Responses to the comments of Reviewers

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 reviewers and editor,

Thank you very much for your careful review and constructive comments on our manuscript, "NortheastChinaSoybeanYield20m: an annual soybean yield dataset at 20 m in Northeast China from 2019 to 2023" (Manuscript ID: ESSD0805). We greatly appreciate the time and effort you have devoted to improving our work.

In response to your suggestions, we have implemented the following substantive revisions:

(1) **Methodological Refinements:**
- Validated the accuracy of Landsat-derived LAI using 94 field samples collected in Northeast China.
- Provide detailed justification for region-specific WOFOST parameter settings.
- 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) **Terminology and Writing Standardization:**
- Replaced potentially misleading terms, for example "scenario" is now "parameter combination", and "time series LAI" is now "two stage-averaged LAI".
- Unified yield unit throughout the manuscript to "kg ha^{-1}".
- Clarified the moisture adjustment procedure to ensure consistency with the 13% benchmark used in statistical data sources.

(3) **Strengthened Literature Support:**
- Added relevant references to substantiate our approach and enhance the methodological credibility, providing stronger support for our research findings.

We also provided a detailed point-to-point responses (highlighted in blue) to each reviewer's comment in the accompanying response document. We believe these revisions significantly enhance the manuscript's rigor, clarity and overall quality. We look forward to your further feedback on our revision.

Best regards

Xin Du

Point-to-Point response

 # Reviewer 1:

After the overall revision of the manuscript, it has met the requirements for publication. All the issues I was concerned about have been thoroughly answered.

Reply: Thank you once again for this encouraging feedback. We greatly appreciate the time and effort you have dedicated to reviewing our manuscript. We're delighted to hear that our revisions have addressed your concerns and enhanced the quality of our work.

 # Reviewer 2:

While the revised manuscript demonstrates improved clarity and structure, the current workflow still lacks the methodological rigor and transparency necessary for a reliable and reproducible data product. Without addressing the following key concerns, the dataset is not yet suitable for public release or long-term citation by the scientific community. Further refinement is needed to ensure its scientific robustness and usability:

Reply: Thank you for your thoughtful assessment and for the opportunity to further strengthen our manuscript.

We fully agree that the methodological rigor and transparency are essential for a widely usable, citable data product.

In response to your concerns, we have undertaken the following major revisions:

(1) Expanded methodological Detail
(2) Reproductivity Enhancements

We believe these changes fully address the need for rigor, transparency, and reproducibility, making the dataset suitable for public release and long-term citation.

1. Lack of region-specific LAI inversion validation. The LAI inversion function used in this study was adopted directly from existing literature based on test sites in Spain and Italy, without any calibration or validation using local field data from Northeast China. Given the differences in crop types, soil backgrounds, and environmental conditions, such region-specific models may introduce systematic bias, undermining the reliability of both LAI estimates and the final yield predictions.

Reply: We fully agree with the original LAI-$NDVI_{RE}$ ($LAI = 5.405 \cdot NDVI_{RE} - 0.114$)

was developed for European test sites and therefore may not directly transfer to Northeast China.

To address this, in the revision, we now have carried out a region-specific validation using 94 independent LAI field plots sampled across our study area, which was acquired from the Common Application Support Platform for Land Observation Satellites of China's Civil Space Infrastructure platform (CAPLOS). For each plot, we paired cloud-free Sentinel-2 images (±2 days of sampling) with ground-measured LAI and re-estimated LAI via the $NDVI_{RE}$ regression. The validation results ($R^2 = 0.81$, RMSE = 0.89, Fig. A3) are now presented in Section 3.3.2 (Lines 341–344). While performance is strong overall, we note a modest overestimation for LAI < 3, which we attribute to local soil-background effects not present in the original European calibration.

We have expanded the manuscript to include:

- A detailed description of the LAI sampling protocol (plot size, instrument, and measurement dates) in Lines 127 – 139.
- A map of the 94 sampling locations in Fig.1
- A discussion (Lines 645 – 650) on why red-edge based VIs (like $NDVI_{RE}$) generally offer more robust transferability across crop types and regions, yet remain sensitive to soil differences under low-canopy conditions.
- The impacts of directly using the LAI-$NDVI_{RE}$ (Lines 650 – 651) on the yield estimation is discussed. In the Section 5.4, we also discussed future works including coupling canopy radiative transfer model with crop growth model to use satellite reflectance directly, which we expect would further improve model accuracy and reduce bias from site-specific LAI regressions (Lines 652 – 656).

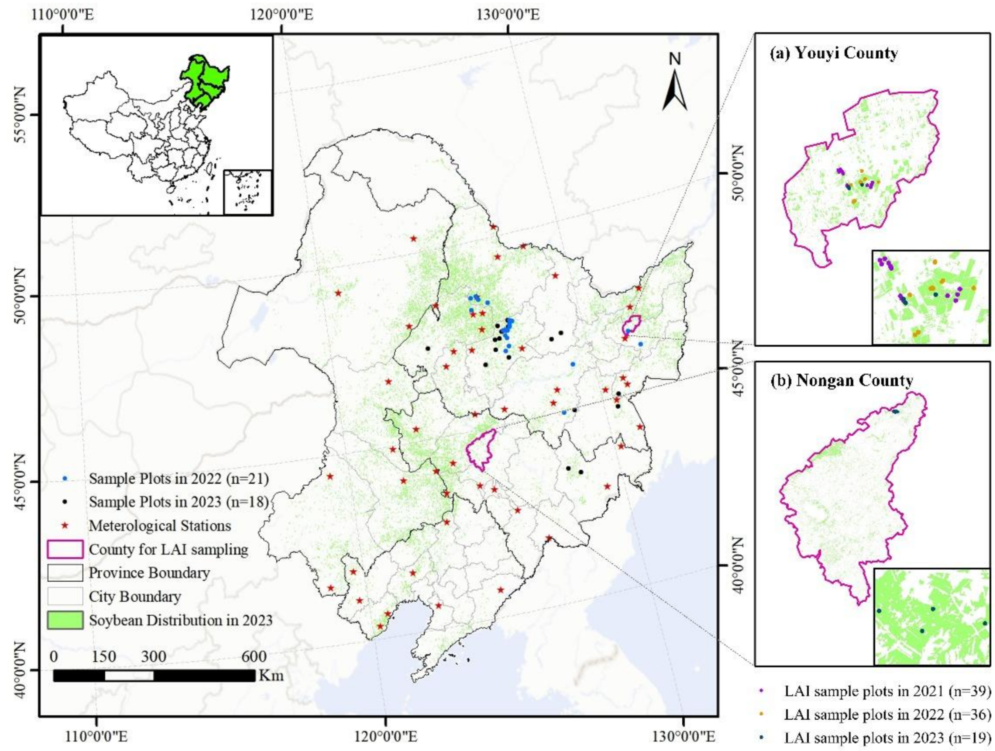

Figure 1: Location of the study area, distribution of sample plots in 2022 and 2023, LAI sampling counties and selected meteorological stations. (a) and (b) display the detailed distribution of LAI sampling plots in Youyi and Nongan counties, respectively. The soybean distribution map was obtained from Zhao et al., (2022), generated using a moment-preserving segmentation method with an overall accuracy exceeding 90% for soybean in 2023 (see Section 2.2.5 for details).

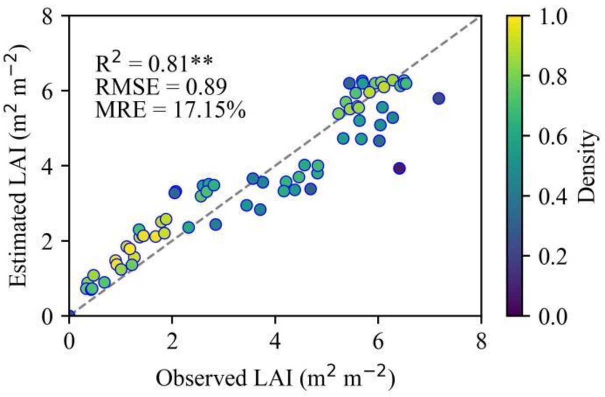

Figure A1: Comparison between estimated and field measured LAI from 2021 to 2023. Error-bars represent one standard deviation indicating the uncertainty of LAI estimations. The dashed lines represent the 1:1 line. ** denotes statistical significance at $p < 0.01$.

2.Absence of model calibration and validation in synthetic training data generation The WOFOST simulations used to generate the synthetic training dataset were built from randomly combined parameters without site-specific calibration or validation against independent observations. This raises concerns about the biological realism and representativeness of the training data, which is especially critical if the final product is to be published as a formal, public dataset. As it stands, the workflow resembles more an experimental modeling framework than a rigorously validated data production pipeline.

Reply: Thank you for raising this important concern.

We fully agree that site-specific calibration and validation are essential to ensure the biological realism and representativeness of WOFOST-based crop growth simulations. In this study, we adopted a careful, literature- and observation-informed parameterization and performed multiple checks to demonstrate the realism of our synthetic dataset:

1. **Region-informed parameter compilation**. Rather than using purely random parameter combinations, we compiled soybean-specific model parameters from established sources relevant Northeast China, including peer-reviewed studies, publicly available field reports, agronomic databases, and WOFOST defaults adapted to the regional context. Parameters were delineated into four groups:

- Meteorological inputs: 42 years (1980–2021) of daily weather data (e.g., temperature, precipitation, and radiation) from 51 representative stations capturing climatic variability across the soybean-growing areas.
- Soil parameters: Four dominant soil types in the study region drawn from regional studies.
- Crop variety parameters: Phenological and physiological values taken from studies on soybean varieties conducted in the same agro-climatic zone.
- Agro-management settings: Four representative sowing dates (20 April, 30 April, 10 May, and 20 May) summarized from literatures to reflect the regional planting window.

2. Extensive but constrained simulation space. These informed parameter sets were systematically combined to create 171,360 various simulations that sample plausible production conditions while remaining grounded in regional practice (see Table 3).

3. Empirical validation of simulated dynamics. To evaluate the realism of the synthetic data, we conducted independent validations:

- LAI validation: WOFOST-simulated LAI trajectories were compared against 94 independent field-measured LAI samples (Fig. 3). With 88% of field observations fell within the simulated range, demonstrating realistic canopy dynamics.
- Yield validation: Simulated yield distributions were compared with multi-source observations (statistical yearbooks, published datasets, and literature-

based field trials), were found to span the observed yield ranges for the region (Fig. 4).

These results confirm that our simulated soybean growth dataset is biologically plausible and representative enough to support GRU model training and data-driven yield estimation. This also indirectly validates our parameterization approach, as the simulated outputs successfully reproduced observed LAI and yield patterns.

4. Transparency and documentation. In revision, we expanded the Section 3.1.1 to detail our parameterization rationale and parameter sources. We also added a discussion of limitations and potential biases of our current workflow in the Discussion section (Lines 603 – 607).

5. Future improvements. In the revised Discussion Section (Lines 607 – 610), we acknowledged the limitations of our current study, and we further outline future studies, such as targeted field-measured and remotely sensed products (e.g., surface soil moisture, SIF, Lidar-based biomass, leaf chlorophyll content) for model calibration, integration of data assimilation (e.g., assimilating SIF or leaf chlorophyll content) and model sensitivity analyses to further increase biological realism.

Taken together, these steps substantially increase confidence that our synthetic dataset is biologically plausible and fit for training the GRU model while transparently acknowledging remaining limitations.

3.Oversimplification of LAI temporal features and limited use of GRU modeling capacity. The entire LAI time series was reduced to just two stage-based mean values (LAImean1 and LAImean2), introducing substantial uncertainty and increasing the risk of non-unique mappings between LAI and yield. Different growth trajectories may produce similar LAI means but result in significantly different yields. Furthermore, such a reduced input structure fails to utilize the temporal modeling advantages of the GRU architecture, calling into question the necessity of using a recurrent model at all.

Reply: Thank you for this important comment that has been raised by the editor as well.

We fully agreed that excessive temporal aggregation can obscure growth dynamics and risk non-unique LAI at key growth stages, therefore resulting in poor crop yield estimation and mapping.

To address this, in the revision, 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).

Among these, the combination LAI_{mean1} + LAI_{mean2} showed the strongest, most consistent correlation with yield., 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 — 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-averaged LAI in Section 5.4 and discus future improvements (Lines 637 – 642).

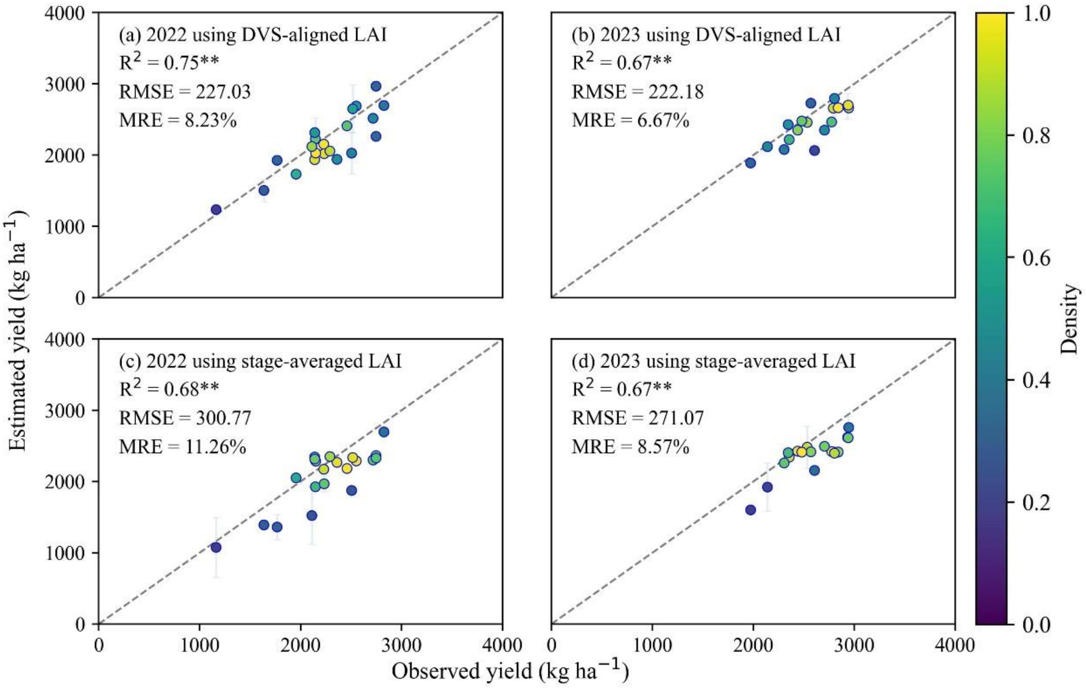

Figure 2: Scatterplots between estimated and observed soybean yield for 2022 and 2023. (a) and (b) show results for 2022 and 2023, respectively, using the full DVS-aligned LAI; (c) and (d) show results for 2022 and 2023, respectively, using two stage-averaged LAI. Error-bars represent one standard deviation indicating the uncertainty of yield estimations. The dashed line represents 1:1 line. ** denotes statistical significance at $p < 0.01$.

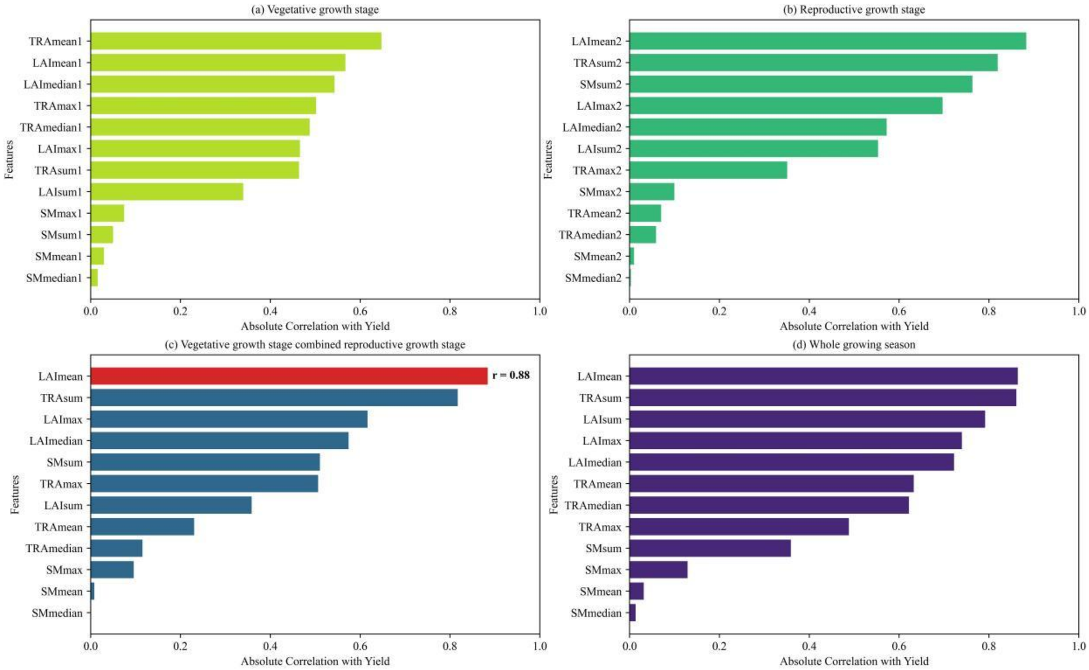

Figure 3: The absolute Pearson correlation coefficients between each candidate feature and simulated soybean yield, grouped by growth stages: (a) vegetative growth stage; (b) reproductive growth stage; (c) vegetative growth stage combined reproductive growth stage and (d) whole growing season, respectively.

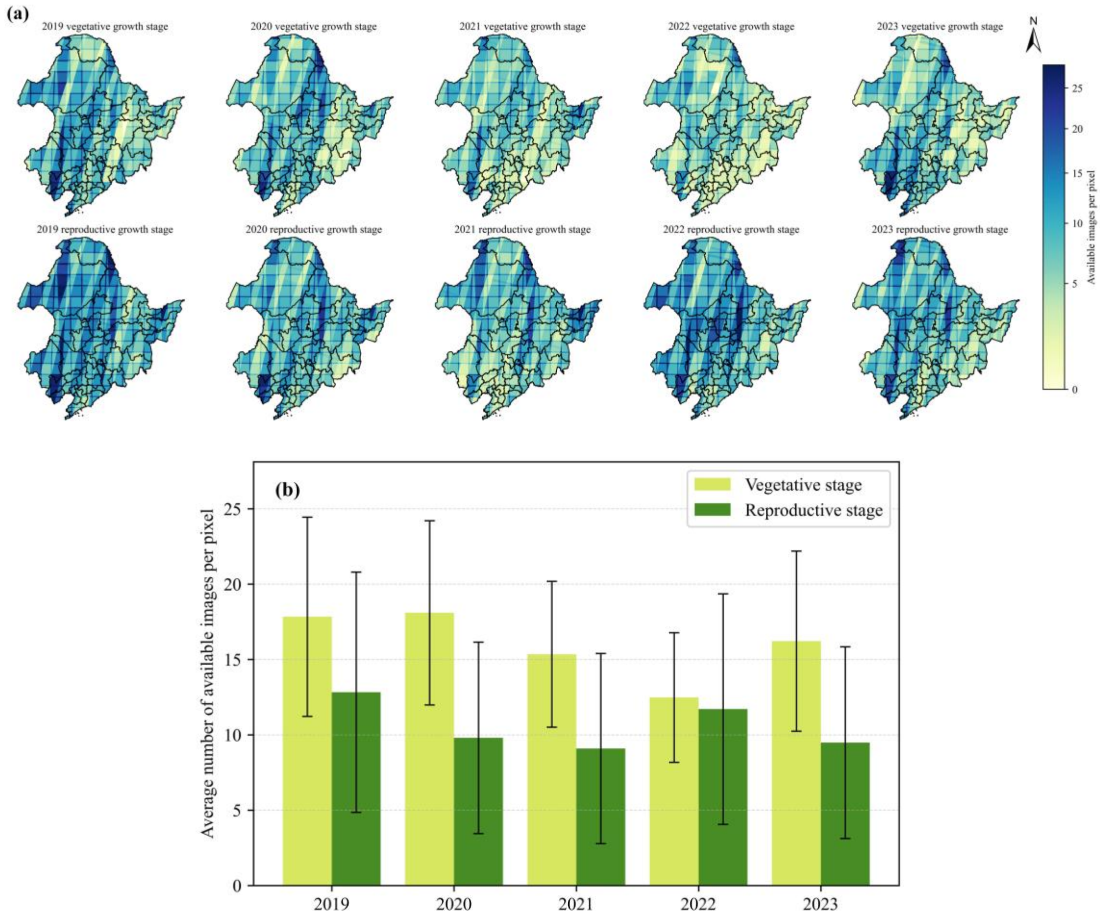

Figure A1: Spatial distribution of the number of available Sentinel-2 images per pixel for each year: vegetative growth stage (top) and reproductive growth stage (bottom) (a) and yearly averages for each growth stage with error-bars representing spatial standard deviation across pixels within the study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

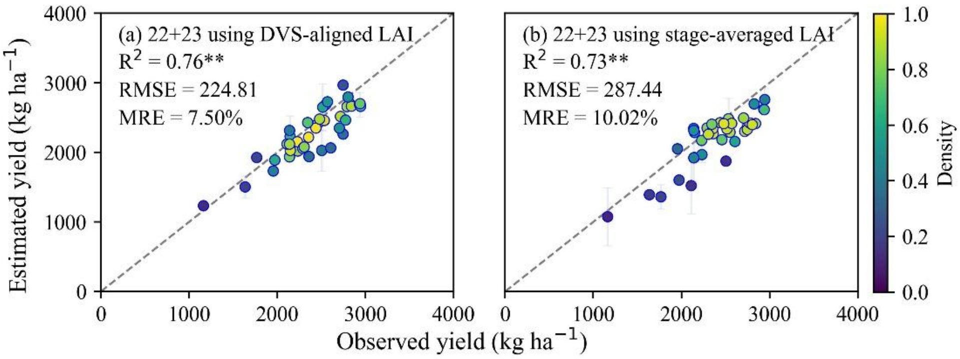

Figure A2: Comparison between estimated and observed yield（2022 + 2023). (a) shows the estimates using the full DVS-aligned LAI and (b) shows the results using two stage-averaged LAI. The error-bars represent one standard deviation indicating the uncertainty of yield estimations. Dashed lines represent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

References supporting the methodology adopted in our study:

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/10.1080/014311602320567955, 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.

Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., Walters, D., Jiao, X., Geng, X., and Shi, Y.: Assessment of red-edge vegetation indices for crop leaf area index estimation, Remote Sensing of Environment, 222, 133–143, https://doi.org/10.1016/j.rse.2018.12.032, 2019.

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.

Nguy-Robertson, A. L., Peng, Y., Gitelson, A. A., Arkebauer, T. J., Pimstein, A., Herrmann, I., Karnieli, A., Rundquist, D. C., and Bonfil, D. J.: Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm, Agricultural and Forest Meteorology, 192–193, 140–148, https://doi.org/10.1016/j.agrformet.2014.03.004, 2014.

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.

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., and Peng, Y.: Comparison of different vegetation indices for the remote assessment of green leaf area index of crops, Remote Sensing of Environment, 115, 3468–3478, https://doi.org/10.1016/j.rse.2011.08.010, 2011.

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.

Specific comments:

1.L196: In Figure 2, the term “time series LAI” may be misleading, as only two stage-averaged LAI values (LAImean1 and LAImean2) were used. I do not think this constitutes a true time series.

Reply: Thank you for pointing this out.

In the revised manuscript, we have replaced “time series LAI” with “two stage-averaged LAI” in the caption of Figure 2 and throughout the text to ensure consistency and avoid confusion. The terminology has been revised to more accurately reflect the nature of the input features used.

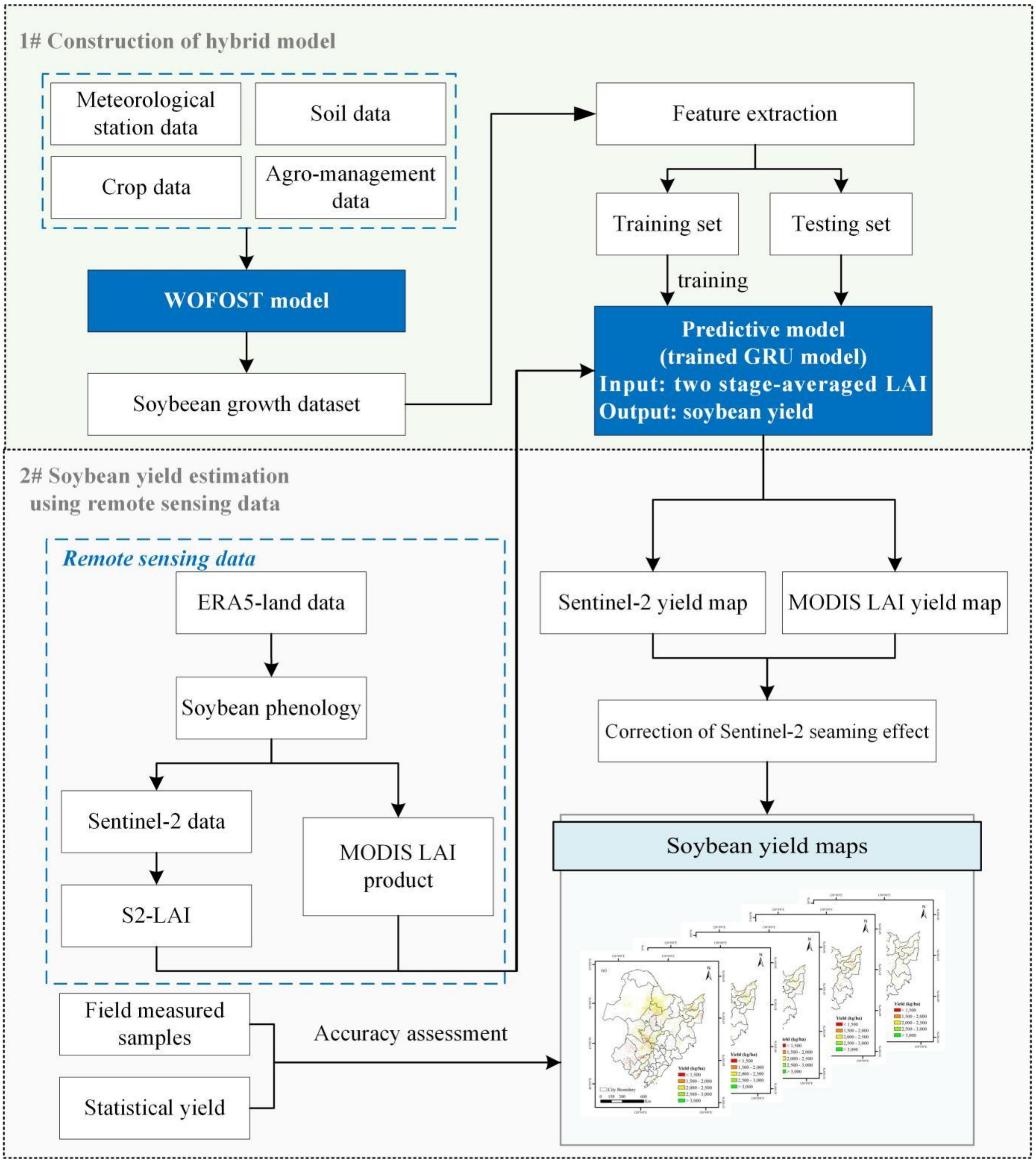

Figure 4: The flowchart of the overall yield estimation methodology in this study.

2.L245: The term "scenario" is misleading in this context. The 171,360 combinations are simply full-factorial simulations used to generate synthetic training data, not meaningful scenario-based experiments. It is suggested to revise the terminology to "simulation runs," "parameter combinations," or "sample pool" for clarity.

Reply: Thank you for the valuable suggestion. In the revised manuscript, we have replaced “scenario” with the more appropriate term “parameter combination” throughout the text (Lines 275 – 281).

3.L335: Were the statistical yields (B) in Figure 4(b) and all subsequent comparisons with statistical yearbook data converted to dry weight? Since WOFOST outputs and field measurements are in dry weight, please clarify whether any moisture adjustment was made.

Reply: Thanks for your suggestion. To ensure consistency across all data sources, all yield records used in the comparisons (including the statistical yields shown in Fig 4.b) were harmonized to 13% moisture content. Because WOFOST outputs and our field measurements were reported as dry matter, they were converted to the 13% moisture basis using: $Yield = Dry\_weight/(1 - 0.13)$.

We have clarified this conversion in the revised manuscript (Lines 351 – 354) to avoid any ambiguity.

4. A minor suggestion: The manuscript uses both "kg ha^{-1}" and "kg/ha" to express yield units. It is recommended to standardize the format throughout the text and figures, preferably using "kg ha^{-1}".

Reply: revised as suggested.