

Dear Referee,

We sincerely thank the referee for providing the second review and taking the effort. We have written our reply in italics, while the referee's comments are in regular text with smaller font size.

We would like to address the referee's comments in two sections: Major and Minor comment.

**Major comment:**

“1. I thank the authors for their response and changes regarding my question about the segmented linear model. I do, however, have two follow-up questions. First, I ask the authors to clarify what they determined manually: only the breakpoints, or also the lines within each segment? If only the breakpoints were chosen manually, how were the segments fitted – e.g. by ordinary least squares (OLS)? If so, this should be stated explicitly. Second, if instead both breakpoints and lines were determined manually, I am uncertain whether the term segmented linear model is appropriate in the statistical sense, and whether it would be clearer to describe these instead as segmented linear functions.”

*Reply: Thank you for the follow-up questions and the attention to detail. We would like to clarify the points you raised. In our analysis, both the breakpoints and the lines were manually determined. We did not use statistical segment fitting and use the second point that referee has mentioned. We apologize for any confusion caused. We agree with referee that the term “segmented linear model” is not suitable for our methodology, so we have replaced it with “segmented linear functions” in L323 and L324.*

**Minor comments:**

1. L167f: Replace “[...] simulated a function [...]” with “[...] is a function [...]” and “[...] which observed [...]” by “[...] who observed [...]”.

*Reply: We have revised this according to your comment in L166 and L167.*

2. L178: Replace “[...] no and high (h) [...]” with “[...] no (l) and high (h) [...]”.

*Reply: We have revised it in L177.*

3. L190: “[...] value of growth rate of dry weight [...]” sound strange. Do you mean the cumulative growth rate expressed as dry weight?

*Reply: Thank you for pointing this out. We have revised the text into “cumulative growth rate of dry weight” in L189.*

4. L216: Replace “[...] stop growing, [...]” with “[...] leaves stop growing, [...]”.

*Reply: We have revised this accordingly in L216.*

5. Eq. 22: Why are the glucose reserves considered when calculating the LAI?

*Reply: Thank you for your question. The inclusion of glucose reserves in the LAI calculation is based on their role as a buffer in leaf dry weight, storing carbohydrates that are not immediately needed for growth. The availability of these reserves directly affects the growth rate, particularly over short time intervals (de Vries et al., 1989) where MATCRO-Soy simulates in hourly. While*

*glucose reserves ( $W_{glu}$ ) are not typically included in standard LAI calculations, we believe their inclusion improves the model accuracy.*

*We added this explanation into L231:*

*“Glucose reserves are added to the leaf dry weight as a buffer, which has an impact on growth by storing carbohydrates that are not immediately required.”*

6. L245 and Eq. 25: Is SR a parameter or a variable? The text says parameter but it also says at harvest time suggesting that SR is generally dependent on time. Please clarify. Also why do you only state that Yield is a function of four variables but do not provide the actual equation or a reference to it?

*Reply: Thank you for your feedback. We apologize for any confusion. SR represents the ratio of harvested yield to harvested pod weight, which is fixed and not time-dependent. We have rearrange the sentences in L244:*

*“SR was derived from experimental datasets in **Error! Reference source not found.** and represents the ratio of yield (seed,  $\text{kg ha}^{-1}$ ) to  $W_{\text{podharvest}}$  at harvest time.*

*About the yield equation, we were unable to provide the full equation due to the need for additional equations and to focus on the yield calculation. However, we referenced Masutomi et al. (2016) for the calculation of the yield along with the water stress factor ( $f_w$ ). Additionally, we updated Eq. (26) by replacing “ $f_{w\text{stress}}$ ” with “ $f_w$ ” and removing temperature ( $T$ ). We did it to clarify the difference with Eq. (27) of the water stress function ( $f_{w\text{stress},t}$ ). Meanwhile, the temperature ( $T$ ) was removed as the current version of the model do not account for heat and cold damage to the yield, as there is insufficient evidence to justify its inclusion for soybean.*

7. L247-250: This sentence is quiet long and complicated. Consider splitting it into two.

*Reply: Thank you for noticing it. We have splitted it into two sentences in L247-250:*

*“The water stress factor ( $f_w$ ) was determined based on the root distribution function and the water stress function at soil layer -i ( $f_{w\text{stress},t}$ ) as a basis.  $f_{w\text{stress},t}$  represents the fraction of available soil water at the soil layer -i ( $FAW_i$ ) relative to crop yield at timestep  $t$  during the crop growth. This approach is based on a previous study on the relationship between the soybean transpiration ratio and transpirable soil water conducted by Ray and Sinclair (1998), given in Eq (26).”*

8. L299: Delete “based on the”.

*Reply: We have revised this accordingly in L300.*

9. L301: “broad planting densities” sounds a bit odd why not use “low “since you use high later in L302.

*Reply: We intended to convey that the row distance in the field is wider, but we agree with the reviewer’s suggestion. We have replaced it from “broad” to “low” in L301.*

10. Sec. 5.1: Are the areas provided by MIRCA comparable to the areas reported to FAO? Even though the authors do not use FAO harvested area data I assume that FAO yield and harvested area data are not independent. Similarly, the GDHY is a product

derived from census and remote sensing data containing assumptions and uncertainties. It could be worth to mention this in the discussion as an additional reason for the discrepancies between MATCRO-Soy and observations unrelated to MATCRO-Soy but uncertainty of the available data.

*Reply: We do not use FAO harvested area data in our analysis. We agree that the FAO yield and harvested area data, used in MIRCA2000, are not independent. While MIRCA and FAO datasets may show similar trends in some cases, there can be differences due to the distinct data sources and estimation methods.*

*We appreciate your suggestion to mention this in the discussion. As you pointed out, GDHY may have inherent biases. GDHY is derived from census and remote sensing data, which contain uncertainties. To address this, we have added the following sentence to Section 6.3 in L696-697:*

*“The simulated yield was compared with GDHY dataset at the grid-cell level, which is derived from census and remote sensing data and may introduce the uncertainties in the evaluation results.”*

11. Fig. 8: The resolution is quiet low making it hard to read the labels.

*Reply: Thank you for pointing this out. We have replaced with new picture with a higher resolution in Figure 8.*

12. L640-646: Here the authors introduce new results which may be better suited for the results and not the discussion section.

*We agree with reviewer, so we have moved “Figure 12” into Section 5.3 and renamed it into “Figure 11” along with the explanation in L536-L42.*

Sincerely,

On behalf of all co-authors