Dear Referees,

before addressing the comments in detail, we would like to thank both reviewers for taking the time to point out the shortcomings of our manuscript and to provide possible solutions to them. We value the meticulous review of our study and genuinely appreciate the efforts to address each of the issues in detail. We believe that the suggested changes significantly improve the quality of the manuscript.

Major Point 1

The authors choose to show global model performances and then only focus on two sites, tropics and Tharandt. (minor comment Tropics should be Amazon, as we don't know how it performs in other sites around the tropics Thank you for pointing this out, we will change it throughout the manuscript.) I would prefer to have a closer look to: a) More fluxnet sites and compare the performance. Don't understand now your global model runs. If you have those, why not comparing those with all fluxnet data? At least use some other sites to see why they deviate from each other, and why they deviate from Michaletz and other sources. Is it the LAI, type of forest, type of climate etc. I do miss a global perspective.)

The reviewer is correct that it may appear somewhat counter-intuitive that we spend a good deal of time discussing two site-level experiments, while the focus of the paper is the large-scale effect of leaf thermoregulation. Here, we did not include the respective sections to provide a holistic model validation, which – as the reviewer correctly pointed out – requires a comparison of numerous sites that are representative of a broad range of climate- and vegetation conditions. The site level comparison was included merely to demonstrate that the model does not reproduce the observed relation between ambient air temperature and leaf-temperature excess, even when the model is forced with the conditions that are observed at the flux-net sites, because the model estimates the average temperature of all leaves (see also response to "Major point 2") and the evaporative cooling effect is substantially reduced in a humid atmosphere and when the plants are subject to water-stress. We felt that we needed to make these points, as it is extremely counter-intuitive that on the global scale the results actually exhibit a positive correlation between ambient temperature and temperature excess (when water stress is included), while the respective formulations that were implemented in the model indicate a negative correlation (see also Fig. 6 of the original manuscript).

In general, we agree with the reviewer, that a comparison of additional sites could help provide a better picture of the model performance. However, for the following reason, we would prefer not to do this using simulations similar to the site level experiments that are already included in the manuscript: The site level simulations were not run with the standard soil/vegetation parameters (which represent a much larger area) and are also run with prescribed, observation-based, atmospheric conditions. Setting up these experiments requires a lot of time and effort, but more importantly these simulations are not necessarily representative of the standard model behaviour – i.e. when using the standard coarse resolution parameters and when coupling JSBACH to the atmospheric model (as you can see for example in the attached figures in the appendix of this letter). Hence, we would rather propose to look at individual grid cells from the coupled (AMIP type) global run and evaluate them using site level observations – even though a good match can not necessarily be expected. Here, the required simulations have already been performed and we would make the comparison for the sites indicated in the appendix to this letter.

b) To use the data available from literature, as shown in Fig.1 but also for instance by Linacre and compare those with your results. I would prefer to extend fig 1 will all your data from the introduction and make a new chapter in which you review more data available in literature. Also include in here the oxygen isotope linear regression line. It now is strange that you plot the data by Linacre but not From Helliker and others. I would prefer to have those data description in chapter 2 and then finally interesting to compare those with the model results

In the introductory chapter, our aim was merely to provide a concise overview over the theory of leaf thermoregulation and in Fig. 1, we simply used the Linacre (1967) data as a visual support when describing the theory (please note that we added the regression lines of the in-situ measurements and the oxygen isotopes, as suggested). However – as our investigation is focused on the simulation of this effect using a large-scale model – we think that a comprehensive literature review (in a separate chapter), including a figure that combines all available observations, is beyond the scope of the present manuscript. Most importantly, the observations cannot be used for a direct comparison with our simulations, because they refer to a different scale (and predominantly to different conditions) than our model. Thus, we included the observation based regression in Fig. 8 and 9 to indicate how the effect works at the scale of individual leaves in comparison to how it works on the canopy scale. For this purpose we think that it is sufficient to show the regression line derived by Michaletz et al. (2016).

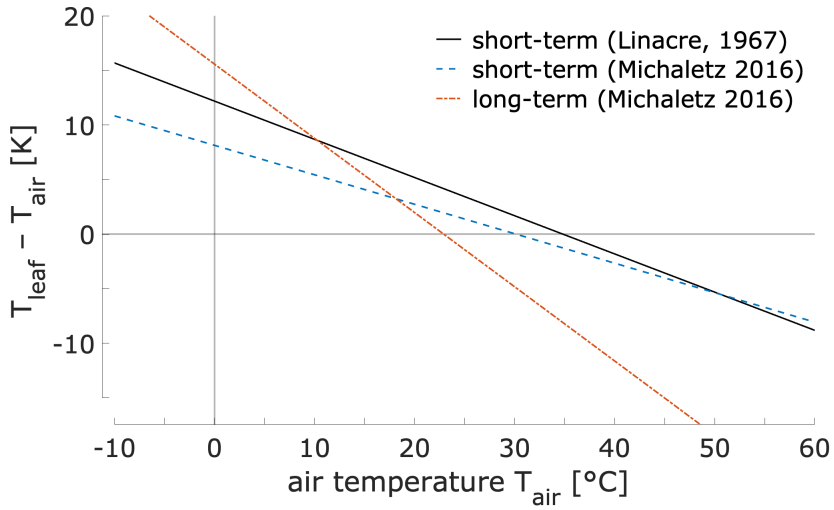

Fig. 1 Relation between leaf temperature excess and air temperature, based on short-term measurements of isolated and water-unstressed sunlit plants from Linacre (1967, black line) and Michaletz (2016, blue line), as well as long-term, photosynthetically weighted estimates based on cellulosic $\delta^{18}O$.

Major point 2

The single big leaf approach has clearly disadvantages. It is not clear to me how sunlet and shaded leaves are distinguished? Moreover, if we really want to understand the Tleaf, then we should better include the role of stomate in here (latent heat, now simply as rc?). The stomata react on T, radiation, but also on a sharp co2 gradient (higher below the canopy, specifically in the morning). I would like to see an analyses on different layers vs single layer approach and if the SBL approach still can be used to assess relialbe Tleaf, that can be verified with measurements.

We agree with the reviewer that the big leaf approach may have certain disadvantages in comparison to a multi-layer canopy scheme. However, it is computationally very efficient and consequently, used by a number of ESM land surface components, including the JS-BACH model. In the big leaf approach sunlit and shaded leaves are only implicitly separated in that the canopy has a surface area (or more precisely area that absorbs radiation) that corresponds (largely) to the sunlit leaves at the top/outside the canopy, but a heat capacity that corresponds to full canopy including the shaded leaves at the bottom/center. Thus, the reviewer is correct, that the big-leaf approach has the disadvantage that the vertical moisture/CO2/temperature structure within the canopy can not be resolved explicitly. As discussed in the manuscript, this disadvantage makes it difficult to compare the model results to observations because the model results represent the average temperatures of all sunlit and shaded leaves, while most observations pertain to exclusively to the sunlit leaves at the outside of the canopy. However, when canopy surface area and heat capacity are well parameterized – as we have spend quite a lot of work to achieve in JSBACH – the big-leaf approach is capable of capturing the overall dynamics at the coarse model resolution sufficiently well.

Unfortunately, we can not comply with the reviewer's request to compare the big-leaf approach to a multi-layer canopy scheme, as this would be a study in its own right. This would not only require implementing an entirely new canopy scheme into JSBACH but also connecting it to the rest of the ESM appropriately, e.g. implementing a vertically resolved wind profile within the canopy, which would most certainly require retuning the entire MPI-ESM. Thus, we would be happy to include a paragraph in the discussion section that addresses the shortcomings of the big-leaf approach in general, but simulations with a new canopy scheme are beyond the scope of the present study.

Small remark

L158 LAL to LAI

LAL here stands for *lowest atmospheric level* (as definied in line 139), not for *leaf area index*.

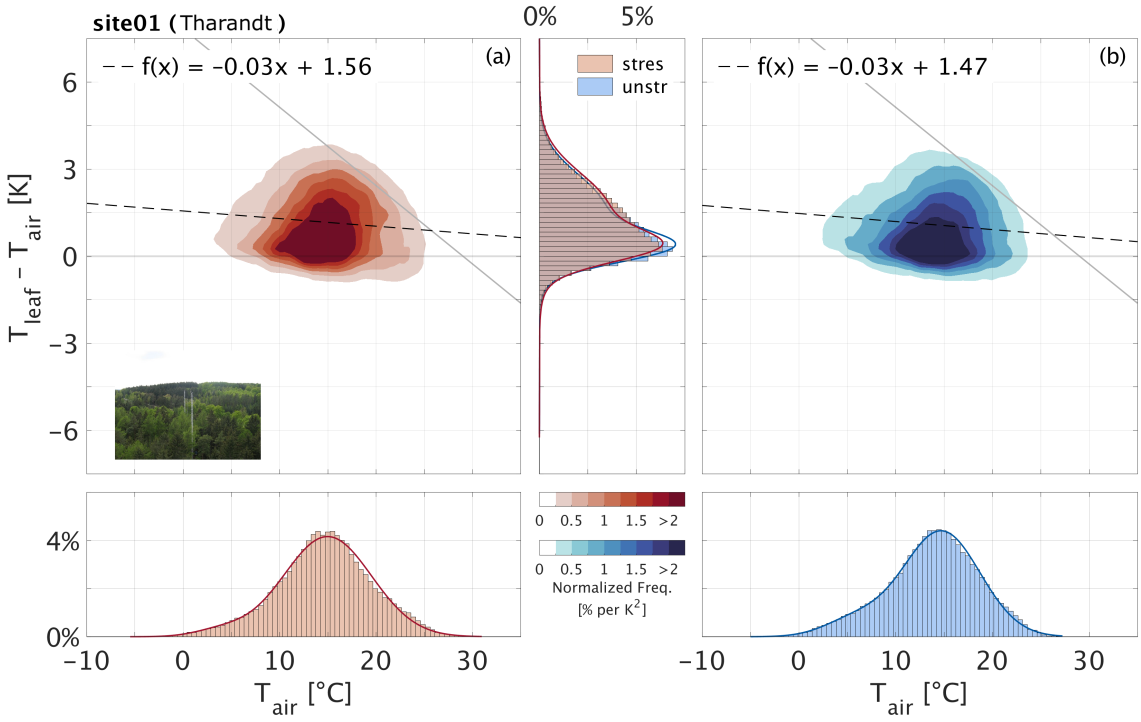

Fig. 2 See Figure 8 of the manuscript, but data derived from the global AMIP experiment (grid cell in which the Tharandt FLUXNET site is located) over 30 years (1979–2008) during growing season (Apr to Sep).

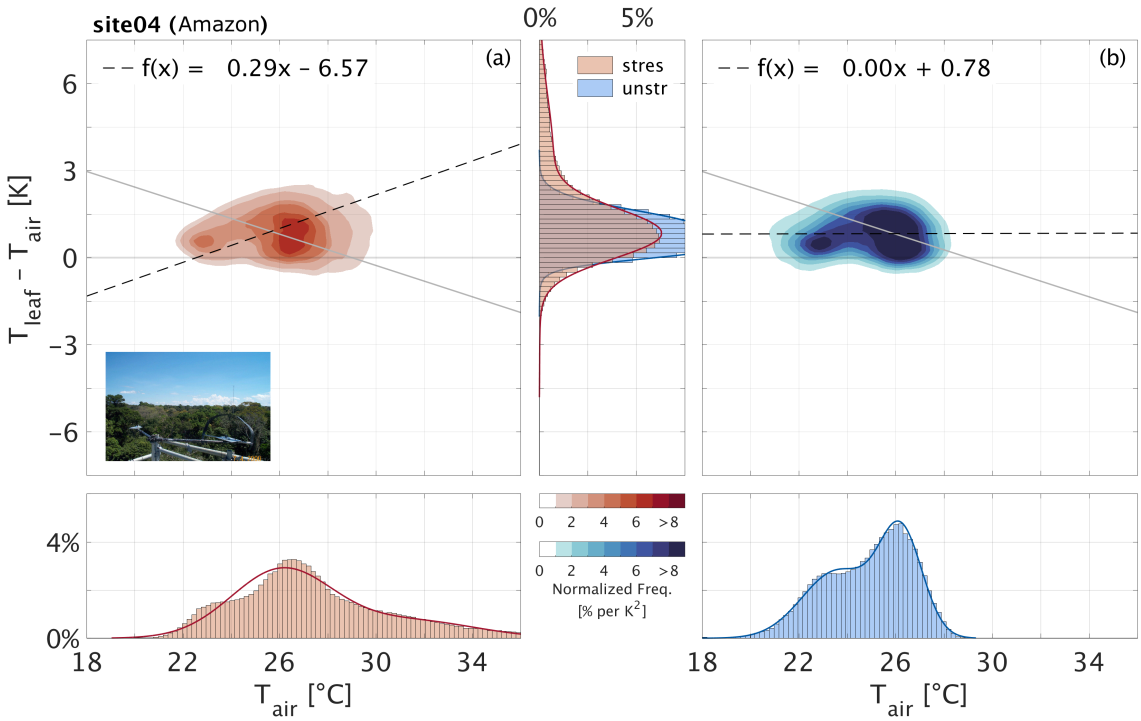

Fig. 3 See Figure 9 of the manuscript, but data derived from the global AMIP experiment (grid cell in which the Amazon FLUXNET site is located) over 30 years (1979–2008) during growing season (Apr to Sep).

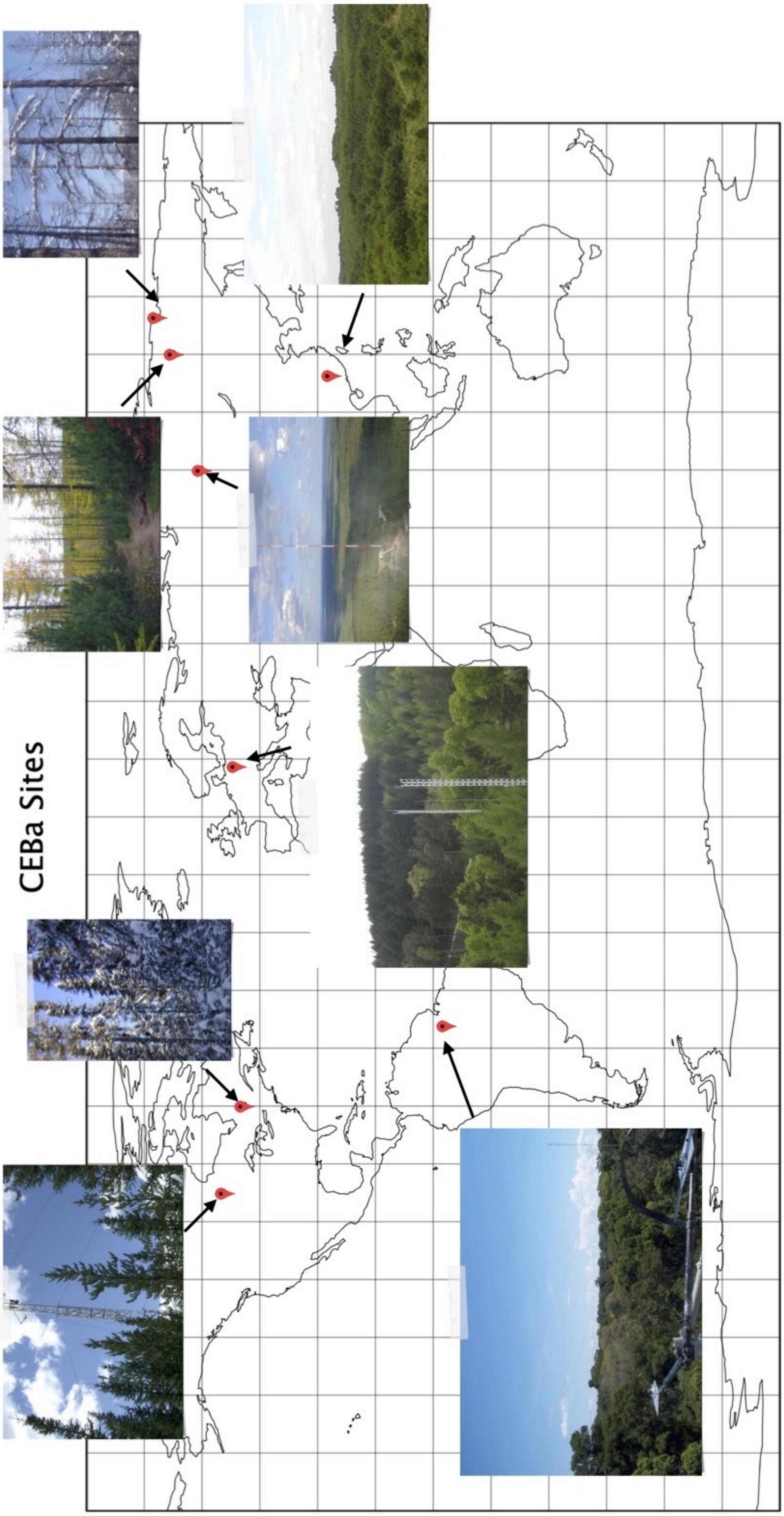

Fig. 4 Global map of forest grid cells that were extracted from the global coupled AMIP run.