We would like to thank the reviewers for their constructive feedback and appreciate their contribution to improving the article. Please see below for answers to the specific points raised in the reviewer comments. Reviewer comments are listed in italics and our responses are shown in bold font.

Reviewer 1 response

The current study shows where the abrupt shift likely to happen but less on how. I am looking for some mechanistic explanations on model difference in abrupt shift (AS) identification – some are concentrated, some scattered, some none? One explanation proposed is that the larger internal variability of model led to scattered distribution of AS, perhaps need to present data to show it? In addition, do we expect to see more/less abrupt shift at regions with larger internal variability at all? Any other potential reasons that can explain inter-model variation.

<u>Reply</u>: We would expect detected abrupt shifts to be localised clusters as neighbouring grid points will experience similar climate changes. More abrupt shifts will be detected for models with higher internal variability, but we attempt to remove as many of these as possible through the final abrupt shift criterion. For models with high internal variability (such as EC-Earth3-Veg), the detected abrupt shifts will depend on the threshold much more than models with low internal variability. We will show in a new figure in the Supplementary Material (Figure S3) how the results depend on the threshold choice specifically for EC-Earth3-Veg.

We will also add the following text to the Results and Discussion to address the possible mechanisms behind the differences in detected abrupt shifts:

> **"Differences between modelled vegetation dieback arise for multiple reasons. Although there is a reduced spread in the CMIP6 model generation, ESMs continue to project different regional climate changes over Amazonia (Parsons, 2020). Even for the same climate change, models produce a range of tropical forest responses, such as different sensitivities to drying (which is affected by assumptions concerning the rootdepth of tropical trees), different responses to warming (controlled through different optimum photosynthesis temperatures), and different representations of climate sensitive disturbance processes (e.g. fires (Table 1)).**
>
> **The assumed optimum temperature for photosynthesis has been highlighted as a particularly important factor in mediating the response of tropical forests to climate change (Booth et al., 2011). The vegetation components of ESMs often also have different responses to a given increase in atmospheric CO_2 (Wenzel et al., 2016).**
> **The direct physiological effects of CO_2 on the rate of plant photosynthesis and on plant water use efficiency typically counteract the negative impact of climate change on tropical forests (Betts et al., 2004). As a result, the extent of CO_2 fertilization is another important difference across the models (Ramming, 2010).**
>
> **Abrupt shifts are driven by stochastic variations in each model, which can be either interannually-generated climate variability or the randomness of disturbance events (such as fire), which is assumed in some vegetation models. Where this stochastic forcing is relatively small, the detected abrupt shifts will tend to be spatially coherent and determined by the underlying large-scale patterns of climate change. However, in models where the stochastic**

forcing is more significant (e.g. EC-Earth3-Veg), detected abrupt shifts tend to be much less spatially coherent. Under these circumstances the detection of an abrupt shift is more dependent on the threshold chosen (see Supplementary Material Figure S3).”

The author demonstrated that diebacks happen at places where there are higher temperature sensitivity of seasonal temperature amplitude – which is further regarded as an early-warning signal (EWS). I am curious that how early could the EWS work, or do we really see higher predictive accuracy of dieback if we use EWS. I am curious whether the ESW is the precursor for the dieback or ESW is caused by the dieback due to climate-vegetation feedback?

Reply: For the purposes of this study, we primarily focus on demonstrating that grid points with higher sensitivities of seasonal temperature amplitude to global warming are more likely to feature a future abrupt dieback, as shown by the tipping risk in Figure 4i. How much forewarning such an indicator could give is a relevant question but is outside the scope of this study. We will therefore tone down our reference to a possible EWS. Climate-vegetation feedbacks will indeed influence the indicator after a dieback event, as can be seen in Figure 3. However, it is important to note that the indicator only uses data up to a doubling of CO_2 and 91% of detected abrupt shifts occur after a doubling of CO_2.

Other comments:

the reason to use 1%. It has been argued it is idealized to use 1% CO_2 simulations, though it is not clear to me how that would be “ideal”

Reply: The 1% CO_2 runs are “idealised” in the sense that they do not take into account socioeconomic factors such as land use changes. For the purposes of our study, this allows us to focus specifically on abrupt changes that are driven by cimate change rather than anthrpogenic deforestation.

Other than dieback, the authors also present other combinations of AS and trendy changes. Though they are less common, I am wondering if the authors need to provide some mechanistic explanations…or I would suggest removing those as they might be distractive.

Reply: We agree and so for clarity and a more focused view of abrupt dieback shifts we will remove the other combinations of AS and trend changes, represented by blue, orange and green points in Figure 1a-h, from the analysis and figures.

Is it possible the key piece of evidence supporting ESW (Fig. 4i) mostly come from one model - TaiESM1. The model has the largest number of valid samples for the analysis, but quite few other models – Samu-UNICO, EC-Earth3, does not such the effectiveness of the ESW. How robust it is if we bootstrap model, or normalize result by valid samples. It linked back to my first major concern that why models show different results.

Reply: Thank you for raising this point. While it is true that some of the models do contribute more to the results shown in Figure 4i than others, it is important to note that for four of the seven models analysed in this study there is a clear threshold in the sensitivity to global warming beyond which we only observe dieback shifts (Figure 4b, d, e, f). Meanwhile, the models which do not exhibit dieback

shifts in the NSA region (MPI-ESM1-2-LR and UKESM1-0-LL) both show low overall sensitivities of the seasonal temperature cycle to global warming. This is demonstrated in the figure below, which will be included in the Supplementary material (Figure S4):

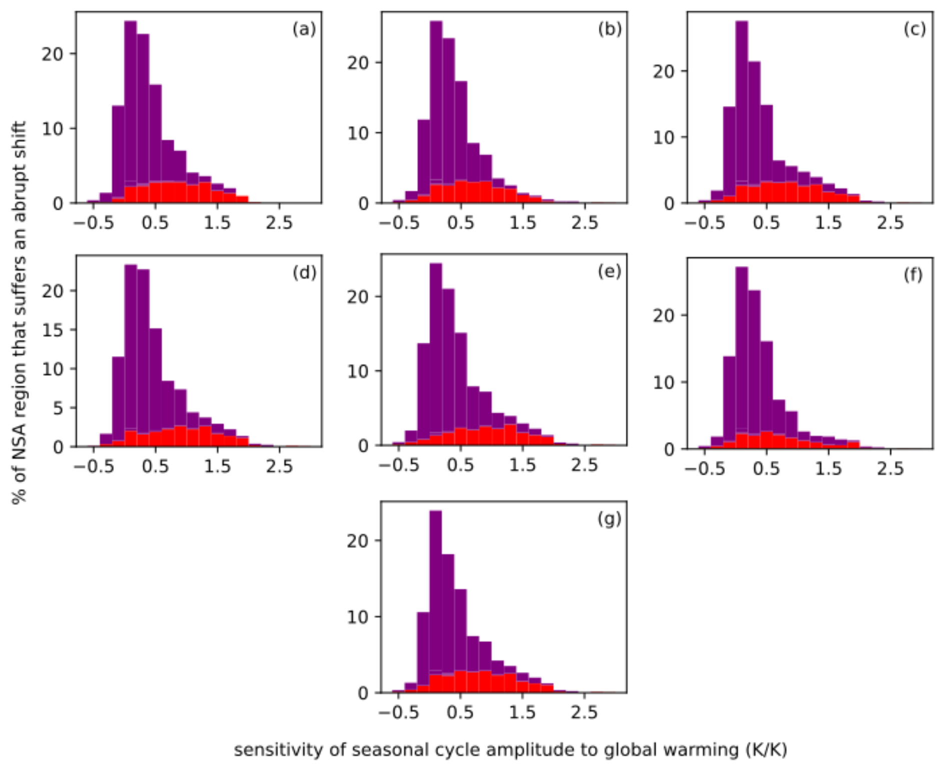

Figure: The compiled model histograms from Figure 4, each with one model removed from analysis to show how this changes the results. Models were removed from each panel as follows: (a) EC-Earth3-Veg (b) GFDL-ESM4 (c) MPI-ESM4-2-LR (d) NorCPM1 (e) SAM0-UNICON (f) TaiESM1 (g) UKESM1-0-LL.

L96. "many abrupt shifts" – perhaps provide a more quantitative statement.

Reply: To provide a more quantitative statement, the sentence on line 96 will be reworded to provide as follows:

"Compared to GFDL-ESM4, EC-Earth3-Veg has approximately 10% fewer abrupt shifts which are more scattered across the Amazon basin."

Figure 3. y axis and caption, what is cVeg? Those are good examples. Is it possible to get a scatter plot of the timings of EWS and dieback for all pixels?

Reply: For clarification we will change the y axis and caption of Figure 3 to read *Vegetation Carbon* instead of *cVeg*. As previously mentioned, we intend to tone down our references to a potential Early Warning Signal.

L115. Regional scale means “region average”?

<u>Reply</u>: Yes, the regional scale refers to the average of the NSA region. For clarity we rewrite the sentence as:

“Interestingly, when inspecting the NSA regional average abrupt changes are not obvious, despite a significant number of local abrupt shifts (see Fig. 2c).”

L175. stufy – study?

<u>Reply</u>: Thanks for spotting this. Correction of ‘stufy’ to ‘study’ on line 175 *.*

Reviewer 2 response

My main issue is in regards to the use of temperature seasonal cycle amplitude as an EWS. It is not clear to me what the criteria for the EWS would be. For instance, how much of a rise in the amplitude needs to occur to indicate an approach to a tipping point? Is the increase significant compared to other localised increases in the signal?

<u>Reply</u>: For the purposes of this study, we primarily focus on demonstrating that grid points with higher sensitivities of seasonal temperature amplitude to global warming are more likely to feature a future abrupt dieback, as shown by the tipping risk in Figure 4i. We do not use the indicator to predict the timing of a tipping point, which would require setting a threshold for rise in amplitude. Future research is required on the applicability of this indicator as an Early Warning Signal (EWS) and therefore we will tone down our references to a new potential EWS in this paper.

Throughout the manuscript samples are used to illustrate behavior (see for instance Figure 1h and Figure 3). It is not explained how these samples are chosen and if they exhibit characteristic behavior of some class of grid points (e.g., those that exhibit a negative abrupt shift). Please consider being more transparent regarding the sample selection and how each sample compares to similar grid points in the respective models. This will help the reader to draw appropriate conclusions (and/or not draw incorrect conclusions) of the general model behavior.

<u>Reply</u>: Thank you for raising this point, the samples mentioned by the reviewer were examples chosen to illustrate several different forms dieback behaviour can take in the models. This will be clarified by editing the figure 3 caption as follows: *Example time series of identified dieback shifts for three model…*

Line 122 mentions a critical threshold of CO_2 but then this is not further discussed in regards to the examples shown. Can you draw any connection to the behavior seen in the models? I would suggest to either make the discussion of this more explicit or leave it out, as it seems out of place currently

<u>Reply:</u> Agreed. As suggested, we will leave out the term to allow a better flow of the discussion.

There are quite a few abbreviations that are not explicitly defined or explained for the reader not familiar with the data and methods. For instance, lines 56 and 57 use terms "1pctCO2" and "PIControl" which although I was able to discern what they probably refer to, it would be better for the reader if these were explained. Also when discussing sensitivity increases the units K/K are used. Is this Kelvin per Kelvin? I don't quite understand the units here.

<u>Reply</u>: We will address these concerns by writing explicitly that 1pctCO2 refers to a run where CO2 is increased by 1% per year and PIControl refers to a control run with a fixed pre-industrial CO_2 concentration. The units K/K are indeed Kelvin per Kelvin and result from the sensitivity being the gradient of a linear regression fit to the amplitude of the temperature seasonal cycle against global warming, both of which have units of Kelvin.

Line 79 - The colors red and purple are mentioned with no reference to a figure.

<u>Reply</u>: We will clarify this in text by referencing Figure 4 on line 79.

Figure 3 caption - Should black squares be black crosses?

<u>Reply</u>: Thank you for spotting this. We will correct the caption for Figure 1 to black crosses instead of squares.

Reviewer 3 response

To give one example, the authors mention that some of the diebacks are mediated by fire, but from reading the manuscript it is unclear which models predicted dieback due to fire, which models predicted dieback due to physiological/hydraulic failure, or even if the models that resulted in no dieback were also the ones that did not have fire.

<u>Reply</u>: Information on which models have fire will be given in Table 1 and be added to the discussion.

> **"Some models have fire present, such as EC-Earth3-Veg, GFDL-ESM4 and MPI-ESM1-2-LR, while others do not (e.g., UKESM1-0-LL). It is interesting to note that UKESM1-0-LL, which experiences no dieback shifts, also has no fires simulated within the model. However, the role that fire plays in inducing vegetation dieback in these models requires further experimentation and work."**

Introduction. There are many studies that investigated the risk of critical transitions specifically in the Amazon, and it may be worth including them to provide a stronger motivation for this work. As a starting point, the recent Amazon assessment report (https://www.theamazonwewant.org/amazon-assessment-report-2021/, chapter 24) has an extensive review on this subject.

<u>Reply</u>: In order to provide a stronger motivation for the work the paragraph starting on line 49 will be expanded to read:

> **"A recent study, using CMIP5 models, determined that Amazon dieback, under the high emissions scenario RCP8.5, is not likely to occur in the 21st century but recognises that an increase in anthropogenic deforestation could bring the Amazon closer to a dieback event (Chai, 2021). Meanwhile, other studies predict the Amazon to have a low resilience of forest to climate change, coinciding with human pressures such as deforestation (Hirota, 2011). There is remaining uncertainty associated with the likelihood of a dieback event occurring, stemming largely from uncertainty in the effects of important factors such as the extent of CO_2 fertilization and soil nutrient limitations (Ramming, 2010; Hirota, 2021). In this paper we look at the projections from the latest CMIP6 Earth System Models for evidence of Amazon dieback and identify a precursor which is based on longer-term temperature records."**

L49-51. This paragraph could be expanded to provide a stronger motivation for this study. The authors could justify why the current analysis is necessary, and how this study contributes to learning something new about the future of the Amazon to climate change.

<u>Reply</u>: As suggested we have expanded this paragraph. Please see the answer above.

Table 1. The authors could expand this table to provide a bit more information of the simulations and models used. For example, they could provide some information about the models (e.g., which ones had fires enabled, which other mechanisms cause mortality and biomass loss), and which variants were used from each model. If citations exist for these models, the authors could add references.

<u>Reply</u>: As suggested, Table 1 will be expanded to include the land surface model, the Transient Climate Response, precipitation change at 2xCO2 and whether fire is represented in the model, as shown below:

Model	Institute	Land Surface Model	Transient Climate Response (TCR) (°C)	Precipitation change to a doubling of CO2 (mm/year)	Fire simulated
EC-Earth3-Veg	EC-Earth-Consortium	HTESSEL	2.6	-139	Yes
GFDL-ESM4	NOAA-GFDL	LM4.1	1.6	-60	Yes
MPI-ESM1-2-LR	Max-Planck-Institut für Meteorologie	JSBACH 3.2	1.6	-57	Yes
NorCPM1	EarthClim	CLM4.0	1.6	-76	-
TaiESM1	AS-RCEC	CLM4.0	2.3	-60	-
SAM0-UNICON	Seoul National University	CLM4.0	2.3	-282	-
UKESM1-0-LL	Met Office Hadley Centre	JULES-ES-1.0	2.6	-196	No

Section 2.3. I found the early warning section disconnected from the introduction. In the introduction the authors describe the theory of critical transitions, critical slowing down and increase in autocorrelation, yet none of these seem to be used in the actual analysis. Was there any reason for not using these established approaches?

<u>Reply</u>: To address this point we will include supplementary material (Figure S1) which focuses on trialing these generic early warning signals for Amazonia in CMIP6 models. This analysis involved calculating the autocorrelation and variance for vegetation carbon for each grid point and found little evidence that these metrics would work as an early warning signal for these models.

Section 3.1. The authors use savannah as the alternate state for forests, and this can be misleading if fire is not the driver for abrupt changes. Also, this section describes the changes across tropical South America, but the authors do not provide any insight on what causes the variability across models. Presumably they also have broad range of predicted climate, presence/absence of fires, and different approaches to simulate drought mortality. Explaining these differences could help us understanding why there was such broad range in dieback responses.

We will add the following text to the Results and Discussion to address the possible mechanisms behind the differences in detected abrupt shifts:

> **"Differences between modelled vegetation dieback arise for multiple reasons. Although there is a reduced spread in the CMIP6 model generation, ESMs continue to project different regional climate changes over Amazonia (Parsons, 2020). Even for the same climate change, models produce a range of tropical forest responses, such as different sensitivities to drying (which is affected by assumptions concerning the rootdepth of tropical trees), different responses to warming (controlled through different optimum photosynthesis temperatures), and different representations of climate sensitive disturbance processes (e.g., fires (Table 1)).**

The assumed optimum temperature for photosynthesis has been highlighted as a particularly important factor in mediating the response of tropical forests to climate change (Booth et al., 2011). The vegetation components of ESMs often also have different responses to a given increase in atmospheric CO_2 (Wenzel et al., 2016). The direct physiological effects of CO_2 on the rate of plant photosynthesis and on plant water use efficiency typically counteract the negative impact of climate change on tropical forests (Betts et al., 2004). As a result, the extent of CO_2 fertilization is another important difference across the models (Ramming, 2010).

Abrupt shifts are driven by stochastic variations in each model, which can be either interannually-generated climate variability or the randomness of disturbance events (such as fire), which is assumed in some vegetation models. Where this stochastic forcing is relatively small, the detected abrupt shifts will tend to be spatially coherent and determined by the underlying large-scale patterns of climate change. However, in models where the stochastic forcing is more significant (e.g., EC-Earth3-Veg), detected abrupt shifts tend to be much less spatially coherent. Under these circumstances the detection of an abrupt shift is more dependent on the threshold chosen (see Supplementary Material Figure S3).”

L115–118. My previous comment applies here too. The remark that models largely disagree is correct, but not very informative. I would not expect the authors to provide details about every model configuration and formulation, but I think they could explore some potential causes by looking at other model output data (at the very least precipitation, some insight on model sensitivity to CO2, and some fire and mortality-related variables if available).

Reply: We will include information on climate sensitivity, precipitation change under global warming and the representation (or otherwise) of fires for each of the models within table 1 (see above) and also further discussion on the potential causes of differences between the models (see response above).

Section 3.3. I am unable to see the causal link between CO2 and the abrupt shift based solely on Figure 3. If anything, in GFDL most of the shift in the amplitude of the seasonal cycle seems to occur after the abrupt transition. Also, the difference in the seasonal cycle is very large across models, with NorCPM1 remaining below 4°C for the entire century, whereas the other models show much higher amplitudes. Is it fair to treat the shifts marked in Fig. 3 the same?

Reply: Figure 3 is used to show that the amplitude of the seasonal cycle increases before an abrupt transition in three examples from different models, with potentially different causes of dieback. This figure is used to motivate our investigation to see if these increases are observed at all grid points or primarily grid points that possess a future abrupt transition (Figure 4). We will make this clearer in our revised manuscript.

Discussion: I support keeping the discussion short, but maybe the current one is a bit too short and narrow in scope. For example, the authors mention that the dieback was present in previous generations *of model but not in CMIP6. Why is this the case? Also, how does this result differ from the analysis by Cox et al. (2013) (https://doi.org/10.1038/nature11882), which had already indicated lower risk of a dieback. Also, the results implied that fire is an important mechanism for dieback, and I think the discussion could emphasise this further, considering that fire activity has significantly increased recently in the Amazon. The mechanistic links between increased CO2 and dieback (and the uncertainty in these links) could be discussed in more depth too. I think addressing some of these aspects would help placing these interesting results from the CMIP6 predictions in a broader context.*

Reply: The analysis by Cox et al. (2013) concerned a proposed emergent constraint on the loss of tropical land carbon with global warming. It did not specifically look at abrupt changes in forest cover. While our study shows little evidence of abrupt large-scale land carbon loss in CMIP6, it does find evidence for multiple localised abrupt shifts. We make this distinction clearer by modifying our title to: "Evidence of localised Amazon rainforest dieback in CMIP6 models".

We also include additional discussion about the reasons for the differences between models, and between model generations in the Results and Discussion section (see above): "Differences between modelled vegetation dieback arise for multiple reasons. Although there is a reduced spread in the CMIP6 model generation, ESMs continue to project different regional climate changes over Amazonia (Parsons, 2020). Even for the same climate change, models produce a range of tropical forest responses, such as different sensitivities to drying (which is affected by assumptions concerning the rootdepth of tropical trees), different responses to warming (controlled through different optimum photosynthesis temperatures), and different representations of climate sensitive disturbance processes (e.g. fires (Table 1)).

The assumed optimum temperature for photosynthesis has been highlighted as a particularly important factor in mediating the response of tropical forests to climate change (Booth et al., 2011). The vegetation components of ESMs often also have different responses to a given increase in atmospheric CO2 (Wenzel et al., 2016). The direct physiological effects of CO2 on the rate of plant photosynthesis and on plant water use efficiency typically counteract the negative impact of climate change on tropical forests (Betts et al., 2004). As a result, the extent of CO2 fertilization is another important difference cross the models (Ramming, 2010).

Abrupt shifts are driven by stochastic variations in each model, which can be either interannually-generated climate variability or the randomness of disturbance events (such as fire), which is assumed in some vegetation models. Where this stochastic forcing is relatively small, the detected abrupt shifts will tend to be spatially coherent and determined by the underlying large-scale patterns of climate change. However, in models where ths stochastic forcing is more significant (e.g. EC-Earth3-Veg), detected abrupt shifts tend to be much less spatially coherent. Under these circumstances the detection of an abrupt shift is more dependent on the threshold chosen (see Supplementary Material Figure S3)."

Minor Points

L9. I am not sure I followed this sentence. Does this mean that an additional 7% of the NSA will experience dieback for every 1°C above 1.5°C (i.e., if the temperature change is 2.5°C, dieback will occur at 14% of the NSA, if the change is 3.5°C, 21% of the NSA will suffer dieback and so on?)

<u>Reply</u>: Regarding the statement on line 9, the reviewer is correct in thinking that this refers to an additional 7% (with an error margin of 5%) of the NSA region experiencing dieback per 1°C above 1.5°C. For clarification we will change the sentence to read:

> Based on the ensemble mean of the detected dieback events we estimate that 7+/-5% of the NSA region will experience abrupt downward shifts in vegetation carbon for every degree of global warming past 1.5°C.

L49. Quantify "fairly short observational records"

<u>Reply</u>: The sentence on line 49 will be changed to include a more quantitative statement as below:

> **"These recent studies (Boulton et al., 2022; Luo and Keenan, 2022) focus on fairly short observational records of less than 60 years."**

L55. Include references that describe both CMIP6 and the 1pctCO2 runs.

<u>Reply</u>: As requested, the references for both CMIP6 and the 1pctCO2 runs will be included:

Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S. and Stevens, B.: Climate Model Intercomparisons: Preparing for the Next Phase, EOS, Transactions American Geophysical Union, 95(9), 77–78, doi:10.1002/2014eo090001, 2014.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9(5), 1937–1958, doi:10.5194/gmd-9-1937-2016, 2016.

L67. "Unforced control run" was not described up to this point. Consider describing it in section 2.1

<u>Reply</u>: To address this point we will clarify the PIControl run as an unforced control run in section 2.1, line 56 as below:

> **"Data from the unforced PIControl runs were also used to determine each model's internal variability."**

L78–81. This seems to be out of place, it reads more like the caption of Figure 4 (which is referred to before Figures 1–3). Perhaps rewrite this to focus on how sensitivity and dieback risk were calculated.

<u>Reply</u>: As suggested, the text in lines 78-81 will be rewritten as follows:

> **"To assess the risk of an abrupt dieback shift occurring the percentage of grid points that experience abrupt dieback out of all grid points with sensitivities within a specified range is calculated. This gives a measure of how likely it is for a grid point with a specific sensitivity to experience a dieback event."**

L87–90. This text repeats what was described in the methods section. I suggest dropping it.

<u>Reply</u>: As suggested, we will drop the repeated text in lines 87 to 90.

L105–111. "Jumps" seems a bit too colloquial, and it is unclear how it differs from "abrupt shifts", which is defined in the methods.

<u>Reply</u>: To address this point we will rewrite the two instances where "jumps" is used to read "abrupt shifts".

L175. "Study" is misspelt.

<u>Reply</u>: Indeed – thanks for spotting. We will correct "stufy" to "study".

Figure 2. In panel (a), I suggest keeping only NSA, as this is the only specific region analysed in this study. Also, make the labels consistent with captions (e.g., use either abrupt shift or dieback shift), and define the acronyms (AS, cVeg) in the caption.

<u>Reply</u>: As suggested, we will change Figure 2(a) to only include the relevant NSA region. The acronyms *AS* and *cVeg* on the axes of Figures 2b,c will be replaced by abrupt shift and *vegetation carbon,* respectively, and *dieback shift* replaced by *abrupt shift* in the caption for consistency.

Figure 3. Why did the authors show different points for each model? Also add "W" after the last 60°.

<u>Reply</u>: The points mentioned by the reviewer were examples chosen to illustrate several different forms that dieback behaviour can take in the models. This will be clarified by editing the figure 3 caption as follows: "Example time series of selected dieback shifts for three model…"

As suggested, the caption will also be corrected to include the W after the final 60^{o}.

Figure 4. I recommend adding a colour legend to the figure.

<u>Reply</u>: We will be removing the green, blue and orange points representing non-dieback abrupt shifts in Figure 1 and changing them to purple points in Figure 4 to allow for clearer analysis. We will also include a colour legend for the red and purple points in Figure 4.

Community Comment Response

It would be great if the authors could make that code publicly available so that the larger community can benefit from it. Alternatively, I would suggest the authors to expand a bit on the exact definition of the AS detection criteria to make them completely unambiguous (e.g. how do you define the variability of the rates of change in the unforced control run? What years do you consider? Etc.). If the code cannot be made available, some illustrative examples could be helpful.

Reply: As suggested we will make the code publicly available on GitHub on the publication of this article. Additionally, the abrupt shift detection criteria will be elaborated on in the supplementary material (Figures S2 and S3) to clarify the mechanics behind the third criterion of the outlined algorithm and how it is dependent on the threshold set.