non-vegetated areas.

IV. Model Evaluation (Simply Comparing to Data) Does Not Go Far Enough: This paper's main point is that new datasets need to be used to confront models and improve them. However, the paper offers nothing to improve the model that is used. Discrepancies are shown but there is no new insight about why, or how the model structure or parameters would best be modified to come to resolution with the data, where appropriate.

Reply: thank you for your suggestion. As you suggested, model improvement is conducted by integrating gridded map of woody residence time (τ_w, years), which is shown to be important parameter for AGB simulation. To do this, we first collected forest plot level τ_w for pan-tropical areas from Galbraith *et al.* (2013). The main reason we focus on pan-tropical area is that it is easier to collect enough forest plot level τ_w in the limited review time. Then, we estimated the spatial map of τ_w for pan-tropical area by a Random Forest method. Third, we integrated the estimated gridded τ_w into IBIS model and simulated AGB for pan-tropical areas and validated by plot observations. As we anticipated, IBIS run with spatial map of τ_w result in best simulation of AGB with least RMSE, compared with baseline run (default τ_w) and calibrated run (calibrated τ_w) (Figs. 3 and 4). These figures and explanations are provided in the revised manuscript.

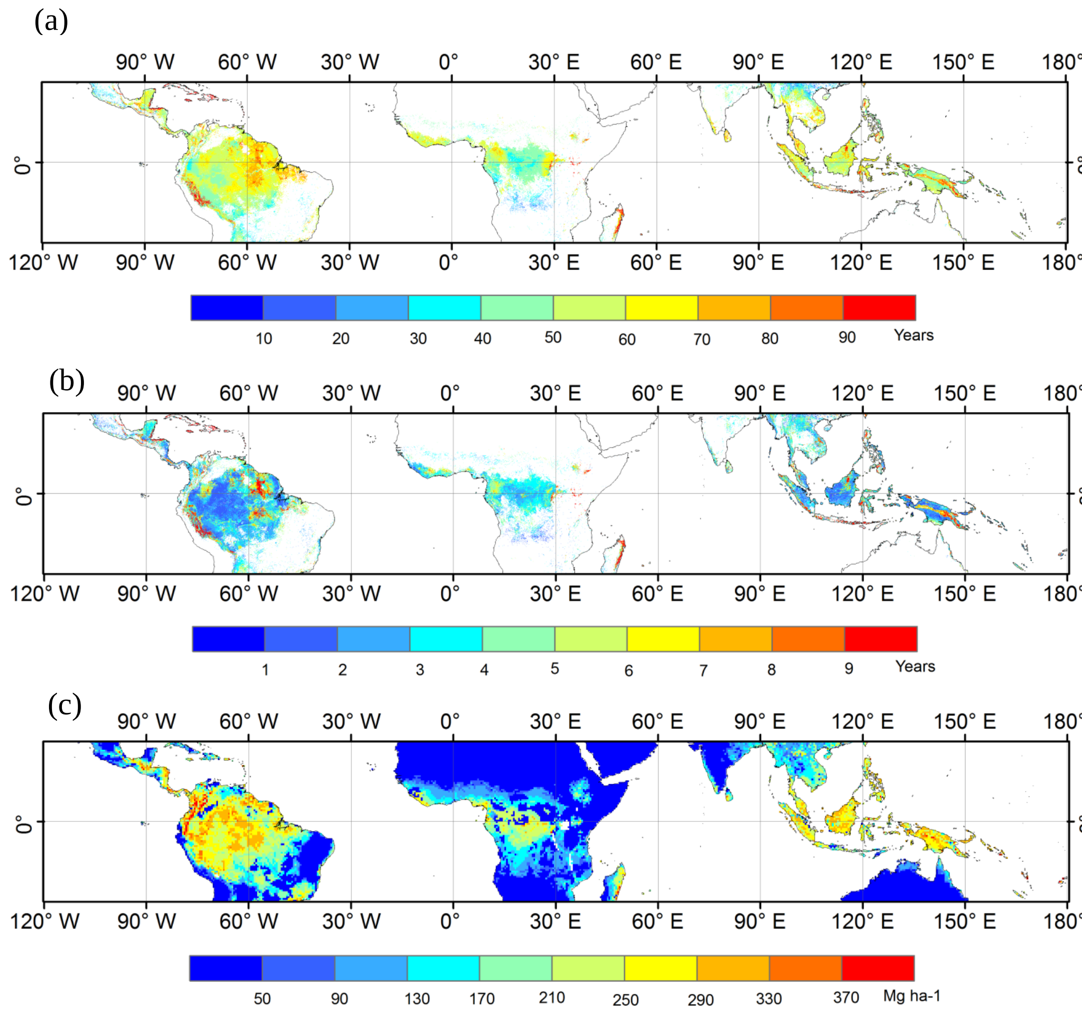

Fig. 3 (a) spatial pattern of woody residence time (τ_w, years); (b) uncertainty of estimated τ_w; (c) simulated AGB by estimated τ_w.

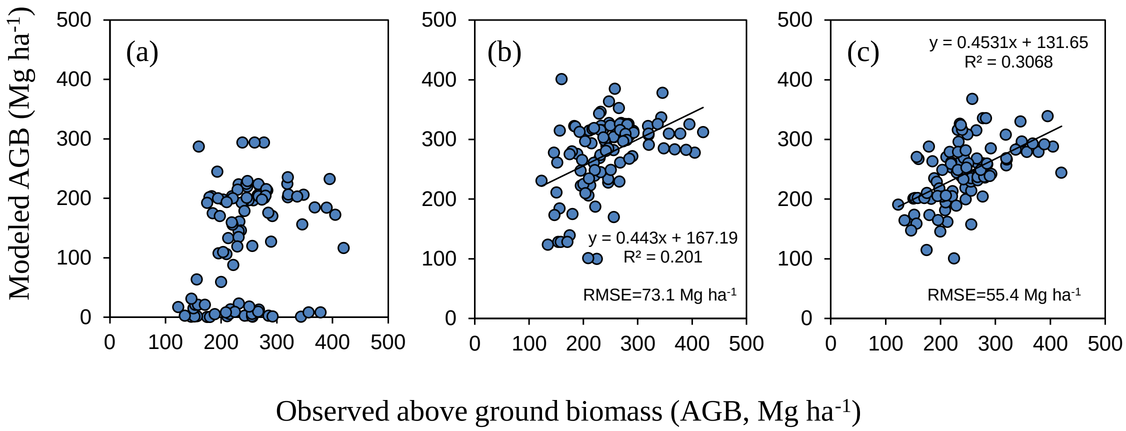

Fig. 4 Comparison of observed and modeled AGB for (a) baseline run with default τ_w; (b) calibrated run with calibrated τ_w and (c) estimated gridded τ_w

Galbraith D, Malhi Y, Affum-Baffoe K et al. (2013) Residence times of woody biomass in tropical forests. Plant Ecology & Diversity, 6, 139-157.

V. Findings and Conclusions Do Not Follow from Results and Do Not Advance Science in a Useful Way: The paper purports to show the following but each is poorly substantiated if at all. 1) Claim: Results of a DGVM can be sensitive to the meteorological driver data that are used but that parameter uncertainties are more important. Concern: This is known, and in fact is not precisely shown here. The paper does not compare sensitivity to parameter values in any way and thus cannot make this claim.
Reply: AGB uncertainties due to τ_w are now provided; see above (Fig. 1).

2) Claim: Bias or error in GPP caused by meteorological data can be transferred to AGB carbon stock. Concern: This is already known, and in fact is not precisely shown here (paper does not show that GPP bias or error relates to AGB bias or error).
Reply: sentenced deleted.

3) Claim: To improve model accuracy, modelers should pay attention to both model parameter calibration and meteorological drivers, with a focus on the former. Concern: This is known, and again, is not evidenced by anything in the present study.
Reply: new results provided. Please see above for our new figures and results.

4) Claim: DGVMs are useful tools for simulation of regional- and global-scale carbon dynamics. Concern: No doubt they are but this is not a conclusion of the study.
Reply: agree; sentenced deleted.

5) Claim: Discrepancies were observed between model-derived and observed spatial patterns of AGB for Amazonian forests, mainly because of the unique parameter set used in the model. Concern: Only a single parameter set was tested so you cannot claim that that is the source of the mismatch. Model structure could be a source of mismatch. Meteorological driver data could too. Nothing presented supports this claim.

Reply: agree; we now deleted figures for Amazonia forests and we focus on pan-tropical areas. Please see above.

6) Claim: The conclusions of our research highlight the necessity of considering heterogeneity of key model physiological parameters in modeling global AGB. Concern: This is already well established and not at all demonstrated by the present study.
Reply: agree; please see above for our new figures and results.

7) Claim: The research also shows that to simulate large-scale carbon dynamics, both carbon flux and AGB data are necessary to constrain the model. Concern: There is nothing here to support this claim. The study does constrain with C flux only, with AGB data only, and then with both to show that both are needed to recover key metrics of carbon dynamics. This claim is another throw away with no substance in the current paper.
Reply: agree; sentence deleted.

VI. It is Unclear Why FLUXNET Upscaled Product Is Newly Estimated: The calibrated model is then compared against a flux-tower upscaled GPP and ET product (Jung et al. 2011) but was actually re-estimated here for some unknown reason, and came up with substantially different results.
Reply: agree; results comparison with Jung *et al*. (2011) is removed in our revised manuscript.

VII. Study Involves a New Phenology Model That is Untested with No Evaluation: When you introduce a new model component such as the phenology model used here it is fitting to evaluate if that model component performs well compared to data. This is, in fact, part of the point of the paper, however the idea seems to have been missed with respect to this paper's new implementation of the phenology component in IBIS.
Reply: thank you very much for your suggestion. We compared the model simulated LAI by the IBIS default (GDD) and GSI phenology models with MODIS values for two forests sites (Fig. 5). Both of the two phenology models can generally reproduce the LAI seasonal variation, even though the lower values in dormancy season were overestimated for the boreal site. For both sites, GDD model results in a longer growing season compared with GSI model and MODIS observations. This may induce overestimated GPP for the model simulations. These results have been added in the revised manuscript.

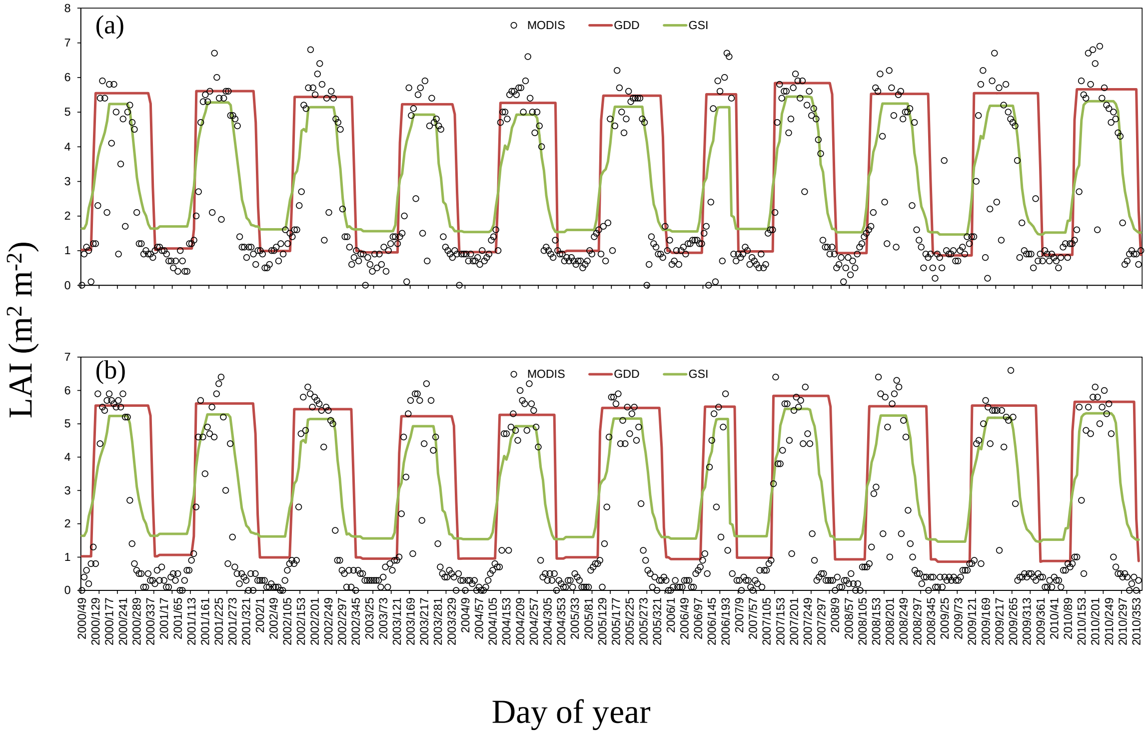

Fig. 5 Comparison of observed (MODIS data) and modeled leaf area index (LAI) for (a) US-Ha1 (temperate broadleaf cold-deciduous forest) and (b) US-WCr (boreal broadleaf cold-deciduous forest).

VIII. This paper does not appear to adhere to the FLUXNET Data Fair Use Policy. It does not cite the appropriate papers and does not include appropriate acknowledgement.
Reply: reference papers are mainly cited in supplementary and we further mentioned them in the acknowledgement.

IX. References missing: e.g. Stockli et al. 2008 is not in the references list, maybe others.
Reply: revised.