Reply to Reviewers

Dear Reviewers,

Thank you for taking the time to review our submission titled "Evaluating Dynamic Global Vegetation Models in China: Challenges in capturing trends in Leaf Area and Gross Primary Productivity, but effective seasonal variation representation" to *Earth System Dynamics*. We are grateful for your feedback and suggestions which have significantly strengthened the manuscript. All the comments have been carefully considered. We believe that the revisions being made have addressed the concerns raised by the reviewers. We hope it meets the standards for publication in *Earth System Dynamics*.

Below, we provide point-to-point replies. Each of reviewer's comment is first presented, followed by our reply and changes in the paper. In the revised manuscript, all the changes are marked in blue. Thank you once again for the time and expertise in reviewing our manuscript.

Sincerely,

Reviewer 1:

Title:

"Evaluating Dynamic Global Vegetation Models in China: Challenges in capturing trends in Leaf Area and Gross Primary Productivity, but effective seasonal variation representation." The title is appropriate, but I would shorten it by removing the phrase "but effective seasonal variation representation." The title modification is a suggestion, primarily to make the title concise. The abstract and text, in several instances throughout the manuscript, emphasize the fact that seasonal variation is well captured by the models.

Reply:

Thanks for the comprehensive and professional comments. We agree with this recommendation and have revised the manuscript title accordingly. The detailed information is as follows:

Changes in the paper:

"Evaluating Dynamic Global Vegetation Models in China: Challenges in capturing trends in Leaf Area and Gross Primary Productivity"

General Comment:

The task taken up in the manuscript, evaluating the DGVMs, is essential for the scientific community to know the strengths and weaknesses of model performance in significant land regions, such as China. The study does a good job of comparing the model outputs with observations. It makes a good argument in assessing the errors in simulation and the drivers responsible for them. The study makes the essential hypothesis of why the observed variation in model estimates is occurring. The discussed impacts, ranging from stomatal conductance, lack of accurate N and P cycles in models, using constant temperature sensitivity curves, to single pft parameters for diverse

ecosystems, are issues other studies can focus on for improving the models and their global applicability.
However, the authors should address a few concerns before the manuscript is ready for publication.

Reply:

Thanks for the comprehensive and professional comments. They are extremely helpful and beneficial for improving our paper. Below, we have carefully prepared point-to-point replies for the comments, particularly about the influences that affect the accuracy of the models. We hope that the updated manuscript can address all your concerns.

Specific Comments(1):

On page 4, starting in line 98, the statement "However, it remains poorly documented what the comparison of between observations and model simulations, leading to significant uncertainty about the application of DGVMs in China," is confusing for the readers. The authors should explain what is lacking in more detail, as this is the main gap the authors are trying to address in this manuscript.

Reply:

Thank you for pointing out this critical aspect. In the revised manuscript, we have clarified this issue by explicitly identifying three primary limitations in previous studies: (1) insufficient long-term observational data for rigorous validation of model outputs, (2) inadequate systematic spatial assessments across China's diverse terrestrial ecosystems, and (3) limited validation considering the specificity of vegetation cover types.

Changes in the paper:

Lines 100-103

"Current evaluations of DGVM applicability in China have predominantly relied on site-scale (Han et al., 2025; Zhu and Zeng, 2024), which lack integration with long-term spatial observational constraints to verify model systematicity. Additionally, despite increasing utilization of remote sensing and multi-source datasets for validation, these validations remain fragmented and inconsistently address vegetation-type-specific model output variables (Yue et al., 2024; Jiao et al., 2024)."

Specific Comments(2):

Overall, the introduction section makes a compelling case that DGVMs require regional scale evaluation using observations, as many studies have found significant errors in simulating regional terrestrial fluxes and vegetation growth. However, this manuscript can include a paragraph in the introduction section (preferably at the end of the section) that explains the broader impacts of the study and how this is beneficial to future regional-scale studies using DGVMs or regional studies in general. This section can elucidate studies that will benefit from assessing various models.

Reply:

We sincerely appreciate the reviewer's valuable suggestion. In the revised manuscript, we have added a concise paragraph at the end of the Introduction, clearly articulating the broader impacts and implications of this study. Specifically, we highlight that evaluating DGVM performance across China will (1) enhance our mechanistic understanding of the unique regional carbon cycle dynamics, (2) provide critical feedback and insights for structural and parametric improvements of DGVMs, and (3) inform regional carbon sink quantification relevant to climate mitigation policies. Collectively, these benefits will facilitate future research efforts focusing on regional ecosystem modeling and carbon management strategies.

Changes in the paper:

Lines 110-114

"This study aims to identify key pathways for improving DGVM structure and parametrization, enhance mechanistic understanding of China's unique carbon cycle dynamics, and provide insights into quantifying regional carbon sinks, thus supporting climate-related policy development and guiding future regional-scale ecosystem modelling studies."

Specific Comments(3):

Equation (1) in the manuscript uses the annual CO_2, temperature, and precipitation all impacted by anthropogenic activities. What does the Xanthropogenic refer to, then? The variables investigated here, like temperature, precipitation, radiation, and vegetation parameters, have high seasonality. Using annual data for the sensitivity analysis might cause a loss of information on vegetation growth and not provide a complete picture of the impact of various drivers. In this analysis, how do you justify using annual temperature, precipitation, and radiation? Since a seasonal relation of monthly LAI and various drivers is also considered in the study, what additional information is being produced from equation (1) and the annual linear regression analysis?

Reply:

We appreciate the reviewer's insightful comment, which has helped us clarify our methodological framework. In response, we have made the following key adjustments and clarifications in the manuscript:

(1)We have removed the ambiguous term $X_{anthropogenic}$ from Equation (1) to prevent confusion, as it was not clearly defined nor essential to our primary analysis.

(2) The annual-scale multiple linear regression (Equation 1) is specifically employed to quantify the interannual sensitivity of vegetation dynamics (LAI) to long-term changes in environmental drivers, thus complementing our seasonal analysis.

(3) For the seasonal-scale analysis, we utilize cross-correlation functions to examine the temporal relationships and lag effects between monthly LAI and environmental drivers, thus effectively capturing seasonal variations that are not discernible at an annual scale.

Changes in the paper:

Lines 215-218

"where LAI and CO_2 are annual average LAI and Carbon dioxide concentration; respectively; pre, tem and rad are the annual average precipitation, temperature, and radiation, respectively; a, b, c and d are regression coefficients, and ε is the residual error term, which amount of influence of anthropogenic on vegetation dynamics. Both the dependent and independent variables were normalized.".

Specific Comments(4):

Different models are identified to be performing well in various sections of the manuscript. For example, DLEM and IBIS in the trend analysis, and CLM5 in simulating the impact of CO_2 on LAI. While discussing the DGVMs challenges in accurately simulating LAI and GPP, authors provide arguments on how DGVMs are missing some processes or do not have diversity in parameters used. Two things are missing in the discussion, which readers might be looking at when referring to this manuscript. (1) What is the reason for individual models not performing well, and what are these models missing regarding processes and parameters? Of the 14 models investigated, only a few are highlighted in the manuscript, which are performing well. Adding one table to highlight the differences in processes and areas where models can improve will greatly benefit the community. (2) The discussion on which models will perform better for the studies on vegetation over China. Provide recommendations supported by the results from this study.

Reply:

We greatly appreciate the reviewer's detailed and constructive suggestions. We have carefully addressed both aspects in the revised manuscript:

1. Reasons for individual model limitations and missing processes/parameters:

We fully agree with the reviewer's suggestion that clearly identifying the reasons behind the varied performance of individual models will significantly enhance the manuscript's value. Therefore, we have added a new subsection (Section 4.3) explicitly discussing the specific limitations in process representation and parameterization among the evaluated DGVMs. To systematically illustrate this, we have included a new summary table highlighting key missing processes, parameterization deficiencies, and recommended areas for improvement for each model. This addition provides readers with a concise overview of where models underperform and how these gaps could potentially be addressed.

2. Recommendations for selecting models for vegetation studies in China:

Our results reveals that none of the 14 models consistently performs well across China l across all aspects of vegetation dynamics (e.g., trend vs. seasonal variations, LAI vs. GPP). For example, while DLEM captures the GPP trend relatively better in some regions, its areas of good performance remain limited (not exceed 50%) and remain poor in simulating LAI trends. Such discrepancies in simulating different variables are common across models. Consequently, we refrain from recommending any individual model as universally reliable for vegetation modeling across China. Instead, we emphasize the effectiveness and importance of utilizing multi-model ensemble (MME) approaches, which collectively capture model strengths and weaknesses, thereby providing more robust and comprehensive assessments. Thus, our recommendation for future regional vegetation studies is the adoption of MME methods rather than reliance on single-model outputs. These descriptions are consolidated in subsection 4.3.

Changes in the paper:

Lines 481-507

"4.3 Challenges of DGVMs for plant physiology

Our analysis highlights several critical limitations in the current DGVMs regarding their ability to accurately represent vegetation responses to environmental drivers and anthropogenic changes. These limitations arise primarily from differences in how individual DGVMs parameterize and simulate critical ecological and physiological processes. For example, Teckentrup et al. (2021) demonstrated significant divergences among DGVMs, particularly in their approaches to modelling responses to elevated atmospheric CO_2 concentrations, land-use and land-cover change (LULCC), and carbon residence times in vegetation pools. Among these physiological processes, plant carbon assimilation mechanisms such as stomatal conductance and maximum carboxylation velocity (Vcmax) play pivotal roles. Previous studies found that stomatal functioning and maximum velocity of carboxylation (Vcmax) are related to elevated atmospheric CO_2 and photosynthesis acclimation, potentially impacting the estimation of vegetation dynamics in DGVMs (Rezende et al., 2016). Many DGVMs currently simplify or inadequately represent these physiological processes. For instance, *several DGVMs estimate Vcmax at the canopy scale rather than the physiologically more appropriate leaf scale, leading to systematic underestimations. Furthermore, nutrient cycling modules integrated into DGVMs often reduce the sensitivity of vegetation growth to elevated CO_2 concentrations (Smith et al., 2014; Zaehle, 2013; Meyerholt et al., 2020). Collectively, these simplifications constitute structural deficiencies that substantially affect DGVM accuracy. To elucidate mechanistic discrepancies in vegetation trend simulations, we synthesize findings from prior studies to summarize the operational frameworks of 14 models across six critical parameters: the presence of stomatal conductance, Vcmax-related leaf nitrogen content, Land-use and Land cover-change (LULCC) responsiveness, nutrient limitation representation, and apparent carbon residence time in vegetation (Teckentrup et al., 2021; Rezende et al., 2016; Lian et al., 2021; Friedlingstein et al., 2022b). A systematic comparison of these parameterizations is presented in Table 1.*

Table1 DGVMs and their main processes of plant physiology

Model	Stomatal conductance	V_{cmax}-related leaf nitrogen content	land-cover and land-use change	nutrient limitations	ACRT in vegetation
CABLE	-	Coupled with leaf N-P ratio	dynamically simulated	Yes	Medium ACRT turnover
CLASSIC	-	Leaf N content determines V_{cmax}	dynamically simulated	Yes	-
CLM5.0	Medlyn et al.(2011)	Leaf N optimization model	dynamically simulated	Yes	constant ACRT
DLEM	-	-	dynamically simulated	Yes	-
IBIS	Collatz et al. (1991)	-	prescribed	No	Dynamic ACRT turnover
ISAM	-	-	prescribed	Yes	-
ISBA	-	-	dynamically simulated	Yes	-
JULES	Collatz et al. (1991)	Linearly related to leaf N	dynamically simulated	No	Long ACRT turnover
LPJ-GUESS	Haxeltine & Prentice	V_{cmax} varies with foliage N concentration and specific leaf area	dynamically simulated	Yes	-
LPX	-	V_{cmax} related to leaf N	dynamically simulated	No	temporary ACRT turnover
OCN	Ball et al. (1987)	Leaf N content determines V_{cmax}	prescribed	Yes	Dynamic ACRT turnover

ORCHIDEEv3	Ball et al. (1987)	V_{cmax} related to leaf N	prescribed	Yes	long ACRT turnover
SDGVM	-	-	dynamically simulated	No	-
VISIT	-	-	dynamically simulated	No	long ACRT turnover

Our results clearly indicate that no single DGVM among the evaluated 14 consistently performs well across all aspects of vegetation dynamics (e.g., trend vs. seasonal variations, LAI vs. GPP) throughout China (Fig 2-3, Fig 7-8). For example, while DLEM performs relatively better in capturing regional GPP trends in certain locations, its successful performance is spatially limited (covering less than 50% of the study area) and remains poor in capturing LAI trends (Fig 2e, Fig 3e). Similar discrepancies exist for other models regarding their ability to simulate different vegetation variables. Thus, we emphasize the effectiveness and importance of utilizing MME approaches, which collectively capture model strengths and weaknesses, thereby providing more robust and comprehensive assessments.

Specific comments(5):

The inlet bar graphs in Figures S2 and S3 show the area percentage of significant decrease, no significant change, significant change, and others. Should they add up to 100%? How is this calculated? Do readers have enough information to understand these and similar figures in the supplement material?

Reply:

Thanks for the suggestions. We sincerely appreciate the reviewer's valuable feedback on improving the clarity of Figures S2–S3. In the revised Supplemental Methods (Section S3.2), we have explicitly clarified the classification logic and summation principles. The five bars (DE, SD, N, SI, IN) represent distinct vegetation trend categories: (1) decrease (slope<0), (2) significant decrease (slope<0, p<0.05), (3) no detectable change (slope=0) combined with non-significant increase (slope>0, p ⩾ 0.05) and non-significant decrease (slope<0, p ⩾ 0.05), (4) significant increase (slope>0, p<0.05), (5) increase (slope>0). The three central bars ,which respectively represent significant decrease (SD), no change/non-significant increase or decrease (N), and significant increase (SI), collectively account for 100% of the study area.

Changes in the paper:

Figures S2

"Figure S2. The observed annual MODIS LAI trend during 2003-2019 in China. Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ⩾0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the significant trend (p<0.05)."

Figures S3

"Figure S3. The observed annual CSIF trend during 2003-2019 in China. Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ⩾0.05)

change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the significant trend (p<0.05).”

Figures S6

“Figure S6. The observed annual spatial precipitation trend during 2003-2019 in China. Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ≥0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the significant trend (p<0.05).”

Figures S7

“Figure S7. The observed annual spatial temperature trend during 2003-2019 in China. Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ≥0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the significant trend (p<0.05).”

Figures S8

“Figure S8. The observed annual spatial radiation trend during 2003-2019 in China. Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ≥0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the significant trend (p<0.05).”

Figures S15

Figure S15. Interannual trends in changes in the impact of land use change on LAI considered by different models (S3-S2 scenario). Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope≠0, p ≥0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN).. The dot indicated the significant trend (p<0.05).

Figures S16

Figure S16. Interannual trends in changes in the impact of land use change on GPP considered by different models (S3-S2 scenario). Pink color represents the percentage of area of decreasing (slope<0) regions (DE), red color represents the percentage of area of significantly decreasing (slope<0, p<0.05) regions (SD), yellow color represents the percentage of area of regions with no significant (slope=0 combined with slope ≠0, p ≥0.05) change (N), green color represents the percentage of area of significantly increase (slope>0, p<0.05) regions (SI), and light green color represents the percentage of area of increase (slope>0) regions (IN). The dot indicated the *significant trend ($p<0.05$).*

Technical corrections(1):

On page 1, line 27, "model's understanding of the CO_2 fertilization effect..." should be "model's representation of the CO_2 fertilization effect...". Please check and change accordingly.

Reply:

Thanks for pointing this mistake out. We revised the sentence to correct the mistake.

Changes in the paper:

Lines 25-27

"We indicate that the main reason for the model's misestimation is that the model's representation of the CO_2 fertilization effect is inadequate, and thus fails to simulate the vegetation response to CO_2 concentration."

Technical corrections(2):

On page 3, line 68, "Medlyn et al. (2015) utilized used..." has two verbs and should be corrected.

Reply:

Thanks for pointing this mistake out. We revised the sentence to correct the mistake.

Changes in the paper:

Lines 67-69

"Medlyn et al. (2015) leveraged empirical data from the Duke and ORNL Free-Air CO_2 Enrichment (FACE) experiments to refine the parameterization of CO_2 fertilization effects in DGVMs, significantly enhancing their capacity to simulate forest responses to elevated atmospheric CO_2 concentrations (eCO_2)."

Technical corrections(3):

On page 9, line 249, "While The..." capitalization of word in the middle of the sentence should be corrected.

Reply:

Thanks for pointing this mistake out. We revised the sentence to correct the mistake.

Changes in the paper:

Lines 253-254

"Although the DLEM model outperforms other models in simulating long-term GPP trends, its accuracy remains constrained below 50% relative to observational benchmarks (Fig. 3e)."

Technical corrections(4):

On page 19, line 346, "he" should be "the". Please check.

Reply:

Thanks for pointing this mistake out. We revised the sentence to correct the mistake.

Changes in the paper:

Lines 349-350

"The CCF analysis revealed statistically significant correlations between observed and simulated LAI and key climatic variables—precipitation, temperature, and solar radiation (Fig. 9-11)."

Technical corrections(4):

On page 23, line 393, the text "This apparent accuracy..." should be "This apparent inaccuracy...". Please check. Correct me if I am wrong.

Reply:

Thanks for the comments. We are grateful for the astute observation regarding this critical wording issue. In the revised manuscript, we have restructured this section to clarify our original intent.

Changes in the paper:

Lines 395-398

"Overall, the observed LAI trends demonstrate reasonable consistency with model simulations within uncertainty bounds (Fig. 1a), indicating that current DGVM frameworks can effectively capture the overall trendy of vegetation dynamics. However, the overall agreement contrasts with substantial spatial discrepancies in trend misestimation, as evidenced by pronounced spatial misestimations in China (Fig. 2)."