Authors' Response to Referee 1 (BGD bg-2021-37)

March 21, 2021

This paper investigates trends in global leaf area index (LAI) and attributes them to drivers (climate, CO_2, land use change) based on factorial simulations with a set of DGVMs and a fully coupled Earth System Model. This is basically a revisiting of a study published earlier (Zhu et al., 2016) that applied the same approach (model-based attribution of drivers) and used the same LAI product (GIMMS3g, based on data from the AVHRR mission; Zhu et al. also used GLOMAP and GLASS to obtain more robust results). Winkler et al. reach conclusions that have potentially high relevance for our understanding of global vegetation dynamics in response to climate change and (in particular) to CO_2. Effects of rising CO_2 remain a major uncertainty in Earth System Model projections, owing to challenges in observing and attributing effects. Hence, deriving new insights from available observational records is needed.
The paper by Winkler et al. is well written and display items are of high quality. The fact that their conclusions directly challenge findings by Zhu et al. (2016) although relying on largely the same method and data, caught my attention. Winkler et al. write in their abstract ("Our results do not support previously published accounts of dominant global- scale effects of CO_2 fertilization.", l. 16) and in their conclusions ("A cause-and-effect relationship between CO_2 fertilization and greening of other biomes could not be established. This finding questions the study by Zhu et al. (2016) that identified CO_2 fertilization as the most dominant driver of the Earth's greening trend.", l. 722), and in the Key Points ("Most models underestimate the observed vegetation browning, which could be due to an excessive CO_2 fertilization effect in the models.")
Strong conclusions require strong evidence. However, I have several strong concerns with how these conclusions were reached. In my view, the evidence presented here does not support this main conclusion (represented by the three citations I refer to above). Although I'm convinced that the analysis itself is diligently carried out and I consider that the paper offers a valuable discussion of the wide and relatively recent literature on the topic, I am concerned that the main conclusion will not meaningfully contribute to advancing the field.

We thank the referee for her/his detailed and constructive review of our manuscript. We appreciate that the referee finds the subject matter of our study highly relevant but we also note the referee's main concerns about one of the key conclusions, i.e. a possibly overestimated CO_2 fertilization effect in terrestrial biosphere models. The revised manuscript will address this specific point in more detail as well as the other referee comments.

[I've reviewed the same manuscript before. As far as I can see (main conclusions are unchanged, figures are identical), the mansucript version under review at here is identical to the earlier version I have reviewed. Therefore, I am posting my previous report here again.]

We acknowledge that the referee already reviewed our manuscript for another journal. There the opportunity to address the referee's comments was not given, so we are glad to do so now. Some of the referee's comments seem to be outdated though as they have been already addressed in the manuscript, e.g., the referee states "Winkler et al. rely on a single LAI product to derive trends. [...]" (Comment 1.1) , yet we have already included analyses comparing a total of five different datasets (see Figure R1-1). More details can be found in our individual responses below.

1 General Comments

1.1 *Winkler et al. rely on a single LAI product to derive trends. Yet, several papers have documented inconsistencies between greening and browning trends between satellite data products. In particular, the product used* *here (LAI3g) is based on data from the AVHRR mission. It has been reported that respective data is affected by orbital drift of the satellite (Tian et al., 2015) and sensor degradation Piao et al. (2019).The MODIS Collection 6 does not support the AVHRR-derived browning trends in several regions (see also Chen et al. (2019)). This affects in particular North American boreal forests. [...]*

We thank the referee for this comment and for emphasizing the limitations of AVHRR-based datasets. The current version of the manuscript addresses this issue in the introductory section:

"*To assess observed changes in vegetation over climatic time scales, we make use of a 37-year record of leaf area index (LAI) satellite observations (1982–2017, GIMMS LAI3g, Section 2.1). The GIMMS LAI3g product is based on the Advanced Very High Resolution Radiometer (AVHRR) sensors, for which there are a number of shortcomings (no on-board calibration, no correction of orbit loss, minimal correction for atmospheric contamination and limited cloud screening; Section 2.1; Zhu et al., 2013; Chen et al., 2019).* ***To address these shortcomings, we also analyze a total of five different remote sensing products that pursue different strategies for dealing with the issues associated with AVHRR data (Section 2.1)****. Due to some inexplicable variations in these datasets (Forzieri et al., 2017) we concentrate on GIMMS LAI3g in our analysis, which is used in most published papers*" (LL59-68).

Additionally, we go into details describing how these technical issues are addressed in the latest versions of GIMMS LAI3gV1 and NDVI in the methods section:

"*The complete time series of LAI3gV1 was generated using an artificial neural network trained on data of the overlap period of the Collection 6 Terra Moderate-Resolution Imaging Spectroradiometer (MODIS) LAI dataset (2000-2017) and and the latest version (third generation) of the Global Inventory Modeling and Mapping Studies group (GIMMS) Advanced Very High Resolution Radiometer (AVHRR) normalized difference vegetation index (NDVI) data (NDVI3g).* ***The latter have been corrected for sensor degradation, inter-sensor differences, cloud cover, observational geometry effects due to satellite drift, Rayleigh scattering and stratospheric volcanic aerosols*** *(Pinzon and Tucker, 2014).*" (LL96-102).

Most of the other AVHRR-based data products analyzed here also rely on MODIS time series, among other measures, to correct for the limitations of the AVHRR sensor mentioned above. As described by the referee, MODIS and AVHRR diverge in their estimates of LAI trends in some regions. We already cite and discuss the study by Chen et al. (2019) who examined the differences between MODIS and AVHRR-based estimates in more detail. **Chen et al. (2019) show that the greening trends in the MODIS record correspond well to regions with intensive land use changes. In this study, we focus on long-term climatic and physiological effects and mask regions with intensive agricultural activities, thus do not address land use changes.** The MODIS time series is still too short to assess long-term changes in the Earth system associated with climate and rising CO_2. AVHRR, on the other hand, now spans nearly 40 years, making it one of the few resources we have to examine long-term land surface changes over time. In contrast, the MODIS record alone cannot provide any information on the state of the vegetation in the 1980s and 1990s. Overall, we agree with the referee that the AVHRR record has important shortcomings that need to be addressed and discussed. We will place more emphasis on these points in the revised version of the manuscript.

1.2 *[...] The AVHRR-based browning seen in this region combined with the apparent failure of models to capture the same trends has been used by Winkler et al. (with a largely over- stretched logic) to argue that CO2 effects were inappropriately represented in models ("Thus, it is important to focus model development not only on a better representation of disturbances such as droughts and wildfires, but also on revising the implementation of processes associated with the physiological effect of CO2, which currently offsets browning induced by climatic changes.", l., 744). This is non-sense both in view of the known lack of robustness of apparent browning trends, and in in the logic of the argument itself. The failure of models to capture a browning trend may also be due to insufficiently sensitive responses to climatic drivers; and in this case (North America) is most likely due to inappropriate representations of disturbance in models Anderegg et al. (2020)).*

We thank the referee for his/her comment on the logic of the argument, however, we disagree that the argument is 'non-sense'. Our argument has two parts as they are stated in the sentence which is quoted by the referee. First, we argue that the underestimation of historic browning trends in the models could be rooted in the mis- or underrepresentation of "disturbances such as droughts and wildfires" in the models. The referee suggests the very same in his/her comment: "*The failure of models to capture a browning trend may also be due to insufficiently sensitive responses to climatic drivers; and in this case (North America) is most likely due to inappropriate representations of disturbance in models*". So, we are in agreement here. Thus, the referee only disagrees with the second part of our argument, which proposes an alternative explanation of why models fail to reproduce vegetation browning: the current implementation of CO_2 fertilization effect in the models could offset the browning induced by climatic changes. We base this statement on our analysis of the factorial simulations of the MPI-ESM as well as the TRENDYv7 models, which simulate changes in LAI with and without certain factors such as CO_2 fertilization and climate change (Please refer to *2.3 Max-Planck-Institute Earth System Model*, *2.4 Land surface models: TRENDYv7* and *2.7 Causal Counterfactual Theory* for details). We find that for global estimates (Fig. 3) and especially for the tropical estimates (Fig. S5), the models reproduce observed decreasing LAI trends only when the CO_2 fertilization effect is being absent, as we explain in the manuscript:

"*TRENDYv7 models show strongly opposing responses of LAI to the different effects of CO_2: LAI decreases when the physiological effect is omitted, but increases when the radiative effect is omitted. MPI-ESM shows qualitatively the same responses, but less pronounced (Figure S5). For the second half of the satellite record, the observed trend switches sign to a strong negative trend ($\sim$ -1.4 % decade^{-1}). The models reproduce this tendency, but the multi-model average of the TRENDYv7 ensemble is still positive. During the same time period, the opposing reactions to CO_2 in the factorial runs are more strongly marked (Figure S5).* ***These results suggest that browning caused by CO_2-induced climate change is compensated by greening affiliated with the CO_2 fertilization effect at the biome level.*** *Based on these findings, we* ***hypothesize*** *that the physiological effect of CO_2 is strong in models and outbalances the negative effect of climate change in the tropical forests (Kolby Smith et al., 2016).*" (LL457-465)

The overarching argument we make, then, is that the failure of models to reproduce vegetation browning may be related to either insufficiently sensitive responses to climate drivers, overly sensitive responses to CO_2 fertilization, or both. Based on the results presented, this argument makes perfect sense.

1.3 *One aspect that distinguishes the study by Winkler et al., from that of Zhu et al. (2016) is their probabilistic driver attribution. As the authors write, the method has been adopted from Pearl (2009) and Marotzke (2019) who applied it to attribute drivers of near-term climate change. However, I consider that the application of this method to investigate drivers of vegetation change is ill-conceived. The usefulness of probabilistic attributions is evident when dealing with systems that are characterized by a substantial inherent stochasticity (deterministically chaotic systems). In such cases, simulated variations are not necessarily forced, by may result from unforced internal variability. This is not the case for vegetation dynamics, where the (simulated) internal unforced variability is typically zero (except for some models, e.g., LPJ-GUESS, that simulate stochastic gap formation, or some relatively small internal variability arising from stand dynamics - which are actually not simulated explicitly at the individual/cohort level in TRENDY models). This actually facilitates driver attibution. All simulated trends are uniquely attributable to drivers using factorial analyses. As a consequence of relying on a probabilistic attribution, Winkler et al., find, e.g., no clear attribution in some semi-arid regions (Africa, South America, AUS) (l. 680-690) due to high interannual variability of green vegetation cover. As I read the paper by Winkler et al., such findings underlie their conclusions (e.g., "We find that CO2 fertilization is an important driver of greening in some biomes, but not dominant globally as suggested previously", l. 126). I would argue that the findings by Winkler et al., do not provide new insights that allow for a revision of findings by earlier studies (e.g., Zhu et al. (2016)), but rather fail to identify drivers (including CO2 effects) due to their application of an inappropriate attribution method. In most other biomes, attributions made here are largely identical with attributions made by Zhu et al. (2016) and also summarised by Piao et al. (2019).*

We thank the referee for his/her critical view of the method Causal Couterfactual Theory, which we will address in three statements. First, we do not agree that its application is ill-conceived for long-term changes in the Earth system. The referee is right, that this method originates from attribution studies of extreme events, e.g. Hannart et al. (2016). Hannart and Naveau (2018) in "Probabilities of Causation of Climate Changes" adapted the method to causal attribution of long-term changes. They successfully applied it to causal attribution of long-term changes in global surface air temperature, i.e., tested whether the warming trend can be causally linked to increasing CO_2. We adapted their approach to the driver attribution problem of long-term vegetation trends and tested whether they can be causally linked to CO_2 or climatic changes. In our approach, we follow the reasoning as explained in Hannart and Naveau (2018):

"***The proposed approach is anchored into causal counterfactual theory** (Pearl 2009), which was introduced recently, and in fact partly used already, in the context of extreme weather event attribution (EA). We argue that these concepts are also relevant to, **and can be straightforwardly extended to, the context of detection and attribution of long-term trends associated with climate change** (D&A)*" (Abstract Hannart and Naveau, 2018).

Second, it is unclear what the referee means with the statement that land surface models' "internal unforced variability is typically zero". It is true that variability in the atmospheric forcing translates into variability in land surface models. However, there are also several ways, besides the stochastic forest gap modeling proposed by the referee, that coupled processes in land surface models can lead to internal variability. There a various feedback loops connecting, for example, processes controlling dynamic vegetation (competition among plant types), biomass accumulation, fire events, nitrogen limitation, soil moisture effects, which can result into temporal and spatial variability. But more importantly, the term variability here refers to a more broader concept of variability, including inter-model variability. To estimate uncertainty / variability in this causal framework we again follow and adapt the approach by Hannart and Naveau (2018) who argue that the overall uncertainty estimates comprises various components, such as climate variability, inter-model variability, and variability in observations (Please read Section "2.7 Causal Counterfactual Theory": "*[...] the overall uncertainty [...] is estimated based on all simulations, comprising factual, counterfactual, and centuries-long unforced (pre-industrial) model runs*"). The intent behind robustly estimating an overall uncertainty is to evaluate the probability of occurrence and magnitude of greening/browning trends over $\sim$ 40-year periods across models and between forced versus unforced systems. By the way, estimation of uncertainty/variability in detection & attribution studies is also a key element in the Optimal Fingerprinting method (e.g., Zhu et al.).

Third, the referee argues that the causal approach to driver attribution of vegetation changes, we present here, does not provide new insights compared to attribution studies that use the conventional Optimal Fingerprinting approach, e.g. in Zhu et al. (2016). As discussed in the manuscript, the Causal Couterfactual Theory-based attribution framework addresses the shortcomings of Optimal Fingerprinting, which mainly relate to the fact that it views observed changes as linear combinations of individual forced signals, is prone to statistical overfitting, and assumes that linear correlation reflects causality (Hannart and Naveau, 2018). For example, a strong correlation between globally increasing CO_2 and the greening signal suggests that CO_2 is the driver, but this is not necessarily the case. The probabilistic causality approach overcomes these issues and allows us to test whether long-term greening/browning trends are due to the effects of rising CO_2 in a probabilistic framework that combines necessary and sufficient causality. Thus, our attribution study and its results are a significant advance over the traditional method. In addition, our attribution study also analyzes a much more recent generation of land surface models (Zhu et. al: TRENDYv3) and an observational dataset that spans an additional decade (Zhu et. al: 1982-2009).

1.4 *I regret that I cannot offer a more positive assessment of this manuscript. However, my review should not discourage authors to use their results for a revised manuscript, where more attention is paid to assessing robustness of greening/browning signals in the context of multiple satellite products, and where caution is applied when reaching conclusions based on absence of evidence following the attribution method ("Causal Counterfactual The-*

ory") applied here, and claiming evidence for an overestimated CO2 effect in the current generation of terrestrial biosphere models.

We thank the referee for his/her critical comments and the encouragement to work on a revision of the manuscript. All the comments will be addressed in the next version.

References

Anderegg, W. R. L., Trugman, A. T., Badgley, G., Anderson, C. M., Bartuska, A., Ciais, P., Cullenward, D., Field, C. B., Freeman, J., Goetz, S. J., Hicke, J. A., Huntzinger, D., Jackson, R. B., Nickerson, J., Pacala, S., and Randerson, J. T. (2020). Climate-driven risks to the climate mitigation potential of forests. *Science*, 368(6497).

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R. R., and Myneni, R. B. (2019). China and India lead in greening of the world through land-use management. *Nature Sustainability*, 2(2):122.

Forzieri, G., Alkama, R., Miralles, D. G., and Cescatti, A. (2017). Satellites reveal contrasting responses of regional climate to the widespread greening of Earth. *Science*, 356(6343):1180–1184.

Hannart, A. and Naveau, P. (2018). Probabilities of Causation of Climate Changes. *Journal of Climate*, 31(14):5507–5524.

Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., and Ghil, M. (2016). Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events. *Bulletin of the American Meteorological Society*, 97(1):99–110.

Kolby Smith, W., Reed, S. C., Cleveland, C. C., Ballantyne, A. P., Anderegg, W. R. L., Wieder, W. R., Liu, Y. Y., and Running, S. W. (2016). Large divergence of satellite and Earth system model estimates of global terrestrial CO2 fertilization. *Nature Climate Change*, 6(3):306–310.

Marotzke, J. (2019). Quantifying the irreducible uncertainty in near-term climate projections. *Wiley Interdisciplinary Reviews: Climate Change*, 10(1):e563.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference.* Cambridge University Press, Cambridge, second edition.

Piao, S., Wang, X., Park, T., Chen, C., Lian, X., He, Y., Bjerke, J. W., Chen, A., Ciais, P., Tømmervik, H., Nemani, R. R., and Myneni, R. B. (2019). Characteristics, drivers and feedbacks of global greening. *Nature Reviews Earth & Environment*, pages 1–14.

Pinzon, J. E. and Tucker, C. J. (2014). A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series. *Remote Sensing*, 6(8):6929–6960.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B. (2013). Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011. *Remote Sensing*, 5(2):927–948.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, 6(8):791–795.

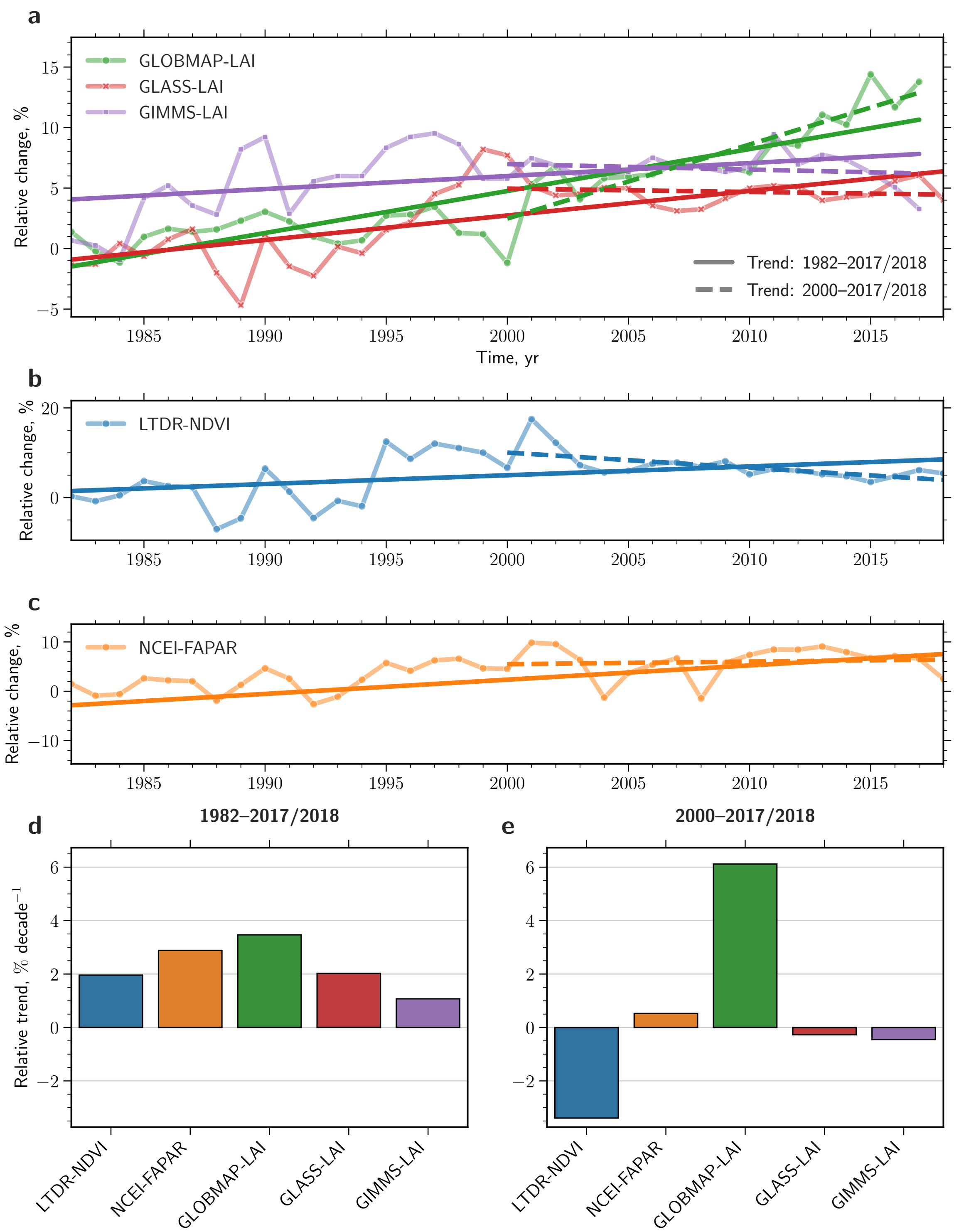

Figure R1- 1: **Five different remote sensing datasets displaying the development of the natural vegetation over the last four decades. a** Time series of changes in LAI relative to the average state from 1982–1984 as depicted in three different datasets (green: GLOBMAP-LAI, red: GLASS-LAI, and purple: GIMMS-LAI; see Materials and Methods section of the main paper for further details). The solid straight line represents the best linear fit for the entire period (1982–2017/2018), the dashed line represents the best linear fit for the second half of the period (2000–2017/2018). **b** as in **a** but for the dataset LTDR-NDVI (blue; see Materials and Methods section of the main paper for further details). **c** as in **a** but for the dataset NCEI-FAPAR (orange; see Materials and Methods section of the main paper for further details). **d** Bar chart comparing relative trends (in % decade^{-1}) in LAI, NDVI and FAPAR from different datasets for the entire period (1982–2017/2018) obtained from the gradients shown in **a-c**, respectively. **e** as in **d** but for the second half of the period (2000–2017/2018).

Authors’ Response to Referee 2 (BGD bg-2021-37)

May 10, 2021

The manuscript by Winkler et al investigates drivers of global LAI trends using a mix of long-term observations from AVHRR combined with Earth System Model sensitivity runs to provide causal attribution. The manuscript is in general well written and the results interesting and of general scientific interest. Please find some comments/questions below.

We sincerely thank Prof. Dr. Christian Frankenberg for his thorough review of our manuscript and his thoughtful comments. We address each comment below.

1 General Comments

1.1 *Title: I am not convinced the title conveys the gist of the paper, in fact I find it somewhat misleading. It reads as if the slow down of greening is driven by a further rise in CO2. The word "instead" instead of "with" would have made more sense but then again, the authors would have to make the topic of the title the key message of their paper, which it isn't (and it is hard to attribute to a weakening CO2 fertilization effect anyhow). For this, the author would have to use their counter-factual theory on the change in LAI changes between the beginning and end of the time period.*

We understand the confusion that may arise from the current title. We are aware that the scientific community in ecology sometimes uses "CO_2 effect" and "CO_2 fertilization" interchangeably, rendering the title counter-intuitive. But from a "Earth system" perspective, rising CO_2 as a forcing agent interacts with various processes in the system, which in turn have an effect on ecosystems. In this paper we investigate the impact of the physiological (PE) versus the radiative effect (RE) of rising CO_2 on leaf area. The title is meant to reflect that the Earth largely greened in the 1980s and 1990s as rising CO_2 had mainly LAI-increasing effects, e.g., by warming high northern latitudes (consequence of RE) and overall more carbon allocation through CO_2 fertilization (consequence of PE). However, as CO_2 continues to rise, the system appears to be entering or has entered a regime in which LAI-decreasing effects are amplified, i.e., climatic changes associated with rising CO_2 become more pronounced and have stronger effects on various ecosystems/biomes (consequence of RE), and possibly plant sensitivity to CO_2 fertilization decreases (as hypothesized in e.g., Wang et al. (2020), as also mentioned by the viewer in comment 1.2). Therefore, we find the title "Slow-down of the greening trend in natural vegetation with further rise in atmospheric CO_2" reflects the key content of the paper.

1.2 *That said, it would be necessary to also discuss the results of Wang et al. (2020) (Recent global decline of CO2 fertilization effects on vegetation photosynthesis) in the current manuscript as it is related to trends in CO2 fertilization as well (especially a reported decline of it, which differs strongly from Trendy).*

We thank the reviewer for pointing us to this recent paper on the global decline of CO_2 fertilization by Wang et al. (2020). We already integrated and discuss this study in the revised version of the manuscript. The authors used linear and non-linear regression methods and observational data ranging from remote sensing to *in-situ* atmospheric CO_2, and obtain the key result that global CO_2 fertilization has decreased. While this result seems to be inline with our findings, we do not argue that the sensitivity of the terrestrial biosphere to the CO_2 fertilization effect has declined, but that the effects of climatic changes rooted in the radiative effect of CO_2 (e.g. precipitation changes or increase in VPD) have strengthened, which probably counteracted the physiological effects of CO_2.

1.3 *One main strong statement of the paper is that it challenges finding by Zhu et al. (2016) (with some shared co-authors!). It sounds like a strong statement early on but if I look at Figure 3, I would say that the CO2 fertilization effect appears to be dominant at the global scale (despite some regional variations). It expands and adds nuance to Zhu et al, but challenges is too strong a word in my mind. There is enough material in this paper to warrant publication and no need to over-emphasize differences wrt to a previous publication.*

We thank the reviewer for this very good comment. We will rephrase this passage so as not to over-emphasize the differences with respect to Zhu et al. (2016)'s study in the revised manuscript. One point we make in the paper is that at the global scale, even the causal attribution technique like the optimal fingerprinting method used by Zhu et al. (2016) points to CO_2 fertilization as the main driver. This aspect motivated the biome-level analysis that led to the point that in many biomes, not all, the CO_2 fertilization effect cannot be identified as the main driver. But there are clear imprints of climatic changes that are obscured in a global analysis.

1.4 *Browning Trend in 2000-2017: When I look at Figure 3B, it appears a lot of the apparent browning trend in the later time-period is driven by a sudden decline in the relative change in years 2015-2017. What happens if you omit these years from the investigated time-period? What might cause such a sudden decline that might be related to the effects of CO2 fertilization or Radiative Effects? If this is related to detector issues or years with strong internal variability, I would remove these years (as long term drivers appear unlikely to suddenly appear). In fact, models and obs seem very consistent with each other between 2000-2014. As far as I can see, most discrepancies might be due to years 2015-2017 but I might be wrong. A critical discussion would be required here. Surprisingly, I couldn't find these strong effects of the last 3 years in the SOM plots, was it specific to some areas only? Can it be checked against MODIS data as well, which could be more reliable now? In fact, the first few years in Figure 3B are also VERY small, so you are fitting a linear trend through a time-period in which both ends are highly unusual. This can heavily bias derived trends, please evaluate and discuss the impact of chosen time-periods for trend analysis critically. https://doi.org/10.31223/X5K89V outlines some concerns I have with respect to AVHRR and the application to look at small changes (beyond pure trends). Please answer all questions in this paragraph.*

We thank the reviewer for this detailed and important comment. To investigate the sensitivity of our results towards the rapid decline in the years 2015–2017, we recalculated the relative trends in global LAI for the time windows for 2000-{2013, 2014, 2015, 2016, 2017} and for comparison for 1982-{1995, 1996, 1997, 1998, 1999}. Figure R2-1a compares the trend sensitivity analysis between the first and the last two decades of the AVHRR GIMMS LAI3g record. Where the relative trend in LAI in the 1980s/1990s is around 5% decade^{-1}, it is between 0-1% decade^{-1} in the 2000s/2010s. The different end years have an effect on the trend calculation, especially in the last two decades, but the differences are rather minor when compared to estimates of first two decades. Accordingly, the slow-down of the trend is also apparent when the sharp decline from 2015–2017 is excluded.

Figure 1b depicts how the global distribution of relative trend changes with varying end-years in the 2000s/2010s (for better readability only three time periods are displayed: 2000-{2013, 2015, 2017}; only time-series which pass the Mann-Kendall trend significance test ($p < 0.1$) are included). With respect to the periods 2000-2013/2015, there is a clear decrease in the pixels count of significant positive trends at the high range (between 10-20% decade^{-1}), a slight increase in the low range of positive trends, and an overall increase in negative trends for the period 2000-2017. Studying the results of the biome-level analysis (Fig. S3–Fig. S16), we find that the apparent rapid shift in the years 2015–2017 is not a global phenomenon, but rather "driven" by the tropical forests. It is currently being investigated whether this rapid decline in recent years could also be a detector problem. We also include here the current Fig. S3, which compares five different remote sensing datasets and how they depict the development of the natural vegetation over the last four decades. NCEI-FAPAR, LTDR-NDVI, and GLASS-LAI do not show this rapid decline in the years 2015–2017 as found in GIMMS-LAI, yet they agree on the slow-down of vegetation greening for the 2000s/2010s. All in all, our results and the overall conclusion of the slow-down of greening are not affected by the singular years from 2015–2017. As suggested by the reviewer, we will discuss the impact of the chosen time-periods for trend analysis based on the material presented here in the revised version of the manuscript.

In the Fig. S3 (Figure R2-2), we now also include MODIS-LAI for natural vegetation only, as suggested by the reviewer. MODIS-LAI depicts a stable moderate greening trend for the time-span of 2000-2019. Since the MODIS record cannot provide any information on the state of the vegetation in the 1980s and 1990s, we cannot assess whether MODIS would also depict a slow-down of the overall greening trend over this time-period. Also, please note that the comparability of relative trends in the long-term remote sensing products (baseline period 1982-1984) and MODIS-LAI (baseline period 2000-2002) is limited. Please see also our response to Comment 1.1 by Reviewer 1 and the discussion on MODIS and AVHRR discrepancies in the manuscript (LL362–369). Since this important issue was raised by both reviewers, we will move Fig. S3 in the main manuscript document and extend the discussion on the various datasets.

The reviewer raised the point that the values of the first few years in Figure 3B are very small. These values are very small be definition, since we are displaying relative changes in % with respect to the baseline period 1982-1984, so the initial values of the time-series are around zero %.

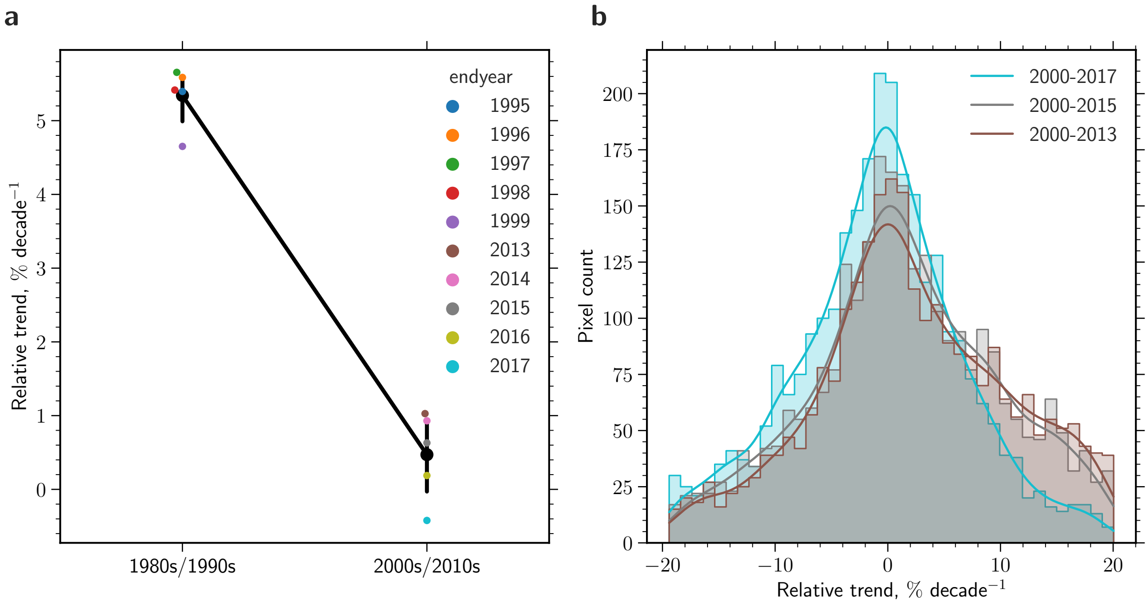

Figure R2- 1: **Estimating the sensitivity in the trend calculation with respect to the selection of the window size. a** Relative trends (in % decade^{-1}) in LAI relative to the average state from 1982–1984, calculated for different end-years, comparing the first two decades (1982–1999) with the last two decades (2000–2017) of the AVHRR GIMMS LAI3g record. The colored dots represent the trend estimates for different end years of the time series. Black dots represents the average value of the five estimates for each period, i.e. 1980s/1990s versus 2000s/2010s, including whiskers which denote their standard deviation. **b** Histograms of relative trends over the last two decades (2000s/2010s) in the AVHRR GIMMS LAI3g record including probability density functions (kernel density estimation) comparing estimates based on varying end-years. Only trends which pass the Mann-Kendall trend significance test ($p < 0.1$) are included.

1.5 A more general question regarding vegetation dynamics and CO2 fertilization, as you mention "as thoroughly equilibrated global carbon cycle" on line 192: What are the time-scales in ESM for CO2 fertilization? At the leaf scale, the gain in GPP is immediate but if you consider LAI, CO2 fertilization might cause a new state, which won't be achieved within a year, especially if species compositions will be affected. I would be curious what time-scales the models predict. E.g. if you changed CO2 suddenly but keep it at a higher level, how long would it take to run the carbon cycle into a new steady-state? I am mostly asking because the CC was certainly not in equilibrium in 1980 as CO2 increase and human land impacts are constantly shifting the needle. How much of the greening effects would have occurred (persisted for a while) even if we had suddenly frozen the CO2 levels at the 1983 mixing ratio and how would these "legacy" effects affect your overall conclusion? This is not a strong criticism but rather scientific curiosity.

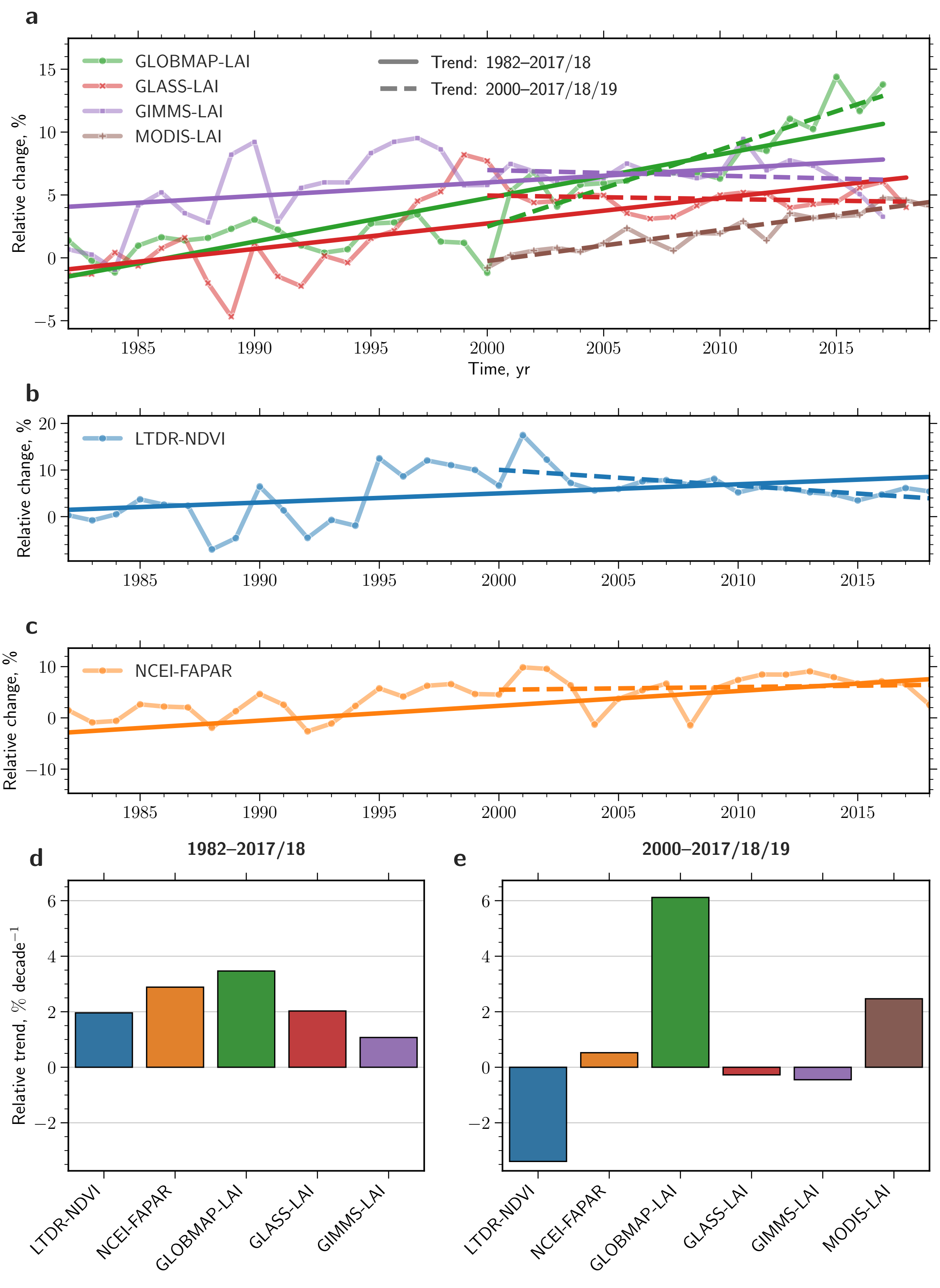

Figure R2- 2: **Five different remote sensing datasets displaying the development of the natural vegetation over the last four decades. a** Time series of changes in LAI relative to the average state from 1982–1984 as depicted in three different datasets (green: GLOBMAP-LAI, red: GLASS-LAI, purple: GIMMS-LAI and brown: MODIS-LAI; see Materials and Methods section of the main paper for further details). The solid straight line represents the best linear fit for the entire period (1982–2017/2018), the dashed line represents the best linear fit for the second half of the period (2000–2017/18/19). **b** as in **a** but for the dataset LTDR-NDVI (blue; see Materials and Methods section of the main paper for further details). **c** as in **a** but for the dataset NCEI-FAPAR (orange; see Materials and Methods section of the main paper for further details). **d** Bar chart comparing relative trends (in % decade^{-1}) in LAI, NDVI and FAPAR from different datasets for the entire period (1982–2017/2018) obtained from the gradients shown in **a-c**, respectively. **e** as in **d** but for the second half of the period (2000–2017/18/19).

We thank the reviewer for raising this interesting discussion point. Bringing an Earth system model (ESM) into general equilibrium can take many thousands of years, especially when deep ocean circulation and slow biogeochemical cycles such as that of nitrogen are included in the feedback network of an ESM (both of which are the case with the MPI-ESM used in this study). Even after the ocean circulation has reached a steady state and all the matter pools have built up, various variables may still exhibit drifts, especially on a regional scale. Thus, it requires expertise and patience to bring an ESM into general equilibrium in all of its subsystems - this is what we meant by a "thoroughly equilibrated" ESM / carbon cycle. For our study, we took the pre-industrial equilibrium of the MPI-ESM prepared by the MPI-M development team for CMIP6.

So, regarding the question ... *if you changed CO2 suddenly but keep it at a higher level, how long would it take to run the carbon cycle into a new steady-state?*: This strongly depends on the magnitude of change, i.e. CO_2 forcing. Many processes respond fairly immediately, such as GPP or radiative forcing, but others respond more slowly, such as ocean heat uptake, dynamical vegetation changes, or the global cycling of carbon. With all the feedbacks between ocean, atmosphere, land, and biosphere, a fixed increase in atmospheric CO_2 of, say, the order of 100 ppm would push the system so strongly that it would again take on the order of a thousand years to reach a new equilibrium for the carbon cycle.

Yes, absolutely, there are legacy effects in the system. Let's say we froze atmospheric CO_2 in the early 1980s, as for many other variables, greening would also continue after CO_2 stopped increasing, e.g. due to slower processes regarding dynamical vegetation. This is a very interesting research question in itself! As described above, our simulations, like the TRENDY simulations, are initialized from a pre-industrial equilibrium (for TRENDY, *near* pre-industrial: year 1900), accordingly these legacy effects are accounted for in this study, and thus the conclusions of this study should not be affected.

1.6 *Causal theory: One caveat that could/should be added is that this is only valid if the models, which are the basis for the sensitivity runs, are representing the truth. E.g. for the browning trend, you would actually find NO causal attribution from models alone, is that right?*

Yes, the causality is based on what the models predict for each counterfactual experiment and region. This is also true for every other method in "Detection and Attribution" using model output, such as the optimal fingerprinting method. We integrate an explicit statement about this caveat in the revised version of the manuscript.

1.7 *Overall, I would recommend revisiting the statements regarding Zhu et al, mention caveats in counter-factual theory using models as surrogate truth, investigate the impact of 2015-2017 on the greening/browning trend in the later time-period.*

We again thank Prof. Dr. Christian Frankenberg for his comments. We will follow his recommendations when preparing the revised manuscript.

2 Specific Comments

2.1 *Line 36: Stomata can even respond at short time-scales when CO2 changes, stomatal density or max conductance takes time to adapt. (you mention "in time").*

Thank you. We have adapted this passage in the manuscript to address the different time scales on which the physiological effects of CO_2 act.

2.2 *Line 88: "not dominant globally". Again, I am having difficulty to not see a similar effect in Figure 3c. In line 421, you even say so yourself. I am a bit lost here.*

We understand the confusion. The effect is dominant when we look at the global-aggregate signal. However, when we look at the regional analysis, we find that the effect is not dominant everywhere (i.e., globally) as the globally-aggregated signal would suggest. We rephrase these statements to be more specific in the revised manuscript.

2.3 Line 449: weaken -> weaker

Thanks, we corrected the typo in the revised manuscript.

2.4 Sections 3.10+: I was just a bit confused as the discussions now move from causal theories to more local descriptions, partially just citing other papers to explain specific events. It also shows the limits of your causal method as the lack of drought legacy effects (e.g. in tropics) can potentially bias your mode sensitivity runs. For some effects that you mention are due to RF, it would actually be interesting to separate out effects of CO2 RF into VPD, temperature and PAR effects (due to cloud cover changes), CO2 RF has various impact factors, which can very regionally in importance...

Yes, we thank and agree with the reviewer, that it'd could be an interesting next step to further decompose the radiative effect of CO_2 into changes in VPD, temperature, and changes in short-wave radiation (PAR / cloud cover). Further, the physiological effect could be decomposed into the stomatal effect and the direct carbon assimilation stimulation effect (RUBISCO). We leave this analysis step for a future study, since this would go beyond the scope of this manuscript.

We show that models are limited in their predictive power in simulating vegetation response to climate change. To address this issue, we rely on the published literature to evaluate evidence in observations that confirm or refute the results based on the causal attribution study.

References

Wang, S., Zhang, Y., Ju, W., Chen, J. M., Ciais, P., Cescatti, A., Sardans, J., Janssens, I. A., Wu, M., Berry, J. A., Campbell, E., Fernández-Martínez, M., Alkama, R., Sitch, S., Friedlingstein, P., Smith, W. K., Yuan, W., He, W., Lombardozzi, D., Kautz, M., Zhu, D., Lienert, S., Kato, E., Poulter, B., Sanders, T. G. M., Krüger, I., Wang, R., Zeng, N., Tian, H., Vuichard, N., Jain, A. K., Wiltshire, A., Haverd, V., Goll, D. S., and Peñuelas, J. (2020). Recent global decline of CO2 fertilization effects on vegetation photosynthesis. *Science*, 370(6522):1295–1300.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, 6(8):791–795.