**Responses to the comments of Editor**

**Article ID:** essd-2024-586
**Title:** NortheastChinaSoybeanYield20m: an annual soybean yield dataset at 20 m in Northeast China from 2019 to 2023
**Authors:** Jingyuan Xu, Xin Du, Taifeng Dong, Qiangzi Li, Yuan Zhang, Hongyan Wang, Jing Xiao, Jiashu Zhang, Yunqi Shen, Yong Dong

Dear Editor,

Thank you very much for your thorough review and valuable comments on our manuscript. Your insightful feedback has significantly contributed to improving the quality and clarity of our work. In response to your suggestions, we have rigorously revised the manuscript. The key modifications include:

(1) **Methodological Refinements:**
- Introduced a comparative analysis between time-series Landsat-derived LAI and stage-averaged LAI at field scale.
- Added a dedicated section on feature selection methodology (Section 5.1).

**(2) Methodological Refinements:**
- Added relevant references to substantiate our approach and enhance the credibility of our methodology, providing stronger support for our research findings.

The detailed point-to-point responses are as follows. Texts in black are the reviewer's comments; those in blue are our responses to the reviewer's comments; and those in *red and italics* are the revised texts appeared in the revised manuscript.

Thanks,

All our best,

Xin Du

The authors developed a deep learning model using a GRU architecture to predict crop yield, utilizing only two predictors: LAImean1 and LAImean2. Given the simplicity of these two predictors, it raises questions about how they can achieve high prediction accuracy. The authors should provide a more detailed explanation of the underlying reasons or mechanisms that enable such effective performance with just these two variables.

**Reply:** Thanks for your suggestion. We have further clarified the reasons behind our use of two predictors ($LAI_{mean1}$ and $LAI_{mean2}$) in the revised manuscript. In addition, we have conducted a field-scale comparative analysis (Section 3.2 and 3.3.2) that directly evaluates the information loss from reducing full LAI time series to two stage-averaged features and the practical trade-offs involved.

**1. Time series vs. two-stages validation**

For each sample plot in 2022 and 2023, we first identified all available Sentinel-2 observation dates and calculated their corresponding Development Stage Values (DVS). We then built two GRU model variants: (a) a full time series GRU that uses DVS-aligned LAI sequences derived from the simulated soybean growth dataset, (b) a simplified GRU that uses two stage-averaged LAI features ($LAI_{mean1}$ (emergence to flowering), $LAI_{mean2}$ (flowering to maturity).

Validation against in-situ yield observations (Fig. 5 and Fig. A4) showed that the time-series GRU model achieved slightly better accuracy (RMSE = 224.81 kg $ha^{-1}$, MRE = 7.50%), while the stage-averaged model remained competitive (RMSE = 287.44 kg $ha^{-1}$, MRE = 10.02%). The difference in MRE is around 3% across both years.

**2. Why two Stage-Averaged LAI remain effective**

In the revised manuscript, we have also added a new section in the Discussion (Section 5.1: Selection of model input features) to elaborate on the design and evaluation of candidate predictors.

We systematically evaluated a broad set of candidate predictors (LAI-based, transpiration-based (TRA), and soil moisture-based (SM) and four summary statistics (mean, max, median, cumulative sum) across vegetative, reproductive, and full-season segments (new Section 5.1), following the methodology present by Ren et al. (2023b).

Notably, the results indicate that the two LAI-derived metrics—$LAI_{mean1}$ and $LAI_{mean2}$—outperformed those derived from either a single phenological stage or the entire growing season in terms of correlation with yield. This supports the findings of Ren et al. (2023b) highlighting the complementary value of combining vegetative and reproductive stage indicators. Conceptually, $LAI_{mean1}$ captures vegetative vigor (establishment and biomass accumulation, Kodadinne Narayana et al., (2024)), while $LAI_{mean2}$ reflects reproductive canopy status — together they summarize the two most

yield-informative phases and mitigate the redundancy present in full sequences.

**3. Practical Constraints on Full Time-Series Feature Use**

Although full LAI sequences outperform stage-averaged inputs at local scale, their application at regional scale is constrained by (a) strong spatiotemporal heterogeneity of Sentinel-2 image availability (Fig. A1), which requires constructing a specific time-series input for every site impartial, and (b) resource-intensive in computational and data-management for many different sequence-patterns. The two stage-averaged design is a phenology-informed compromise that preserves most predictive power while ensuring scalability and robustness to data gaps.

**4. Exclusion of TRA and SM Features**

While transpiration and soil moisture are relevant agronomic variables, they were ultimately excluded due to:

- Lack of high-resolution, high-frequency remote sensing products (especially for TRA); and
- Weak or inconsistent correlations with yield in our dataset, possibly due to indirect or stage-specific effects.

In summary, our comparative experiment, expanded feature evaluation, and the discussion on practical limitations demonstrates that (a) the two stage-averaged LAI features are a computationally efficient choice for regional yield mapping, and (b) full time-series inputs offer modest accuracy gains that are best exploited at local scales or in contexts with dense, regular observations. We have added these results and the associated discussion in Section 4.2, Section 5.1, and Figures 5, A4, and A1. We also acknowledged the limitations on the two-stage average LAI in Section 5.4 and discus future improvements (Lines 637 – 642).

We believe these additions and clarifications, supported by relevant literature, strengthen the methodological transparency and scientific rigor of our study. By selecting $LAI_{mean1}$ and $LAI_{mean2}$, we aimed to balance physiological interpretability with practical feasibility, ensuring that the model remains efficient and scalable for regional to continental applications using remotely sensed data.

Thank you again for your valuable feedback. We look forward to your further comments and suggestions.

Below is a part of the revised manuscript for your reference:

***3.2 Development of the Grated Recurrent Unit model (GRU)***

*Trained on the simulated dataset, the GRU constructed based on TensorFlow 2.6 linked simulated environmental inputs to yield outputs. For field scale yield estimation,*

*we first identified all available Sentinel-2 observation dates for each sample plot based on its corresponding Sentinel-2 tiles in 2022 and 2023 (Table 1), and computed the development stage value (DVS) for each date (Section 3.3.1). As LAI is a key biophysical indicator of soybean photosynthetic capacity and productivity (Malone et al., 2002; Shi et al., 2025), we extracted LAI values at the corresponding DVS from the soybean growth dataset to construct DVS-aligned LAI time-series. These DVS-aligned LAI were served as inputs to the GRU model, with simulated yield used as the target variable.*

***Table 1 DOY of available Sentinel-2 images for LAI extraction.***

| *Sentinel-2 tiles* | *DOY* |
|---|---|
| *Available Sentinel-2 data in 2022* | |
| *51UYP* | *118, 128, 138, 143, 158, 213, 218, 228, 253, 263, 268* |
| *52TCT* | *105, 128, 138, 143, 158, 170, 193, 213, 218, 228, 235, 238, 245, 253, 263, 270* |
| *52TDQ* | *105, 117, 130, 140, 160, 222, 245, 250, 265, 270* |
| *52TES* | *102, 110, 117, 130, 140, 162, 172, 187, 202, 207, 220, 240, 245, 252, 257, 262, 270* |
| *52TFR* | *104, 117, 129, 139, 159, 167, 187, 222, 229, 252, 257, 262, 269* |
| *52TGS* | *107, 114, 119, 129, 161, 187, 222, 229, 232, 252, 257, 262, 269* |
| *Available Sentinel-2 data in 2023* | |
| *51TWN* | *98, 103, 108, 113, 118, 121, 126, 138, 143, 148, 153, 163, 193, 211, 218, 226, 231, 248, 253, 263, 268* |
| *51TYM* | *98, 103, 113, 123, 128, 138, 143, 148, 193, 218, 248, 253, 258, 263, 273* |
| *52TCT* | *98, 103, 125, 130, 138, 143, 173, 183, 193, 218, 235, 245, 250, 255, 263* |
| *52TDP* | *102, 122, 127, 142, 165, 170, 187, 230, 245, 250, 255, 272* |
| *52TDQ* | *112, 117, 130, 142, 165, 245, 250, 257, 272* |
| *52TDT* | *105, 130, 137, 142, 147, 167, 177, 232, 245, 250, 255, 265, 272* |

*However, the spatiotemporal variability of Sentinel-2 image availability across regions posed challenges for regional-scale yield modelling, as constructing separate GRU models for each date combination demands considerable computational and storage resources. Accounting for the computational efficiency of the model in large areas, two stage-averaged LAI include $LAI_{mean1}$ (mean value of LAI during vegetative growth: emergence to flowering) and $LAI_{mean2}$ (mean value of LAI during reproductive growth: flowering to maturity), were calculated as inputs, while simulated yields acted as outputs. Boken and Shaykewich, (2002) has demonstrated the feasibility of estimating crop yield using average features derived from a specific phenological stage or from the entire growing season. To evaluate the effectiveness of the two stage-averaged LAI in soybean yield estimation, we applied the GRU model trained by the two stage-averaged LAI to yield estimation at the field scale for comparison with model based on DVS-aligned LAI time-series, and further extended its application to the regional scale.*

### *3.3.2 Model estimations of soybean yield*

*DVS-aligned LAI values derived from available Sentinel-2 data (Table 1) were firstly used as input of the GRU model for yield estimation at field scale. Meanwhile, averaged LAI values during the vegetative ($LAI_{mean1}$) and reproductive ($LAI_{mean2}$) growth stages were calculated and used as model input for estimations at both field and regional scales.*

### *4.2 Yield estimation at field scale*

*The field scale performance of GRU models using full DVS-aligned LAI and two stage-averaged LAI was validated against in-situ measurement from 2022 and 2023 (Fig. 5). The estimated yields exhibited strong agreement with observed yield, with $R^2 > 0.65$ ($p < 0.01$) in all scenarios. Validation results (Fig. 5 and Fig. A4) showed that the DVS-aligned GRU model achieved slightly better accuracy (RMSE = 224.81 kg ha$^{-1}$, MRE = 7.50%), while the stage-averaged model remained competitive (RMSE = 287.44 kg ha$^{-1}$, MRE = 10.02%). The difference in MRE was around 3% across both years, suggesting that the simplified approach using two stage-averaged LAI was a feasible alternative for yield estimation.*

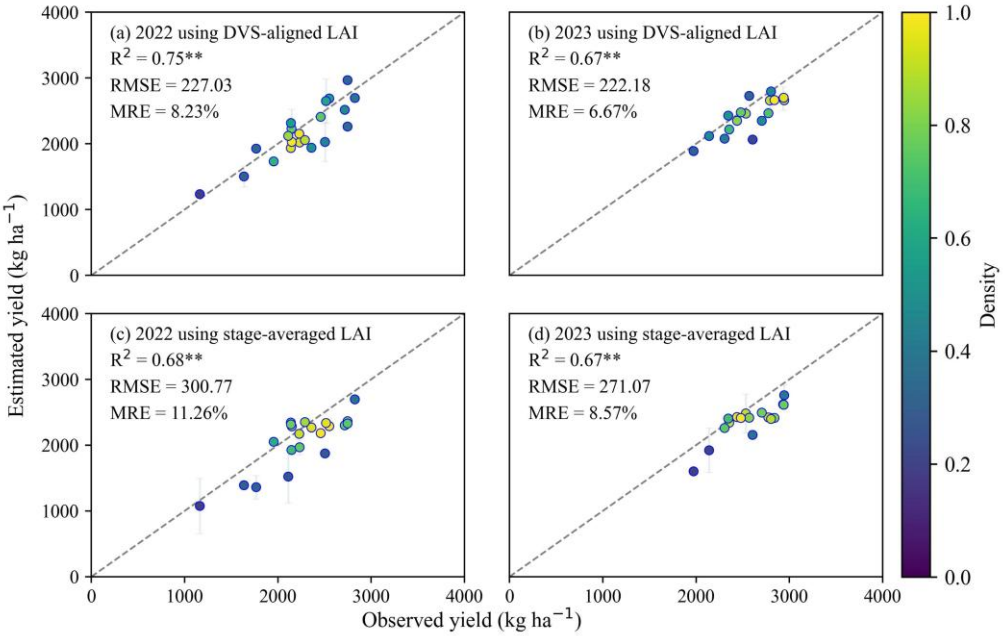



***Figure 1: Scatterplots between estimated and observed soybean yield for 2022 and 2023. (a) and (b) show results for 2022 and 2023, respectively, using the full DVS-aligned LAI; (c) and (d) show results for 2022 and 2023, respectively, using two stage-averaged LAI. Error-bars represent one standard deviation indicating the uncertainty of yield estimations. The dashed line represents 1:1 line. ** denotes statistical significance at $p < 0.01$.***

## *5.1 Selection of model input features*

*In this study, two stage-averaged LAI ($LAI_{mean1}$ and $LAI_{mean2}$) were selected as alternative input features to DVS-aligned LAI for soybean yield estimation. Although full LAI sequences yielded higher accuracy at local scale (Fig. 5 and Fig. A4) , their regional application was limited by (a) strong spatiotemporal heterogeneity of Sentinel-2 image availability (Fig. A1), which required constructing a specific time-series input for every site impartial, and (b) resource-intensive in computational and data-management costs associated with model training and maintaining models for many different sequence-patterns. The two stage-averaged LAI features are a computationally efficient solution for regional yield mapping, while full time-series inputs offer modest accuracy gains that are best exploited at local scales or where dense, regular observations are available.*

*For further analysis, we systematically evaluated a broad set of candidate features derived from LAI, transpiration (TRA), and surface soil moisture (SM) extracted from the simulated soybean growth dataset. To develop a unified model suitable for long-*

*term and large-scale soybean yield estimation, we employed statistical summaries of these features rather than time-series features tied to specific image acquisition dates (as done by Du et al., (2025)). For each variable, four statistical descriptors—including mean, maximum, median, and cumulative sum—were calculated separately for the vegetative growth stage, the reproductive growth stage, and the whole growing season, following the approach of Ren et al., (2023b).*

*As shown in Fig. 11, LAI-derived features exhibited consistently strong correlations with simulated yield (r = 0.54 to 0.88), reflecting the role of LAI as a critical proxy for canopy development, light interception, and biomass accumulation (Cao et al., 2025; Shi et al., 2025). The multi-spectral retrieval of LAI therefore effectively characterizes both structure and physiological status of the crop canopy, supporting its dominant predictive power in our feature set.*

*Notably, the two stage-averaged LAI ($LAI_{mean1}$ and $LAI_{mean2}$) exhibited stronger correlations with yield than features derived from either single phenological stage or the entire growing season, which is consistent with Ren et al., (2023b). Conceptually, $LAI_{mean1}$ captures vegetative vigor (establishment and biomass accumulation, Kodadinne Narayana et al., (2024)), while $LAI_{mean2}$ reflects reproductive canopy status. These two features jointly summarize the two most yield-informative phases and mitigate the redundancy present in full sequences. Among the candidate features, mean-based features outperformed maximum, median, and cumulative counterparts. This likely due to their lower sensitivity to extreme values and day-to-day fluctuations, making them a more stable and representative indicator of canopy conditions across the two growth periods.*

*While some TRA-based features (e.g., $TRA_{sum}$) showed relatively high correlation with yield, they were excluded owing to practical constraints. Current TRA retrieval methods primarily rely on thermal-infrared remote sensing, which typically has coarse spatial and temporal resolution (Hou et al., 2018; Zhang, 2003) limiting its utility for high-resolution, regional mapping. Similarly, SM-related features showed weak or inconsistent correlations with yield across growth stages in our simulations, indicating a limited direct influence on soybean yield production under modeled conditions.*

*In summary, to optimize model inputs for efficient, large-scale applications, and to facilitate the generation of soybean yield dataset, the two stage-averaged LAI features ($LAI_{mean1}$ and $LAI_{mean2}$) were selected. This selection balances physiological relevance and temporal specificity with strong predictive performance and practical feasibility, enabling competitive yield estimation using only two interpretable, remotely sensed retrievable predictors.*

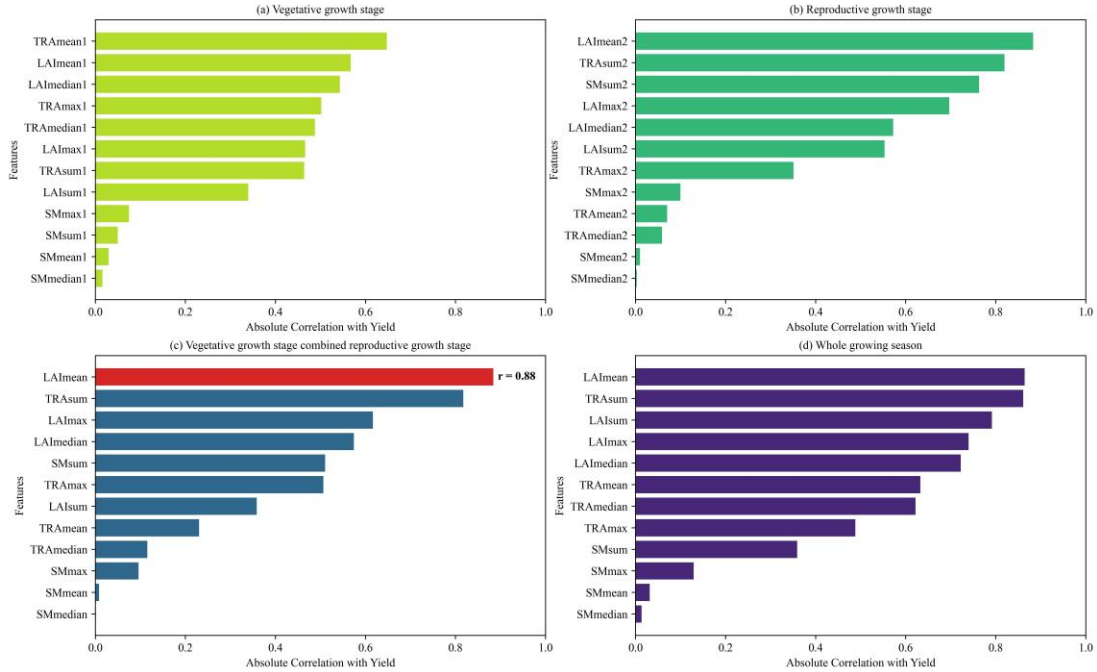
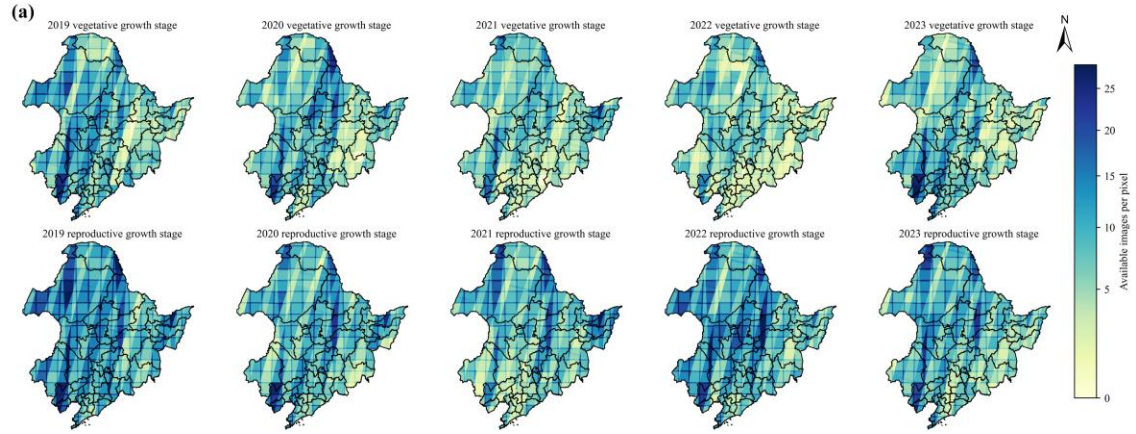



***Figure 2: The absolute Pearson correlation coefficients between each candidate feature and simulated soybean yield, grouped by growth stages: (a) vegetative growth stage; (b) reproductive growth stage; (c) vegetative growth stage combined reproductive growth stage and (d) whole growing season, respectively.***

## *5.4 Limitations and future developments*

*Third, using the two stage-averaged LAI introduces additional sources of uncertainty in yield estimation. Excessive temporal aggregation inevitably obscures growth dynamics. Different growth trajectories can produce similar stage-based LAI yet correspond to different yields, increasing the risk of non-unique LAI for spatially yield mappings. This simplification also limits the modelling capacity of GRU architectures, which are specifically designed to exploit sequential dependencies in time series inputs. Future work can explore hybrid approaches that combine stage-based summaries with higher-frequency or full-season time series of vegetation indicators to improve both interpretability and yield prediction robustness.*



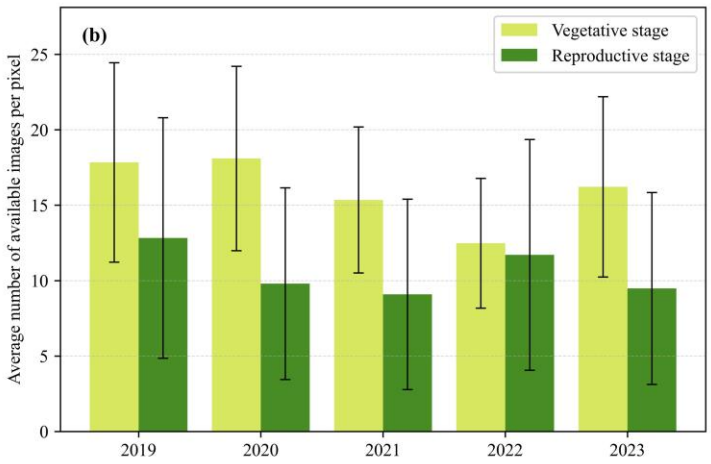



***Figure A1: Spatial distribution of the number of available Sentinel-2 images per pixel for each year: vegetative growth stage (top) and reproductive growth stage (bottom) (a) and yearly averages for each growth stage with error-bars representing spatial standard deviation across pixels within the study area (b).***

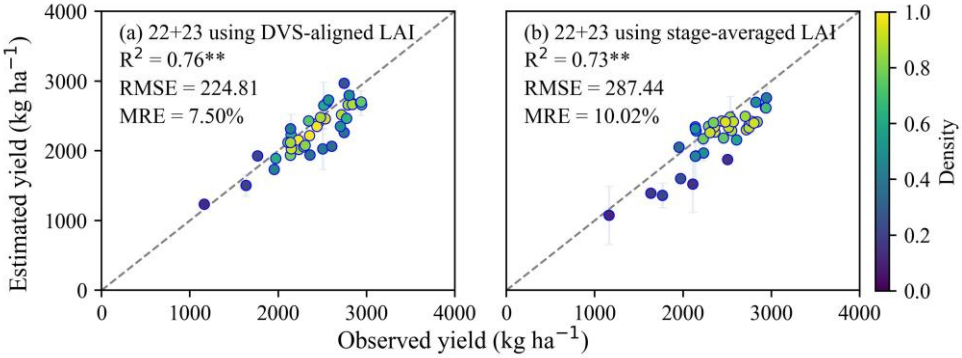



***Figure A2: Comparison between estimated and observed yield (2022 + 2023). (a)shows the estimates using the full DVS-aligned LAI and (b) shows the results using two stage-averaged LAI. The error-bars represent one standard deviation indicating the uncertainty of yield estimations. Dashed lines represent 1:1 line. ** denotes statistical significance at $p < 0.01$.***

*Boken, V. K. and Shaykewich, C. F.: Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index, International Journal of Remote Sensing, 23, 4155–4168, https://doi.org/10.1080/01431160232056795 5, 2002.*

*Cao, H., Zhao, R., Xia, L., Wu, S., and Yang, P.: Trends in crop yield estimation via data assimilation based on multi-interdisciplinary analysis, Field Crops Research, 322, 109745, https://doi.org/10.1016/j.fcr.2025.109745, 2025.*

*Du, X., Zhu, J., Xu, J., Li, Q., Tao, Z., Zhang, Y., Wang, H., and Hu, H.: Remote sensing-based winter wheat yield estimation integrating machine learning and crop growth multi-scenario simulations, International Journal of Digital Earth, 18, 2443470, https://doi.org/10.1080/17538947.2024.2443470, 2025.*

*Hou, M., Tian, F., Zhang, L., Li, S., Du, T., Huang, M., and Yuan, Y.: Estimating Crop Transpiration of Soybean under Different Irrigation Treatments Using Thermal*

*Infrared Remote Sensing Imagery, Agronomy, 9, 8, https://doi.org/10.3390/agronomy9010008, 2018.*

*Kodadinne Narayana, N., Wijewardana, C., Alsajri, F. A., Reddy, K. R., Stetina, S. R., and Bheemanahalli, R.: Resilience of soybean genotypes to drought stress during the early vegetative stage, Sci Rep, 14, https://doi.org/10.1038/s41598-024-67930-w, 2024.*

*Malone, S., Ames Herbert, D., and Holshouser, D. L.: Relationship Between Leaf Area Index and Yield in Double-Crop and Full-Season Soybean Systems, ec, 95, 945–951, https://doi.org/10.1603/0022-0493-95.5.945, 2002.*

*Ren, Y., Li, Q., Du, X., Zhang, Y., Wang, H., Shi, G., and Wei, M.: Analysis of Corn Yield Prediction Potential at Various Growth Phases Using a Process-Based Model and Deep Learning, Plants, 12, 446, https://doi.org/10.3390/plants12030446, 2023b.*

*Shi, B., Guo, L., and Yu, L.: Accurate LAI estimation of soybean plants in the field using deep learning and clustering algorithms, Front. Plant Sci., 15, https://doi.org/10.3389/fpls.2024.1501612, 2025.*

*Zhang, R.: Determination of regional distribution of crop transpiration and soil water use efficiency using quantitative remote sensing data through inversion, Sci China Ser D, 46, 10, https://doi.org/10.1360/03yd9002, 2003.*