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,

Anzhou Zhao

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:

Line 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:

Line 110-114

"This work aims to identify priority pathways for DGVM structural improvements and enhance understanding of carbon cycle applicable to China and other regions sharing similar ecological characteristics. Specifical, our analysis seeks to identify key pathways for improving DGVM structure and parametrization, advance 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:

Line 220-223

"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:

Line 491-517

"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 and nutrient cycle. 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 Vcmax are related to elevated atmospheric CO2 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, dynamic natural PFT coverage and nutrient cycle (Teckentrup et al., 2021; Rezende et al., 2016; Lian et al., 2021; Friedlingstein et al., 2022b; Sitch et al., 2024). A systematic comparison of these parameterizations is presented in Table 2.*

Table2 DGVMs and their main processes of plant physiology

Model	Stomatal conductance	V_{cmax}-related leaf nitrogen content	Dynamic natural PFT coverage	N cycle	P cycle
CABLE	-	Coupled with leaf N-P ratio	N	Y	N
CLASSIC	Ball et al. (1987)	Leaf N content determines V_{cmax}	Y	N	N
CLM5.0	Medlyn et al. (2012)	Leaf N optimization model	Y	Y	-
DLEM	-	-	N	Y	N
IBIS	Collatz et al. (1991)	-	Y	N	N
ISAM	-	-	N	Y	N
ISBA	-	-	Y	Y	N
JULES	Collatz et al. (1991)	Linearly related to leaf N	Y	Y	N
LPJ-GUESS	Haxeltine and Prentice (1996)	V_{cmax} varies with foliage N concentration and specific leaf area	Y	Y	N
LPX	Ball et al. (1987)	V_{cmax} related to leaf N	Y	Y	N
OCN	Ball et al. (1987)	Leaf N content determines V_{cmax}	N	Y	N

ORCHIDEEv3	Ball et al. (1987)	V_{cmax} is prescribed	N	Y	N
SDGVM	Ball et al. (1987)	Leaf N content determines Vcmax	Y	Y	N
VISIT	-	-	Y	N	N

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:

Line 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:

Line 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:

Line 258-259

"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:

Line 354-355

"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:

Line 400-403

“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).”

Reviewer 2:

Overview

The work by Zhao and colleagues focuses on assessing the ability of a set of Dynamic Global Vegetation Models (DGVMs) to capture the observed trend and seasonality in Leaf Area Index (LAI) and Gross Primary Productivity (GPP) in China. Moreover, the study highlights possible sources of discrepancies between models and observations. Consequently, this study provides an insightful assessment and suggestions for future model developments. However, some extra sources of differences between DGVMs need to be referred to in the manuscript before being ready for publication.

General comment (1)

DGVMs simulate dynamically the terrestrial biogeochemical cycles, accounting for carbon pools and fluxes. However, these pools need to be spun up. For this reason, spin-up simulations are run. Are the considered DGVMs applying the same spin-up protocol? Do they show relevant differences at the beginning of the simulations?

Reply:

We thank the reviewer for this insightful comment. In the TRENDY project, all DGVMs adopt the same spin-up duration. Each model initiates simulations from standardized baseline conditions representative of the year 1700 (e.g., atmospheric CO_2 concentration of 276.59 ppm and prescribed crop/pasture distributions), while employing model-specific plant functional types. To reach carbon equilibrium, models apply pre-industrial climate forcing (1901–1920) repeatedly, either through cyclic repetition or stochastic sampling, depending on the model's configuration. The spin-up duration was sufficiently long (typically 500–1500 years) to allow major ecosystem carbon and water pools to reach a near-equilibrium state.

Despite this standardized protocol, relevant differences do exist at the beginning of the transient simulations. These discrepancies arise due to inherent differences in model structure, such as the treatment of soil and litter carbon pools, vegetation dynamics and turnover rates, phenology schemes, and PFT parameterizations and spin-up convergence criteria. As a result, the initial values of variables such as LAI and GPP can differ among DGVMs, even when they have followed the same spin-up forcing and duration. These differences can propagate into early transient simulations and may influence model–data comparison in the initial decades.

Changes in the paper:

Lines 144-147

"Within the TRENDY project, all Dynamic Global Vegetation Models (DGVMs) employ uniform spin-up durations. Initialization of simulations utilizes the 1700 baseline conditions with model-specific plant functional types (PFTs). Pre-industrial climate forcing data (1901–1920) is recycled through model-dependent cycling or stochastic sampling until a carbon equilibrium state is achieved (Sitch et al., 2024)."

General comment (2)

Land-use changes in DGVMs are based on reference global products such as LUH2, as stated in the manuscript. Do all DGVMs use the same land-use change forcing dataset? Does the dataset used in the DGVMs present relevant differences compared to observations?

Reply:

We thank the reviewer for this important question regarding the land-use change (LUC) forcing used in the DGVM simulations. All participating DGVMs in our study adopted the same standardized land-use forcing dataset, LUH2, as required by the TRENDY protocol. While LUH2 is not a direct observational product, it is a harmonized and downscaled dataset that structurally adheres to observations by integrating three empirical sources: HYDE 3.3 historical land reconstructions, FAO wood harvest statistics, and LUH2-GCB transition matrices. Its core design explicitly separates anthropogenic drivers from natural processes, ensuring methodological alignment with observational principles.

Although absolute agreement with observations at all spatial scales is unattainable for any global product, LUH2's CMIP-optimized architecture maximizes consistency with the constraints required by DGVMs. This design facilitates inter-model comparability and long-term scenario analysis. That said, LUH2 may still differ from high-resolution regional remote sensing or national inventory datasets in aspects such as spatial resolution (~0.25°), treatment of shifting cultivation or pasture expansion, and sub-grid land-cover representation. Such discrepancies are particularly notable in regions with complex or rapidly changing land-use histories.

Changes in the paper:

Lines 147-148

"Concurrently, for land-use change representation, all participating DGVMs adopt LUH2 — developed to standardize land-use forcing for models and enhanced for DGVM compatibility—as the unified land-use change dataset."

Lines 477-481

"In TRENDY project, most of DGVMs use the HYDE and LUH2 database for land use and land cover change (LULCC) data input variables (Klein Goldewijk et al., 2017; Hurtt et al., 2020; Sitch et al., 2024). However, while these LULCC datasets are common across models, their implementation varies according to the land-use processes and functionalities inherent to individual DGVMs. In this study, we employ Table 1 to describe differences in the performance of individual DGVMs, which adapted from Sitch et al. (2024)."

Line 488

Table 1 DGVMs and their main processes relevant to LULCC emissions

Model	Fire	Wood Harvest	Shifting Cultivation/ Subgrid transitions	Crop harvest
CABLE	N	Y	Y	Y
CLASSIC	Y	N	N	Y
CLM5.0	Y	Y	Y	Y
DLEM	N	Y	N	Y
IBIS	Y	Y	N	Y

ISAM	N	Y	N	Y
ISBA	Y	Y	Y	Y
JULES	Y	N	N	Y
LPJ-GUESS	Y	Y	Y	Y
LPX	Y	Y	N	Y
OCN	N	Y	N	Y
ORCHIDEEv3	N	Y	N	Y
SDGVM	Y	N	N	Y
VISIT	Y	Y	Y	Y

General comment (3)

A table showing the main scheme, forcing, and parameterization used in each DGVM in simulating GPP and LAI may be helpful to the readers to understand the main differences among them.

Reply:

Thanks for the comments. We have incorporated the reviewer's suggestion by introducing Table 2 in the revised manuscript. This table synthesizes inter-model variations in five key elements governing vegetation simulations—stomatal conductance schemes, leaf nitrogen content (linked to maximum carboxylation velocity, Vcmax), dynamic natural PFT coverage and nutrient cycle—as documented in prior studies. By systematically comparing these mechanistic and input differences, Table 2 provides a structured framework for understanding the origins of divergent vegetation dynamics simulations across DGVMs.

Changes in the paper:

Line 508

Table2 DGVMs and their main processes of plant physiology

Model	Stomatal conductance	V_{cmax}-related leaf nitrogen content	Dynamic natural PFT coverage	N cycle	P cycle
CABLE	-	Coupled with leaf N-P ratio	N	Y	N

CLASSIC	Ball et al. (1987)	Leaf N content determines V_{cmax}	Y	N	N
CLM5.0	Medlyn et al. (2012)	Leaf N optimization model	Y	Y	-
DLEM	-	-	N	Y	N
IBIS	Collatz et al. (1991)	-	Y	N	N
ISAM	-	-	N	Y	N
ISBA	-	-	Y	Y	N
JULES	Collatz et al. (1991)	Linearly related to leaf N	Y	Y	N
LPJ-GUESS	Haxeltine and Prentice (1996)	V_{cmax} varies with foliage N concentration and specific leaf area	Y	Y	N
LPX	Ball et al. (1987)	V_{cmax} related to leaf N	Y	Y	N
OCN	Ball et al. (1987)	Leaf N content determines V_{cmax}	N	Y	N
ORCHIDEEv3	Ball et al. (1987)	V_{cmax} is prescribed	N	Y	N
SDGVM	Ball et al. (1987)	Leaf N content determines V_{cmax}	Y	Y	N
VISIT	-	-	Y	N	N

General comment (4)

Given the availability of monthly data, it would also be interesting to evaluate the average annual trend of the maximum LAI and GPP value in observations and DGVMs, as part of the seasonal variation section. In this way, the ability of DGVMs to correctly capture the observed peak LAI value and greening could also be assessed.

Reply:

We thank the reviewer for this valuable suggestion. While our study evaluates DGVM performance in simulating interannual trends and seasonal cycles of LAI and GPP, the assessment of peak seasonal values and associated phenological greening dynamics represents an important but distinct research direction that lies beyond the scope of the present study. Nevertheless, we fully acknowledge the significance of this perspective and have included a discussion highlighting the need to assess peak-value simulations in future work. We agree that such analyses would further advance models' capabilities in capturing vegetation-climate interactions, and we intend to prioritize this theme in subsequent studies.

Changes in the paper:

Lines 538-539

"As with the above research, studies on peak seasonal values and phenological greening dynamics hold significant research importance, and future work will strengthen research efforts on this thematic area."

Specific comment (1)

Line 88: "(2021)suggested" space missing

Reply:

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

Changes in the paper:

Lines 87-89

"Song et al. (2021) suggested that errors in the land-use and land-cover change (LULCC) dataset used as input data for many DGVMs likely led to inaccurate estimations of vegetation biomass changes in China."

Specific comment (2)

Line 134: "employed14" space missing

Reply:

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

Changes in the paper:

Lines 140-141

"In this study, we employed 14 DGVM models (CABLE-POP, CLASSIC, CLM5.0, DLEM, IBIS, ISAM, ISBA-CTRIP, JULES, LPJ-GUESS, LPX, OCN, ORCHIDEEv3, SDGVM, and VISIT),"

Specific comment (3)

Lines 135-136: "selected for their superior performance in simulating gridded monthly LAI and GPP, to explore vegetation dynamics." provide a reference for this statement or drop it.

Reply:

Thanks for pointing this mistake out. We have added references related to this sentence.

Changes in the paper:

Lines 140-142

"In this study, we employed 14 DGVM models (CABLE-POP, CLASSIC, CLM5.0, DLEM, IBIS, ISAM, ISBA-CTRIP, JULES, LPJ-GUESS, LPX, OCN, ORCHIDEEv3, SDGVM, and VISIT), selected for their superior performance in simulating gridded monthly LAI and GPP, to explore vegetation dynamics (Zou et al., 2023)."

Specific comment (4)

Line 145: Which temporal resolution is provided by the DGVMs?

Reply:

Thanks for pointing this mistake out. We confirm that all DGVMs participating in the TRENDY project provide output data at a standardized resolution of 0.5° × 0.5° spatially and monthly temporally, as now explicitly stated in the revised manuscript (Line 156).

Changes in the paper:

Lines 156-157

"In TRENDY project, all DGVMs provide monthly outputs at a standardized spatial resolution of 0.5° × 0.5°."

Specific comment (5)

Lines 182-184: Non-vegetated area selection is described twice.

Reply:

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

Changes in the paper:

Lines 189-194

"The image resolution was initially resampled to 0.5°× 0.5° using the majority rule, Evergreen coniferous forest, evergreen broadleaf forest, deciduous coniferous forest, deciduous broadleaf forest, mixed forest, woody savanna, and savanna were then combined into a single category termed forested land. Closed shrubland, open shrubland, grassland, and permanent wetland were collectively classified as grassland. Farmland and agricultural land (natural vegetation) were grouped together as cropland. Urban and built-up land, permanent snow and ice, and unutilized land were also categorized as non-vegetation."

Specific comment (6)

Figure 1: Add in the caption the description of the black bar on MME values.

Reply:

Thanks for pointing this mistake out. We added description of the error line for MME.

Changes in the paper:

Lines 252-253

"The error line for MME is the standard deviation of the 14 modeled trends."

Specific comment (7)

Lines 233,240,241: you can use fewer digits in the percentages, such as 60,3% instead of 60,305%

Reply:

Thanks for pointing this mistake out. We have standardized all percentage values throughout the manuscript to two significant figures.

Specific comment (8)

Line 402: "One the other [..]" should be "On the other […]"

Reply:

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

Changes in the paper:

Lines 411-412

"On the other hand, the spatial patterns of dominant environmental and anthropogenic drivers affecting observed LAI responses differ markedly from those affecting simulated responses (Fig. S13)."

Editor revisions:

Overview

Thank you for your detailed replies to the Reviewers' comments and the edits that you have implemented in the manuscript. Before your submission may be accepted for publication in Earth System Dynamics, I would encourage you to consider the minor points below.

General comment (1)

In one of their comments, Reviewer #1 suggests to: "explain the broader impacts of the study and how this is beneficial to future regional-scale studies using DGVMs or regional studies in general." In response to this, you added the following text: "This work aims to identify priority pathways for DGVM structural improvements, advance mechanistic understanding of China's unique carbon cycle dynamics. 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."

I would suggest avoiding the repetition "This work" – "This study". Perhaps for the second reference to your work you could use "Specifically, our analysis aims to" or something similar. You may also consider highlighting the applicability of your results to DGVM studies beyond China, as ESD has a focus on studies whose relevance extends beyond a specific region.

Reply:

We appreciate the editor's constructive suggestions. We have made the following improvements to address the comments.

Changes in the paper:

Line 110-114

"This work aims to identify priority pathways for DGVM structural improvements and enhance understanding of carbon cycle applicable to China and other regions sharing similar ecological characteristics. Specifical, our analysis seeks to identify key pathways for improving DGVM structure and parametrization, advance 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."

General comment (2)

In their General Comment #1, the second Reviewer asks: "DGVMs simulate dynamically the terrestrial biogeochemical cycles, accounting for carbon pools and fluxes. However, these pools need to be spun up. For this reason, spin-up simulations are run. Are the considered DGVMs applying the same spin-up protocol? Do they show relevant differences at the beginning of the simulations?" In your reply, you provide a clear answer to the first question of the Reviewer, but never address the second question, namely whether the DGVMs show relevant differences at the beginning of the simulations.

Reply:

We thank the Reviewer and Editor for this insightful question. All DGVMs considered in our study followed a harmonized spin-up protocol, as defined by the TRENDY simulation experiment design. Specifically, each model was spun up using repeated cycles of early 20th-century climate (e.g., 1901–1920 CRU-NCEP data), with fixed atmospheric CO_2 concentrations (e.g., 276.59 ppm), and under potential natural vegetation without land-use changes. The spin-up duration was sufficiently long (typically 500–1500 years) to allow major ecosystem carbon and water pools to reach a near-equilibrium state.

Despite this standardized protocol, relevant differences do exist at the beginning of the transient simulations. These discrepancies arise due to inherent differences in model structure, such as the treatment of soil and litter carbon pools, vegetation dynamics and turnover rates, phenology schemes, and PFT parameterizations and spin-up convergence criteria. As a result, the initial values of variables such as LAI and GPP can differ among DGVMs, even when they have followed the same spin-up forcing and duration. These differences can propagate into early transient simulations and may influence model–data comparison in the initial decades.

Changes in the Reply to Reviewers:

We have now incorporated the above responses into our reply to Reviewer #2-General comment (1).

General comment (3)

In their General Comment #2, the second Reviewer asks: "Land-use changes in DGVMs are based on reference global products such as LUH2, as stated in the manuscript. Do all DGVMs use the same land-use change forcing dataset? Does the dataset used in the DGVMs present relevant differences compared to observations?" In your reply, you provide a clear answer to the first question of the Reviewer, but never address the second question, namely to what extent the datasets used in the DGVMs adhere to observations.

Reply:

We thank the reviewer for this important question regarding the land-use change (LUC) forcing used in the DGVM simulations. All participating DGVMs in our study adopted the same standardized land-use forcing dataset, LUH2, as required by the TRENDY protocol. While LUH2 is not a direct observational product, it is a harmonized and downscaled dataset that structurally adheres to observations by integrating three empirical sources: HYDE 3.3 historical land reconstructions, FAO wood harvest statistics, and LUH2-GCB transition matrices. Its core design explicitly separates anthropogenic drivers from natural processes, ensuring methodological alignment with observational principles.

Although absolute agreement with observations at all spatial scales is unattainable for any global product, LUH2's CMIP-optimized architecture maximizes consistency with the constraints required by DGVMs. This design facilitates inter-model comparability and long-term scenario analysis. That said, LUH2 may still differ from high-resolution regional remote sensing or national inventory datasets in aspects such as spatial resolution (~0.25°), treatment of shifting cultivation or pasture expansion, and sub-grid land-cover representation. Such discrepancies are particularly notable in regions with complex or rapidly changing land-use histories.

Changes in the Reply to Reviewers:

We have now incorporated the above responses into our reply to Reviewer #2-General comment (2):

Specific comment

You have implemented all of the technical corrections suggested by the Reviewers, but there are still a number of typographical errors and formulations that should be corrected. I list some examples below, but this is by no means a comprehensive list. I would recommend a thorough proof-reading of the text.

Reply:

We sincerely appreciate the meticulous review of our manuscript. All typographical errors and suboptimal formulations identified in the editorial comments have been systematically corrected in the revised manuscript. Additionally, we have conducted a full linguistic audit of the text, with all technical corrections implemented and tracked via highlighted text.

Technical corrections(1):

l. 100 "site-scale(Han et al" missing space

Changes in the paper:

Line 110

"Current evaluations of DGVM applicability in China have predominantly relied on site-scale (Han et al., 2025; Zhu and Zeng, 2024)"

Technical corrections(2):

Fig. 1 caption: "error line" --> "error bar"

Changes in the paper:

Fig.1 and the Supplementary figure captions

"The error bar for MME is the standard deviation of the 14 simulated trends."

Technical corrections(3):

Supplementary figure captions: "The dot indicated the significant trend" --> "Dots indicate significant trends".

Changes in the paper:

Supplementary figure captions

"Dots indicate significant trends ($p<0.05$)"

Technical corrections(4):

Fig. 9 caption: "The values over the red line indicated" --> "The values over the red line indicate"

Figs 1, 4, 6, 9, 10, 11, and the Supplementary figure captions

Revise tense truthfulness in sentences.