We are thankful to all reviewers for their valuable feedback which helped us to improve the manuscript. In response, aside from several minor corrections, we have introduced the following main changes to the paper:

- We have increased the amount of models considered in this study from 8 to 12.
- The ELI is now calculated with the soil moisture averaged over the top meter, which better represents effective water availability for terrestrial evaporation, as opposed to total column soil moisture.
- Hot spot region “NAS” has been moved northwards slightly and extended eastwards, as the regional pattern of largest changes in temperature excess has shifted slightly following the inclusion of additional models in the analysis.

As a result of these changes, the figures and main conclusions are even more pronounced or remain similar, which reflects the robustness of the methodology.

Using a small ensemble of CMIP6 simulations, the authors show that areas with increasing ecosystem water limitation tend to feature stronger warm season maximum temperature trends (compared to mean temperature changes). While the mechanisms behind this have long been known, most analyses focus on past changes and it is an interesting, well-designed study that I consider to be relevant for a broad audience. Nevertheless, I list a few suggestions below that could be helpful in further improving the manuscript.

Main comments

1.) I am not convinced by the choice of “mrso” to indicate root-zone soil moisture. “mrso” is simply the total column soil moisture, and the actual depth that is represented varies from model to model and can easily exceed 2 meters (Qiao et al., 2022). In the Supplementary, it becomes clear that you use ERA5-Land soil moisture down to 100 cm (first 3 layers), and I think this is a good choice as the bottom layer extending to nearly 3m depth is arguably more uncertain. However, it would probably make sense to use the very same definition for the CMIP6 models, and not rely on the column soil moisture. 1m soil moisture could be calculated by using all layers within 100 cm and adding a fraction of the respective lowermost layer (e.g., 0.5 if it extends from 80 to 120 cm).

We agree with the reviewer here. We have recomputed the ELI and remade all figures with soil moisture from layers averaged over the top meter of the soil. Using this root-zone soil moisture is more representative for the water availability that ecosystems experience. As a result of these changes, the figures and main conclusions are even more pronounced or remain similar, which reflects the robustness of the methodology.

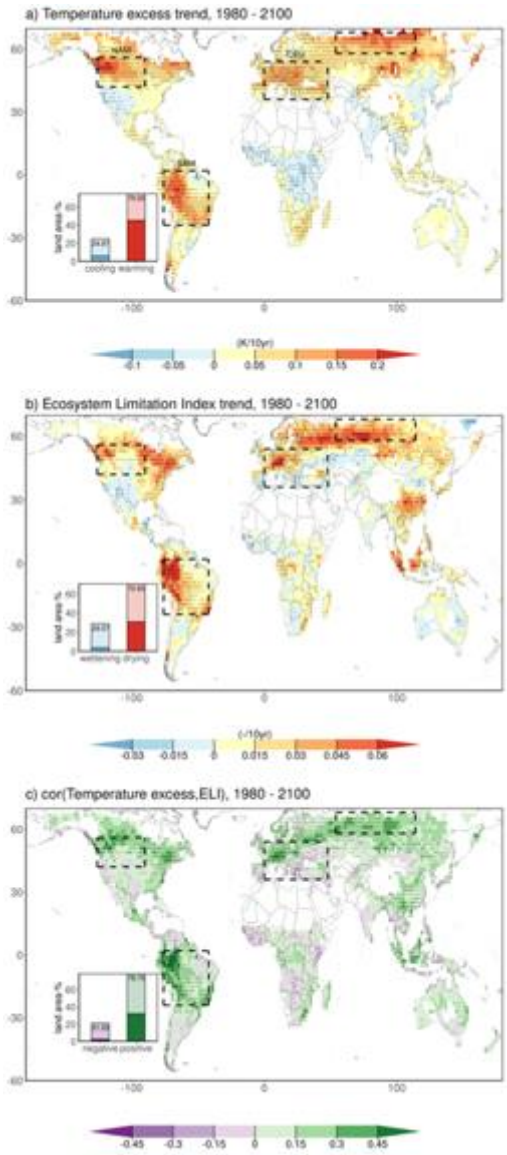


Figure 1. *Similarity of global patterns of change in temperature excess and ecosystem water limitation. Multi-model means of trends based on decadal time series per respective CMIP6 model of a) temperature excess) and b) Ecosystem Limitation Index (ELI). c) Multi-model means of Kendall's rank correlation coefficient between model-specific time series of ELI and temperature excess. The insets display the fraction of the warm land area with positive or negative trends or correlations, respectively (at least 8 out of 12 models agreeing on the sign of the trend or correlation are hued darker). Stippling indicates that at least 8 out of 12 CMIP6 models agree on the sign of the trend or correlation. All trends and correlations are calculated over the warm season and are only displayed if at least 8 CMIP6 models have full time series available, such that white areas denote regions with no or insufficient data. The dashed boxes indicate regions of interest, which are regions where temperature excess increases are particularly rapid and spatially coherent: North and South America (NAM and SAM), Central Europe (CEU) and Northern Asia (NAS).*

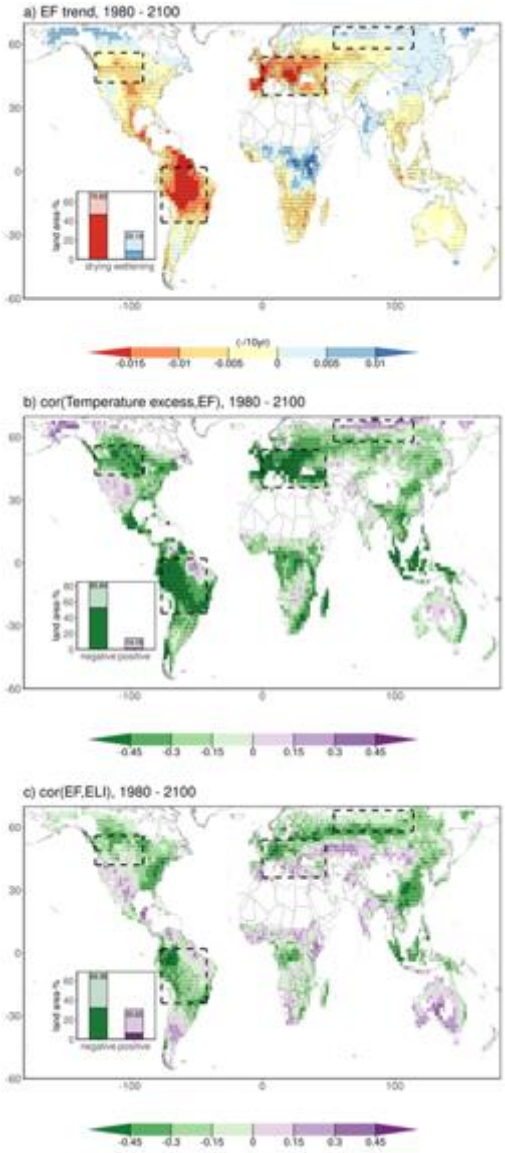


Figure 2. *Global multi-model mean distribution and trends of Evaporative Fraction (EF). Multi-model mean of trends based on decadal time series per respective CMIP6 model of a) EF and b) Ecosystem Limitation Index (ELI). c) Multi-model mean of Kendall's rank correlation coefficient between model-specific time series of ELI and temperature excess. The insets display the fraction of the warm land area that with positive or negative trends or correlations, respectively (at least 8 out of 12 models agreeing on the sign of the trend or correlation are hued darker). Stippling indicates that at least 8 out of 12 CMIP6 models agree on the sign of the trend or correlation. All trends and correlations are calculated over the three hottest months-of-year, defined as the 3 months–of-year which have the highest average temperature over 1980 - 2100. The dashed boxes indicate regions of interest.*

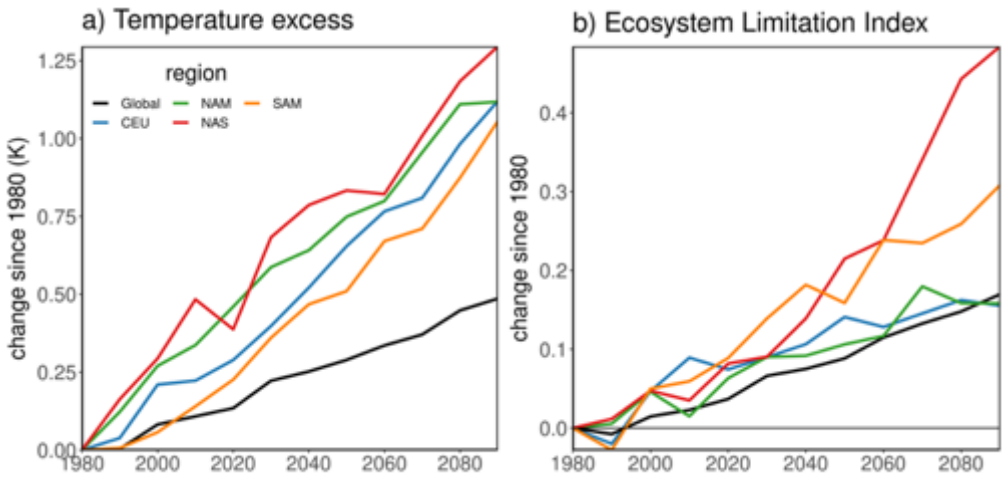


Figure 3. *Changes in global and regional temperature excess with increasing ecosystem water limitation. Temporal evolution of a) temperature excess and of b) Ecosystem Limitation Index (ELI) globally and for the regions of interest. Solid lines depict multi-model mean time series. Global and regional averages are calculated over land grid cells that have complete time series for all models and variables and are weighted according to the surface area per grid cell.*

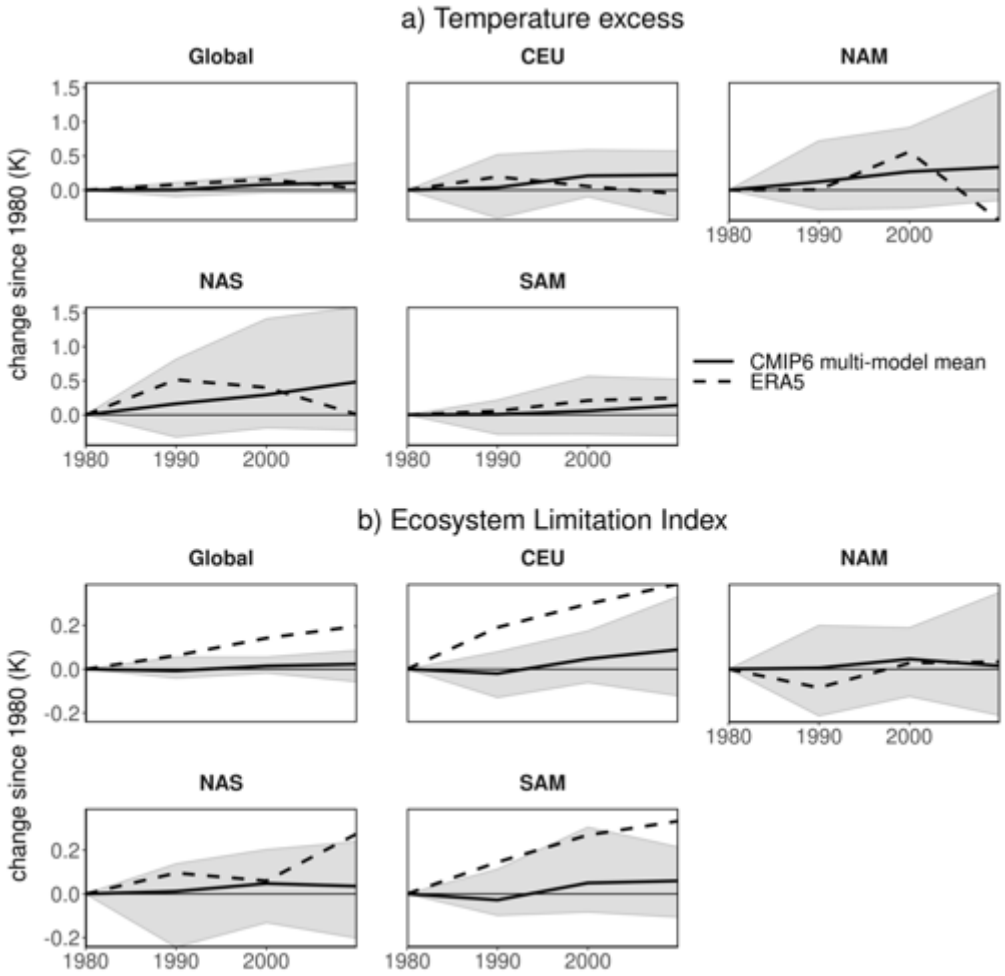


Figure 4. *Changes in global and regional temperature excess in concert with increasing ecosystem water limitation from CMIP6 models and ERA5-Land. Temporal evolution of a) temperature excess and of b) Ecosystem Limitation Index (ELI) globally and for the regions of*

interest. The black solid lines depict global and regional time series from the CMIP6 models, while the black dashed line represents ERA5-Land. The grey ribbon displays the envelope which encapsulates all the CMIP6 results. Global averages are calculated over land grid cells that have complete time series for all models and variables and are weighted according to the surface area per grid cell. The same mask is applied for CMIP6 models and ERA5-Land.

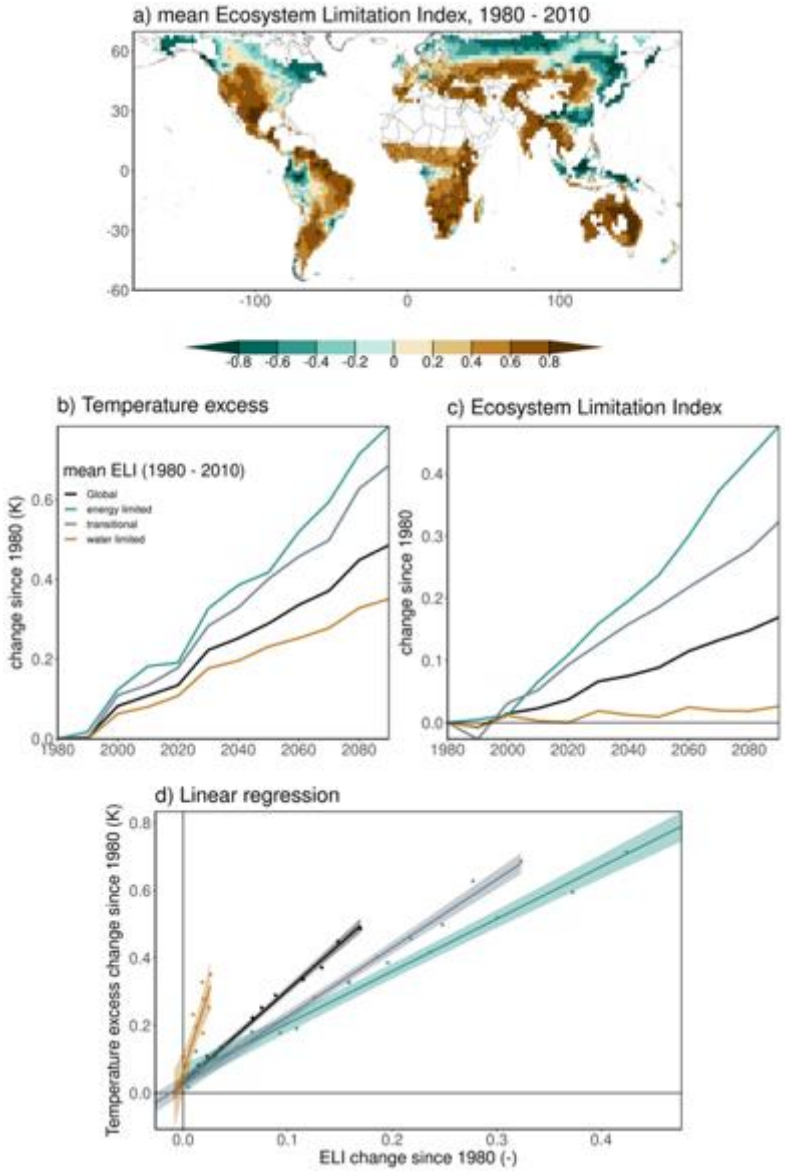


Figure 5. *Relation between temperature excess and ecosystem water limitation. a) Multi-model mean Ecosystem Limitation Index (1980 - 2010). Solid lines depict the time series of multi-model means inferred from globally (black) and regionally (colored) decadally averaged model simulations for b) temperature excess and c) Ecosystem Limitation Index. The classification is defined based on the model-specific mean ELI over 1980 - 2010 (Supplementary Figure 9): Energy limited (ELI < -0.2), transitional (-0.2 < ELI < 0.2) and water limited (ELI > 0.2). d) Points denote the global (black) and regional (colored) decadal multi-model means of ELI (x-axis) and temperature excess (y-axis), expressed as change since 1980. The lines denote linear regressions, with a shaded colored 95% confidence interval. Land grid cells that do not have complete time series for all models are excluded (white*

regions, Methods). Global and regional averages are weighted according to the surface area per grid cell.

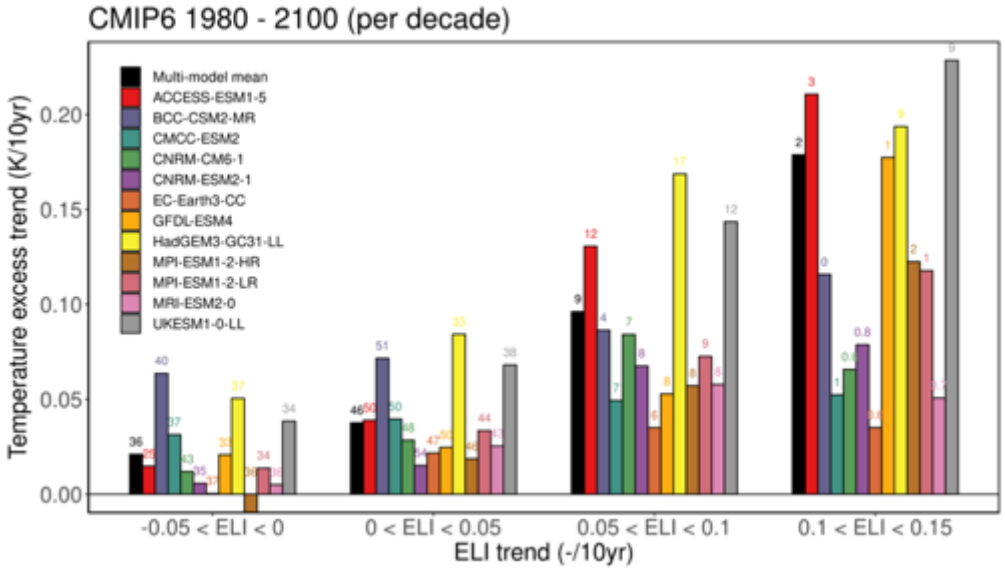


Figure 6. *Temperature excess trends increase with stronger trends in ecosystem water limitation. The bars denote the multi-model mean and model-specific temperature excess trends (y-axis) binned according to their respective ELI trends (x-axis) for the multi-model mean trends (black) and all individual models (colors). The numbers display the fraction of warm vegetated land area in which respective temperature excess and ELI trends occur. These area fractions may not add up to 100%, because values outside of the defined bins on the x-axis are possible.*

2.) I am quite surprised to see how few models seem to have all the required variables, especially since you only need them in monthly resolution. I get at least 40 different models (not simulations, as for some models such as, e.g., CanESM5, MPI-ESM-LR or MIROC6, there are dozens of initial condition ensemble members) for each variable, and while I did not check the overlap for all variables, I am absolutely sure that far more than 8 models remain. It should be close to or even more than 30...
I would also like to point out that according to Qiao et al. (2022), the BCC-CSM2-MR model constitutes a rather unfortunate “choice”, as it does not perform well with regards to soil moisture. Moreover, to quote Qiao et al. (2022), “For deep soil moisture, the top-five best-performing models are CESM2, MPI-ESM1-2-LR, ACCESS-ESM1-5, CESM2-WACCM, and CNRM-ESM2-1, [...]”, of which only CNRM-ESM2-1 is used here. While such evaluations are particularly challenging for variables that are hardly observed/measured and notoriously spatially inhomogeneous, I still think it is a pity that a) only few models were used in the first place, and b) that state of the art models such as CESM2 with plant hydraulics (see, e.g., Zhao et al., 2022) are not included. I thus
strongly encourage the authors to check an alternative data source if they cannot obtain the required variables for more than the 8 models used thus far.

Retrieving data from the Earth System Grid Federation (https://aims2.llnl.gov/search/?project=CMIP6/) instead of the Google cloud CMIP6 public data has led to a larger sample of 12 CMIP6 models that could be retrieved. These are the only models that meet the criteria described in the methodology (see lines below). The biggest

bottlenecks that prevented obtaining an even larger number of CMIP6 models were the unavailability of total water content per soil layer (mrsol), which excluded CIESM, HadGEM3-GC31-MM, INM-CM4-8, INM-CM5-0 and MIROC-ES2H, and/or unavailability of maximum daily temperature (tasmax), which excluded CESM2, CESM2-WACCM, CMCC-CM2-SR5, EC-Earth3-Veg and EC-Earth3-Veg-LR. Further, amongst the selected models, we have increased the amount of models with a better representation of deep soil moisture.

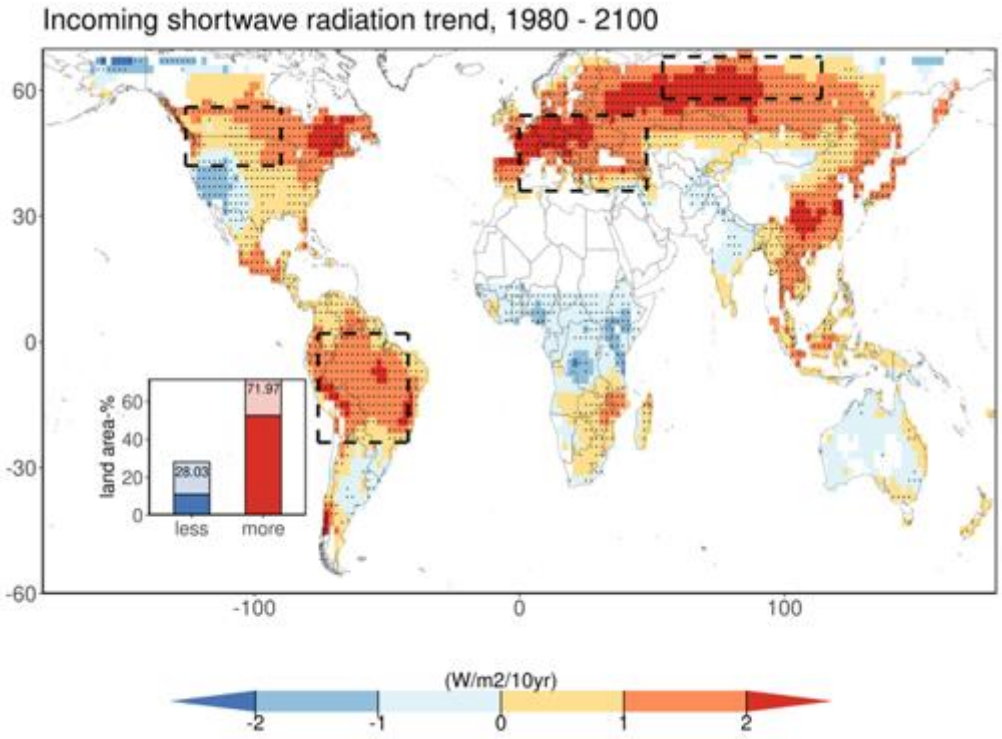
"We only selected models that provide i) historical (1980 - 2015) and "worst-case" SSP5-8.5 (2015 - 2100 (O'Neill et al., 2016)) simulations, ii) the necessary variables (Table 1) and iii) sufficient spatial (2˚x2˚ or finer grid cell resolution) and temporal (monthly) resolutions."

3.) I appreciate that the authors state that land–atmosphere coupling does not necessarily account for all of the "temperature excess", but it also makes me wonder what else could contribute to stronger maximum than mean temperature trends. I agree that (changes in) advection could play a role, but I think there is another, perhaps even more important mechanism at play: in several regions around the world, aerosol emissions have decreased substantially and are projected to decrease further in the ongoing century. This results in more shortwave radiation reaching the surface compared to past decades due to higher atmospheric transmission, which noticeably alters the surface energy budget and hence near-surface temperatures (e.g., Nabat et al., 2014), particularly in the warm season when incoming shortwave radiation is typically highest. Maximum temperatures tend to occur between noon and late afternoon and are arguably closer related to incoming shortwave radiation than mean temperatures, which, during nighttime, are primarily governed by the longwave radiation budget (which is directly altered by anthropogenic greenhouse gas emissions and water vapor feedbacks). The study of Qian et al. (2011) supports this rationale by reporting that aerosol-related temperature effects mostly occur through (daytime) maximum temperatures. I would thus not be surprised if shortwave radiation changes — which can, of course, also be mediated by changes in cloudiness and not just aerosol absorption (although at least for central Europe, this aspect has been far less important since 1980; see, e.g., Wild et al., 2021) — also contributed to the temperature excess patterns shown in Fig. 1a. In some regions such as, e.g., China (Qian et al., 2011), India and central Africa, shortwave radiation has decreased in the last decades, so my example provided above should not be generalized. Showing downward shortwave radiation trends (rsds) for all models could be helpful to understand why areas where the sign of temperature excess and ELI trends is inconsistent.

We agree with the reviewer. We have inserted the multi-model mean incoming shortwave radiation trends in Figure 1b and show model-specific incoming shortwave radiation trends in Supplementary Figure 4. We have elaborate on incoming shortwave radiation trends in the following lines in the results section.

"There is a widespread increase in incoming shortwave radiation in about 71% of the warm vegetated land area, with high inter-model models agreement (Supplementary Figure 4), which can directly affect near-surface temperature through the surface energy balance. These trends could result from projected decreases in aerosol emissions (Nabat et al., 2014), or from changes in cloud cover. As daily maxima of incoming shortwave radiation roughly co-occur with daily temperature maxima, increased incoming shortwave radiation links more strongly to increased in maximum temperatures rather than mean temperatures (Qian et al., 2011), which are more strongly governed by the longwave radiation budget."

"Further deviations from a positive relationship between temperature excess and ELI might result from alternative processes such as (changes in) advection of warm air masses through large-scale circulation patterns and changes in incoming shortwave radiation (Supplementary Figure 4)."



Supplementary Figure 4: *Multi-model mean trend in incoming shortwave radiation based on decadal time series per respective CMIP6 model. The insets display the fraction of the warm land area with positive or negative, respectively (at least 8 out of 12 models agreeing on the sign of the trend are hued darker). Stippling indicates that at least 8 out of 12 CMIP6 models agree on the sign of the. All trends are calculated over the warm season and are only displayed if at least 8 CMIP6 models have full time series available, such that white areas denote regions with no or insufficient data. The dashed boxes indicate regions of interest, which are regions where temperature excess increases are particularly rapid and spatially coherent: North and South America (NAM and SAM), Central Europe (CEU) and Northern Asia (NAS) (see Figure 1).*

Additional comments
- Some citations should be double-checked; e.g., "(Eyring et al., 2016))" comes with an additional right bracket.

All double brackets were checked and removed if possible.

- L. 85: I recommend changing "[...] please refer to Denissen et al. (Denissen et al., 2022)" to "please refer to Denissen et al. (2022)". Same thing for "from Teuling et al. (Teuling, 2018)" on L. 321.

We have done as the reviewer suggested.

- L. 167 onwards: “Moreover, ET is generally significantly correlated with both temperature excess and ELI, respectively, establishing the physical link between these quantities”. The authors acknowledge themselves later on in the manuscript that their correlative analysis cannot establish causal links, so perhaps something like, e.g., “[...] , suggesting a physical link [...] ” would be more appropriate.

We have done as the reviewer suggested.

- L. 200 onwards: ERA5-Land is an offline land surface model simulation that does not assimilate any observations. The meteorological forcing provided by ERA5 does indeed make use of data assimilation, but this is largely restricted to “classic” variables such as 2-meter temperature and humidity. Surface soil moisture data from scatterometers is also assimilated, but this only affects the top soil layer and does not help much with regards to root-zone soil moisture.

We have adjusted the discussion accordingly:

“Note that ERA5-Land is only indirectly supported by data assimilation, as meteorological forcing from ERA5 assimilates observations only for 2m temperature, relative humidity and surface soil moisture. Therefore, temperature excess benefits more directly from data assimilation than ELI, which is based on ET and (root-zone) soil moisture which are not readily observed across the globe.”

- L. 315: “[...] increased entrainment of dry air above the atmospheric boundary layer”, I think rephrasing this to “[...] increased entrainment of dry air from above the [...]” or similar would be a good idea, the current version could be a bit confusing.

We have done as the reviewer suggested.

References

Pierre Nabat, P., Somot, S., Mallet, M., Sanchez-Lorenzo, A. & Wild, M (2014): Contribution of anthropogenic sulfate aerosols to the changing Euro-Mediterranean climate since 1980. Geophys. Res. Lett. 41, 5605–5611. doi=10.1002/2014GL060798

To cite this article: Qian, Y., Leung, L. R., Ghan, S. J. & Giorgi, F. (2011): Regional climate effects of aerosols over China: modeling and observation. Tellus B Chem. Phys. Meteorol. 55, 914–934, doi=10.3402/tellusb.v55i4.16379

Qiao, L., Zuo, Z. and Xiao, D. (2022): Evaluation of Soil Moisture in CMIP6 Simulations. J. Clim. 35, 779–800, doi=10.1175/JCLI-D-20-0827.1

Wild, M., Wacker, S., Yang, S., Sanchez-Lorenzo, A (2021): Evidence for Clear-Sky Dimming and Brightening in Central Europe. Geophys. Res. Lett. 48, e2020GL092216. doi=10.1029/2020GL092216

Zhao, M., A, G., Liu, Y. et al. (2022): Evapotranspiration frequently increases during droughts. Nat. Clim. Chang. 12, 1024–1030. doi=10.1038/s41558-022-01505-3