Review #1

This manuscript presents a computationally efficient canopy-temperature emulator derived from a complex energy-balance model and implemented in LPJmL. The topic is highly relevant because canopy temperature is increasingly recognized as a key variable for representing heat stress processes in crop models, while computational constraints have limited its adoption in large-scale simulations. The manuscript is generally well written, the model description is clear, and the code and parameterization are made openly available. The proposed emulator reproduces the parent EBSC model with high accuracy and is integrated into a widely used global crop model. Overall, the study represents a useful methodological contribution that fits well within the scope of GMD. However, several aspects require clarification and further justification before publication.

Thank you very much for the positive evaluation and the constructive and detailed feedback. We appreciate the recognition of the topic's relevance and usefulness. Below we provide responses to the individual points and suggested changes to improve the manuscript.

Major comments

1. A central conclusion of the manuscript is that simulated maximum canopy temperatures outperform 2 m air temperatures because they agree better with ERA5 skin temperatures. However, ERA5 skin temperature represents the uppermost land surface temperature of the grid cell, which integrates crops, natural vegetation, bare soil, and other land-cover components. It is not a direct observation of crop canopy temperature. The authors themselves acknowledge potential mismatches in vegetation structure, LAI, land cover, and water status between ERA5 and LPJmL. Consequently, the comparison may not provide strong evidence that the emulator improves crop canopy temperature realism. The manuscript would benefit from a more cautious interpretation of these results and a clearer discussion of the limitations of using ERA5 skin temperature as a surrogate for crop canopy temperature. The current conclusions appear stronger than the validation evidence supports.

Thank you for this valuable comment. In the ERA5 comparison, we aim to reduce the mismatch by only selecting cells with certain minimum shares of the respective crops being evaluated. Nevertheless, we introduce an error that is weakening the strength of the evaluation. Therefore, the comparison with ERA5 data indeed only provides indicators supporting the results from the site-specific validation. In our view, this comparison is still a valuable addition to the site-specific evaluation with experimental data. Particularly, as there is no large-scale observation data set available that is directly usable for a global evaluation. We agree that the evaluation with ERA5 data is not sufficiently discussed in the current version of the manuscript and we will provide a clearer discussion and a more cautious interpretation of the respective results in the revised version.

2. The main motivation is that canopy temperatures should improve representation of heat-sensitive processes and ultimately crop growth and yield responses. However, the manuscript stops at evaluating canopy temperature itself and does not assess whether the new temperature estimates alter phenology, heat stress, biomass accumulation, or yield outcomes in LPJmL. As a result, the practical significance of the emulator remains uncertain. Even a limited sensitivity analysis comparing crop responses under air-temperature versus canopy-temperature forcing would greatly strengthen the study and demonstrate the value of the new module beyond reproducing temperatures.

While we agree that assessing effects of maximum canopy temperature usage in heat sensitive processes compared to maximum air temperature inputs is very interesting, we would like to point out that high temperature responses in LPJmL are not the objective of this paper. We think that including such an assessment would be beyond the scope of this study, particularly as this would require more model description regarding recent LPJmL developments. We are already working on a follow up paper that is meant to address such model developments including the comparison of canopy vs. air temperature effects. But this paper is on how canopy temperatures can be simulated in a global-scale crop model without introducing a significant computation time penalty.

3. The final emulator contains 84 terms for the lower limit and 56 terms for the upper limit, including numerous higher-order interactions. While cross-validation statistics are excellent, the manuscript provides little discussion of model complexity, parameter stability, extrapolation behaviour, or risks of overfitting. The authors note that the emulator may produce artifacts outside the training domain and therefore introduce threshold constraints. This suggests that the statistical model may have limited robustness under novel climate conditions, particularly under future warming scenarios where extrapolation is likely. More justification is needed regarding why this highly parameterized formulation was preferred over simpler machine-learning or reduced-form approaches and how robust it remains under climates beyond the training range.

Thanks a lot for this valuable feedback. The use of a machine learning approach is unfortunately not compatible with LPJmL, at least not in the version we use in our study due to technical reasons, and might be a barrier for the applicability in other crop models. That is why we decided to use a statistical model. We acknowledge that a simpler form might indeed simplify the applicability and the interpretation of the emulators. However, there are two main reasons why we think the complexity should not be reduced: First, the equations, on which the EBSC model is based on, indicate that the identified 5 and 6 regressors are important to describe the upper and lower canopy temperature limits sufficiently. Second, higher order and interaction terms are important to represent the nonlinearity of canopy temperature dependency on weather conditions and crop status and the interaction between the variables. Though, when using higher order polynomials, overfitting has to be avoided, as you pointed out. Considering this trade-off, we think that a polynomial of third order is a good solution. We will include a more comprehensive explanation in the revised manuscript.

The production of artifacts outside the training domain is a drawback of models that have been trained on data, regardless whether these are statistical or machine learning models. In order to reduce this risk and to increase robustness, we trained the emulators on a very wide range of climatic conditions by using a global dataset including both historic and projected future conditions. For the latter, we decided to use climate data from an SSP5-8.5 scenario including a broad spectrum of conditions. Nevertheless, we included and recommend to use the given thresholds to make sure the simulated canopy temperatures are reasonable. In order to assess the robustness of the emulator's results within the training domain, we will evaluate, whether and under which conditions the used thresholds are surpassed in the revised manuscript. Additionally, we will add a more detailed description and discussion about emulation via statistical models and possible drawbacks of the approach.

4. The maximum-temperature emulator assumes that canopy temperature peaks simultaneously with maximum air temperature and that radiation follows a sinusoidal diurnal pattern with a fixed maximum at noon. These assumptions may be reasonable for large-scale applications but could introduce systematic biases, especially under cloudy conditions, advection events, or water-limited environments. The manuscript would benefit from

quantitative evidence demonstrating that these assumptions do not substantially affect emulator performance.

Thank you for this helpful comment. We will conduct a sensitivity analysis to test the robustness of our assumptions and add a respective paragraph in the revised manuscript.

5. The emulator is presented as a general solution for twelve crops, yet most detailed evaluations (field experiments, canopy-air temperature differences, and discussion) are conducted only for wheat. The global ERA5 comparison includes other crops, but this remains an indirect evaluation. The authors should clarify the extent to which performance can be generalized across crops with substantially different canopy architecture, aerodynamic properties, and transpiration characteristics. Additional crop-specific validation or a stronger discussion of this limitation is warranted.

Indeed, we only compare wheat canopy temperatures with experimental data as we do not have access to respective data of other crops. That is why we will add information on transpiration characteristics of the other crops and compare them with those of wheat. Additionally, we will discuss the results more thoroughly.

Minor comments

1. Please provide additional information on computational performance. Reporting absolute runtimes and hardware specifications would help readers assess the practical value of the claimed 3% runtime increase.

We will add more information, so that the reader can better understand the meaning behind the 3% runtime increase.

2. The rationale for selecting 90,000 training samples and 45,000 validation samples should be explained. Was emulator performance insensitive to sample size?

We will add the respective information in the *EBSC input and modeling protocol* and *Evaluation details and reference data section*.

3. More information should be provided regarding the spatial distribution of training samples and whether certain climatic regions are over- or under-represented. Figure B1 is useful but deserves more discussion.

We will explain and discuss Figure B1 more comprehensively.

4.Please clarify whether the emulator coefficients are expected to remain stable across future LPJmL versions or whether retraining may be required when crop parameterizations change.

We will add the respective information in the revised manuscript.