

Response of review “Seasonal Cycle Biases in DGVM Simulations of Double- Cropping Systems: A Case Study in the Huang-Huai-Hai Plain”

Dear Reviewer,

Thank you very much for your time and insightful suggestions. Your in-depth discussion has been especially valuable for improving this paper. We have also provided detailed responses to address your concerns. Although some of the results might be expected from a theoretical perspective, the goal of this paper is to quantify current deficiencies and thereby draw attention to the need for improving the representation of multi-cropping regions in DGVM development. This is of great importance for the advancement of Earth system models. Below are our point-by-point responses to your comments.

Thanks again.

Tiexi CHEN and co-authors

General remarks

This paper compares LAI outputs from seven TRENDY DGVMs against three remote sensing products over the Huang-Huai-Hai Plain, a major double-cropping region in eastern China. The authors find that remote sensing consistently reveals a bimodal seasonal LAI cycle (peaks in April and August, corresponding to winter wheat and summer maize), whereas none of the DGVMs reproduce this pattern. They also report that DGVMs tend to overestimate LAI magnitude and underestimate the cropland contribution to regional greening. The paper calls for incorporating explicit double-cropping mechanisms into DGVMs.

I appreciate the authors' effort to draw attention to the limitations of current DGVMs in representing intensively managed croplands. This is an important and, I think, underappreciated problem, particularly as we increasingly rely on these models for carbon budget attribution and greening studies. The Huang-Huai-Hai Plain is a well-chosen study area, and the use of multiple remote sensing products alongside multiple models adds robustness. I do think the core message of this paper is worth communicating. But at the same time, I do have several concerns that prevent the paper from reaching its potential in its current form.

Response

Thank you for your generally positive evaluation of our manuscript. We fully agree that this is an important and challenging issue.

On the one hand, Dynamic Global Vegetation Models (DGVMs) are powerful tools for identifying the drivers of global greening and attributing their relative contributions. On the other hand, current DGVMs still lack adequate representation of managed lands, particularly in regions characterized by intensive

agricultural practices such as double-cropping systems. In situations where discrepancies arise among model outputs, comparative or ensemble-based analyses are commonly adopted; however, rigorous performance evaluation against observations remains essential.

With the increasing global demand for food, cropping intensity has risen substantially through multiple-cropping practices, among which double-cropping plays a crucial role. It accounts for approximately 17.42% (2.62×10^6 km²) of global cropland area (Zhang et al., 2021), and therefore cannot be neglected in large-scale assessments.

Why, then, has limited progress been made in addressing this model deficiency? We argue that one key reason is the lack of broad consensus within the research community regarding the magnitude and implications of this issue. This gap highlights the need for systematic case studies in representative regions to explicitly quantify its impacts.

In this study, although we are unable to fully disentangle all underlying processes across the seven DGVMs analyzed, we aim to clearly demonstrate the significant biases associated with the absence of double-cropping representations. This is achieved by quantifying discrepancies between remote sensing observations and model simulations. More importantly, we hope this work will help raise awareness among model developers and stimulate future improvements in the representation of agricultural management processes.

Major comments

First of all, the introduction needs substantial restructuring. As it stands, the logic chain is fragmented and the narrative wanders through several loosely connected themes without building a coherent argument toward the study's objective. The introduction opens with a broad discussion of vegetation change, greening, and remote sensing (L23–33), which is fine as context-setting. But then it takes a detour into the specific quantitative findings of Zhu et al. (2016) — reporting that 25–50% of vegetated land shows increased LAI, that CO₂ accounts for 70% of greening, nitrogen deposition 9%, and so on (L36–41). This level of numerical detail belongs in the Discussion when interpreting your own results, not in the Introduction. The important point from that work — that DGVMs attribute most greening to CO₂ fertilisation — can be stated in one sentence. The transition at line 42 ("However, managed lands—particularly croplands—have been shown by observations to make a substantial contribution to global greening") is actually the key motivating observation for the study, but it is buried rather than highlighted. And then the transition at line 47 is jarring: the paper suddenly shifts from "croplands challenge DGVM greening attribution" to "it is essential to first validate that DGVMs can accurately capture the fundamental seasonal characteristics of vegetation." The logical connection — that if DGVMs get the seasonal cycle of croplands wrong, their estimates of cropland productivity and greening contribution are unreliable — is never stated. The reader has to infer it. Lines 47–52 then discuss phenology in quite general terms (SOS advancement, EOS delay, peak greenness enhancement, global warming and northern ecosystem productivity), which reads like textbook material that does not

connect specifically to the problem of double-cropping. By the time the reader reaches the study objective at line 68, they have passed through global greening drivers, CO₂ fertilisation percentages, generic phenology concepts, and global multiple cropping statistics — none of which flow naturally into each other. I would suggest reorganising the introduction around a cleaner logic chain.

Response

Thank you very much for your insightful comments. We recognize that the lack of a clear logical structure in the Introduction has led to confusion, including Reviewer 1's concern regarding the scientific significance of this study.

To better highlight the core research question, we have substantially condensed the Introduction and clarified its logical flow. Specifically, we now emphasize that, given the strong reliance of current research on DGVMs as a powerful tool, and the substantial contribution of croplands to global greening, it is essential to rigorously evaluate the ability of DGVMs to simulate vegetation dynamics at the regional scale.

The revised Introduction is provided below:

Introduction

Vegetation change and its climatic feedback are central to current global change research, as terrestrial ecosystems remain highly sensitive to anthropogenic and natural drivers (Houghton, 1995; Esau et al., 2016). Over recent decades, a significant global "greening" trend—characterized by an interannual increase in leaf area index (LAI) or greenness—has been widely documented through satellite-based observations (Myneni et al., 1997; Huete et al., 2002; Piao et al., 2019). To attribute these changes, Dynamic Global Vegetation Models (DGVMs) have become indispensable tools, generally identifying the CO₂ fertilization effect as the primary driver of global vegetation growth (Zhu et al., 2016; Keenan et al., 2013).

However, a significant discrepancy exists between model attributions and observational evidence regarding managed lands. While DGVMs emphasize CO₂ fertilization (Zhu et al., 2016), recent remote sensing studies highlight that intensive land management, particularly in croplands, contributes substantially to global greening. For instance, the greening observed in China and India is predominantly driven by agricultural intensification rather than natural factors (Chen et al., 2019). This contradiction suggests that current DGVMs may overestimate the role of CO₂ while underrepresenting the impact of human-induced agricultural dynamics, highlighting an urgent need to evaluate and refine how models simulate agricultural systems.

The reliability of long-term greening attributions in DGVMs depends fundamentally on their ability to capture short-term seasonal dynamics. If a model cannot accurately simulate the seasonal cycle of vegetation—including phenological shifts and peak productivity—its estimates of cumulative biomass

and long-term trends remain questionable. While natural vegetation adapts to climate change through physiological adjustments (Huang et al., 2017), croplands exhibit unique phenological traits dictated by human management. This is most evident in regions characterized by multiple cropping systems, where intensive agricultural activities create complex seasonal patterns—such as multiple greenness peaks—that differ starkly from the unimodal cycles of natural ecosystems.

Despite the fact that multiple cropping systems cover approximately 12% of global croplands and are vital for cereal production (Waha et al., 2020; Wu et al., 2018), these processes are rarely or inadequately represented in current DGVMs. Most models simplify agricultural land use or fail to account for the intermittent fallow periods and shifting cultivation intensities inherent in these systems (Liu, 2021). Consequently, there is a critical knowledge gap in whether DGVMs can accurately replicate the intensified seasonality of highly managed landscapes. In this study, we focus on the Huang-Huai-Hai Plain, a representative region dominated by a stable winter wheat–summer crop double-cropping system since the 1980s. The objective of this research is to evaluate the capacity of state-of-the-art DGVMs to simulate the vegetation seasonality driven by intensive double-cropping, thereby identifying the specific structural limitations of these models in representing human-managed ecosystems.

Moving to the core analysis, my main concern is that the paper stays largely at the descriptive level. The paper documents that models fail but does not explore why. Most TRENDY models were never designed to simulate double-cropping, so the mismatch is expected. This does not make the paper pointless as systematic documentation has value, but it needs a diagnostic dimension. I would suggest adding a table summarising each model's cropland representation (crop functional types, phenology scheme, management features, ability to simulate sequential cropping). This information is available from published model descriptions and would immediately contextualise the results. The three-type classification of seasonal patterns (L178–183) is interesting but should be linked to model design choices rather than left hanging. The Discussion similarly needs to move from restating results to interpreting them: which models already have infrastructure for double-cropping? What would need to change in the TRENDY protocol?

Response

We agree with the reviewer's professional judgment that the core analysis in the original manuscript remains largely at the descriptive level of model-observation mismatches, without in-depth exploration of the model structural causes underlying the biases, resulting in a lack of diagnostic value and mechanistic interpretation of the study. Although most models in the TRENDY ensemble were not originally designed to incorporate double-cropping simulation modules, and this mismatch is to a certain extent expected, the systematic diagnosis of such biases is precisely the core academic value that this study should supplement. In the revised manuscript, we will upgrade the analytical content from three dimensions:

First, we will add a table to systematically summarize the core structural characteristics of the 7 models used in this study, specifically including the classification scheme of crop functional types, phenology simulation schemes, agricultural management module settings, and the capability to simulate sequential cropping and double-cropping. All information will be derived from the official published description documents of each model, to provide complete structural background support for the interpretation of the results.

Second, we will establish the correlation between simulation results and model design choices, conduct a corresponding analysis between the three-type classification of LAI seasonal patterns in the original manuscript and model structural characteristics, and clarify the intrinsic relationship between the seasonal simulation bias characteristics of different models and their phenology schemes and cropland management representation capabilities, thus achieving the upgrade from "*describing bias phenomena*" to "*interpreting the causes of biases*".

Third, we will comprehensively restructure the Discussion section, abandon repetitive restatements of the results, and focus on the mechanistic interpretation of three core issues: which models already have the basic framework for double-cropping simulation, and the core reasons why this function is not activated under the TRENDY protocol; the core adjustments needed in the model structure and simulation protocol of the TRENDY multi-model ensemble for regional cropland simulation; and the cascading effects of the lack of double-cropping mechanisms on DGVM simulations of regional carbon cycles and climate feedbacks. This will echo the scientific questions raised in the introduction and strengthen the scientific significance of the study.

Table 3. Overview of the simulation capabilities for Cropland of the selected models from Trendy v12

Dataset Name	Cropland module	Crop Functional Type	Phenology	Irrigation module	Fertilization Module	Single/multiple cropping
ED	NAN					
ISAM	YES	Maize (C4), Wheat, Rice, Soybean (C3)	Defined by accumulated temperature	YES	YES	Single
LPJ-GUESS	YES	Separate crop phenology schemes on a daily time-step are present for simulations with and without N-limitation. Cropland without N-limitation is represented by eleven crop PFTs (temperate cereals, rapeseed, pulses, sugarbeet, maize, soybean, tropical cereals, sunflower, peanut, cassava and rice), simulated separately (without inter-PFT competition) and two grass PFTs (competing C3 and C4 grass) as cover crop between harvest and sowing. The same grass PFTs are used to represent pastures. Currently, two crop PFTs are defined for N-limited simulations (wheat and maize).	Defined by accumulated temperature and precipitation	YES	YES	Single
LPJml	YES	In LPJmL4, 12 different annual CFTs are simulated (Table S10), similar to Bondeau et al. (2007) with the addition of sugar cane. The basic idea of CFTs is that these are parameterized as one specific representative crop (e.g. wheat, <i>Triticum aestivum</i> L.) to represent a broader group	The phenological development of crops in LPJmL4 is driven by temperature through the accumulation of growing degree days and can be modified by vernalization requirements and sensitivity to daylength (photoperiod) for some CFTs and some varieties.	YES	YES	Single

		of similar crops (e.g. temperate cereals)				
LPXbern	YES	The farmland and the pasture share the common herbaceous PFT, without distinguishing specific crop types such as wheat, corn, or rice, and there is no dedicated crop PFT.	The core physiological process framework is consistent with the natural herbal PFT.	NAN	YES	Single
ORCHIDEE	YES	Define two specific agricultural PFTs: Agricultural C3 grass and Agricultural C4 grass	ORCHIDEE assumes that all C3 agriculture is a perennial prairie and does not account for harvest.	NAN	NAN	Single
VISIT	YES	The global farmland is divided into three core functional types: 1. C3 crop type, represented by crops such as wheat, which adopts the parameterization scheme of C3 plant photosynthetic physiology; 2. C4 crop type, represented by crops such as corn in tropical/subtropical regions, which is adapted to the photosynthetic and water utilization characteristics of C4 plants; 3. Rice field type specifically designed for the rice-growing areas in the Asian monsoon region. In addition to the basic crop physiological simulation, it additionally couples the seasonal flooding hydrological process and the exclusive module for rice field CH ₄ emissions.	In response to the agricultural production characteristics of different climate zones, differential planting systems and phenological-driven schemes are adopted. The core rules are as follows: 1. In temperate regions: A one-year cropping system is implemented, with the growth cycle driven by temperature thresholds: ◦ The core indicators for initiating and ending the growing season are the monthly average temperature of 5°C as the critical value; ◦ When the growth season begins, a fixed amount of carbon is added to the crop biomass reservoir to simulate the initialization process of sowing/transplanting; ◦ When the surface temperature is below the critical value of 5°C, the crop harvest process is triggered. 2. In tropical regions (average annual temperature > 20°C): A continuous annual cropping system is adopted, without setting seasonal start and end periods for growth. Sowing and harvest occur continuously at a constant rate each month.	NAN	NAN	Single

The land-use mask inconsistency is a real problem. The analysis uses RESDC (1 km) to define cropland, but TRENDY models are forced with HYDE3.3 (0.5°). Figure 8 shows these datasets disagree substantially. This means part of the model–observation mismatch could reflect differences in what land is classified as cropland, not just the absence of double-cropping mechanisms. A sensitivity test such as applying the HYDE mask to the remote sensing data and checking whether the bimodal pattern persists would resolve this. The HYDE data are already available to the authors from Figure 8, so this requires no new data.

Response

We acknowledge that the original manuscript failed to fully disclose the application differences between the RESDC and HYDE3.3 land use datasets, nor did it verify the potential impact of such differences on the core conclusions, which constitutes an omission in methodological rigor. In the revised manuscript, we will strictly follow your suggestions to complete two core revisions and supplementary analyses:

On the one hand, we will supplement a sensitivity test of the land use mask. Using the HYDE3.3 0.5°

cropland mask, which is fully consistent with the forcing data of the TRENDY models, we will re-perform mask extraction and spatiotemporal change analysis for both remote sensing and simulated LAI data. We will systematically compare the differences in the bimodal seasonal characteristics of LAI and interannual greening trends derived from remote sensing data under the two cropland masks, quantify the contribution of inconsistent cropland classification standards to the model-observation mismatch, and verify the robustness of the core conclusions of this study. The relevant results will be supplemented to the Results section of the main text, and the corresponding figures will be included in the supplementary materials.

On the other hand, we will clearly define the application boundaries and applicable scenarios of the two datasets in the Methods section. We will clarify that HYDE3.3 is the official global-scale land use forcing data adopted by the TRENDY models, while RESDC is a regional-scale remote sensing-based land cover product that is more consistent with the actual spatial distribution of croplands in the Huang-Huai-Hai Plain. Meanwhile, we will objectively describe the advantages and limitations of the two datasets, revise the one-sided statements about the HYDE dataset in the original manuscript, and resolve the issue of inconsistent methodological descriptions.

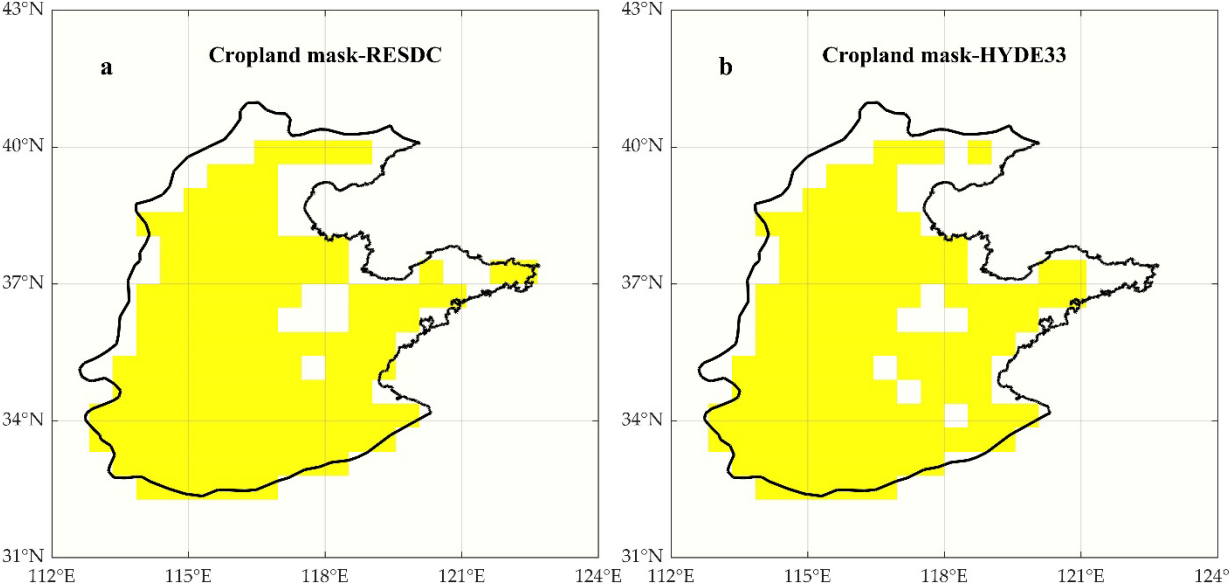


Figure 1: Cropland Mask produced using RESDC and HYDE3.3 Land-Use/Cover Datasets

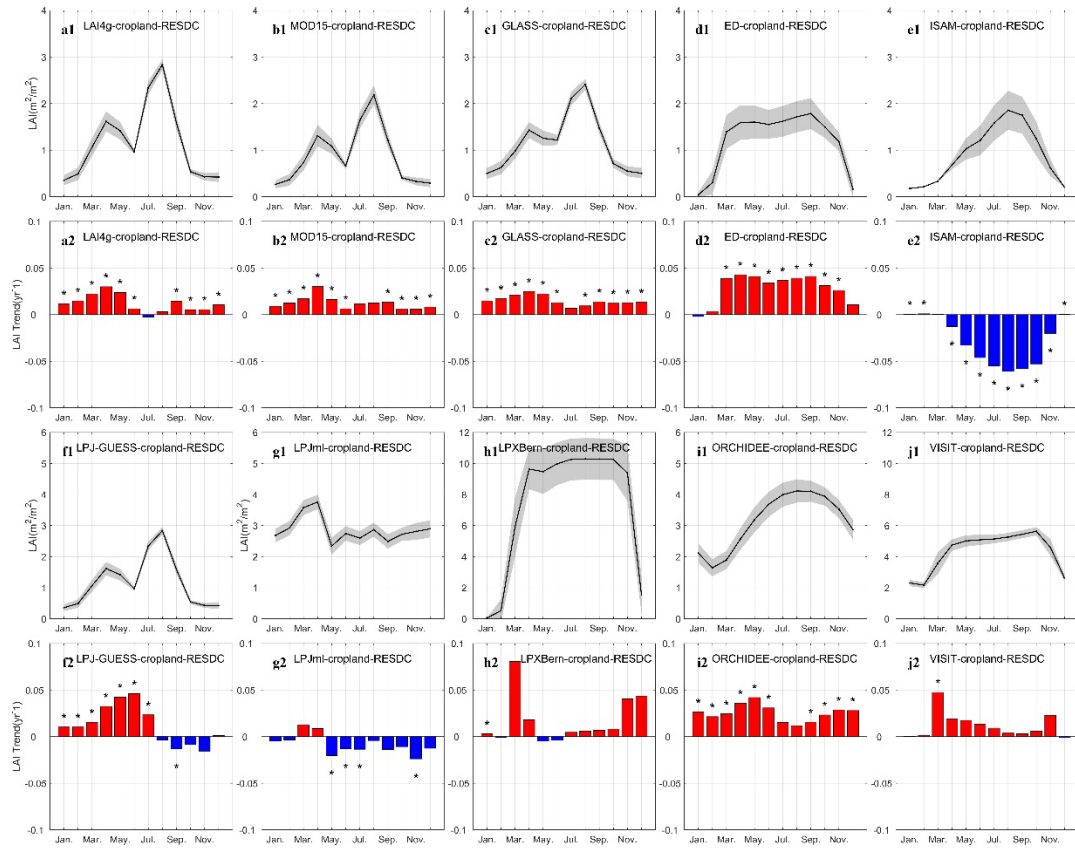


Figure 2: Seasonal cycles and linear fitting trends of monthly LAI for cropland masked by RESDC. The dataset names are abbreviated in the subheading; The line graph represents the monthly average of LAI and the bar graph represents linear trend of average monthly LAI, Red represents a positive trend and blue represents a negative trend; The asterisk(*) on the top of bar means the linear trend pass the significance test($p < 0.05$).

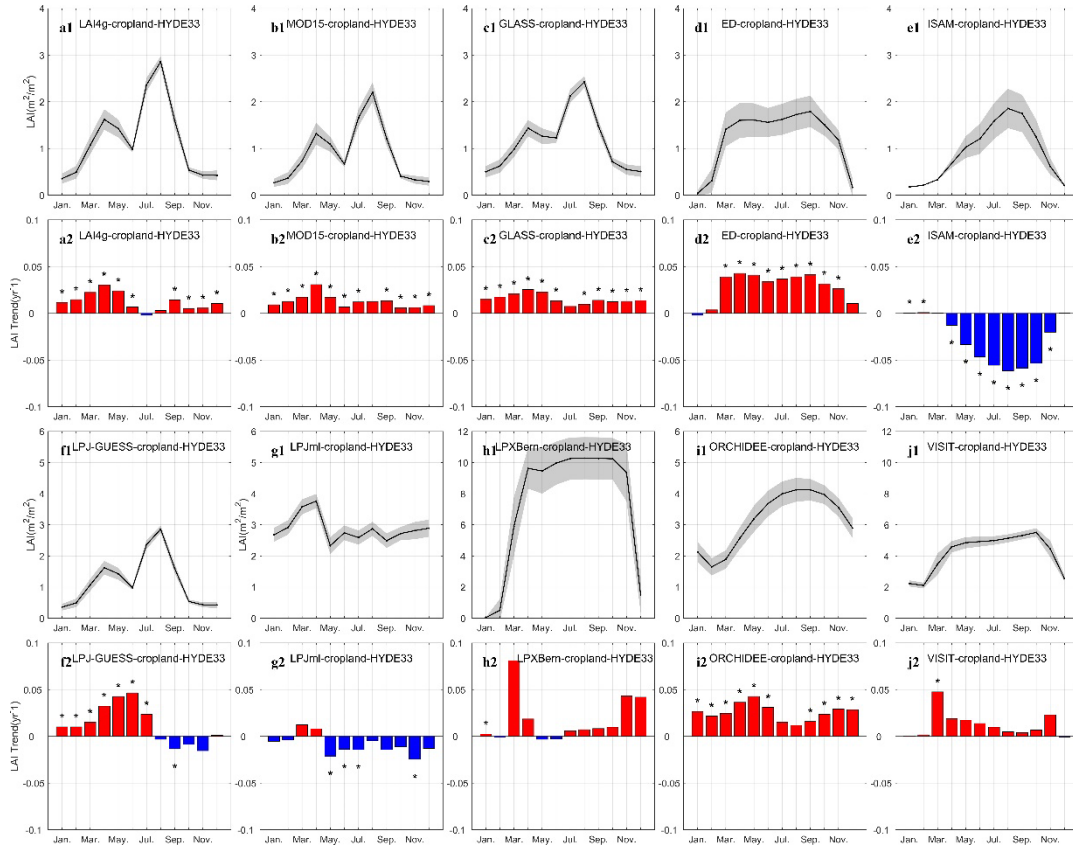


Figure 3: Seasonal cycles and linear fitting trends of monthly LAI for cropland masked by HYDE3.3. The dataset names are abbreviated in the subheading; The line graph represents the monthly average of LAI and the bar graph represents linear trend of average monthly LAI, Red represents a positive trend and blue represents a negative trend; The asterisk(*) on the top of bar means the linear trend pass the significance test($p < 0.05$).

Mask contrast and sensitivity tests revealed that differences in the Cropland masks produced by the two datasets are very small, and they do not have a significant impact on the bimodal structure of LAI in the farmland. We believe that the mask is not the cause of the seasonal differences in remote sensing and simulated LAI.

Data processing methods are missing. The three remote sensing products have very different native resolutions and are all aggregated to 0.5°/monthly, but the aggregation procedure is not described. Monthly means vs. maxima, the order of masking vs. upscaling, and MODIS quality flag handling can all affect peak amplitude and timing — exactly what the paper analyses. A brief methods subsection would address this.

Response

We acknowledge that the comprehensive disclosure of the entire data processing workflow is a fundamental prerequisite for ensuring the reproducibility of research results and the reliability of conclusions; we sincerely apologize for the significant lack of detail regarding the aggregation and quality control procedures for multi-source remote sensing data in the initial version of this manuscript. In this revision, we will introduce a dedicated subsection titled "Spatio-temporal Aggregation and

Standardization of Multi-source Data" within Section 2.2 (Remote Sensing Data) of the paper to provide a complete and lucid account of the entire processing workflow, which clarifies the operational details for the three core questions raised by the reviewer one by one to ensure full reproducibility of the method:

Temporal aggregation method: For the 8-day resolution data from MODIS and GLASS, we first performed linear interpolation to daily resolution, and then calculated the arithmetic mean of all valid daily observations within each calendar month to derive the monthly LAI value. For the semi-monthly resolution GIMMS data, the monthly LAI value was represented by the arithmetic mean of the first and second half-month observations in the same month.

This method can smooth the random noise of single-period data while fully retaining the characteristic bimodal seasonal dynamics of the double-cropping system, and avoids the over-amplification of transition period signals caused by the maximum value composite method.

Spatial aggregation method: For the high-resolution raw data of MODIS (500 m) and GLASS (0.05°), the area-weighted averaging method was used to aggregate the data to the target 0.5° grid, rather than simple arithmetic averaging. This method eliminates the bias caused by the pixel area proportion of different resolution data, and ensures that the value of the 0.5° grid fully represents the mean cropland LAI of the corresponding region, which is completely matched with the grid scale of the TRENDY model outputs.

Implementation sequence of cropland masking: We adopted the workflow of aggregation followed by cropland masking. Mixed pixels are an unavoidable issue in scale conversion. It is difficult to find 0.5° grids with pure cropland cover in both model results and real-world remote sensing images. We therefore performed grid screening based on the 0.5° aggregated land use/cover percentage data. A grid was classified as a cropland grid in a given year when its cropland coverage exceeded 50%. If the grid was consistently classified as cropland throughout the study period, we considered it to have stable land use with no conversion, and included it in the cropland mask.

Through this method, we aim to extract the seasonal characteristics of croplands while retaining sufficient sample representativeness and preserving cropland information to the greatest extent possible. This workflow is applicable to both the RESDC land cover dataset and the HYDE land use dataset. We have also included the masking results based on the HYDE dataset in the Discussion section as a sensitivity analysis to test whether the HYDE-based cropland mask alters the observed bimodal pattern in the remote sensing data.

The manuscript is redundant in many sections. The same three conclusions appear at L215–222, in the Discussion, and in the Conclusions. The paragraph at L219–222 essentially repeats L215–218 with different wording. I would remove the end-of-Results summary entirely and add a Limitations subsection instead.

Response

We agree with the reviewer's suggestions that the original manuscript has issues of repetitive expression of core conclusions and content redundancy, and lacks the discussion of study limitations, which is essential for academic papers and does not comply with the academic norms of international journals. In the revised manuscript, we will complete two core revisions:

First, we will thoroughly streamline the redundant content of the manuscript, delete the repetitive summary paragraphs at the end of the Results section, condense the expression in the Discussion and Conclusion sections, and remove content with synonymous repetition. We will achieve a structural division where the Results section only presents objective analysis results, the Discussion section focuses on mechanistic interpretation, and the Conclusion section extracts core scientific findings and research prospects, ensuring that the three sections perform their respective functions without overlapping or redundant content.

Second, we will add an independent subsection entitled Study Limitations at the end of the Discussion section, to comprehensively and objectively disclose the core limitations of this study, specifically including the potential biases from the selection of land use datasets, the scale matching issue of the 0.5° grid, other potential influencing factors for model-observation mismatches, and the limitations of extrapolating the findings from the study region to other areas. This will ensure that the expression of the study conclusions is objective and rigorous.

The greening trend analysis deserves more attention. This is arguably the most novel part of the paper. The finding that ISAM and LPJmL show browning where all observations show greening is striking, what might explain it? Even a tentative hypothesis based on published model descriptions would add value. The seasonal pattern of greening (strongest in spring, consistent with enhanced wheat productivity) could be connected more explicitly to agricultural drivers like cultivar change or fertilization.

Response

We sincerely thank the reviewer for recognizing the research value of the greening trend analysis section. The original manuscript had insufficient in-depth exploration and elaboration of this part, and failed to fully release the innovative value of the study. In the revised manuscript, we will focus on deepening the analytical content of this section, and specifically supplement two core modules:

First, we will systematically analyze the potential causes of divergent greening trends among models. Based on the official published description literature of each model, we will put forward mechanistic tentative hypotheses for the browning trends simulated by the ISAM and LPJmL models, which are completely opposite to the greening signals observed in all remote sensing products. We will focus on comparing the core differences between these two models and other models in four dimensions: the parameterization scheme of crop CO₂ fertilization effects, the design of cropland phenology modules, the

representation of nitrogen cycle processes, and the representation of agricultural management measures. We will systematically analyze the potential driving mechanisms of the simulated browning trends, filling the gap in the original manuscript where only the phenomenon was described without exploring the causes.

Second, we will establish the correlation between the seasonal characteristics of greening trends and regional agricultural driving factors. We will link the finding of "*the strongest greening trend in spring*" in this study to the changes in agricultural management practices in the main winter wheat producing areas of the Huang-Huai-Hai Plain, supplement empirical data and literature support on agricultural driving factors in the region including crop variety improvement, increased fertilizer application, irrigation management adjustment, and agricultural machinery application, interpret the contribution of agricultural management measures to regional cropland greening, and improve the regional pertinence and mechanistic explanatory power of the analysis.

Specific Comments:

The section heading at L22 has a typo ("Introduction"). Missing spaces before parenthetical references occur throughout. The Figure 1 caption has a placeholder ("at xx resolution"). "LPX-Bern"/"LPXBern" is inconsistent. The second results subsection (L200) is mislabelled "3.2" instead of "3.3." "MOD15A12H" at L355 should be "MOD15A2H".

Response

For the detailed issues of format, spelling, and terminology consistency throughout the manuscript pointed out by the reviewer, we will complete a comprehensive and item-by-item revision in the modified manuscript, which specifically includes: correcting the typo in the section heading, changing "Introduction" to "Introduction"; uniformly revising the formatting issue of missing spaces before parenthetical references throughout the manuscript, in strict compliance with the citation format specifications of the target journal; revising the placeholder in the caption of Figure 1, and supplementing the native resolution information of the remote sensing products; unifying the nomenclature of the model name throughout the manuscript, standardizing it as "LPX-Bern" to eliminate the inconsistency in abbreviation; correcting the numbering error of the subsection heading in the Results section, changing "3.2" at line 200 to "3.3"; correcting the typo of the remote sensing product name, changing "MOD15A12H" to "MOD15A2H". Meanwhile, we will complete a full read-through and check of the entire manuscript to eliminate all detailed errors in spelling, punctuation, and formatting, to ensure the academic standardization of the manuscript.

Figures 5 and 6 would benefit from a summary panel overlaying remote sensing curves with a model ensemble envelope.

Analysis code should be archived for reproducibility.

Response

In response to your suggestions regarding the figures, we will optimize the visualization of Figures 5 and 6 in the revised manuscript. Meanwhile, we will introduce a summary panel that overlays the mean curves and upper and lower envelopes of both the multi-source remote sensing LAI data and the TRENDY model ensemble. This approach will intuitively illustrate the overall characteristics of the discrepancies between the models and observations, thereby enhancing the information density and readability of the figures and enabling readers to quickly grasp the core results.

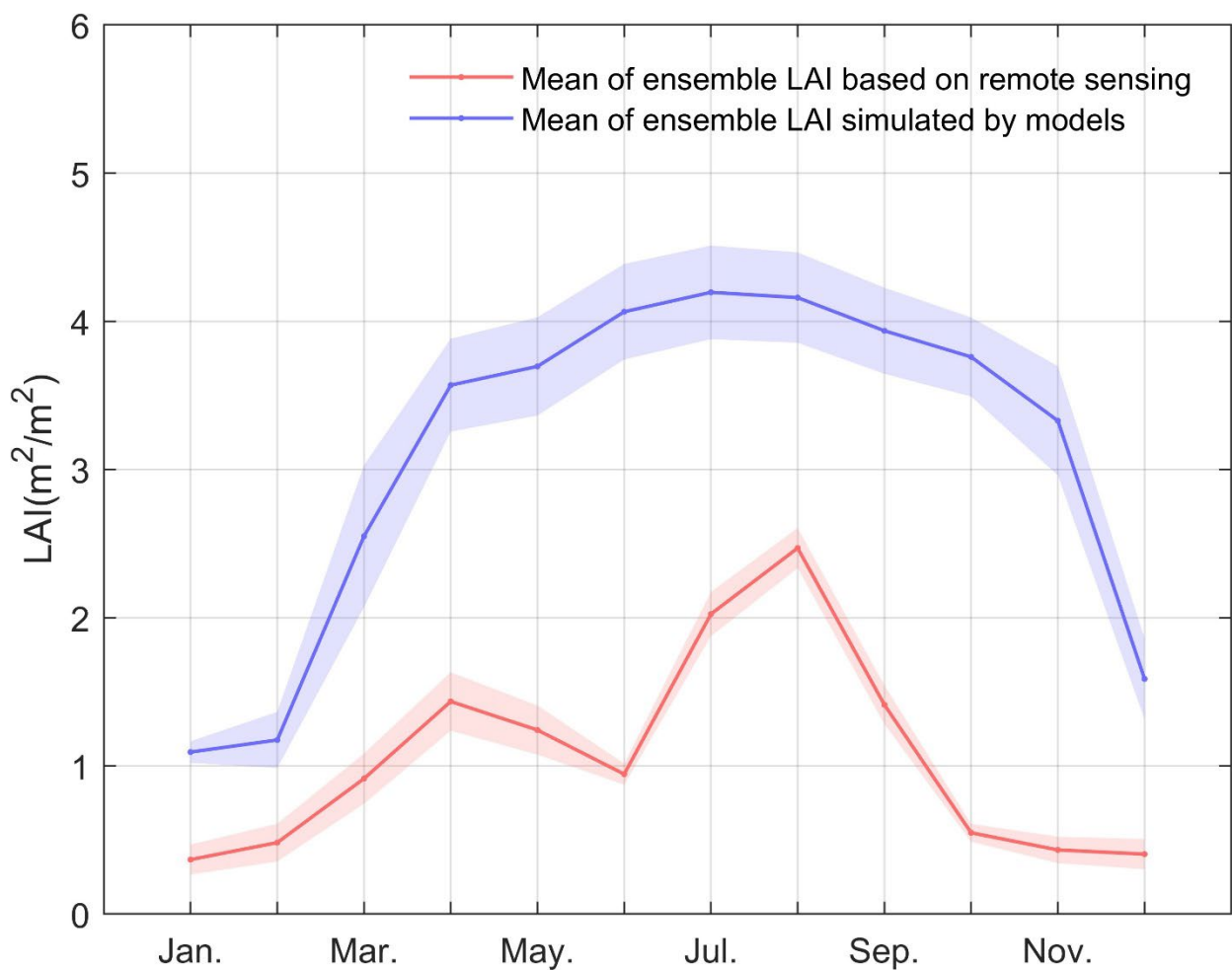


Figure 4: Seasonal cycles of monthly ensemble LAI for cropland masked by RESDC. Red represents LAI based on remote sensing, while blue represents the LAI simulated by the model. And the shaded area represents plus or minus one standard deviation