

## 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*

*The manuscript compares simulation results from the TRENDY model ensemble against remote sensing products for the Huang-Huai-Hai Plain, focusing on double-cropping systems. While the paper is well written, I am sorry to say that I have fundamental concerns about the scientific contribution that lead me to recommend rejection.*

*My primary concern is that the paper's central finding (TRENDY models fail to reproduce bimodal LAI patterns characteristic of double-cropping) provides limited scientific insight. These models do not implement double-cropping mechanisms. Comparing model results against data for phenomena they do not simulate is neither surprising (expected outcome confirmed) nor novel. Waha et al., (2025), whom the authors cite, already established this limitation globally. Since the model structures are not discussed it is also not of diagnostic relevance (nothing is learned on how to improve the models specifically).*

*In my opinion, the comparison therefore lacks the scientific value necessary for publication.*

*I have considered whether a comparison of the LAI patterns from the different remote sensing products might have standalone value independent of the model comparison. However, the manuscript does not establish that this regional documentation is novel, which would need to be verified. Furthermore, this would in my opinion need the connection of the patterns to broader implications appropriate for Biogeosciences (e.g., biogeochemical cycles or Earth system science). Additionally, several methodological issues (detailed below) would need to be resolved.*

*The manuscript also presents analyses of greening trends (Sections 3.1-3.2, Figures 3-7) that could potentially provide scientific insight, as models should capture vegetation responses to climate and management changes. However, this analysis suffers from similar issues: The models are grouped based on how their results compare against the remote sensing data but structural differences that might*

*explain divergent behaviors are not explored, and the analysis remains descriptive rather than diagnostic.*

*I do not believe major revisions can address these issues. The paper would need to be reconceived as either: (a) a diagnostic analysis of models that attempt to simulate double-cropping (requiring different models than the TRENDY ensemble), (b) if it is not too similar to existing works, a purely observational documentation of remote sensing products (requiring different focus), or (c) a diagnostic analysis of what distinguishes models showing greening versus browning trends (requiring systematic structural comparison). A combination of b and c would also be possible. However, all of these would constitute a substantially different study rather than a revision.*

*I hope that the specific comments below, which detail the technical and methodological issues, help the authors in reconsidering how to present this work.*

## **Response**

We greatly appreciate the reviewer's thorough and insightful evaluation of our manuscript, particularly the systematic and in-depth comments on the core rationale and scientific contribution of this work, rather than focusing solely on technical details. Our research team is an experienced user of Dynamic Global Vegetation Models (DGVMs), with a long-term focus on eco-climatic process modelling, including the estimation of vegetation productivity and attribution analysis of vegetation greening. In recent years, we have paid special attention to the modelling and optimization of agricultural land management processes. The core objective of this manuscript is to identify the specific biases introduced into DGVM simulations by the lack of explicit representation mechanisms for double-cropping processes.

As the reviewer correctly points out, the finding of this study is **an "expected" result**, which may lead to an insufficient scientific contribution. However, a critical question we seek to address is: **why has there been no significant progress in resolving this model limitation to date?** We argue that one core reason is the lack of a broad consensus in the research community on the severity of this issue, which necessitates a systematic case study in a representative region to concretize the impacts of this limitation. Since remote sensing information can extract the planting intensity of crops, and the potential for multiple cropping based on accumulated temperature, along with real-world constraints, can theoretically be used to assess cropping systems, we have noticed that some models have already considered multiple crop types. For example, the LPJ-guess model classifies over a dozen crop types. This means that the differences between crop types are gradually gaining attention. However, for multiple cropping, no systematic improvements have been made so far. Moreover, we believe that improvements in planting intensity may be at least equally important as crop classification.

The simulation accuracy of DGVMs is fundamental to multiple core research directions in terrestrial ecosystem modelling, which are detailed in the following three aspects:

**1. Attribution of vegetation change:** Represented by the classic studies of Piao et al. (2015) and Zhu et al. (2016), the core methodological framework in this field relies on multi-scenario simulations using DGVMs, and quantifies the contribution of different driving factors via the differential analysis of scenario outputs. This is also one of the landmark achievements of the TRENDY project.

**2. Uncertainty sources of vegetation greening:** As stated in Chapter 3 of the IPCC AR6 Working Group I Report (Masson-Delmotte et al., 2021), land management is a key source of low confidence in the global attribution of vegetation greening, as it has become the dominant driver over CO<sub>2</sub> fertilization in some regions. [Original quote: "The main driver of the observed increase in the amplitude of the seasonal cycle of atmospheric CO<sub>2</sub> is enhanced fertilization of plant growth by the increasing concentration of atmospheric CO<sub>2</sub> (medium confidence). However, there is only low confidence that this CO<sub>2</sub> fertilization has also been the main driver of observed greening because land management is the dominating factor in some regions. Earth system models simulate globally averaged land carbon sinks within the range of observation-based estimates (high confidence), but global-scale agreement masks large regional disagreements. {3.6.1}"]

**3. Requirements for coupled Earth System Model (ESM) simulations:** The IPCC Seventh Assessment Report (AR7) is currently in the drafting phase, and the simulation accuracy of ESMs is highly dependent on the robust representation of terrestrial vegetation systems by DGVMs to establish reliable land-atmosphere coupling relationships. The lack of double-cropping process representation is a critical flaw in current DGVM cropland simulations. **Double-cropping occupies 17.42 % (2.62×10<sup>6</sup> km<sup>2</sup>) of global croplands which could not be ignored (Zhang et al., 2021).**

This issue is particularly pronounced in regions with extensive multiple-cropping systems, most notably during the summer harvest-sowing transition period. Models fail to capture the rapid shift from peak vegetation cover to bare soil/seedling states, which introduces large biases into simulations of land surface biogeochemical cycles and energy balance. Furthermore, as global warming expands the climatically suitable areas for multiple cropping toward higher latitudes, accurate representation of multiple-cropping processes has become an urgent unresolved issue in climate change impact research.

It should be noted that not all DGVMs are open-source. We are currently working on the improvement of two models: the open-source Community Land Model version 5.0 (CLM5.0), and the VEGAS model accessed via academic collaboration (with related findings published in Zhou et al., 2022, same first author).

In this revised manuscript, we will further refine the presentation of the patterns used in this paper in accordance with the reviewers' suggestions.

As the reviewer notes, the three recommendations (a)-(c) have clarified our core future research directions, and our team is already advancing work on several of these aspects. The core positioning of this study is as a foundational preliminary work for this research direction, with the central goal of systematically revealing the key limitations of current models – this is the core scientific value of our

work. We sincerely hope that this research positioning can be recognized by the reviewer and the journal editors. Meanwhile, based on the reviewer's suggestions on the scientific contribution of this work, especially Major Comments 2 and 3, we have provided more systematic and in-depth elaboration in the revised manuscript to comprehensively enhance the scientific contribution of this study.

#### *Major comments*

*1、Missing assessment of model capabilities. The manuscript lacks an overview of TRENDY model structures and the modelling protocol. It is my understanding that such an overview would immediately reveal that the majority of TRENDY models—with the exception of LPJmL and certain versions of ORCHIDEE—cannot simulate double-cropping systems. Furthermore, even these two models do not appear to activate double-cropping functionality under the TRENDY protocol.*

*This is a fundamental issue for the paper's framing: the authors compare model outputs to observations of a phenomenon the models do not simulate, then interpret the mismatch as model failure. The finding that TRENDY models fail to reproduce the bimodal LAI pattern (L176-184) is therefore neither surprising nor informative and only confirms what a review of model documentation would reveal. The authors cite Waha et al. (2025), who specifically highlight that multiple cropping "is hardly accounted for" in global land use models.*

*However, this is not picked up when interpreting their results. This creates a misleading impression that models attempted but failed to capture double-cropping, rather than acknowledging they were not designed to do so.*

*A table summarizing each model's cropland representation (single/multiple crops per year, crop types simulated, phenology scheme) would make this limitation explicit. However, it would also show that the comparison is of limited diagnostic value. If a mechanism is entirely absent from a model, the amount that can be learned about parameterisation and model processes is limited. This represents a fundamental conceptual issue rather than a missing detail.*

*Additionally, the introduction (lines 63-66) mentions that "cropping intensity is generally expressed on a scale of 1-3" but it is not clear to me whether this refers to observational datasets, model inputs, or model capabilities. This ambiguity extends throughout the manuscript. As far as I can tell, the authors never explicitly state that TRENDY models do not simulate double cropping for this region, leaving readers to infer this important limitation. The authors should also clarify what cropping intensity the HYDE land-use forcing (used by TRENDY) assigns to the Huang-Huai-Hai Plain and how this is handled by TRENDY.*

#### **Response**

We sincerely apologize for the ambiguity in the original manuscript that caused misunderstanding for the reviewer. The core intent of this study is not to argue that "models attempted to simulate double-cropping but performed poorly", but to reveal the specific simulation biases arising from the lack of explicit

representation of double-cropping processes in current DGVMs, and further explore how models can effectively represent multiple-cropping systems (cropping intensity 2-3). This core positioning was stated in the original discussion section with the sentence: "This discrepancy arises because current DGVMs lack explicit mechanisms for double-cropping systems."

In the revised manuscript, we highlight and explicitly clarify this core research positioning in accordance with the reviewer's comments. For model selection, the TRENDY multi-model ensemble is the most widely used framework in global vegetation dynamic modelling, with the most complete publicly available datasets, and was therefore selected as the core analysis object for this study.

In the revised manuscript, we have explicitly stated this positioning at the outset of the objective subsection in the Introduction, and clarified that cropping intensity information in related studies is extracted from remote sensing phenological cycles. The revised text is as follows:

"Intensifying multiple cropping is considered an important means of increasing food production without expanding cropland area (Wu et al., 2018). As global warming expands the climatic potential for multiple cropping toward higher latitudes, understanding and modeling these systems becomes increasingly urgent. **However, processes related to multiple cropping are rarely represented in current DGVMs.** Cropping intensity is generally expressed on a scale of 1-3, corresponding to single, double, and triple cropping, respectively, and can be determined by extracting phenological cycles from remote sensing data. Yet, because multiple cropping systems are often complex and interspersed with intermittent fallow periods, many such areas are not stably cultivated over long timescales (Liu, 2021). **Consequently, most DGVMs still simulate croplands under a single-cropping assumption, leaving a critical gap in our ability to assess the biogeophysical and biogeochemical impacts of agricultural intensification.** To address this, we selected the Huang-Huai-Hai Plain, a highly representative double-cropping region, as our study area. Since the 1980s, this region has been dominated by a winter wheat–summer crop double-cropping system. The objective of this study is to quantify the biases introduced by the single-cropping assumption in current DGVMs when simulating vegetation seasonality under intensive double-cropping management, thereby highlighting the urgent need to better represent agricultural practices in land surface models."

In addition, we added the following model information in table 2 as:

Table 2. Spatial Resolution, Temporal Resolution and Temporal Coverage of chosen Leaf Area Index datasets. The abbreviations represent the dataset names that appear later in the text.

Dataset Name	Spatial Resolution	Temporal Resolution	Temporal Coverage	Abbreviation	cropland included	Single/multiple cropping
GIMMS-LAI4g	1/12°	half monthly	1982-2020	LAI4g		
MOD15A2H	500m	8 day	2000.feb-2023	MOD15		
GLASS-LAI	0.05°	8 day	2000.feb -2022	GLASS		
ED	0.5°	monthly	1700-2023	ED	NAN	NAN
ISAM	0.5°	monthly	1700-2023	ISAM	YES	Single
LPJ-GUESS	0.5°	monthly	1700-2023	LPJ-GUESS	YES	Single
LPJml	0.5°	monthly	1700-2023	LPJml	YES	Single
LPXbern	0.5°	monthly	1700-2023	LPXbern	YES	Single
ORCHIDEE	0.5°	monthly	1700-2023	ORCHIDEE	YES	Single
VISIT	0.5°	monthly	1860-2023	VISIT	YES	Single

Meanwhile, we have compiled the structural characteristics of cropland simulation for each model in the revised manuscript, and added Table to summarize the representation capacity of each model for cropland ecosystems. Additionally, regarding the current issue of Multi-Season-Cropping farmland, we conducted further literature research and presented our findings in the Discussion section.

We investigated each model, including both the model development papers and the papers citing the models, as detailed in the list. Unfortunately, we are currently unable to fully clarify how each model handles cropland information in detail. We can only determine whether they attempt to include multiple cropping. Based on current understanding, whether multiple cropping is included is the main reason for the limited simulation capability in this region, rather than whether crop types are refined.

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
ED	NAN				
ISAM	YES	Maize (C4)、Wheat、Rice、Soybean (C3)	Defined by accumulated temperature	YES	YES
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
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 of similar crops (e.g. temperate cereals)	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
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
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
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 <sub>4</sub> 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	NAN	NAN

			initialization process of sowing/transplanting; <ul style="list-style-type: none"> <li>◦ When the surface temperature is below the critical value of 5°C, the crop harvest process is triggered.</li> </ul> 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.		
--	--	--	---	--	--

2、 *Overgeneralization of model characteristics. The authors do not acknowledge the structural differences in cropland representation between the TRENDY models. For example, Figures 5- 6 present all model outputs together without differentiating by model structure. These structural differences likely explain some of the inter-model spread in Figures 5-6 but remain unexplored. Models differ in fundamental ways that could influence LAI seasonal patterns:*

*Phenology schemes: Some models use grass-like phenology for crops (baseline ORCHIDEE) while others implement crop-specific phenology with heat unit accumulation and photoperiod responses (ORCHIDEE-CROP, LPJmL).*

*1) Crop functional type representation: Models range from generic C3/C4 categories to multiple crop-specific functional types with distinct parameterizations.*

*2) Management representation: Some models lack explicit management routines—crops follow natural phenology without prescribed sowing dates, fertilization schedules, or harvest triggers (baseline ORCHIDEE)—while others include these processes (ORCHIDEE-CROP, LPJmL).*

*3) Land use models—including TRENDY models—"primarily assume monocropping" (Waha et al., 2025), meaning they simulate only a single crop per grid cell per year. While some models like LPJmL and LPJ-GUESS can represent multiple crop types within a grid cell, TRENDY simulations do not implement sequential cropping (two crops harvested from the same field within one year). This fundamental limitation means models should not reproduce the bimodal LAI pattern characteristic of winter wheat-summer maize double-cropping in the study region. If they would reproduce these patterns, they would be right for the wrong reasons.*

*In my opinion, the analysis needs to differentiate models depending on these structural characteristics and document their capability to simulate double cropping. Otherwise, the analysis cannot meaningfully interpret inter-model differences or explain why models collectively fail to capture the observed seasonal pattern.*

**Response**

We thank the reviewer for this valuable comment. As you mentioned, these models have begun to focus on the vegetation changes brought about by phenology (such as LPJ-GUESS and LPJmL), achieving this

through crop responses to the climatic response cycles. However, multiple cropping itself disrupts the natural cycles. Whether multiple cropping can be realized depends on whether the water and heat conditions throughout the year support it, i.e., the potential for multiple cropping estimation. Currently, the Trendy model primarily assumes single-season cropping. The issue we wish to highlight is the lack of clarity introduced by single-season cropping at the regional scale. This is particularly important because these models are currently used as key data support and tools for calculating global greening and its attribution. The significant gaps in multiple cropping regions imply considerable errors, yet these errors are seldom addressed.

We hope to clearly highlight this issue in the paper, namely that the problems arising from the requirement of single cropping are highly significant at the regional scale and should be given due attention in the design.

*3、 Insufficient analysis of divergent greening trends. The manuscript documents substantial disagreement among models regarding greening trends (Figures 3-4, L145-155), with some models showing strong greening (LPX-Bern, ED) while others show browning (ISAM, LPJmL). I believe this assessment could have real merit and provide substantial scientific insight. However, the analysis only reports that the ensemble collectively “fails” to reproduce these trends without trying to understand why models disagree. Yet this could increase understanding of which processes or parameterizations are most important for simulating agricultural vegetation dynamics under climate change. The paper mentions potential drivers (CO<sub>2</sub> fertilization, nitrogen deposition, climate change, land management; L39-41, 239-240) but does not systematically evaluate how these are represented in different models. It would be interesting to know if for example models that show browning lack CO<sub>2</sub> fertilization effects on crops. Or if crop-specific versus grass-like phenologies are associated with better trend representation. Additionally, the conclusion that “DGVMs underestimate the contribution of croplands to regional greening trends” (L 283) appears to mix models with opposite behaviours together. Similar to major comment 2, I think an analysis of the structural differences between models (see the table suggested in major comment 1) in relation to greening performance would be beneficial. I would suggest grouping models at least by their trend direction and investigating what distinguishes “greening” from “no- greening” models. Such an analysis might help identify which model features are critical for capturing observed agricultural responses to global change. However, it would require the systematic structural comparison outlined in major comment 1. In my opinion this is a different study focused on diagnosing model differences rather than documenting failure.*

## **Response**

First, we would like to thank you for your insightful observations and analysis. The global vegetation change trends and their attribution are very important issues, and DGVMs seem to be the only reliable methods capable of simultaneously identifying driving factors and quantifying attribution. The directional differences in trends (both positive and negative) are an unexpected outcome that we did not anticipate. However, since we are currently only able to conduct validation analysis on the published

data, the model structure can only be obtained from surveys. Therefore, we are uncertain how to approach a deeper exploration of the mechanisms behind these trends, and can only present the issue as it stands. The statement in L283 was inaccurate, so we have made adjustments and specifically addressed this in the discussion, hoping to draw attention to it. We are unsure whether similar response issues exist in other regions.

In the revision, we have removed the sentence: *"DGVMs underestimate the contribution of croplands to regional greening trends."* At the same time, the following content was added in the discussion section:

*"There are significant discrepancies in the simulated trends, even showing contradictory growth and decline trends, while remote sensing LAI shows a relatively consistent growth trend. We are currently unable to provide a clear conclusion on why such large discrepancies exist in the trends, and which specific factors contribute to this. The absence of multiple cropping is just one of the potential influences. Given that DGVMs are important tools for attributing and quantifying global greening, the current inconsistency issue should receive sufficient attention."*

BTW, stepping out of the scope of this paper, this is an open topic. Model development is a very challenging task, and the development teams of these models have made significant contributions. However, at the same time, objective conditions limit our ability to easily access model codes or participate in the design of comparison project plans. As a data-user researcher, our current research on model performance is still primarily focused on data-related issues. We also hope to be more involved in model modifications, but as mentioned earlier, the work we can participate in is mainly through open-source models.

*4 、 Insufficient description of data processing methods. While the manuscript provides detailed descriptions of the remote sensing products (Section 2.2, L 90-103), it does not explain the aggregation methods used to standardize these datasets. The products have very different native resolutions—MODIS at 500m/8-day, GLASS at 0.05°/8-day, and GIMMS at 1/12°/semi-monthly—yet all are aggregated to 0.5°/monthly for comparison. The authors state only that datasets were "standardized" (line 101) without specifying:*

*(1) Temporal aggregation: How were 8-day or semi-monthly values converted to monthly? (e.g., mean, maximum, integrated LAI-days?)*

*(2) Spatial aggregation: How were fine-resolution pixels aggregated to 0.5°? (e.g., area-weighted mean, simple average?)*

*(3) Cropland masking: Was aggregation performed before or after applying the cropland mask? This matters substantially when cropland covers only a fraction of 0.5° cells. I am concerned that these choices could significantly affect the bimodal pattern the authors analyze. For example, monthly means vs. monthly maxima could alter the prominence of growth peaks. This could make a substantial difference but without the description of their methods the reader cannot know. I would recommend*

*adding a subsection describing aggregation procedures, and a justification of the authors' choices.*

## **Response**

We thank the reviewer for the rigorous and detailed professional comments on the data processing methods. Full disclosure of the standardized aggregation methods for multi-source data is a core prerequisite for ensuring the reproducibility of research results and the reliability of conclusions. We acknowledge the omission of method description in the original version, and fully agree with the reviewer's professional judgment that the choice of processing workflow can significantly affect the analysis of the bimodal LAI pattern. We have completed the supplementation and revision of this section in response to the comments, with details as follows:

We have added a dedicated subsection Spatiotemporal Aggregation and Standardization of Multi-source Remote Sensing Data under Section 2.2 Remote Sensing Products in the manuscript, 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.

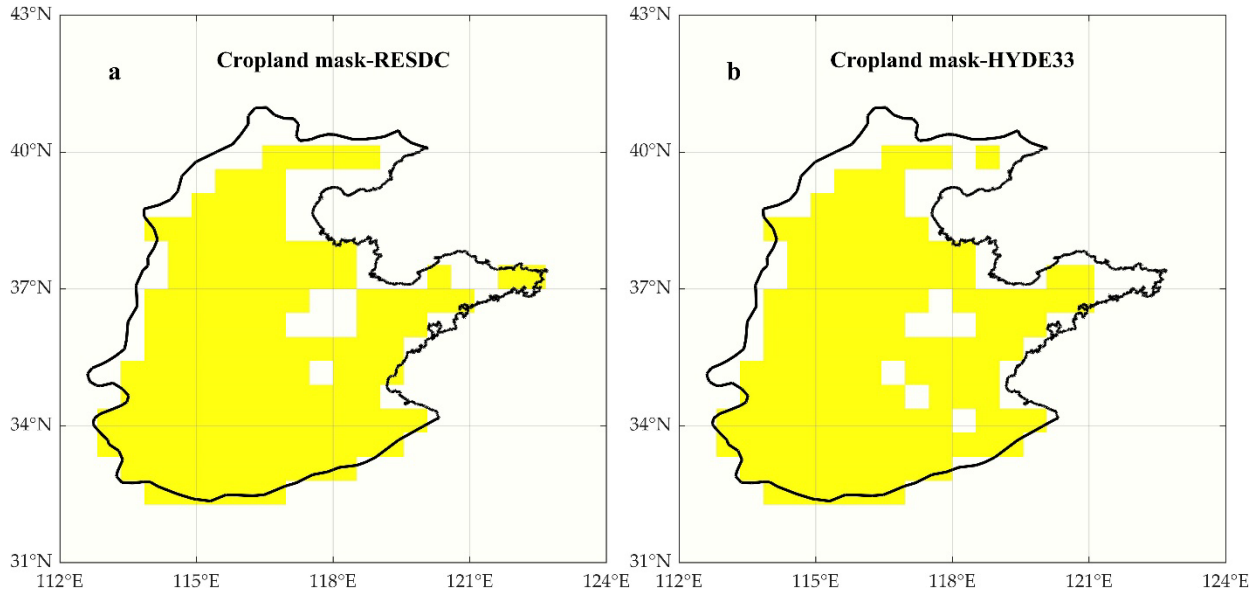
This method can smooth the random noise of single-period data while fully retaining the characteristic bimodal seasonal dynamics of the double-cropping system, and avoids the over-amplification of transition period signals caused by the maximum value composite method.

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.

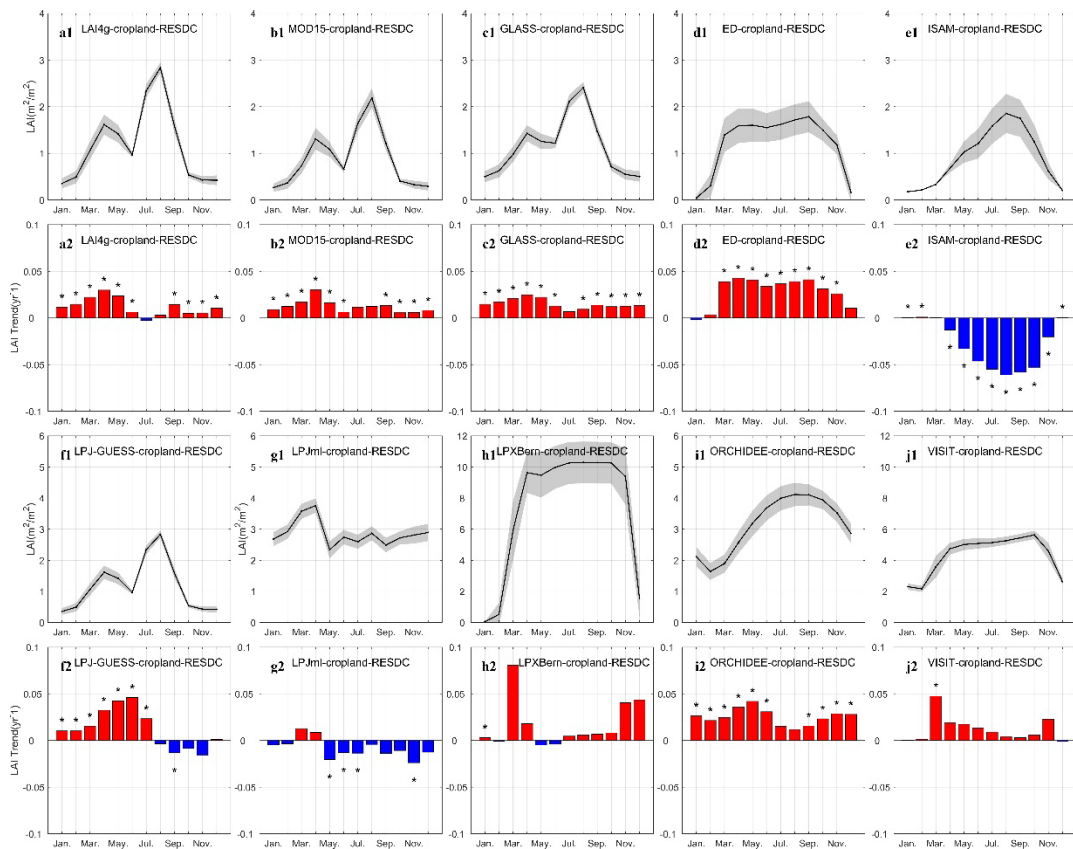
Implementation sequence of cropland masking: We adopted the workflow of aggregation followed by cropland masking. Mixed pixels are an unavoidable issue in scale conversion. It is difficult to find 0.5° grids with pure cropland cover in both model results and real-world remote sensing images. We therefore 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.

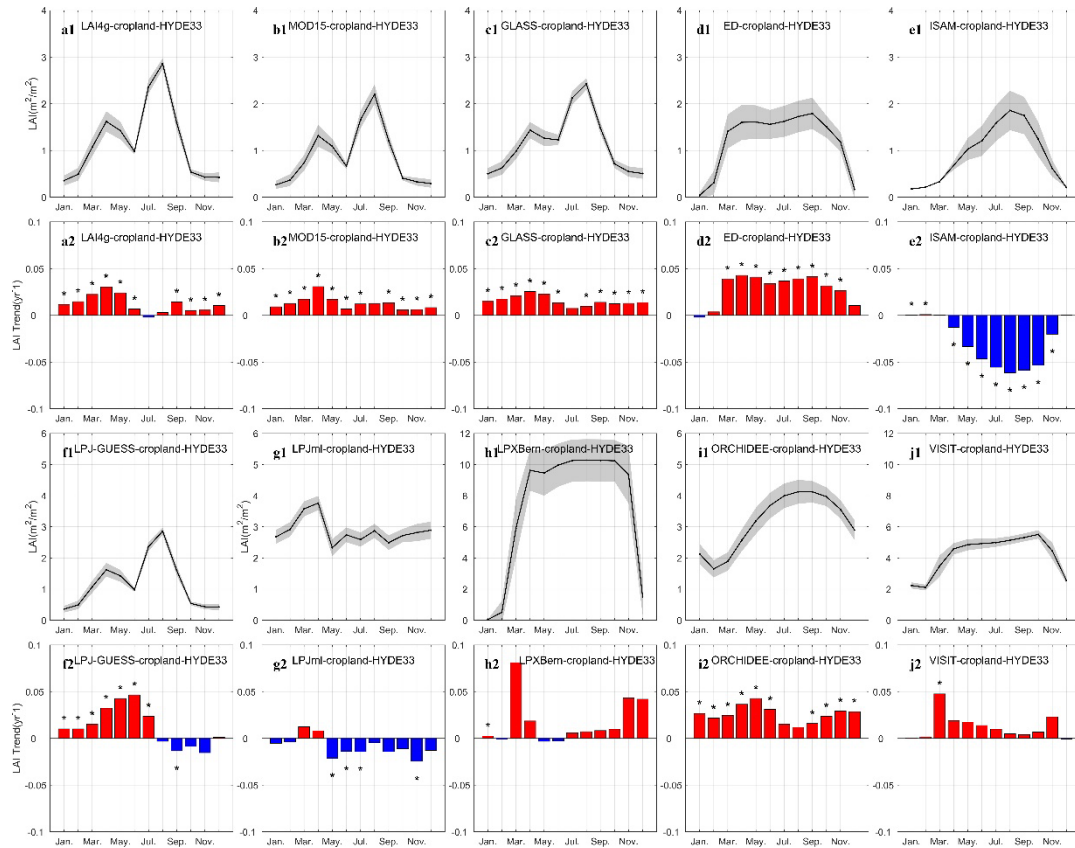


**Figure 1: Cropland Mask produced using RESDC and HYDE3.3 Land-Use/Cover Datasets**



**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( $p < 0.05$ ).**

The 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.

*5. Missing limitations and methodological inconsistencies. The manuscript lacks any discussion of study limitations. Most critically, it appears that the authors use different land-use datasets for their analysis than the models they evaluate. They apply a cropland mask based on RESDC (1km resolution, lines 104-109) to define their study region, while TRENDY models use HYDE3.3 (0.5° resolution) for land-use forcing. Their own Figure 8 demonstrates these datasets disagree substantially in both spatial patterns and temporal dynamics. Yet the authors criticize TRENDY for using HYDE (lines 249-254) but do not acknowledge that comparing model outputs forced with one land-use dataset against observations masked with another is an inconsistency. Additional unacknowledged limitations include: (i) scale mismatch and homogeneity assumptions—while the authors note that 0.5° grid cells may obscure*

*dynamics (line 255), they apply this only as a model limitation without acknowledging that their own analysis treats aggregated 0.5° cells as homogeneous cropland; (ii) lack of explanation of aggregation methods (see major comment 4); (iii) no consideration of alternative explanations for model-data (beyond lack of bimodal pattern) mismatch beyond missing double-cropping (e.g., incorrect phenology parameters, cultivar choices, or sowing dates that could be corrected without implementing full double-cropping). In my opinion a limitations section is needed to address these or other issues. It would also be interesting to test whether using HYDE-based cropland masks changes the observed bimodal pattern in remote sensing data. This would give insight on how much the land-use dataset inconsistency affects the main conclusions.*

## Response

We greatly appreciate the reviewer's comments, as the land cover mask has always been one of the sources of error. We will first discuss the logic of our analysis and then make the corresponding modifications based on the reviewer's suggestions.

In our analysis, we used the RESDC data as the true value, which was developed specifically for China and is widely applied. Indeed, there are certain discrepancies between the RESDC and HYDE data. Therefore, our analysis logic is as follows:

Examining the difference between remote sensing observations and simulations on actual cropland. This introduces two potential uncertainties: first, the inconsistency between the model's cropland mask and the actual land cover; second, the mixed pixel issue at the 0.5° resolution, where pixels are not purely cropland, as shown in Figure 8.

Let's first address the mixed pixel issue. At this resolution, mixed pixels are inevitable, and we can only determine whether a pixel is cropland based on a certain proportion threshold. We chose 50%, and the bimodal structure also indicates that the dominant signal is cropland, with no such signal for natural vegetation.

Next, we compared the 0.5° mask of RESDC and HYDE with the 50% threshold. As shown in the figure below, the mask ranges of both datasets are relatively consistent, with the differences being smaller than the total cropland area. This is because the proportion of cropland in the 0.5° grid point of HYDE is smaller than in the RESDC data. Further analysis also indicated that the results did not show significant changes, as shown in the results figure below.

Given the limited potential impact of HYDE, we have revised the discussion section as follows: we suggest removing the following sentence:

*"Inappropriate parameterizations may also lead to systematic overestimation of grid-scale LAI values."*

The following content has been revised: **Change** "Moreover, the TRENDY project employs the HYDE global historical environment database (Klein Goldewijk et al., 2017) as a proxy for cropland dynamics.

Yet, while global products are useful at large scales, regional-scale land use/cover change (LUCC) datasets are often more accurate. The RESDC LUCC dataset, derived from remote sensing and validated at local scales, provides a more realistic representation of land cover in the Huang-Huai-Hai Plain. A comparison of HYDE3.3 and RESDC (Fig.8) shows that HYDE unrealistically exaggerates interannual variability in cropland area, despite the relative stability of land use in this region, and that its spatial distribution of croplands deviates from actual conditions." to "Moreover, the TRENDY project uses the HYDE global historical environment database (Klein Goldewijk et al., 2017) as a proxy for cropland dynamics. In contrast, we used the RESDC data, which is more widely applied in China, for the actual farmland coverage analysis. There are certain discrepancies between the two datasets, as shown in Figure 8. Although land use in this region is relatively stable, HYDE exaggerates unrealistic interannual variability in cropland area and its spatial distribution of croplands deviates from actual conditions. Both datasets face the mixed pixel problem at a 0.5° resolution, so a threshold is needed for analysis. We used a 50% cropland proportion as the threshold. At this threshold, the spatial distribution of both datasets at 0.5° resolution becomes more consistent, as the actual cropland proportion in the coarse-resolution pixels is higher in the RESDC data. We also analyzed the results under the HYDE mask, and the findings showed no significant changes."

At the same time, the paper focuses more on the impact brought about by multiple cropping information, so we have removed the following paragraph. Although the meaning is correct, it is not directly related to the issue addressed in our results: "To improve the accuracy and applicability of DGVMs, future research should extend analyses to larger regions to investigate double-cropping and fallow systems at the global scale, with a particular focus on three priorities: (1) incorporating explicit mechanisms for cropping systems; (2) optimizing regional parameterizations; and (3) integrating locally validated LUCC and crop calendar datasets."

## Reference

- [1] Piao S , Yin G , Tan J ,et al. Detection and attribution of vegetation greening trend in China over the last 30 years[J].Global Change Biology, 2015, 21(4).DOI:10.1111/gcb.12795.
- [2] Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneeth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viogy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and its drivers, Nat. Clim. Change, 6, 791–795, <https://doi.org/10.1038/nclimate3004>, 2016.
- [3] Masson-Delmotte et al. 2021. Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, 2(1), 2391.

[4] Zhou S, Chen T, Zeng N, et al. The Impact of Cropland Abandonment of Post-Soviet Countries on the Terrestrial Carbon Cycle Based on Optimizing the Cropland Distribution Map[J]. *Biology*, 2022, 11(5): 620.