Dear Reviewer and Editor,
Thank you very much for your effort in handling this manuscript.
In the revised manuscript, we made minor changes to the paramaters value, unit, and terms that was not based on reviewer comments. This correction improves the clarity of the paper, model's outcome, and does not affect the main conclusions of the study.
Moreover, we also revised the result section using new parameterization based on reviewer comment (RC2, point 1) on the overestimation value.

We refer the revised sentence in line XXX as LXXX in the revised version of manuscript (without track changes) "egusphere-2025-453-manuscript.pdf".

Review RC1

Dear RC1,
Thank you very much for your comments!

"1. Why the model only considers the NPP? It should be the GPP first, then the respiration, then the NPP."

Reply: You are correct. Pardon for our lack of explanation, this model firstly calculates GPP and respiration to calculate NPP. We revised the clearer context in Line 85 as follows: (now **L84**) "The photosynthesis module initially estimates gross primary production (GPP) and respiration at the leaf-level using the Farquhar model (Farquhar et al., 1980), and extends net primary production (NPP) estimation to the canopy level following the concept of de Pury and Farquhar (1997).".

"2. Please add other human management information, like the planting day and density."

Reply: I understand, thank you for your concern.
We put the information in the **Table 3** and paragraph body **(L299)** "*Furthermore, there are different farming practices based on the across countries. Soybeans are planted between May and June in the United States, China, and Japan, while planting starts in October or November in Brazil. The experimental data also showed broad planting density in China and Japan, while soybeans are typically grown at higher planting densities in the United States and Brazil*.".

"3. Why don't compare the yield in the site scale."

Reply: We compare the yield in the site-scale to confirm before running them in global. Our apology that we didn't mention the comparison clearly.

We added the following sentences to **L334**: "*The independent dataset was used for evaluating the calibrated model at the point-scale level. After removing the calibration data, the simulated yield in the site scale showed a correlation coefficient of 0.68 and significancy (p value < 0.001) with observed data (Supplementary Figure S2). This agreement is also applied for the aboveground biomass weight, pod weight, and leaf area index with correlation coefficient of 0.60–0.90*"

Review RC2

Dear RC2,
We greatly appreciate the reviewer for the detailed feedbacks that have helped us to significantly improve the paper. We will write the reply for 2 sections (major and minor comments).

About major comments,

"1. The authors' reparametrize an existing crop model for rice to extend its usage to soybean. However, soybean is a legume and therefore capable of symbiotic nitrogen fixation. It is my understanding that neither symbiotic nitrogen fixation nor nitrogen uptake are explicitly simulated in the model but are implicitly captured by the nitrogen limitation (SLN see also major comment 2). Even if symbiotic nitrogen fixation is captured implicitly reducing nitrogen stress, the respiratory costs of symbiotic nitrogen fixation are not considered. Yet, these costs are essential when determining the pathway of nitrogen uptake a plant follows depending on environmental conditions. For example, Fischer et al. (2006) show that the simulation of N uptake requires the optimization of the respiration costs of different uptake paths to reduce respiration losses. Neglecting these respiratory costs may lead to an overestimation of NPP and in turn crop yield under nitrogen limited conditions. When assessing Fig. 5 and 6 a), MATCRO-Soy indeed shows a systematic overestimation of FAO yields. It would be interesting to assess whether this overestimation is stronger in countries with lower fertilization rates which could indicate that the missing representation of the cost of N fixation contributes to the overestimation.

Reply:

According to your comment, we investigated causes of the overestimation and found that we had overestimated the growing degree days (from the harvest time input). We use a different formulation for growing degree days (GDD). GDD for the input data was calculated using daily time steps, while GDD in the simulated phenology was calculated using hourly time steps. Hence, we recalculated the growing degree days on hourly basis for the input files to avoid the overestimation in the simulated growing degree days. This revision has been updated in the revised manuscript and result to lower yield value in Argentina, Brazil, Canada, and Italy as the growing period become lower compared with the previous version. Thank you for pointing out this overestimation and improve our results.

In both previous and updated simulation results, the model tended to overestimate crop yields in China, where nitrogen fertilizer application is relatively high. In contrast, it produced more accurate yield simulations in the United States, where nitrogen input is lower. We had also run some simulations using DSSAT that consider the biological nitrogen fixation and found that inclusion or not of the nitrogen fixation resulted in minimal changes in yield under the tested conditions. The same result has also been reported by simulation conducted by La Menza et al. (2017). These results may suggest that the omission of explicit respiration costs for nitrogen fixation have limited effect on yield outcomes in certain conditions.

However, we agree about the importance of respiratory costs in symbiotic nitrogen fixation, particularly under nitrogen-limited conditions. In MATCRO-Soy, both nitrogen fixation and uptake are implicitly constrained through specific leaf nitrogen (SLN). This simplification may indeed contribute to the overestimation of yields under low N input countries, particularly at Fig. 6 (e.g. Bolivia, Paraguay, Russia). Those countries have nitrogen fertilizer input less than 25 gNm^{-2} for the growing season.
We acknowledge this limitation in the revised manuscript (**L669**):
"*While nitrogen fixation and uptake are implicitly constrained by the SLN parameter, an approach of carbon costs economics explicitly represents the respiratory cost due to different nitrogen uptake pathways (Fischer et al., 2010). MATCRO-Soy simplified the nitrogen fixation mechanism, which may have contributed to yield overestimation in low nitrogen input countries (e.g. Bolivia and Russia). However, this model still presented relatively small bias in countries with high nitrogen fertilizer application (e.g. China), as well as in countries with low nitrogen fertilizer input (e.g. the United States).*"

Additionally, I would ask the authors to explain why they do not consider respiratory costs of biological nitrogen fixation and the implications of this assumption and also considers this when discussing the limitations of their model. What would be potential solutions to solve this issues in future model versions? "

Reply:

We acknowledge that the respiratory costs of biological nitrogen fixation process are not represented in the current version of MATCRO-Soy due to model simplification. Quantifying respiration costs of BNF at a global scale require detailed input data and parameters tested across different cultivars/environments. Until now, introducing this component without well-supported parameters may introduce more uncertainty than insight for MATCRO-Soy in my understanding.

We might consider incorporating this mechanism in the future into the model structure with further examination for its effect. A potential solution is to identify and validate BNF-related parameters that are widely applicable for global scale.

We have added the limitation in the revised manuscript for discussion section (**L673**).

"*This highlights an opportunity for future model development to incorporate variable of respiratory costs associated with different nitrogen uptake pathways. While limited empirical data across cultivars, environments, and management systems poses a challenge at the global scale, addressing this would improve understanding of the physiological mechanisms under nitrogen-limited conditions.*"

"2. The explanations of Eq. 8, 9 17, 18 and 19 are incomplete. For example the explanation of eq. 8 and 9 is missing the explanation of $SLNY0$, $SLNY1$, $SLNY2$ and $SLNY3$, $SLNY3_{,h}$, $SLNY3_{,l}$ and $Nfert_{,max}$. All of these seem to be parameters that are listed in tables later on but this is not clearly explained. In addition I think that the assumption that biological nitrogen fixation (BNF) is captured through SLN needs a thorough explanation. Do the authors assume that BNF is captured by $Nfert$?. If so this should be stated and its implications need to be discussed. This is also related to major comment 1 regarding the assumption to not represent the respiratory costs of BNF. Similar in eq. 17, 18 and 19 $Pleaf0$, $Pleaf1$, $Pleaf2$, $Ppod1$ and $Ppod2$ are only listed in the tables but not explained."

Reply:

Thank you for pointing this out. In the current model version, we implicitly assume that biological nitrogen fixation (BNF) was captured in the empirical nitrogen concentration in the leaf (SLN) across developmental stages, based on experimental data in the study by Menza et al. (2023). Figure S1 below presents the segmented linear function of SLN derived from literature, reporting soybean leaf nitrogen content under both zero and high nitrogen fertilizer treatments. This function enables the model to simulate SLN under varying nitrogen availability, including the contribution of BNF reflected in the observed data.

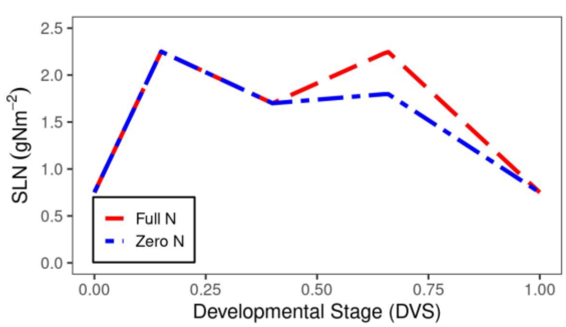

Figure S1. The specific leaf nitrogen (SLN, in g(N)m-2) toward the developmental stages (DVS) under zero nitrogen fertilizer (blue dotted-dashed line) and full nitrogen fertilizer (red long-dashed line) adopted from Menza et al. (2023). In the absence of observed data, values were derived based on model assumptions.

We added this figure in Suplementary file Figure S1 and added the following explanation in the manuscript **(L165):**

"*SLN was determined by nitrogen supply (including nitrogen fertilizer, soil mineral nitrogen, and biological nitrogen fixation) and by plant demand. The changes in SLN over the growing period in MATCRO-Soy simulated* *a function derived from La Menza et al. (2023), which observed SLN under both low and high nitrogen fertilization conditions (See Supplementary file Figure S1). It shows the specific leaf nitrogen value toward the crop growth period where higher nitrogen fertilizer results in a higher leaf nitrogen content. In the absence of empirical data for initial growth stages, the model assumes low nitrogen content"*

Regarding Equations (8) and (9), we added **(L176)**: "*While $SLNY0$, $SLNY1$, $SLNY2$, $SLNY3,h$, $SLNY3,l$ represent the SLN at the time of initial stage, early decline, pre-flowering increase, subsequent decline during the reproductive stage under high (h) and low (l) nitrogen inputs with the value of 0.75, 2.25, 1.7, 0.75, and 1.8, respectively. $Nfert,high$ refers to the high nitrogen fertilizer input used in the model for parameterization, as described in Table 2. Y denotes the observed gap function in specific leaf nitrogen under high and low nitrogen fertilizer treatments (g(N) m-2) in Supplementary file Figure S1.*"

Regarding Equations (17)–(19), we added **(L216)**: "*$Pleaf0$, $Pleaf1$, $Pleaf2$ represent the glucose partitioning ratio of leaf toward shoot at the time of initial stage when leaf growth starts to decline (leaf0), stop growing (leaf1), and at maturity (leaf2), respectively. While $DVSpod1$ and $DVSpod2$ indicate the developmental stage values at which glucose partitioning of pod to the shoot begins to increase and eventually saturates. Figure 2 in Section 3.2 illustrates the glucose partitioning ratio during the crop growth as calibrated in this study (see Figure 2a for leaf partitioning and Figure 2b for pod partitioning).*

"3. ***Eq. 25 and 26: Water stress is a factor to reduce the yield at harvest. It is not clear whether only the value for the water content of each soil layer at the day of harvest is used to calculate water stress or if this is integrated over the entire growing season but only applied at harvest. In both cases only applying water stress at harvest has its limitations (e.g. such an approach misses propagation effects of early season water stress as well as drought mortality) which should be discussed."***

Reply:

In this model, we calculate the water stress to be integrated over the entire growing season, not only in the harvest time. Then we believe it still captures the effects of the early season of water stress. I added the timestep index (t) to the equation to explicitly represent temporal resolution and added "[…] *over crop yield in timestep t during the crop growth.*" **(L248, Eq. 26)**.

"*4. The authors use segmented linear models to estimate several parameters in different developmental stages (DVS). However, they do neither provide the software they used to create these models nor how the models were trained.*

Reply:

Thank you for pointing out this paragraph. We apologize for the lack of clarity which may have led to a misunderstanding. The segmented linear lines were manually determined by visual inspection of the plots, due to the difficulty of nonlinear optimization under multiple constraints. This explanation has been included in the revised manuscript in Section 3.2. There is no software we used hence we have clarified this method in the following sentences.

(**L323**): "*The segmented linear models for glucose partitioning were manually determined by visual inspections of the plot. This approach was chosen due to the challenges of applying nonlinear optimization under multiple constraints.*"

Therefore, I think the answers to the following questions need to be added to section 3.2: How where the initial breakpoints estimated? What kind of optimization method was used?"

Reply:

The initial breakpoints were determined based on assumptions based on phenological growth characteristics. For example, the leaf starts to stop growing in the seed filling stage while the pod start growing after the flowering stage. The segmented functions were constructed by visually fitting straight-line segments to follow calibration data. This study does not include optimization algorithms. In the early growth stage where data is lacking in the leaf, it is assumed that approximately 40% of glucose allocation supports leaf and shoot growth. Meanwhile beyond that point, segmentation follows observed data trends. We also explained about it in point 23.

We added this sentences in **L325**:

"*Breakpoints in the developmental stage were determined based on assumed growth characteristics, such as leaf development declines after the seed-filling stage, while pod formation starts after flowering. We assumed increasing trend of glucose allocation to leaf and shoot development during the early stage when data were unavailable, with subsequent segments aligned with observed data trends.*"

"*5. Detrending: It looks like the authors apply linear detrending which removes both the slope and intercept of the linear regression from the observations and simulations respectively. This should at least be explained when the term "detrending" is introduced. In addition, I think the authors should highlight that the detrended comparison is useful to evaluate interannual variability and sensitivity to climate variability but mention that linear detrending removes important signals from the data.*"

Reply:

Thank you for the helpful suggestion. Yes, you are correct. We added the description of the detrending in the following sentences (**L407**): "*Detrended yield represents the time-series yield data for both simulated and observed values after removing the linear trend by subtracting the slope and intercept of the fitted linear regression (long-term yield trend). This approach enables the separation of short-term yield fluctuations from systemic long-term shifts.*"

In addition, I think the authors should highlight that the detrended comparison is useful to evaluate interannual variability and sensitivity to climate variability but mention that linear detrending removes important signals from the data."

Reply:

We agree with reviewer that this detrended comparison is also removes important signals from the data, hence we also added in the following sentences **(L444)** in section 5.1: "*Detrended yield is the value after yield is reduced by its long-term trend from the original yield data. It isolates the variability primarily driven by climate fluctuations to evaluate interannual variability independent of long-term trends. However, it also removes longer-term signals related to genetic improvements, cultivar and management changes, or increased CO_2 effects.*"

"*6. L622-627: The authors describe the role of $CO2$-fertilization effects for simulations results and hypothesize that this is the reason for the positive bias compared to observations. I would appreciate a more thorough explanation how the authors come to this conclusion. What is their explanation for the models ability to capture the temporal trend of the yield increase well but systematically overestimate yields? If this is all explained by $CO2$- fertilization, this indicates that the model captures the temporal development of the effect well (slope) but not the overall magnitude of the role of $CO2$ (bias) for yield formation. Additionally, while the $CO2$ -fertilization is one possibility, I do not think that this is the only possible explanation that is supported by their results. As stated in major comment 1 and 3, underestimation of respiratory costs for symbiotic nitrogen fixation and water stress are also likely explanations that should be discussed.*"

Reply:

Thank you for the critical point. We agree that CO_2 fertilization is not the sole explanation for the observed yield overestimation. In our revised results, while the model successfully captured the temporal trend (slope) in yield increase, it is now not systematically overestimated absolute yield levels (bias). The previous version has significantly overestimated absolute yield levels due to overestimation in growing degree days (which I concluded due to CO_2-fertilization effect, my apologies for missing explaining that it is only one possibility in the manuscript and lead to confusion). This recalculation improved yield accuracy in countries such as Argentina, Bolivia, and Canada.

We also agree that other factors, such as underestimated respiratory costs of symbiotic nitrogen fixation and insufficient water stress representation are also part of contributors.

We revised clear point of CO_2 fertilization effect (**L677, previous paragraph)**:

(**L677**): *"The simulated yield increases throughout the year driven by the positive effects of increased atmospheric CO_2, a phenomenon known as the CO_2 fertilization effect, has been observed in studies by Long et al. (2005) and Sakurai et al. (2014). The CO_2 fertilization response may become a more prominent source of overestimation in future projections if the model overestimates the crop response to elevated CO_2. Compared with simulations using statistical radiation use efficiency (Ai and Hanasaki, 2023), process-based models have this tendency because of the greater effect of CO_2 on the photosynthesis process. Therefore, further investigation is needed into the CO_2 sensitivity of MATCRO-Soy and other process-based models, particularly the potential downregulation of photosynthesis under elevated CO_2 conditions observed in the measurements (Ainsworth et al., 2002; Zheng et al., 2019). This is especially important for adaptation studies, as reliable yield projections are critical for designing effective adaptation strategies under future climate scenarios."*

We revised the discussion to reflect this respiratory cost and water stress (**L686, later paragraph**):

(L686) "*MATCRO-Soy simulations showed that MSD component of SB was the dominant contributor in the global and country-level yield error. It indicates the bias was in the over or underestimation of average yield, rather than in variability of discrepancy in the year-to-year yield pattern (Figure 8). These results highlighted the model uncertainty in simulating mean yield for improvement in major soybean-producing countries with large cultivation areas. The model overestimated the long-term trend in some countries. Inaccurate representation of CO_2 fertilization effect may have contributed to the mean yield bias. Other possible contributing factors for the bias are the simplified assumption of no respiratory costs for symbiotic nitrogen fixation and insufficient representation of water stress responses. The accuracy of data input may partly reflect the inherent gap between field experiment data and national average yields, which are influenced by local farming practices. While these discrepancies between the country and global levels are insightful, it provides a valuable opportunity for model improvement."*

About minor comments,

Recommendations from RC2 of grammar change for point 1-3, 5-8, 12-13, 15-19, 24, 26-27, and 29 each are adopted in the revised version (of total 29 points).

While other points are replied below:

- Point 4: ***I believe it should be "AgMIPs efforts" instead of "AgMIP efforts"***
  Reply:
  I personally think the acronym of AgMIP is robust for "Agricultural Model Intercomparison Project" and I don't know how to use it in sentences like AgMIPs. Hence I will change the phrasing from "*AgMIP efforts have demonstrated*" to "*AgMIP has demonstrated*" (**L48**) rather than using "AgMIPs efforts". If you think AgMIPs is better, please let me know.

- Point 9: ***Is Masutomi et al., 2019 referring to the ozone implementation? If yes this is not clear from the current sentence structure***
  Reply:
  Yes, it is about the ozone implementation in the model. Apologies for the confusion.
  Here we revised the text (**L63**):
  "The mechanisms that consider photosynthesis and stomatal conductance to assess the impact of greenhouse gases on carbon and water fluxes have been incorporated into MATCRO-Rice. Masutomi et al. (2019) described the implementation of ozone effects within these mechanisms, indicating the model's capability to account for the environmental stressor."

- Point 10: ***It is not clear to me what the authors mean by carbon allocation driven by photosynthetic activity. To me carbon allocation is the distribution of carbon between different plant organs – here called carbon partitioning – while photosynthetic activity drives carbon assimilation. Please clarify***
  Reply:
  Sorry for the confusion. You are correct, photosynthesis activity drives carbon assimilation.
  revised:
  (**L91**): "*The photosynthesis and carbon partitioning modules are closely linked, as carbon assimilation from photosynthesis process is subsequently allocated to different plant organs*."

- Point 11: ***I do not understand what is meant by "The phenology module serves as a time dimension***
  Reply:
  Sorry for the confusion, we refered to the phenology module simulates crop phenological development over time. Hence, we revised: (**L81**) *"The phenology simulates crop phenological development over time"*

- Point 14: ***Is de Vries et al., 1989 the reference to "school of De Wit" or MACROS? Also the acronym MACROS has not yet been introduced.***
  Reply:
  de Vries et al. (1989) is not the original reference for "school of De Wit" but using its concept and described parameterization for MACROS (Modules for an Annual CROp Simulator). Pardon for the confusion, I deleted MACROS and revised as follows:
  (**L92**): "*The carbon partitioning module distributes the glucose into each organ (i.e. leaf, stem, root, and storage organ) following the method derived from the school of de Wit by simulating biosynthetic processes (de Vries et al., 1989).*"

- Point 17: ***Similarly why are b, h and o used for minimum, maximum and optimum? Why not use min, max and opt which are far more intuitive?***
  Reply: Sorry for the confusion, I fixed them into base, optimum, and highest temperature.
  Revised:
  The parameters Tb, To, and Th are crop-specific and represent the base, optimum, and highest temperatures for crop development, respectively **(L127)**.

- **Point 20: I think it needs to be "Vcmax,top" instead of "Vcmax".**
  Reply: Sorry for the confusion in the text of previous manuscript, the definition of Vcmax in here is the maximum Rubisco capacity that depends on leaf nitrogen within the canopy, not only the top.
  We added following sentences **(L155)**:" $Vcmax$ is the maximum Rubisco activity within the canopy (mol(CO2) m^{-2} s-1) limited by the exponential value of vertical distribution of leaf nitrogen (Kn), leaf area index (LAI, in m^2 m-^{2}), and maximum Rubisco activity at the top of canopy (Vctop, in mol(CO2) m^{-2} s^{-1})."

- **Point 21: How can the developmental stage (DVS) be negative? I would assume it starts at 0.**
  Reply:
  Yes, it starts at 0. Sorry for the confusion and thank you for noticing them. We revised the Eq. 17

- Point 22: ***I am a bit confused by the specific leaf weight parameter, this seems to be the same as leaf mass per area (LMA) trait. If so why not use the common definition? Or this not per leaf area but per agricultural area? If this is the case this should be mentioned. There also seems to be a discrepancy regarding the unit …***
  Reply:
  The unit is per leaf area. This term is commonly used in crop modeling (at least in our understanding, e.g. de Vries et al. 1989 used this term too).
  We put this sentences in **(L230)** "*LAI is calculated from the estimated leaf dry weight (Wleaf, in kg ha-1) and glucose reserves in leaves (W_glu, in kg ha-1) divided with specific leaf weight (SLW, in kg ha-1). SLW indicates leaf dry weight per unit leaf area.*",
  with the unit of SLW in Table 2 is kgha^{-1}. Now, we revised the discrepancy regarding the unit. Thank you for noticing them.

- Point 23. *Fig. 2a): **How is the first segment fitted to the data if there are only values for DVS of around 0.2 or larger?***
  Reply:
  In the early growth stage where data is lacking, it is assumed that approximately 40% of glucose allocation supports leaf and shoot growth and then increased (before 0.2). Meanwhile beyond that point, segmentation follows observed data trends. This reply have also we explained in point 4.
  We added this sentences (**L327**): "*We assumed increasing trend of glucose allocation to leaf and shoot development during the early stage when data were unavailable, with subsequent segments aligned with observed data trends*."

- Point 25: ***Why does the model have a higher accuracy for countries with high production levels? Are fertilization levels in these countries higher and water stress levels lower? If so this should also be discussed in the context of N-fixation (major comment 1) and water stress (major comment 3).***
  Reply:
  Thank you for your critical question. We consider the top producer's countries to have high accuracy primarily due to the better data quality, well-managed production systems, and calibration data from these countries.
  We added following sentences (**L692**): "The accuracy of data input may partly reflect the inherent gap between field experiment data and national average yields, which are influenced by local farming practices"

  However, we couldn't find proof that the country with higher accuracy has higher fertilizer and lower water stress due to limited data availability for parameterization. Among the top producers of 10 countries, 3 of them (China, India, Italy) have relatively high N fertilization (>30 gNm^{-2} per growing season). In contrast, most top producers (US, Brazil, Argentina) rely more on the biological nitrogen fixation (represented by Specific Leaf Nitrogen in the model), and do not have lower soil moisture. We also couldn't prove this relationship in the revised result.
  We revised as in **L672**: "However, this model still presented relatively small bias in countries with high nitrogen fertilizer application (e.g. China), as well as in countries with low nitrogen fertilizer input (e.g. the United States)."

- Point 28: ***This paragraph is very general and I am missing the connection to the MATCRO- Soy's strengths and application.***
  Reply:
  We were rewriting the paragraph by first highlighting the strength of the model by describing the specific important parameters and the importance of such a model (**L623-626**), then demonstrating application (**L626-631**) and one possible solution to the limitation (**L631-634**).

  Revised: **(L623)** "The strength of MATCRO-Soy lies in its ability to simulate key physiological processes of soybean growth (e.g. photosynthesis, phenology, and biomass partitioning), under varying climatic conditions. Its process-based structure allows for sensitivity analysis for further environmental impacts evaluation, such as effects of elevated CO2 and temperature stress. It has been shown to reasonably capture the temporal dynamics of yield formation. In addition to climatic factors, variations in yield may be attributed to technological advancements, shifts in agricultural practices, and changes in crop management strategies outside the scope of model can further improve the accuracy at the local scale. For example, including pest and crop interaction may enhance the model's capability to reflect local yield response to climate-driven pest dynamics (Chen and Mccarl, 2001). The integration of crop models with remote sensing data will enhance its capability for monitoring and predicting crop productivity at finer spatial scales (Basso et al., 2001). However, it is important to acknowledge the limitations of the model, particularly its ability to predict yield variations under extreme or rapidly changing climatic conditions. Continuous updates of the experimental dataset are necessary to maintain its relevance and accuracy in predicting future soybean yields."

References

La Menza, N. C., Monzon, J. P., Specht, J. E., & Grassini, P. (2017). Is soybean yield limited by nitrogen supply?. Field Crops Research, 213, 204-212.