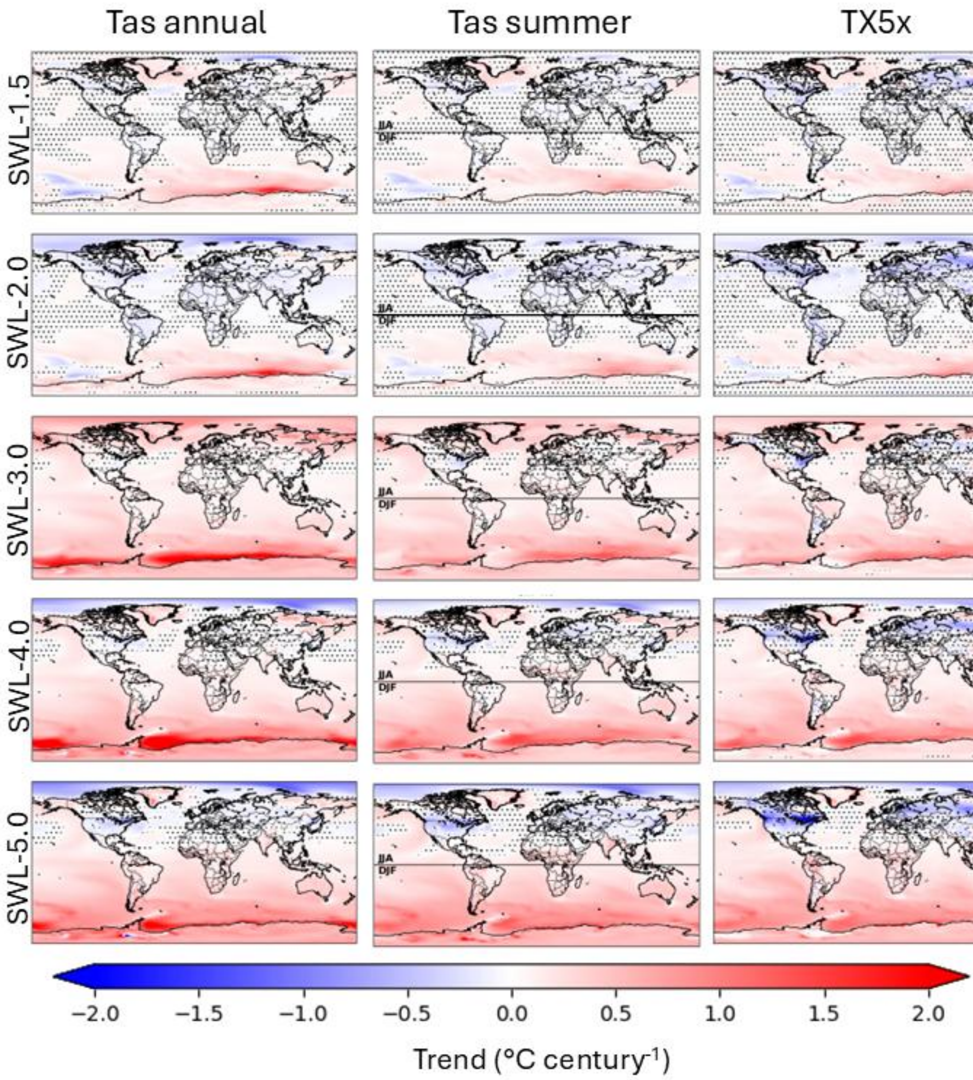
We thank the reviewer for the insightful comments and feedback. Below are our replies, with the original comments in red and the replies in black.

Major comment:

1. Figure 2 requires a statistical significance test for all the plots.

   Statistical significance has been added to all plots in figure 2, with the stippling indicating areas where trends are not significant to the 95% confidence level of a two-sided test ($p>0.05$).



2. To derive the dynamical component in the anomalous temperature pattern, the authors apply simple pattern scaling (Equation 3). I am concerned that the pattern changes with the temperature gradient, and this contribution is grouped into the thermodynamic component.

   The methodology we apply here is different from a simple pattern scaling approach, as we do not scale a time-mean pattern with different levels of global mean warming, as is typically understood by pattern scaling. Instead, we reconstruct the pattern found in the forced run by drawing individual circulation analogues from the piControl, with an analogue for each target event (TX5x, a five-day event), repeated 30 times. Therefore, the mean pattern is reconstructed from analogues for each individual 5-day event. The 'scaling factor' beta used in equation 3 is thus not a scaling factor in the usual pattern scaling sense, but a "correction factor" accounting for differences between the analogue and target patterns. As discussed lines 88-89, all the data from the net zero projections used in the constructed circulation analogues analysis are detrended and their seasonal cycle removed, thus removing the contribution from external forcing agents on temperature and geopotential height.

To improve the clarity of the method's description and the use of β, we added the following at lines 111-114:

As explained in Deser et al., (2014) and Merrifield et al., (2017), β *effectively represents an optimal set of weights used to reconstruct the target zg500 pattern from the corresponding field in the analogues.*

Nevertheless, the interpretation of the thermodynamic component is based solely on the local energy budget, and the thermodynamically derived effect of anomalous advection is not discussed.

We added more details on the limitations of the constructed circulation analogues approach used here at lines 122-130. Here we explain how the method might fail at accurately considering all relevant dynamical processes, such as advection, as the analysis is based on one atmospheric variable. We then explain how we try to overcome this limitation by comparing the results with other indicators of dynamical and thermodynamic processes.

The limitations of the constructed circulation analogues approach are thoroughly discussed in Horowitz et al., (2022). In the context of our study, the main simplification derives from the assumption of a linear relationship between the pressure and temperature fields. Thus, there is a possibility that some of the forcing component attributed to the residual might in fact derive from non-linearities in the pressure-temperature relationship. Moreover, the pressure fields used to extract the dynamical component are not complete descriptors of the state of the atmosphere, so they might be biased in the representation of some relevant processes like horizontal advection. It is also important to note that, while conceptually distinguishable (Deser et al., 2014), it is difficult to fully separate the contribution of dynamical and thermodynamic processes to heat extremes, as they constantly interact and affect each other (Fischer et al., 2007a Miralles et al., 2019). Thus, to corroborate the results of this analysis we address *them in the context of the projected trends in the pressure fields (i.e., dynamical component) and surface energy fluxes (i.e, thermodynamic component).*

3. The centrally important process in this study is the transition from needleleaf evergreen to broadleaf evergreen vegetation. Nevertheless, some fundamental information is missing in the manuscript: a) how the simulated baseline vegetation is compared to the observations, b) how the model determines whether the grid cell is covered by needleleaf evergreen and broadleaf evergreen vegetation, c) what property differences (albedo, root depth, snow masking depth, etc.) are specified for these two vegetation types, and d) how a)-c) played roles in the results. The authors do not need to provide a full description of the model, but rather sufficient detail to understand the dynamic vegetation response to climate change and the climate response to altered vegetation. Descriptions of the model resolution and the land surface model are also desired.

   We thank the reviewer for this comment and agree that more details on the dynamic vegetation model will be helpful for the readership. To address this, we made the following additions to the original text:

   a) Lines 296-303: *The model seems to capture well this migration pattern, with all regions experiencing an increase in BdlEvg vegetation bordering areas where these plants are present at the beginning of the simulation (Sellar et al., 2019).*
   Lines 333-344*: An additional source of bias can be found in the model baseline representation of global PFTs distribution. While previous studies have shown how JULES-4.2 offers a good representation of the baseline land carbon cycle and distribution of global plant functional types at a global level (Harper et al., 2018; Sellar et al., 2019), some non-negligible biases against observational records appear at the regional level (Harper et al., 2018). In particular, the model overestimates the needleleaf evergreen vegetation cover in mid-latitude and boreal forests at the expenses of broadleaf deciduous vegetation. As shown in Harper et al., (2018), this bias is present in all the regions addressed in this study, with NE North America, NE Asia, southern Chile and New Zealand all having an overestimated needleleaf vegetation cover. This bias might have influenced vegetation dynamics throughout the ramp-up and stabilised emissions runs: with a different baseline PFT to compete with (i.e., broadleaf deciduous instead of needleleaf evergreen), the migration of BdlEvg vegetation from tropical to midlatitude regions could have unfolded differently both temporally and spatially, potentially leading to different regional climatic impacts. Nonetheless, it is*

worth noting that significant uncertainties also apply to observational estimates of PFT distribution (Hartley et al., 2017) (Hartley et al., 2017).

b,c,d) Lines 251-261: *Although some association with changes in mean temperature seems apparent, other factors must be playing a role in driving the biome shifts, as the affected regions are characterized by substantially different temperature regimes. This non-linearity is consistent with the design of the dynamic vegetation component of UKESM1.2 (Clark et al., 2011). Competition between plant functional types (PFTs) is controlled by canopy height, which is ultimately influenced by air temperature since this controls the plants growth rate through the photosynthetic activity. As described in Harper et al., (2016), the parameterization scheme in JULES allows BdlEvg to achieve higher photosynthetic rates than needleleaf species when the temperature regime allows their development. This is likely due to the higher maximum rate of carboxylation assigned to BdlEvg species and their higher photosynthetic temperature regime, which favour from the warmer climate conditions found in swl-4.0 and swl-5.0. Nevertheless, several non-linear relationships control growth and expansion dynamics within and across PFTs based on climate conditions. This non-linearity hinders a straight-forward derivation of an absolute temperature threshold for vegetation changes, with more in-depth examinations laying outside the scope of this work (Cox, 2001; Harper et al., 2016).*

We also added more details on the model resolution at lines 49-50:

The model horizontal resolution in the atmosphere is approximately 135 km in midlatitudes (1.25° × 1.875°) with 85 vertical levels, while the ocean component has 1° resolution and 75 vertical levels.

4. Please clarify how the intraseasonal “tasmax” variance trend becomes more negative when the Evaporation Fraction increases. It is easy to imagine how the mean trend becomes more negative, but additional physical explanation is desired for the change in variance.

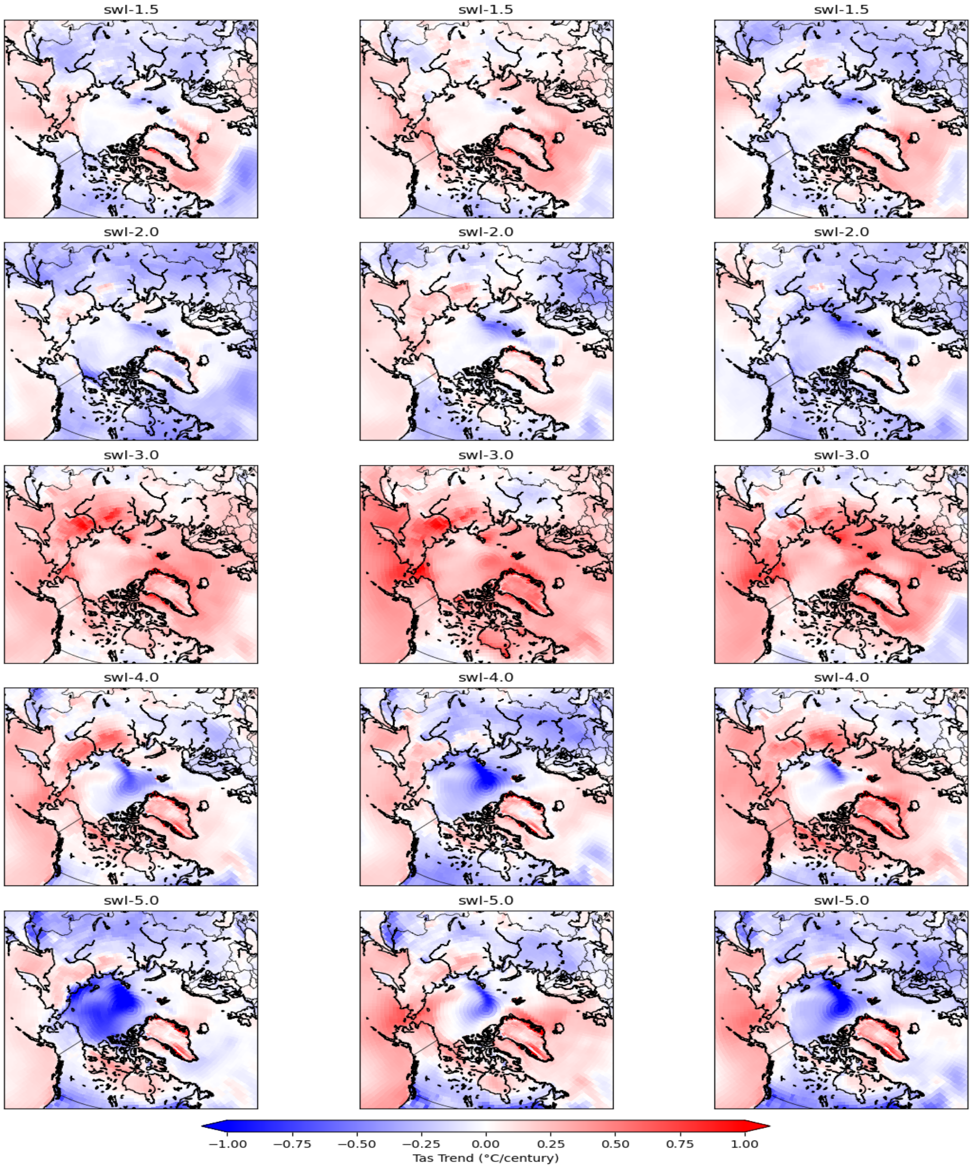
The reduction in tasmax variance is thus related to the accentuated effect of the vegetation shift on the temperature extremes compared to the mean. In terms of the mechanisms explaining this trend, we expanded the text to provide more insights into this process at lines 327-332:

The effects of the vegetation shift on the temperature variance and distribution align with the stronger cooling effect projected for the annual hot extremes than the mean *shown in Figure 2. This is likely because the higher evapotranspiration potential leads to an accentuated cooling effect during very hot days with a higher demand (Lemordant et al., 2016).*

Minor comment:

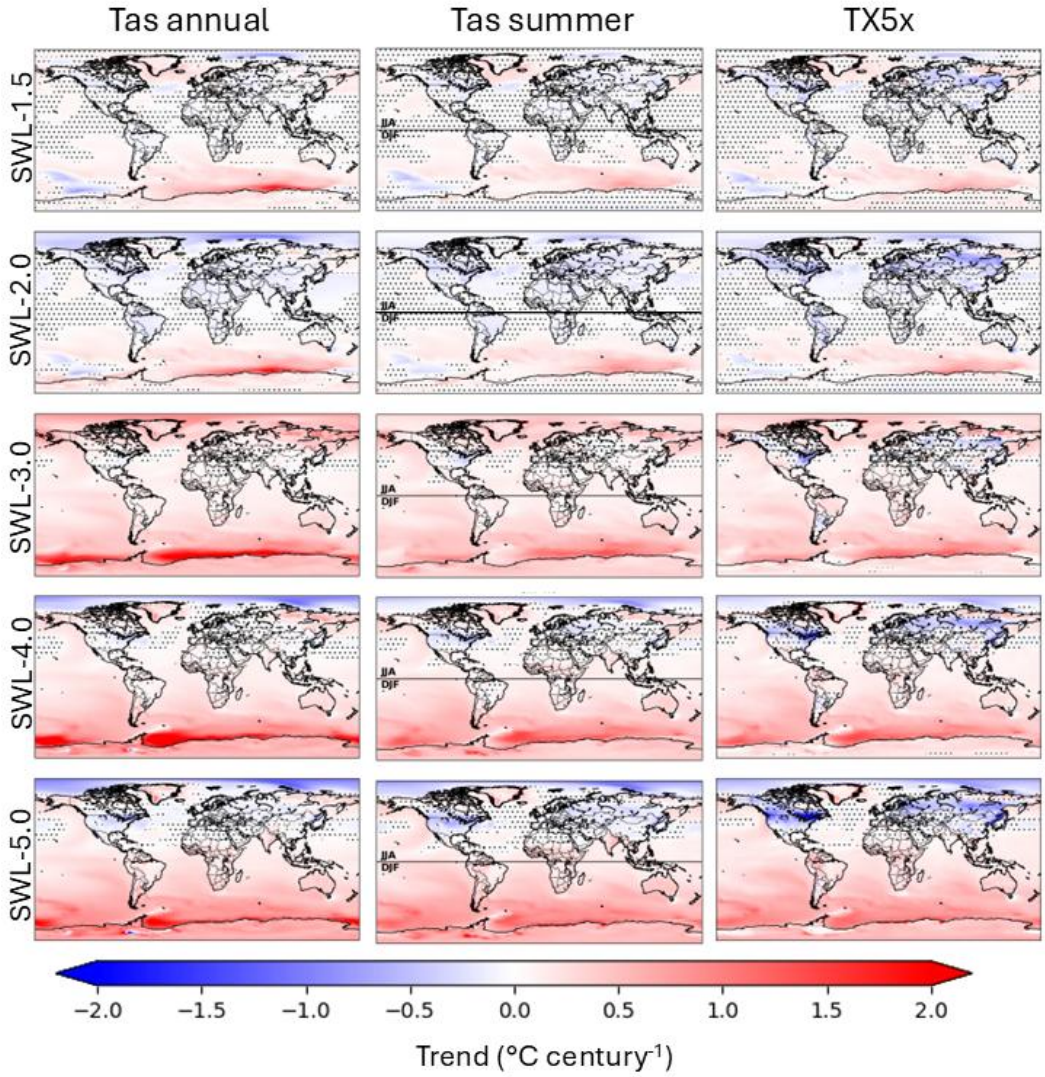
1. The trend in the central Arctic is puzzling: it is negative in SWL-2.0, -4.0, and -5.0, but positive in -3.0. Some explanation is needed here.

We agree with the reviewer about the surprising difference in temperature trends in the Arctic across scenarios. Investigating the mechanisms causing the warming seen in swl-3.0 is beyond the scope of this study. The temperature trends in the Arctic at warming levels <3.0 are likely to be sensitive to the trend length (250 years vs. longer or shorter periods) due to a relatively large role of internal variability in this region. As shown in the figure below, the strong influence of internal variability on Arctic temperature trends also leads to considerable variability among ensemble members. In ongoing separate work led by colleagues who are not involved in the present study, preliminary results suggest that the albedo feedback changes sign for warming levels greater than 3 degrees. Below 3 degrees, it is likely that the trends are more sensitive to the interplay between internal variability, clouds and sea-ice concentrations and timing of sea ice loss. It is likely that Arctic sea ice loss plays an important role in the temperature trends for the 3-degree SWL, though we cannot fully explain the Arctic temperature trends seen across warming levels with the sea-ice timeseries alone. We refer the reader to further upcoming work where radiative feedbacks and drivers of global mean temperature patterns will be analysed in more detail.



2. 143: Describing the difference between the left column and the right column in Fig. 2 is a bit awkward because the left column is for the annual mean while the right column is for the summer. It is unclear what the “substantial differences” mean when comparing the different seasons.

We agree with the reviewer that the previous version of the figure could be misleading. We now added an additional column with trends averaged for the boreal (JJA) and austral (DJF) summers in the respective hemispheres.



3. Please clarify how the decreased 2m temperature is explained by increased transpiration associated with plants, given that transpiration-derived cooling over plant leaves likely occurs above 2m.

Several studies have documented how a vegetation-induced increase in evapotranspiration rates produces a cooling effect on air surface temperatures (see, for example, Chahine (1992) or Zen et al., (2017)). As we explain at lines 214-215, transpiration produces a cooling effect on surface temperatures by increasing the fraction of radiative forcing dispersed upwards through evaporation. This effect involves both the land surface and the near surface air temperatures as their radiative balance is highly coupled. For instance, Alkama and Cescatti (2016) shows how changes in forest cover produce highly correlated and very similar responses in the global land and air surface temperatures.

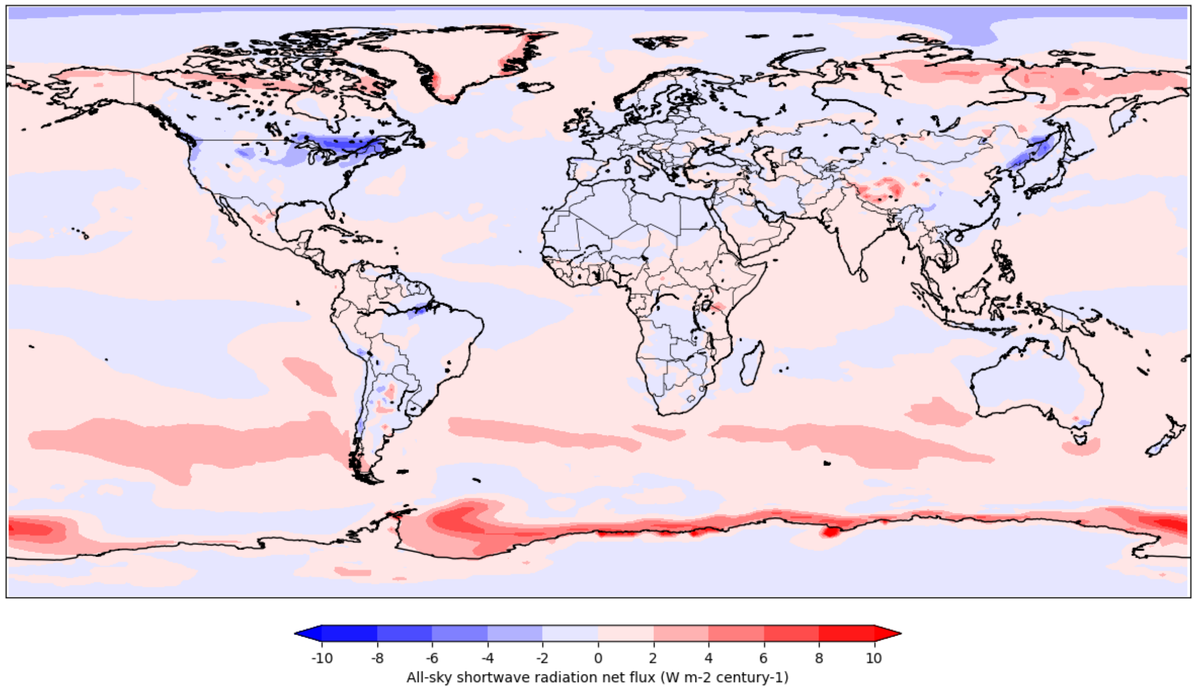
Alkama, R., & Cescatti, A. (2016). Biophysical climate impacts of recent changes in global forest cover. *Science*, *351*(6273), 600–604.

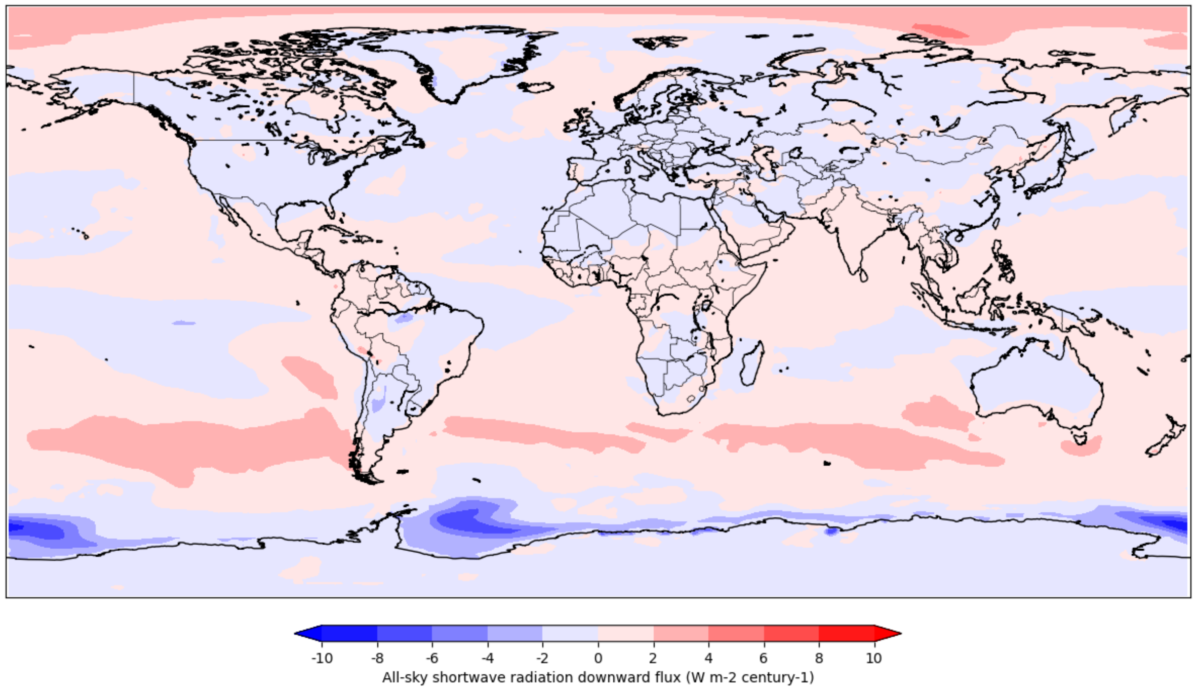
Chahine, M. T. (1992). The hydrological cycle and its influence on climate. *Nature*, *359*(6394), 373–380. https://doi.org/10.1038/359373a0

Zeng, Z., Piao, S., Li, L. Z. X., Zhou, L., Ciais, P., Wang, T., Li, Y., Lian, X., Wood, E. F., Friedlingstein, P., Mao, J., Estes, L. D., Myneni, R. B., Peng, S., Shi, X., Seneviratne, S. I., & Wang, Y. (2017). Climate mitigation from vegetation biophysical feedbacks during the past three decades. *Nature Climate Change*, *7*(6), 432–436. https://doi.org/10.1038/nclimate3299

4. The analysis relies mostly on the partition between sensible and latent heat fluxes. There are other energy fluxes, including radiation affected by clouds and water vapor. Could the authors be more thorough on this aspect?

   We thank the reviewer for this comment. As we show in Figure A2, the local surface energy budget is also affected by changes in albedo due to the vegetation shift resulting in a lower fraction of shortwave radiation absorbed. As shown in the figures below, this effect seems to be dominant over changes in cloud cover and water vapour, as the surface net all-sky radiation flux pattern largely mirrors the clear-sky pattern shown in figure A2 and no substantial changes in the downward shortwave radiation are projected. We decided to show the clear-sky pattern in the manuscript to highlight the effect on surface albedo.


-10
-8
-6
-4
-2
0
2
4
6
8
10
All-sky shortwave radiation net flux (W m-2 century-1)


-10
-8
-6
-4
-2
0
2
4
6
8
10
All-sky shortwave radiation downward flux (W m-2 century-1)