Response to Reviewer 2

Main comments

First, I fail to see what is new here compare to previously published studies: the current study uses only one model which does not seem to perform better than the TRENDY models used in Zhu et al. (2016) according to the results p. 8 l. 24-27 and p. 9 l. 22-25. As referee #1 mentions, this paper reads like a model evaluation and new scientific insights should be brought forward. If version 5 outperforms version 4 as mentioned p. 8 l. 25-27, the authors should consider using it instead.

R: There are a number of new features of our study. Firstly, it uses a new model, SSiB4/TRIFFID, which was not included in the TRENDY model inter-comparison. SSiB5/TRIFFID (as described on p. 8 l. 25-27) is a significantly updated version of SSiB4/TRIFFID. It is a new model which has not completed its testing. It will become publicly available next year. We have taken the sentence out in the text to avoid the confusion.

Secondly, and most importantly, our study focusses specifically on the impact of the climate regime shift during the 1980s. Previous DGVM studies (including TRENDY models used in Zhu et al, 2016), which investigated the effect of climate change on vegetation, only focused on the period after the 1980s. In our study, the contribution of the primary drivers on the ecosystem trends for each regime is identified. We estimate large-scale trends in terms of carbon fixation (GPP), vegetation growth (LAI), and expansion (vegetation fraction) rather than focus only on one aspect such as LAI trends in (Zhu et al, 2016).

Model inter-comparison exercises are excellent ways to assess common features of different model projections, and to estimate uncertainties. However, analysis of the responses to given climate anomalies is more effectively carried-out through detailed analysis of a single model, and this is precisely what we undertake in this paper. We have added text to the paper to make these innovative aspects of our study clearer.

Second, why is the study limited to the years 1958-2007? Considering the increasing availability of EO since 2007, extending the study period to nowadays would help address the "global vegetation variability" using satellite data as the title and the introduction (p. 3 l. 17).

R: We have downloaded three versions of Princeton meteorological dataset with the ending year of 2007 (v1x), 2010 (v1) and 2014 (v2.2), respectively. v1 had merged the data v1x plus the data from 2008-2010. However, when we compared the two versions (i.e. v1x and v1), we found that although v1x and v1 are generally consistent before 2007, however, there was an abrupt shift in some variables (such as wind speed) after 2007 (See Response Figure 1). To ensure the consistence and minimize the uncertainties could be involved by the meteorological forcing data, we decided to stop the simulation at 2007. The v2 data, which starts to be available in later 2016, is quite different from the v1 data (Response Figure 1, blue line) for a number of variables. Since by the time we have finished most of our work, we have kept with on the v1x data.
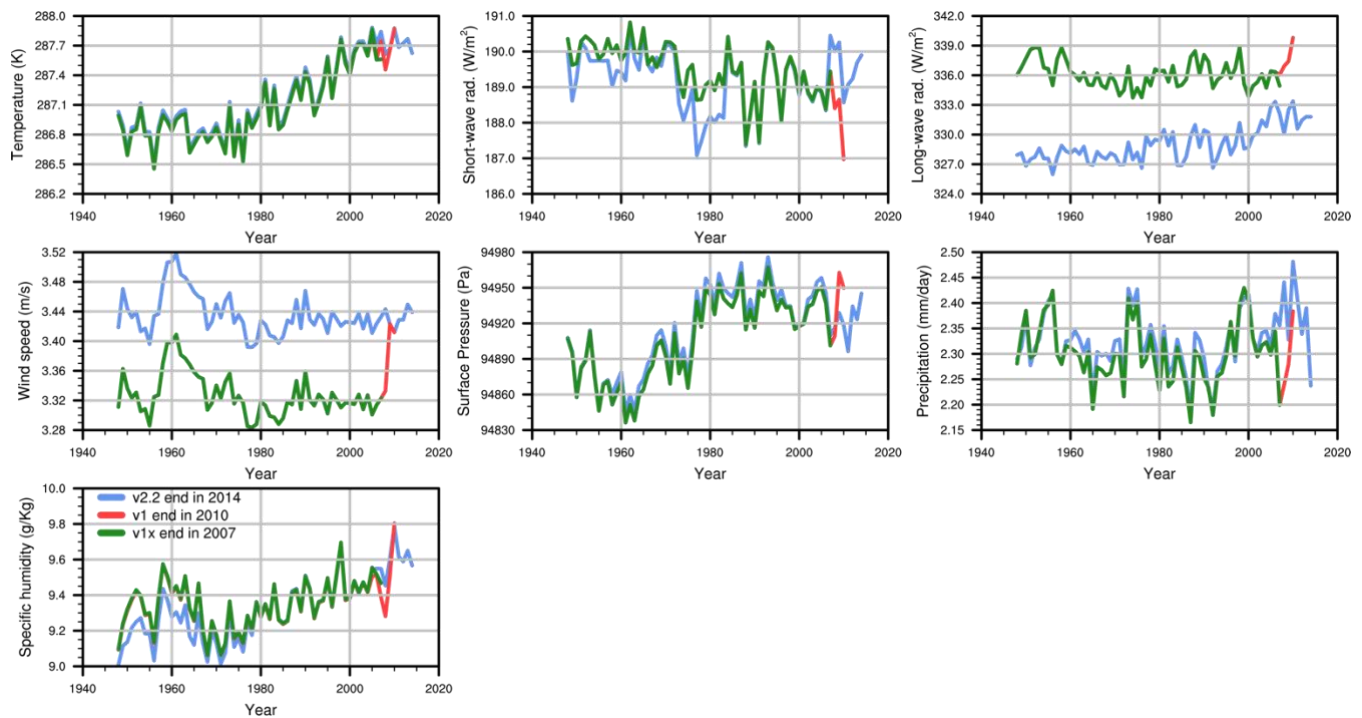
Response Figure 1. Comparison between different version of Princeton meteorological datasets over global land (-180o W,180o E, -60o S, 75o N)

5 Third, there is a lack of consistency between p-values reported, see for example p.9 l. 23 which points to possible cherry-picking from the authors.

R: Thanks for pointing out this. A consistent p-value of 0.05 is be used in the revised manuscript.

Minor comments

p. 2 l. 6 Can you support this statement with a reference?

10 R: This sentence refers to carbon budget reported in Ballantyne et al., 2012 and Le Quéré et al., 2013, LAI trend reported by Piao et al., 2011, Piao et al., 2015 and Zhu et al., 2016, NDVI trends reported by Myneni et al., 1997, Ichii et al., 2013, Los, 2013, and Piao et al., 201, GPP trends reported by Nemani et al., 2003 and Anav et al., 2015, and vegetation fraction change reported by Piao et al., 2005, Donohue et al., 2009, McDowell et al., 2015.

p. 2 l. 9 Leaf area "per unit of ground" area

15 R: Corrected to "per unit ground area" according to Referee 1, Referee 2, and book "Plant Factory" Chapter 9: Photosynthesis and respiration.

p. 2 l. 25 Consider citing Zhu et al. 2013 as an example of dataset covering the period 1980 to present

R: The citation was added.

p. 2 l. 30 Please cite articles that support this 'general consensus'

R: Ichii et al 2013; Piao et al 2015; Zhu et al 2016, 2017 have been added in the revised manuscript.

p. 3 l. 15 See my main comment about the study period

R: Please see the response for the second main comment.

p. 3 l. 18 'apportioned' is perhaps more correct than 'attributed'

R: In this study, the experiments were designed to identify and quantify the contribution of three external forcings on the ecosystem trends. These drivers are the cause of these ecosystem trends. As a matter of course, previous studies on this subject also used the term "attribution", for instance Ichii et al 2013; Piao et al 2015; Zhu et al 2016, 2017. We feel the word "attribute" may be more familiar to readers.

p. 4 l. 19 Please define SPOT (Satellite Pour l'Observation de la Terre), indicate what type of sensor VEGETATION is and a what resolution these data were available.

R: We have added a definition of SPOT in the revised manuscript. The VEGETATION sensor has four spectral bands.. The spectral bands are blue (437–480 nm), red (615–700 nm), near-infrared (772–892 nm) and short-wave infrared (1600–1692 nm). It has been detailed in the revised manuscript.

p. 4 l. 26 To my knowledge GIMMS is also derived from AVHHR data.

R: Yes. The GIMMS LAI was generated using the overlapping AVHRR GIMMS NDVI3g data and best-quality MODIS LAI, then generating the full temporal coverage GIMMS LAI3g data using AVHRR GIMMS NDVI3g. The GIMMS LAI provides observation at 15-day temporal resolution and 1/12-degree spatial resolution for the period from July 1981 to December 2014. Proper modification was added in the revised manuscript.

p. 5 l. 3-7 As the study aims to use satellite data, why not using MODIS GPP/NPP (or GIMMS-based NPP from Kolby-Smith et al. 2016).

R: This is a good suggestion. The observation-based datasets are used in this paper to evaluate the model simulation whenever the reference data available. The MODIS data set starts from 2000. The period is too short for this study. In fact, comparing the two observation-based GPP datasets (MODIS and FLUXNET MTE GPP), Anav et al. (2015) found they are in similar range of the global average and inter-annul variability, as well as in similar spatial pattern. FLUXNET-MTE is found to be more climate representativeness. Moreover, model simulations presented in Anav et al (2015) show higher spatial and temporal correlations against FLUXNET-MTE GPP than that against MODIS GPP.

p. 5 l. 20 onwards Can you summarize the experiments in a table?

R: A table is added in the revised manuscript.

p. 6 l. 8 Transient simulations are usually performed from a steady-state obtained under past conditions. Using the average conditions of the period 1948-2007 may reduce the model's sensitivity to the warming that occurred during that period. It would have been better to use the first ~10 years of driving data for this procedure.

R: Thank you, this is a good idea. We decided to use the average conditions of the period 1948-2007 in the quasi-equilibrium simulation was based on the practical consideration. Many experiments have been conducted to test the model performance under different meteorological condition. For instance, using the first year (1948) meteorological forcing, and the first 10/20

years' average. We obtain the best and the most stable results when the model was driven by average of the whole period 1948-2007, and also excluded the first 10-year results as spin-up period in the analysis.

p. 6 l. 15: Is this checked at pixel level, or only globally? Have you checked whether fluxes and initial stocks were at equilibrium? see e.g. Exbrayat et al. (2014) for the importance of initial stocks on transient simulations.

R: It is checked at reginal scale. We have checked the quasi-equilibrium status at different regions across the world, with particular attention on the transition areas between major climate zones. Per referee's concern, we draw the spatial distribution of absolutely relative changes of the last 10-year simulation for each PFT, which is defined as

$$average\left(\left|\frac{fraction - mean\_fraction}{mean\_fraction} * 100\%\right|\right),$$

where mean_fraction is the averaged from the last 10-year simulation, please see Response Figure 2.

As defined in the manuscript, the quasi-equilibrium status is reached when the fraction change is less than 2% of the mean vegetation fraction. Therefore, only fraction time series are shown here. We also checked other variables such as LAI and GPP at the equilibrium simulation, and both reach the quasi-equilibrium status. SSiB4/TRIFFID is a water, carbon and energy balanced model. When several key variables involved in water, carbon and energy cycle reach a steady-state systematically, other variables should follow. In the current version of SSiB4/TRIFFID, soil organic carbon mentioned in Exbrayat et al (2014) is a diagnostic variable as a result of plant litter and soil microbial process; no feedback to vegetation growth.
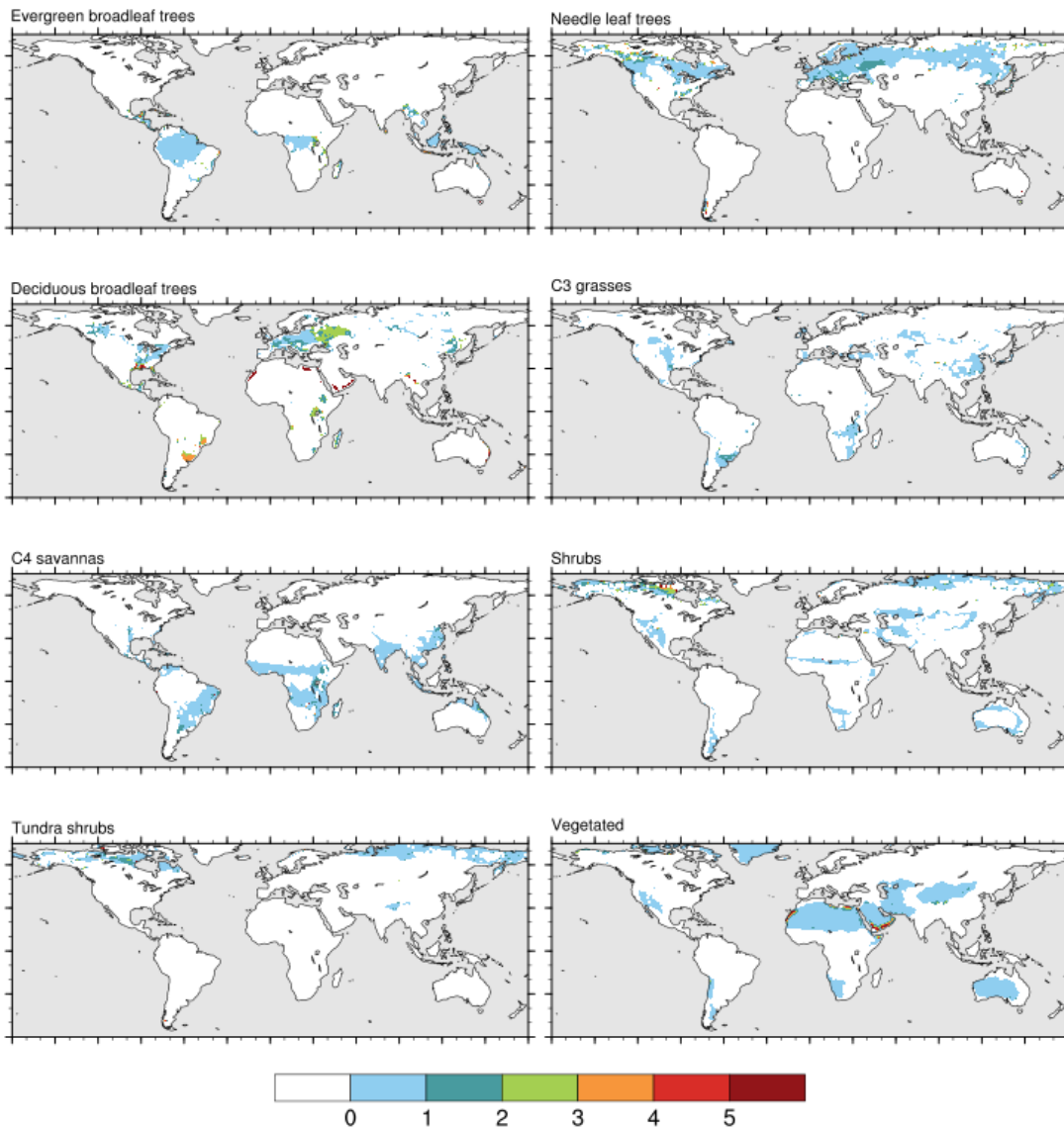
Response Figure 2. Mean absolutely relative change of the last 10-year simulation in quasi-equilibrium simulation for each PFT

p. 6 l. 22-25 Are these sentences referring to the model or GLC?

R: It was referring to the comparison between simulation and GLC. This sentence has been removed in the revised manuscript.

5 p. 7 l. 1 This part is very specific to the model used here. Readers who are not familiar with TRIFFID need a bit of context to understand how the LSD coefficient is used, and the impact of increasing its value ten-fold.

R: This is a valid point. LSD (γ_v) is a parameter in TRIFFID describing the rate of vegetation loss (units: yr^{-1}) caused by large scale disturbance such as fire, flooding and insect outbreaks. The change in plant functional type (PFT) fraction (v) is controlled by the PFT competition (the first term of the right-hand side of the equation) and the disturbance (the 2nd term).

$$C_v \frac{dv}{dt} = \lambda \Pi v \left\{ 1 - \sum_j c_{ij} v_j \right\} - \gamma_v v C_v,$$

where v is the vegetation fraction for each PFT, Cv is the carbon content in the plant, γ_v is the large-scale disturbance which results in vegetated area loss at the prescribe rate. It was set to 0.004/yr^{-1} for trees and 0.100/yr^{-1} for grasses. Those values were chosen largely by model calibration in offline tests. We increased LSD in the tree and grass mixed areas as the consideration of more fire occurrences in those areas. We are developing the fire module to more realistically simulate that disturbance in the future.

p. 7 l. 21 You can also cite Poulter et al. (2014)

R: Done. Thanks.

p. 8 l. 19 Please clarify whether you are referring to global average LAI.

R: This sentence is regarding the LAI spatial distribution. The spatial correlation coefficients comparing observations are presented for both global and North Hemisphere.

p. 8 l. 25-27 This statement raises an important question: why do you use model version 4 when you know that model version 5 outperforms it?

R: As response to the main common 1, the SSiB5/TRIFFID mentioned in p. 8 l. 25-27 is an updated version of current SSiB4/TRIFFID. It is still under development. SSiB4/TRIFFID works well for the current study. We have removed this sentence in the revision.

p. 9 l. 22-23 Please consider rewording... correlations of 0.35 cannot be described as matching the reference data closely.

R: Thanks for pointing out. We have modified this sentence and only indicate it is a significant correlation in the revised manuscript.

p. 10 l 19 Once again p-value...

R: Correction is made in the revised manuscript with consist p-value of 0.05.

p. 14 l. 29 I have not been able to access the data using this link, please check.

R: We have uploaded the data to a University server.