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.

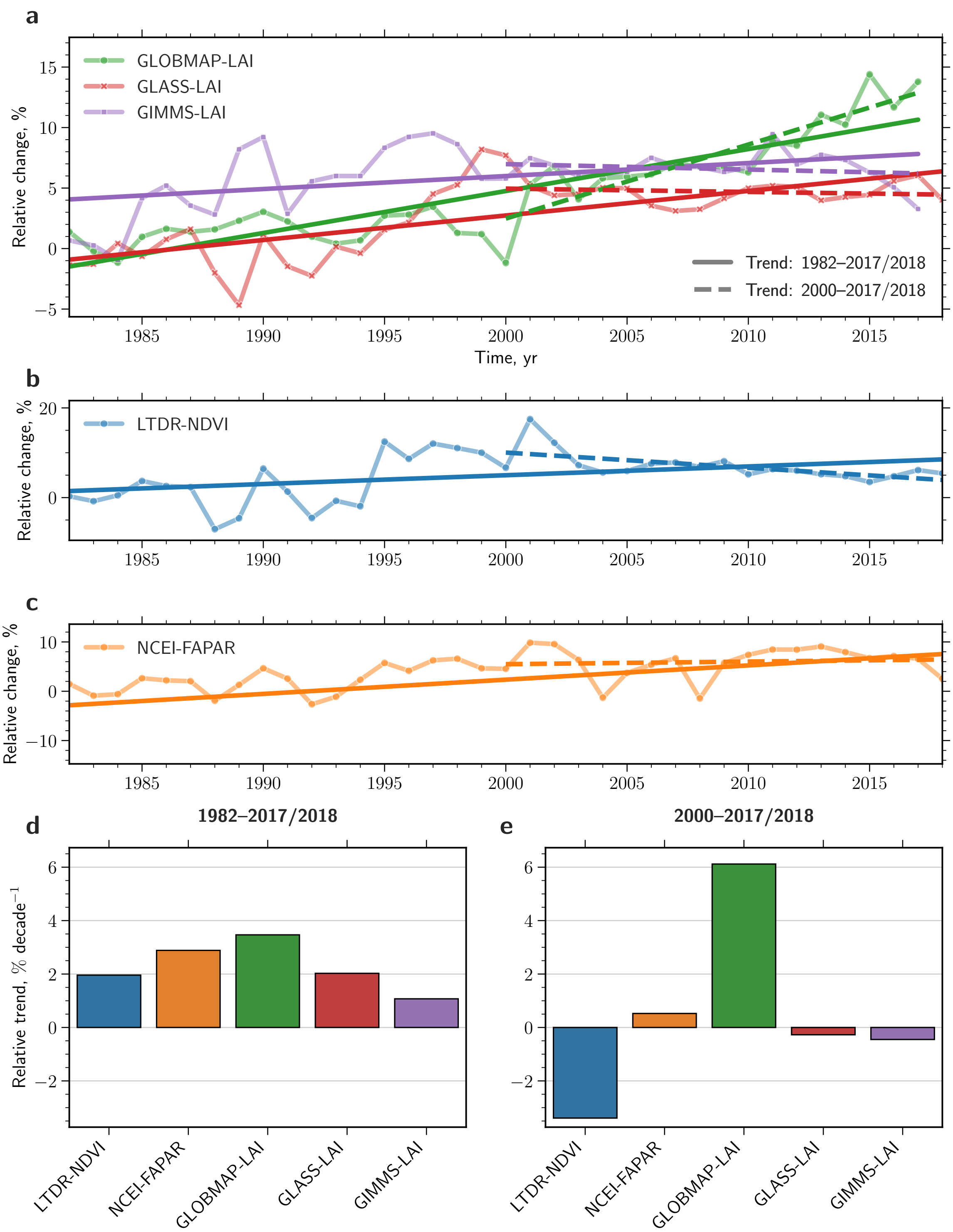

Figure R1- 1: **Five different remote sensing datasets displaying the development of the natural vegetation over the last four decades. a** Time series of changes in LAI relative to the average state from 1982–1984 as depicted in three different datasets (green: GLOBMAP-LAI, red: GLASS-LAI, and purple: 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).