Sept. 16th, 2022
Dear Biogeosciences Editor and Reviewers,

We appreciate your review and comments on “Exploring the impacts of unprecedented climate extremes on forest ecosystems: hypotheses to guide modeling and experimental studies”. Your time spent on this peer-review process is appreciated. Below is a point-by-point response to the reviews, and a description of how the manuscript has been revised accordingly.

Both reviewers had concerns related to comparing model results with site measurements. Such an approach would be logical for a more classic model study, such as testing model performance and subsequently using the model for temporal forecasting. However, the intention of our study was not to conduct a classic model experiment but rather a review of how models currently capture unprecedented climate extremes (UCEs) on forest ecosystems by using a handful of novel model results to guide that discussion. Our hypothesis testing approach is because data on UCEs, which would be needed for a classic model application are rare or non-existing, when extreme events have occurred, they are poorly monitored, or the extreme events that have been tested experimentally have been relatively moderate. Nevertheless, our hypothesis is that UCEs may in fact have the potential to not only have dramatic direct effects on forest ecosystems but also determine future growth and structure of such ecosystems. Therefore, we applied well tested models that have already been applied and validated at the sites used in this study. These models currently serve as our best description of the ecosystems, and therefore are well suited to test their responses to the UCEs as a kind of “hypotheses” in order to evaluate the results and to guide the need for future experiments and/or monitoring data. Our purpose is not to claim that the models are correct - most likely they are not simply because there have been no experiments or data to train and test them against UCEs.

We think many of the reviewers' concerns have now been addressed by clearly stating at the beginning of the manuscript these rationales for the study, and that this study was not intended to solely be an original research paper, but rather a road map and review paper. In order to generate an informed perspective on model responses to unprecedented climate extremes, we had to generate some original model results, which enabled our novel insights.

Best,
Jennifer Holm and co-authors

**R1 expressed concerns about:**

1. **R1: The novelty of the paper; and whether this paper was a review paper or a research paper.**

Answer: The novelty of our study is to exemplify unprecedented climate extremes (UCEs) and how models predict their impacts. In the revised manuscript, we improved the text about tighter linkages between concepts, hypotheses, and model outcomes. For example, we now emphasize that a goal of this paper is to demonstrate how to use these vegetation demographic models (VDMs) to help generate future hypotheses about and experimentally test UCEs. Therefore, we used the models and sites as conceptual "experimental" tools to investigate the given hypotheses.
We also revised the manuscript to state more clearly at the beginning that this is more of a review and "guidebook" manuscript, but based on original and well tested and published research to enable exploration of novel forest responses to such unprecedented climate extremes.  Because of the "unprecedented-ness" of the extreme, they and the models have for good reasons not been tested. We believe this combination of original research with review of current limitations of models is a strong novelty in and of itself.

We would like to clarify that the original "research" aspects of this study were the new UCE model results, and specifically the integrated C-loss with sensitivity to different climate change treatments which has not previously been done with these models, and is a knowledge gap that was filled by this study. The discussion section and detailed tables are designed to be a "review" and to help guide future research, for example in terms of future needs for experiments and observations to help improve model descriptions in this context.

New text at the beginning of the discussion to make clear this is more of a review paper:
"Vegetation demographic models (VDMs) allowed us to uniquely explore two hypotheses regarding a range of modeled response of terrestrial ecosystems to unprecedented climate extremes (UCEs), and setting the stage for the following perspectives to help guide future research."

2. **R1: While one paper can only tackle so much, R1 thinks this study would benefit from additional simulation experiments.**

Answer: The scope of our paper was *not* to do many, in-depth, detailed model experiments, but rather to exemplify how models currently capture UCEs in extreme durations/lengths and interactions with climate change. Furthermore, there is little if any data available to perform such a test. Therefore, we do not feel that there would be a strong benefit to adding additional simulation experiments, which would be a large, costly endeavor and make the already very long manuscript even longer.

Also, and importantly, additional simulations (as suggested by R1) might not lead to the novelty and clarity that R1 seeks. We revised the manuscript to state that outstanding modeling perturbations and experiments would be useful outcomes of future studies. The initial modeling investigation here highlights how VDMs (as opposed to typical LSMs) can be used to answer hypotheses and guide future studies.

In the summary paragraph of the manuscript, we include the following text: "Our study takes some initial steps to identify and assess model uncertainties in terms of mechanisms and magnitudes of responses to UCEs, which can then be used to inform and develop field experiments targeting key knowledge gaps as well as to prioritize ongoing model development (Table 3). Our intention was not to do an exhaustive list of UCE simulation experiments, and outstanding modeling perturbations and experiments would be useful outcomes of future studies."

3. **R1: Lack of model validation at the two sites, and site-level observations to inform model parameterization.**

Answer: We agree that model comparison to some site-level observations is a worthy improvement to the paper. We now include site-level benchmarking observations such as biomass and leaf area index for basic model validation. We also mention that both models have already been run and validated at these sites in previously published papers (Xu et al., 2016; Medlyn et al., 2016), thus making model application possible with a "built in" reasonable degree of validation. These references were included in the original manuscript, but we failed to make a clear connection that they were publications that validated the models at the two sites.

In the methods section (Section 2.2) we added in some additional site-level observations to help the reader gain a better understanding of the site characteristics. For example, included new text such as: "stand basal area is 29.2 (± 8.1) $m^2$ ha, stem density of 64 (± 12) trees $ha^{-1}$….." and "…...leading to a strong seasonality in LAI ranging from 3 to 4.5, but can get as low as 1.2 $m^2$ $m^{-2}$ (Kalacska et al., 2005). "

In the results section we added this new text: As a basis for the treatment results presented here, we compared the baseline simulations (prior to drought or climate change treatments) of the two VDMs to observations at both sites for biomass and LAI (Table S2, Fig. S1). Both models had similar biomass compared to observations at Palo Verde (10.4 - 11.7 vs. 11.0 kgC $m^{-2}$), and at EucFACE biomass matched well in LPJ-GUESS (12.1 vs. 12.7 kgC $m^{-2}$) but was low in ED2 (5.6 kgC $m^{-2}$). Both models also had similar LAI to observations at Palo Verde (3.3 – 4.5 vs. 3.8 (± 1.06) $m^2$ $m^{-2}$), and at EucFACE LAI matched well in ED2 (1.6 vs. 1.7 $m^2$ $m^{-2}$), but was high for LPJ-GUESS (3.2 $m^2$ $m^{-2}$). At EucFACE LAI ranged from 1.2 to 2.1 over a 28-month measurement period (Duursma et al., (2016), but LPJ-GUESS had very large fluctuations in annual LAI outside of these ranges (Fig. S1). These models are well documented and investigated VDMs, with many studies that have looked into parameter uncertainty (see Supplemental Text A for select references that explore model/parameter sensitivity).

In the supplements we added site-level observations and model validation for both models, and both sites:

**Table S2.** Comparison of *in situ* observations and baseline model simulations from ED2 and LPJ-GUESS for the two example study sites, Palo Verde in Costa Rica (Kalacska et al., 2005; Xu et al., 2016) and EucFACE in Australia (Medyln et al., 2016; Duursma et al., 2016). Mean and ± standard deviation.

| | Palo Verde Costa Rica | EucFACE Australia |
|---|---|---|
| Obs. Biomass (kgC $m^{-2}$) | 11.0 (5.2) | 12.7 (4.5) |
| ED2 Biomass (kgC $m^{-2}$) | 11.7 (0.3) | 5.6 (0.3) |
| LPJ-GUESS Biomass (kgC $m^{-2}$) | 10.4 (0.2) | 12.1 (0.2) |
| Obs. LAI ($m^2$ $m^{-2}$) | 3.8 (1.06) | 1.7 (0.6) |
| ED2 LAI ($m^2$ $m^{-2}$) | 3.3 (0.1) | 1.6 (0.2) |
| LPJ-GUESS LAI ($m^2$ $m^{-2}$) | 4.5 (0.1) | 3.2 (1.3) |

4. **R1: Reporting uncertainties in model parameters.**

Answer: The ED2 and LPJ-GUESS models are well documented and investigated VDMs, with many previous studies that have looked into parameter uncertainty. For example, we know that parameters related to plant hydraulics and non-structural carbohydrate storage have large uncertainties and thus included these parameters in the overview Table 1. We updated the supplemental material (Supplemental Text A: “Review of Model Parameter Uncertainty”) to now reference additional papers that have done more in-depth investigations of specific model parameters with the largest uncertainties within each of the models, and list which processes or mechanisms were evaluated.  We updated the Supplements to include these references: Oberpriller et al., 2022; Zaehle et al., 2005; Pappas et al., 2013; Jiang et al., 2012; Shiklomanov et al., 2020; Viskari et al., 2019 for more description of parameter uncertainty.

In Section 4 we more clearly state that large losses in carbon are likely linked to uncertainties in how we currently represent plant hydraulics (section 4.1.2), non-structural carbohydrate storage (section 4.1.4), or phenology diversity (section 4.1.1). We also revised the manuscript to emphasize that a goal of this paper is to demonstrate how to use these VDMs in order to help generate future hypotheses about UCEs.

Therefore, we used the models and sites as conceptual “experimental” tools to investigate the given hypotheses.

The aim of the paper was not to do in-depth, detailed model experiments with tuned parameters specific to each site, but rather to set up a general modeling framework so that hypotheses about unprecedented climate extremes can be investigated, and to provide understanding on how model behavior of physiological and ecological processes might be lacking in state-of-the-art ecosystem models in order to capture extremes. In the discussion we provide a review of where/how models could be improved in future studies, in which we also guide where future studies could look into better process representation and parameter uncertainty.

We were glad to hear that R1 thought the paper was a “very informative read” and that “the presented framework is practical and logical”. As well as “the manuscript is in itself coherent, well-written, the structure is easy to follow with the aid of straightforward visualizations and tables.”

**R2 expressed concerns about:**

1) **R2: It is not clear whether the main focus of this study is to introduce the newest model development or review what is missing in the current VDMs and suggest future directions, or even the combination of both.**

Answer: We revised the manuscript to clarify that our goal is to highlight what is missing in current VDMs, even after taking into account cutting edge model developments. In order to review what is lacking in current VDMs, we needed to describe the existing model frameworks and latest model developments (which might have been a little confusing). We clarified this in the revision.
We also revised the manuscript to state more clearly at the beginning that this is more of a review and “guidebook” manuscript, but informed by some original research that allowed us to address novel forest responses to climate extremes. We believe this combination of original research with a review of current limitations of models is a strong novelty in and of itself.

In the abstract we added this new text: “Here, we present a road map of how two dynamic vegetation demographic models (VDMs) can be used to investigate hypotheses surrounding ecosystem responses to UCEs (e.g., unprecedented droughts).”

2) **R2: Lack of model validation and comparisons with the observation/data, and validating against some historic drought events at these sites.**

Answer: We agree that model comparison to some site-level observations is a worthy improvement to the paper. As discussed above, we now include site level benchmarking observations such as biomass and leaf area index for basic model validation. With regards to validating the models against historic drought events at these sites, as also pointed out by reviewer #2, we agree that this would be a logical and excellent way of testing the model outputs in another paper. It is not easy to track the effects of a drought event in reality when "extreme" events rarely coincide with field campaigns, and do not allow control of other potentially interacting factors. The goal of this paper was to explore how models represent ecosystem responses to *extreme* droughts, and identify how they are different from less intense droughts, as the responses can be nonlinear.

In section 2 of the manuscript the following text is included: "Since field data needed to evaluate UCE responses are, by definition, unavailable, we do not perform model-data comparisons. Rather, we use the model results and conceptual framework as a road map to explore our hypotheses and illustrate their implications for ecosystem responses under UCEs, not historical drought events."

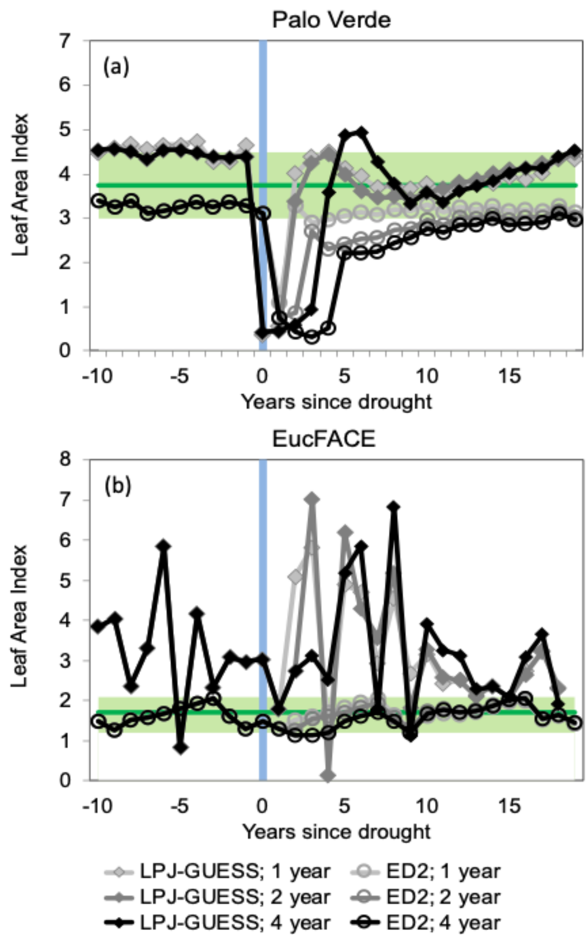
**Table S2.** Comparison of *in situ* observations and baseline model simulations from ED2 and LPJ-GUESS for the two example study sites, Palo Verde in Costa Rica (Kalacska et al., 2005; Xu et al., 2016) and EucFACE in Australia (Medyln et al., 2016; Duursma et al., 2016). Mean and ± standard deviation.

| | Palo Verde Costa Rica | EucFACE Australia |
|---|---|---|
| Obs. Biomass (kgC $m^{-2}$) | 11.0 (5.2) | 12.7 (4.5) |
| ED2 Biomass (kgC $m^{-2}$) | 11.7 (0.3) | 5.6 (0.3) |
| LPJ-GUESS Biomass (kgC $m^{-2}$) | 10.4 (0.2) | 12.1 (0.2) |
| Obs. LAI ($m^2$ $m^{-2}$) | 3.8 (1.06) | 1.7 (0.6) |
| ED2 LAI ($m^2$ $m^{-2}$) | 3.3 (0.1) | 1.6 (0.2) |
| LPJ-GUESS LAI ($m^2$ $m^{-2}$) | 4.5 (0.1) | 3.2 (1.3) |

3) **R2: LPJ-GUESS showed large swings in LAI at EucFACE site, and R2 is wondering if this is reasonable?**

Answer: In the revised manuscript, we provided observational data of leaf area index (LAI) at the EucFACE site and compared against the large swings from the LPJ-GUESS model (Figure S1a,b). We found that the large swings in the model are *not* typical of what has been measured at the Australian field site, and are a potential source of model biases/uncertainty. Since this paper is more of a guidebook for how we can improve models, this is exactly the type of discrepancy between model and observations we would like to uncover, and fix in future studies.

We updated the discussion to include: "This capability of large swings in LAI (5.8 to 0.8) by LPJ-GUESS could contribute to model uncertainty and the considerable mortality response at EucFACE.  Modeled LAI was the largest source of variability in another ecosystem model, CABLE, when evaluating the simulated response to $CO_2$ fertilization (Li et al., 2018)."




4) **R2: Question about why these two specific study sites are selected? (And why no site level tuning was done).**

Answer: We choose these two forested sites in Australia and Costa Rica mainly because the models have already been run at these sites in previously published papers (Xu et. al., 2016; Medlyn et. al., 2016), thus making model application possible with a "built in" reasonable degree of validation. These sites also contain multiple measurements and information that allowed for previous papers to validate the models. These sites were also chosen because they span an interesting range of vegetation types (a temperate-subtropical transitional forest in EucFACE and a seasonally dry tropical forest at PaloVerd), and are in warm, seasonally dry climates that are more likely to experience droughts in the future. In the revised manuscript we include in Section 2.2 a description of why the sites were chosen and reference the publications where model validation has already occurred. We also emphasize that the purpose of this paper was not to do site comparisons among many different sites, but rather to do some specific hypothesis testing at the best sites available.

New text: "To exemplify how VDMs can be tools to explore new hypotheses related to UCEs we applied the models at two field sites, that were chosen due to being extensively studied and the models used here have already been run at these sites and previously benchmarked against field data (see Xu et al., 2016; Medlyn et al., 2016; Medvigy et al., 2019 for model-data validation). The purpose of this paper was not to do a large multi-site comparison, but rather just select a few for hypothesis testing. In addition, the two sites span a range of vegetation types and are in warm, seasonally dry climates that are more likely to experience droughts in the future (Allen et al., 2017)."

"The two models were previously tuned for each site (Xu et al., 2016; Medlyn et al., 2016), and no additional site-level parameter tuning was conducted here due to evaluating responses from hypothetical UCEs."

5) **R2: Would like to see a tighter connection between the simulation results and discussion section which describes what is lacking in VDMs, by providing more detailed descriptions of the two models.**

Answer: In the revised manuscript, we improved the text about tighter linkages between concepts, hypotheses, and model outcomes. We renamed section 2.3 in the methods to be: "2.3 Linking concepts, hypotheses, and model outcomes", where we describe the concepts to explore how UCEs are modulated by climate change in the model results.

In addition, in Section 4 we more clearly state that large losses in carbon are likely linked to uncertainties in how we currently represent plant hydraulics (section 4.1.2), non-structural carbohydrate storage (section 4.1.4), or phenology diversity (section 4.1.1). We also revised the manuscript to emphasize that a goal of this paper is to demonstrate how to use these VDMs in order to help generate future hypotheses about UCEs. Therefore, we used the models and sites as conceptual "experimental" tools to investigate the given hypotheses. In the supplemental material (Supplemental Text A: "Review of Model Parameter Uncertainty") we now reference additional papers that have done more through investigations of specific model processes within each of the models, and list which processes or mechanisms were evaluated.

We updated the Supplements to include these references: Oberpriller et al., 2022; Zaehle et al., 2005; Pappas et al., 2013; Jiang et al., 2012; Shiklomanov et al., 2020; Viskari et al., 2019 for more description of the two models.

We are glad to hear that R2 thought the paper was “clearly written and will be an interesting topic to the readers of Biogeosciences.” As well as, “In the discussion, what is missing in the current VDMs are thoroughly reviewed”, which, as we listed above in our answer to concern #1, is the goal of this paper.