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 reviewer 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: essd-2024-586). We greatly appreciate the time and effort you have devoted to improving our work.

In response to your comments, we have carefully revised the manuscript throughout. The main revisions include:

(1) improving the clarity and consistency of figures and maps;

(2) polishing the language and correcting grammatical, spelling, and formatting issues throughout the manuscript;

(3) updating and strengthening the reference support;

(4) clarifying several methodological descriptions and data sources; and

(5) refining the discussion to better present the limitations and possible future improvements of the proposed framework.

We also provided a detailed point-to-point responses (highlighted in blue) to each 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

This manuscript presents a valuable soybean yield dataset for Northeast China based on a hybrid WOFOST–GRU framework. Overall, the study is meaningful, technically sound, and potentially useful for the remote sensing and agricultural modelling communities. The manuscript is generally well organized, and the dataset appears to be of relatively high quality, with good spatial resolution, multi-year temporal coverage, and promising application value for crop monitoring and agricultural analysis. There are some suggestions as follows, which can be considered for further improvement of the manuscript.

Reply: Thank you very much for the suggestion. As suggested, we have carefully gone over the entire manuscript and made corresponding revisions.

1. Some of the references cited in the manuscript are relatively old. The manuscript would benefit from including more up-to-date studies.

Thanks for your suggestion. We have conducted a more comprehensive literature review and incorporated updated references to strengthen the theoretical foundation.

Anderson, M. C., Kustas, W. P., Norman, J. M., Diak, G. T., Hain, C. R., Gao, F., Yang, Y., Knipper, K. R., Xue, J., Yang, Y., Crow, W. T., Holmes, T. R. H., Nieto, H., Guzinski, R., Otkin, J. A., Mecikalski, J. R., Cammalleri, C., Torres-Rua, A. T., Zhan, X., Fang, L., Colaizzi, P. D., and Agam, N.: A brief history of the thermal IR-based Two-Source Energy Balance (TSEB) model – diagnosing evapotranspiration from plant to global scales, Agricultural and Forest Meteorology, 350, 109951, https://doi.org/10.1016/j.agrformet.2024.109951, 2024.

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.

Gaso, D. V., Paudel, D., De Wit, A., Puntel, L. A., Mullissa, A., and Kooistra, L.: Beyond assimilation of leaf area index: Leveraging additional spectral information using machine learning for site-specific soybean yield prediction, Agricultural and Forest Meteorology, 351, 110022, https://doi.org/10.1016/j.agrformet.2024.110022, 2024.

Ko, J., Shin, T., Kang, J., Baek, J., and Sang, W.-G.: Combining machine learning and remote sensing-integrated crop modeling for rice and soybean crop simulation, Front. Plant Sci., 15, 1320969, https://doi.org/10.3389/fpls.2024.1320969, 2024.

Radočaj, D., Plaščak, I., and Jurišić, M.: Phenology-Based Maize and Soybean Yield Potential Prediction Using Machine Learning and Sentinel-2 Imagery

Time-Series, Applied Sciences, 15, 7216, https://doi.org/10.3390/app15137216, 2025.

Xin, M., Zhang, Z., Han, Y., Feng, L., Lei, Y., Li, X., Wu, F., Wang, J., Wang, Z., and Li, Y.: Soybean phenological changes in response to climate warming in three northeastern provinces of China, Field Crops Research, 302, 109082, https://doi.org/10.1016/j.fcr.2023.109082, 2023.

2. The quality of some distribution maps should be improved. In particular, the basemaps used in Figure 1 and Figure 9 do not appear sufficiently clear or fully consistent in style. In Figure 8 and Figure A5, panels (a)–(e) seem to represent different years, but this is not explicitly stated in the figure titles or captions.

Reply: Thank you for this helpful comment. We have revised the distribution maps to improve both clarity and stylistic consistency. Accordingly, Figures 1 and 9 have been redrawn using a simplified white-background basemap with unified cartographic elements, including consistent boundary styling, font settings, north arrows, and scale bars. This revision improves readability and ensures a more standardized presentation across figures. In addition, the captions of Figures 8 and A5 were revised to explicitly state that panels (a)–(e) correspond to the years 2019, 2020, 2021, 2022, and 2023, respectively (Line 430, Line 655).

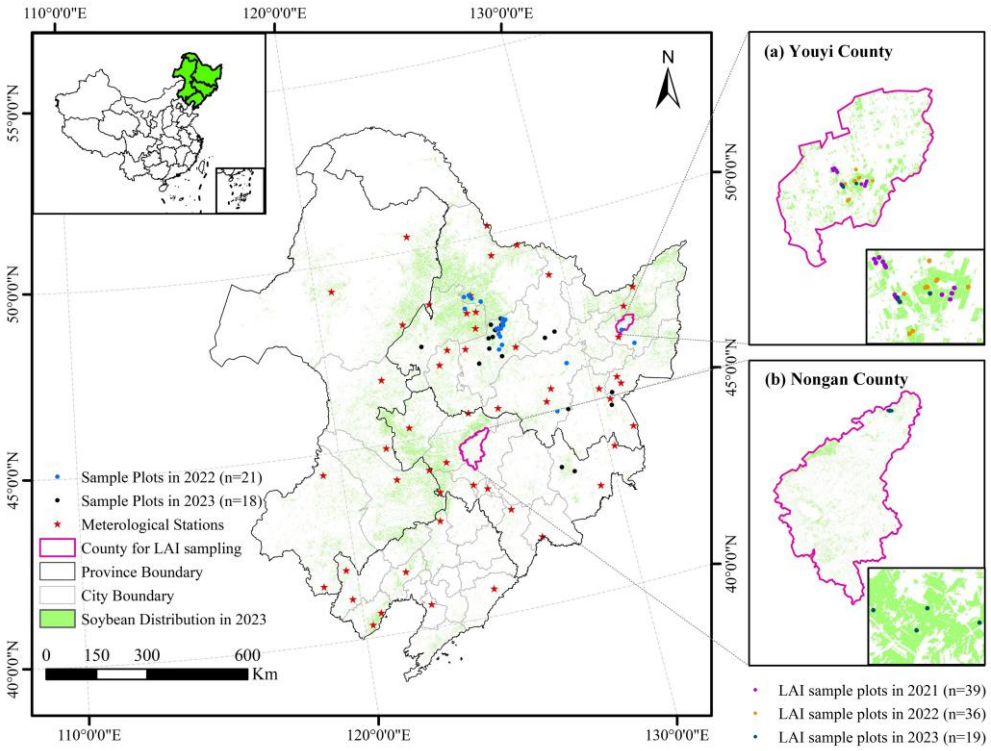

Figure 1: Location of the study area, distribution of sample plots in 2022 and 2023, LAI sampling counties and selected meteorological stations. (a) and (b) display the detailed distribution of LAI sampling plots in Youyi and

Nongan counties, respectively. The soybean distribution map was derived from Zhao et al. (2022), generated using a moment-preserving segmentation method with an overall accuracy exceeding 90% for soybean in 2023 (see Sect. 2.2.5 for details).

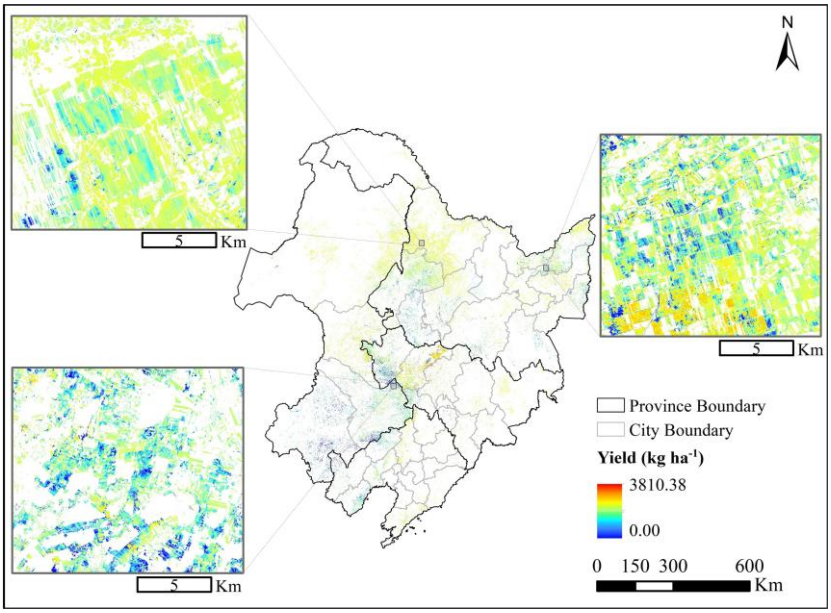

Figure 2: An example of a yield estimate for 2023 used to showcase detailed local estimates.

3. The manuscript should pay more attention to consistency of expression and formatting throughout. For example, the use of hyphenation should be standardized, such as “20 m” versus “20-m”.

Reply: Thanks for your suggestion. We have thoroughly checked the entire manuscript and revised the text to improve consistency in expression and formatting throughout. In particular, the use of hyphenation has been standardized across the manuscript. We also carefully reviewed similar issues in terminology, capitalization, spacing, and figure/table formatting to ensure a more uniform presentation.

4. Line 195-200: The description is not fully clear. The flowchart contains two main steps: (1) construction of the hybrid model and (2) soybean yield estimation using remote sensing data. However, in the text, step (1) is further broken down in a way that is not fully aligned with the figure.

Reply: Thanks for your suggestion. We have revised the relevant text to make the workflow consistent with the figure. Specifically, the methodology is now described as consisting of two main steps: (1) construction of the hybrid model and (2) soybean yield estimation using remote sensing data. In the revised text, the generation of the WOFOST-based training dataset and the GRU model training are clearly presented as

two subcomponents of Step 1, rather than as separate parallel steps. This revision improves the consistency between the figure and the text (Line 198–203).

5. Line 300: the source of "thermal zone divisions and regional adapted cultivars" should be clearly specified.

Reply: Thank you for pointing this out. We have revised the manuscript to explicitly indicate the sources of "thermal zone divisions" and "regional adapted cultivars". Specifically, the thermal zone divisions were derived using the same method as Wang et al. (2022), and the regional adapted cultivars were based on Qu et al. (2023). The corresponding sentence has been revised for clarity (Line 297–298).

Wang, H., Liu, D., Chen, P., Li, Y., Han, X., and Hao, X.: Distribution of maturity types of maize based on accumulated temperature rezone in Northeast China, Chinese Journal of Agricultural Resources and Regional Planning, 43, 102–112, 2022.

Qu, H., Li, X., Zhu, H., Wang, L., Qu, B., Wang, Q., Lv, J., Ji, Y., and Jiang, L.: Effects of combination of low temperature and excessive precipitation at seedling stage on soybean yield in high-latitude cold region, Chinese Journal of Ecology, 1–10, 2023.

6. The manuscript demonstrates the practical value of the proposed framework under conditions of uneven Sentinel-2 image availability across regions. At the same time, some model architectures, such as GRU-D and Transformer-based approaches, are specifically designed to handle irregular or missing sequential observations. I would encourage the authors to briefly discuss such possible model extensions in the Discussion section as a direction for future improvement.

Reply: Thank you for this helpful suggestion. In the revised Discussion section, we have included a brief discussion of this potential direction for future research. Specifically, we now highlight that methods such as GRU-D and Transformer-based approaches for incomplete time series may offer greater flexibility in handling uneven Sentinel-2 image availability and irregular temporal observations across regions. We also added corresponding references to support this discussion. This revision helps place the present study in a broader methodological context and clarifies a promising avenue for future improvement (Line 599–602).