Responses to the comments of Referee #1

Article ID: essd-2026-284

Title: NortheastChinaMaizeYield10m: A 10-m Resolution Maize Yield Dataset for Northeast China (2019–2024) Generated via a Mechanistically Interpretable, Label-free Framework

Authors: Jingbo Hu, Xin Du, Qiangzi Li, Yuan Zhang, Hongyan Wang, Jiansong Luo, Jingyuan Xu, Yachao Zhao, Zhaoming Zhang, Yong Dong, and Yunqi Shen

Dear Reviewer,

Thank you very much for your thorough review and constructive feedback on our manuscript. We have carefully addressed each comment and suggestion to refine our work, enhance its clarity and strengthen its scientific contribution. The key revisions include:

(1). **Terminological refinement.** We have replaced the term "label-free" with "field-label-free" throughout the manuscript, and added a dedicated statement in the Introduction to precisely define its meaning and distinguish it from conventional supervised and unsupervised learning paradigms.

(2). **Enhanced data preprocessing descriptions.** We have expanded the technical details for Sentinel-2 imagery preprocessing in the Data Sources section, and clearly documented the sources and acquisition of the official statistical datasets.

(3). **Clarification of LAI as the core feature.** We have added a dedicated discussion to elucidate the physiological rationale for selecting LAI as the central input feature, articulating its role as an integrated proxy for canopy structure, photosynthetic capacity, and environmental stress responses, as well as its advantages for large-scale crop monitoring.

(4). **Strengthened discussion on simulation–reality discrepancies.** We have supplemented Section 4.5 to explain the model's capacity to simulate waterlogging-induced stress, and significantly extended Section 4.6 to enumerate the key uncertainties and limitations arising from input data deviations and systematic differences between simulated and satellite-derived LAI.

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.

This manuscript presents a well-structured and comprehensive framework for large-scale maize yield estimation by integrating process-based modeling, remote sensing data, and a GRU deep learning model. The overall framework is well designed and addresses the common challenge of limited field yield observations in large-scale agricultural monitoring. The generation of a 10 m resolution maize yield dataset covering Northeast China from 2019 to 2024 further demonstrates the practical value of the framework. The evaluation based on both in-situ measurements and government statistical data provides strong validation. The study is relevant to the scope of the journal and provides useful insights into the integration of process-based models and deep learning for yield estimation.

Reply: Thank you for your positive feedback and recognition of our work. In the revision, we have carefully addressed your thoughtful comments and suggestions to improve our manuscript.

Specific Comments:

1 The term "label-free" is used frequently in the article, from the title to the main text. However, the study still relies on WOFOST-simulated labels during model training. It would be helpful for the authors to clarify the exact meaning and scope of "label-free" in the introduction, particularly how it differs from conventional supervised and unsupervised learning methods.

Reply： Thank you very much for your insightful and constructive comment.

We completely agree with the reviewer that using the absolute term "label-free" without precise contextual boundaries could lead to conceptual confusion, as the Gated Recurrent Unit (GRU) model does rely on the yield outcomes simulated by the WOFOST model as target labels during its training phase.

To eliminate any potential ambiguity and ensure strict terminological rigor, we have thoroughly revised the manuscript. Specifically, we have consistently replaced the term "label-free" with "field-label-free training" throughout the title, abstract, introduction, and Methods section. This revision explicitly clarifies that our framework is completely free of real-world / in situ ground yield annotations during the model training loop.

In addition, following your valuable suggestion, we have added a dedicated statement in the Introduction to clearly define the exact meaning and scope of “field-label-free” and to clarify its essential differences from conventional machine learning paradigms.

To address these challenges, this study proposes a field-label-free training framework for maize yield estimation that couples mechanistic model with deep learning.

To address the aforementioned challenges, this study constructs a field-label-free training framework for large-scale maize yield estimation.

From a practical application perspective, this study not only constructs a low-cost, *scalable, mechanism-driven, field-label-free training estimation solution…*

Unlike conventional supervised learning, which relies on real-world ground-truth labels, or unsupervised learning, which does not involve target variables, the proposed framework employs process-based crop model simulations to generate synthetic target labels.

2 The dataset (field measurements, meteorological/soil data, satellite imagery, crop distribution maps, and statistics) is comprehensive and representative. However, some data preprocessing steps require more detailed technical descriptions. For example, the pre-processing for Sentinel-2 imagery is only briefly introduced, and the source and acquisition of the statistical datasets are not clearly documented.

Reply： Thank you for your recognition of our dataset and your constructive feedback.

We have significantly expanded the technical descriptions in the Data Sources section (Section 2.2), specifically addressing the two points you raised.

2.2.1 Remote Sensing Data

Sentinel-2 imagery acquired by the Multi-Spectral Instrument (MSI) was selected as the primary remote sensing data source for retrieving leaf area index (LAI) in this study. Sentinel-2 provides relatively high spatial and temporal resolution data (10 m and 20 m spatial resolution with a five day revisit interval) and includes three red-edge spectral bands, which offer distinct advantages for monitoring vegetation conditions during the middle and late stages of crop growth (Drusch et al., 2012) . Using the Google Earth Engine (GEE) cloud computing platform, Sentinel-2 Level-2A surface reflectance products covering the study area were collected for the maize growing seasons from 2019 to 2024, spanning from April to October. These products have been radiometrically calibrated and atmospherically corrected. Notably, because the time series reconstruction algorithm employed in this study can utilize all valid pixels, no cloud filtering was performed at the initial collection stage. To eliminate interference from clouds and shadows, we applied cloud masking to all images using the QA60 band. The 60-m bands were excluded due to their low spatial resolution and limited relevance for field-scale yield estimation. In contrast, the 10-m bands (B2: Blue, B3: Green, B4: Red, B8: Near-Infrared) and key 20-m red-edge bands (B5 and B8A) were retained.

2.2.6 Statistical data

Official annual maize yield statistics at the county level from 2019 to 2024 were collected from the Statistical Yearbooks publicly released by the Statistical Bureaus of Heilongjiang Province (http://tjj.hlj.gov.cn), Jilin Province (http://tjj.jl.gov.cn), Liaoning Province (https://tjj.ln.gov.cn), and the Inner Mongolia Autonomous Region (https://tj.nmg.gov.cn). However, because the Statistical Yearbooks data were not published for certain regions in some years, the available data cover only a subset of *regions in those specific years. These statistical data are primarily used to validate the results of this study, and are also combined with the published yield records to assess the reasonableness of the simulated yields produced by our approach.*

3 The current input features are almost exclusively centered around the LAI. Some important factors affecting yield formation, such as water stress and extreme temperature conditions, are not explicitly considered. It is suggested that the authors further elucidate the rationale behind ultimately selecting LAI as the core feature for modeling. Especially, the authors may discuss whether LAI can be regarded as an integrated proxy of crop canopy growth and photosynthesis accumulation, as well as the advantages of using LAI for large-scale yield estimation.

Reply: Thank you for this exceptionally profound and constructive comment.

We fully agree that water stress and extreme temperature events are critical factors affecting crop yield formation, and we acknowledge that they are not explicitly included as direct input features in our current modeling framework. We have added a dedicated discussion in the Discussion (4.3 Physiological Interpretation of LAI Features) to explain why LAI was selected as the core feature.

As the core input feature, the LAI is not merely a static structural variable; rather, it serves as a comprehensive characterization of both canopy architecture and photosynthetic capacity, which directly determines the fraction of intercepted photosynthetically active radiation and consequently drives biomass accumulation (Parker, 2020). The temporal dynamics of LAI throughout the entire growing season can reflect the cumulative impacts of water stress and extreme temperature conditions, as any environmental disturbances will inevitably leave distinct imprints on the seasonal LAI sequence through inhibited leaf expansion, accelerated senescence, or changes in canopy greenness (Chen et al., 2025; Zhang et al., 2025). Furthermore, LAI can be stably retrieved from high-resolution satellite imagery such as Sentinel-2, making it suitable for large-scale, high-frequency crop monitoring, whereas observational datasets for moisture and temperature are often difficult to acquire at comparable spatial and temporal scales. Therefore, selecting LAI as the core feature in this study successfully balances mechanistic interpretability with data feasibility.

4 The discussion regarding the differences between simulated and real-world observations could be further strengthened. While Figures 4 and 5 demonstrate the overall representativeness of the simulated data, some factors in real agricultural systems are not explicitly represented in the WOFOST simulations. Additional discussion of these potential limitations would improve the rigor of the manuscript.

Reply: Thanks for your suggestion.

We acknowledge that the fundamental premise of our framework—training a deep learning model entirely on simulated data and transferring it to real-world satellite observations—relies on the representativeness of the WOFOST simulations. While our validation results confirm the feasibility of this transfer, several inherent gaps between the simulated and real domains warrant explicit discussion. Accordingly, we have supplemented the manuscript with Section 4.5, which explains the model's capacity to simulate waterlogging-induced stress, and have significantly extended Section 4.6 to enumerate the key discrepancies between simulated and real domains.

4.5 Simulation and Model Response to Extreme Precipitation Stress

In this study, the WOFOST model was operated under the water-limited production mode, which is not equivalent to simulating only drought stress. The WOFOST soil water balance module is driven by actual daily precipitation and dynamically tracks changes in daily soil moisture content (van Diepen et al., 1989). Although the WOFOST model does not explicitly simulate the full range of physiological processes associated with anaerobic root conditions under waterlogging, this simplification does not prevent the model from capturing waterlogging-induced yield losses. When cumulative precipitation exceeds the soil drainage and runoff capacity, the model inherently triggers excess-water stress responses—suppressing root oxygen availability, limiting canopy transpiration and photosynthesis, and ultimately reducing LAI and biomass accumulation(H. et al., 1998; Liu et al., 2020). Therefore, our training dataset inherently includes yield reduction scenarios caused by excessive soil moisture, which originate directly from the 30-year historical meteorological records (1995–2024). Sentinel-2 satellite observations are capable of capturing such LAI anomaly signals, and the GRU model has learned, during the training phase, the general mapping relationship between depressed LAI trajectories and reduced final yields under various environmental stresses. When applied to real satellite imagery, the model does not explicitly detect flood events; instead, it infers yield losses through the observed LAI decline as a comprehensive stress signal.

4.6 Uncertainties and Limitations

Although this study achieved relatively high yield estimation accuracy, the model still carries certain uncertainties due to data source limitations and simplified assumptions. The gap between crop growth model simulations and reality, along with input data uncertainties, are the primary sources constraining the reliability of crop yield estimates.

The input layer of the WOFOST simulations, including both cultivar parameters and agronomic management settings, inevitably deviates from the complexity of real world conditions. In this study, several key physiological parameters governing phenological development were assigned based on literature values and the model's default settings, rather than calibrated against site-specific field trials for each cultivar. In terms of agronomic management, our framework assumes non-limiting nutrient conditions and discretizes the sowing window into only four fixed dates. In reality, smallholder farmers *across Northeast China employ flexible sowing schedules and highly diverse management strategies (Chuanwei et al., 2024). These two types of input deviations may jointly cause the simulated LAI trajectories to diverge to some extent from actual growth curves in specific fields. Nevertheless, we argue that the full-factorial parameter combination strategy adopted in this study constructs an exhaustive training set that covers a broad physiological and management space. The consistency between estimated yields and statistical yearbook data (Figure 11), together with the robust multi-year independent field validation performance (Section 3.4), demonstrates that the simulated data generated under the current input settings are sufficiently reliable to support accurate yield estimation.*

A systematic discrepancy also exists between the WOFOST-simulated LAI and the satellite- derived LAI. The simulated values represent a deterministic, noise-free "green leaf area index" at the point scale under idealized atmospheric conditions (H. et al., 1998). In contrast, the satellite-retrieved LAI is subject to multiple factors, including residual atmospheric correction errors, cloud-gap filling algorithms, and mixed-pixel effects. As noted by Lu et al. (2025), satellite-derived vegetation parameters (e.g., LAI) are not free from uncertainty, particularly in areas with high LAI or dense canopies, where saturation or underestimation may occur. Moreover, different data processing workflows, such as cloud masking and filtering algorithms, can introduce systematic errors. Consequently, input uncertainties can propagate through the prediction model when a model trained on clean simulated data is generalized to noisy satellite observations (Xie et al., 2025). However, the gating mechanism of the GRU provides inherent robustness to such input noise by dynamically weighting the contribution of each time step (Cho, 2014). The stable performance of the model in independent validation across 2022–2024 (Table 7) further confirms that the proposed framework can effectively mitigate the adverse effects of uncertainties in satellite-derived LAI under the current data quality.

5 The discussion regarding the ability of the simulated dataset to represent extreme conditions may be somewhat overstated. Although the results from the 2023 disaster year demonstrate the potential robustness of the framework under adverse conditions, they may not be sufficient to prove comprehensive coverage of all extreme scenarios. A more cautious interpretation of these findings is recommended.

Reply： Thank you for your valuable comments.

We agree that a more cautious interpretation is needed regarding the model's performance under extreme conditions. Accordingly, we have added a dedicated discussion in Section 4.5 ("Simulation and Model Response to Extreme Precipitation Stress") to more carefully address this issue and provide a balanced assessment of the model's capabilities and limitations.

4.5 Simulation and Model Response to Extreme Precipitation Stress

In this study, the WOFOST model was operated under the water-limited production mode, which is not equivalent to simulating only drought stress. The WOFOST soil water balance module is driven by actual daily precipitation and dynamically tracks changes in daily soil moisture content (van Diepen et al., 1989). Although the WOFOST model does not explicitly simulate the full range of physiological processes associated with anaerobic root conditions under waterlogging, this simplification does not prevent the model from capturing waterlogging-induced yield losses. When cumulative precipitation exceeds the soil drainage and runoff capacity, the model inherently triggers excess-water stress responses—suppressing root oxygen availability, limiting canopy transpiration and photosynthesis, and ultimately reducing LAI and biomass accumulation(H. et al., 1998; Liu et al., 2020). Therefore, our training dataset inherently includes yield reduction scenarios caused by excessive soil moisture, which originate directly from the 30-year historical meteorological records (1995–2024). Sentinel-2 satellite observations are capable of capturing such LAI anomaly signals, and the GRU model has learned, during the training phase, the general mapping relationship between depressed LAI trajectories and reduced final yields under various environmental stresses. When applied to real satellite imagery, the model does not explicitly detect flood events; instead, it infers yield losses through the observed LAI decline as a comprehensive stress signal.