Responses to Reviewer #1's Comments

This manuscript examines how long-term trends and interannual variability differ among NDVI, VOD, and model-derived GPP across the Asia-Oceania region. The combined analysis of these temporal characteristics could provide useful insights into indicator-dependent assessments of vegetation change. However, the current analysis does not adequately distinguish ecological divergence from differences arising from temporal aggregation, product uncertainty, or model structure. I believe that substantial revisions to both the analysis and its interpretation are needed before the manuscript can be considered for publication.

Response:

We thank the reviewer for the careful and constructive assessment. We agree that the original manuscript did not sufficiently separate differences caused by temporal aggregation, product-specific uncertainty, and model structure from differences among the vegetation indicators themselves. In response, we have conducted additional sensitivity analyses using matched temporal windows, revised the statistical treatment of interannual variability and trends, re-examined the VISIT sensitivity experiments using undetrended component changes, and performed an independent comparison of VISIT GPP with eddy-covariance-based GPP. In the revised manuscript, we will incorporate these analyses and moderate interpretations that cannot be directly supported by the data.

Major comments:

1

The indicators are not compared with over equivalent temporal supports. NDVI is averaged over a fixed growing-season mask, VOD is sampled in a single month selected from the month of maximum long-term mean NDVI, and GPP is averaged over the full year. These choices can themselves generate different trends and interannual variability, especially in monsoon, boreal, and seasonally dry regions. The authors should repeat the main comparisons using matched temporal definitions, for example a common growing-season mean for all indicators, annual values for all indicators where appropriate, and several alternative VOD windows such as a three-month peak-season mean or the VOD seasonal maximum. Sensitivity to a fixed versus annually varying growing season could also be tested. Without these controls, it is difficult to interpret the reported disagreement.

Response:

We agree that the use of different temporal supports was an important limitation of the original analysis. To evaluate the extent to which the reported disagreement was caused by temporal aggregation itself, we repeated the comparisons using the same growing-season (GS) definition for NDVI, VOD, and VISIT GPP. For VOD, we additionally tested four temporal aggregation methods: the original single month defined by the long-term NDVI peak month, a three-month mean centered on that peak month, the seasonal maximum within the GS, and the mean over the common GS. To ensure a fair comparison among the four VOD treatments, we also evaluated them using the same set of valid years at each pixel.

The sensitivity analysis shows that temporal aggregation substantially affects the NDVI–VOD IAV comparison. Across Asia–Oceania, the mean signed NDVI–VOD correlation increased from 0.310 with the original single-month treatment to 0.448 with the common-GS treatment, while the area showing a significant positive correlation increased from 37.59% to 52.17%. In the common-year comparison, the mean correlations were 0.311, 0.367, 0.363, and 0.456 for the single-month, three-month, seasonal-maximum, and common-GS treatments, respectively; the corresponding significant-positive fractions were 37.78%, 42.49%, 41.96%, and 52.95%. These results demonstrate that the original temporal definition contributed appreciably to the NDVI-VOD disagreement, although substantial disagreement remained after temporal matching.

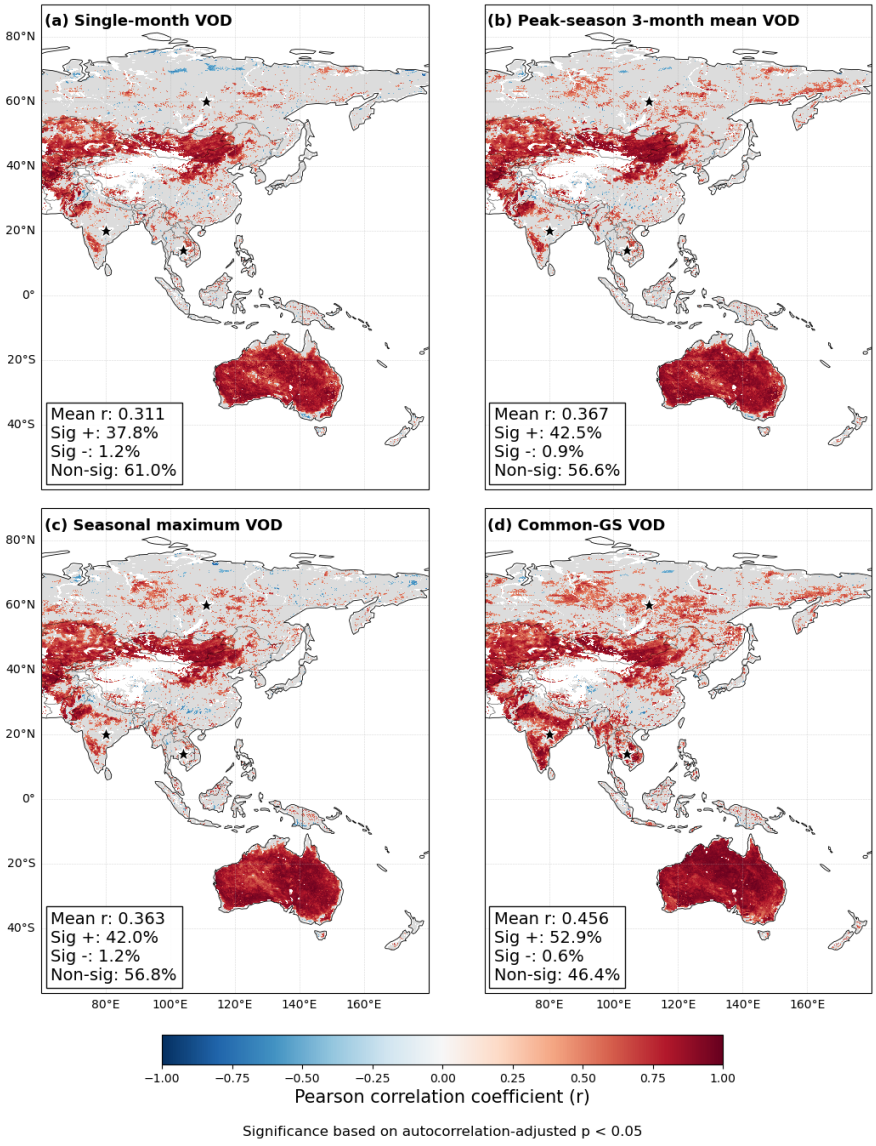


Figure R1. Sensitivity of NDVI–VOD interannual variability to the temporal aggregation of VOD.
Spatial distributions of the signed Pearson correlation coefficient between detrended SNU NDVI and VOD for (a) the original single-month treatment based on the long-term NDVI peak month, (b) a three-month mean centered on the NDVI peak month, (c) the seasonal maximum VOD within the growing season, and (d) the mean VOD over the common growing season. To ensure comparability among the four treatments, correlations were calculated using the same set of valid years at each pixel. Statistical significance was evaluated using an effective sample size accounting for lag-1 temporal autocorrelation. Across Asia-Oceania, the mean correlations were 0.311, 0.367, 0.363, and 0.456 for panels (a)–(d), respectively.

For NDVI–GPP, applying the same GS definition had a much smaller effect. The mean signed correlation changed only slightly from 0.374 with the original treatment to 0.366 with the common-GS treatment, and 96.38% of pixels retained the same correlation-significance classification. Thus, unlike the larger sensitivity

observed for NDVI–VOD, the mismatch between GS-averaged NDVI and annual-mean GPP appears to have only a limited influence on the overall NDVI–GPP IAV relationship at the domain scale.

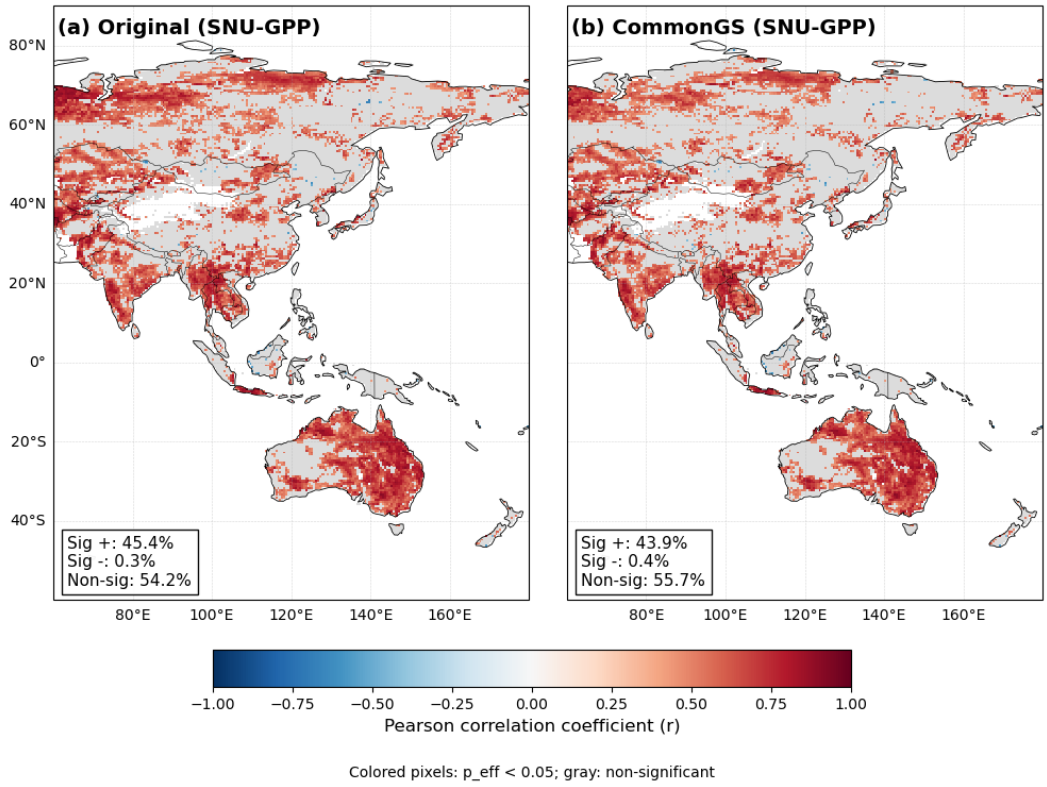


Figure R2. Sensitivity of NDVI–GPP interannual variability to temporal-window matching.
Spatial distributions of the signed Pearson correlation coefficient between detrended SNU NDVI and VISIT GPP for (a) the original treatment, in which NDVI was averaged over the growing season and GPP over the full year, and (b) the common-GS treatment, in which the same growing-season definition was applied to both indicators. Correlation significance was evaluated using an effective sample size accounting for lag-1 temporal autocorrelation. Across Asia-Oceania, the mean signed correlation changed only slightly from 0.374 to 0.366, and 96.38% of the pixel-wise significance classes were unchanged.

These sensitivity tests were conducted using the fixed GS definition employed in the submitted analysis. The GS was defined separately for each grid cell and calendar month using thresholds of long-term mean air temperature and NDVI and was held constant across years. The analyses shown here therefore evaluate the effect of applying the same temporal window across indicators; they do not evaluate interannual changes in growing-season timing.

We agree that further tests are useful for evaluating the robustness of these conclusions. We have not yet completed two additional sensitivity analyses suggested by the reviewer: an annual-mean comparison using

annual values for all indicators, including annual-mean VOD, and a comparison using an annually varying GS. We will include both in the revised manuscript. The annual-mean analysis will test the consequences of using a fully annual temporal support, while being interpreted cautiously because seasonally varying data availability and residual snow- or freeze–thaw-related effects may affect VOD at high latitudes. The annually varying GS analysis will evaluate whether interannual shifts in growing-season timing materially affect the trend and IAV relationships.

In addition, following the reviewer's comment on the reference period used to define the GS, we will recalculate the fixed GS definition using monthly air temperature and NDVI during the 2000–2021 study period and update the GS-dependent analyses accordingly. In the revised manuscript, the resulting common fixed-GS comparison will be used as the primary temporally matched analysis for NDVI, VOD, and VISIT GPP. The original indicator-specific temporal definitions, alternative VOD windows, annual-mean comparison, and annually varying GS will be presented as sensitivity analyses.

We will revise the Methods and Results accordingly so that differences attributable to temporal processing can be distinguished explicitly from disagreements that persist after temporal matching. Importantly, disagreement remaining after temporal matching will not be interpreted by itself as evidence of intrinsic ecological divergence, because observational and model uncertainties may also contribute.

2

Statistical analysis requires more attention. The coefficient of determination is not an adequate measure of agreement because squaring the correlation removes its sign. A strong negative relationship can therefore appear as high agreement. The maps and regional summaries should report signed correlations and significance based on an effective sample size that accounts for temporal autocorrelation. In addition, the threshold $R2 < 0.3$ is arbitrary and needs justification, and the large number of pixel-wise tests requires multiple-testing treatment. Similar concerns apply to the trend-agreement maps. The three categories in Figs. 3 and 6 sum to 100%, but there is no category for non-significant trends or for cases in which only one product has a significant trend. Classifying the signs of very small, non-significant Sen slopes as greening, browning, or disagreement can greatly inflate the reported percentages. The authors need to distinguish significant positive agreement, significant negative agreement, opposite significant trends, mixed-significance cases, and no significant trend, while accounting for serial correlation in the Mann-Kendall tests. Indeed, trend magnitudes from NDVI, VOD, and GPP also cannot be compared directly because they have different units and scales (Fig. 1). Standardized or relative trends would be needed for such statements.

Response:

We agree and have revised the statistical framework accordingly. For IAV, we now use the signed Pearson correlation coefficient (r) rather than r^2, so that negative relationships are not interpreted as agreement. Statistical significance is evaluated using an effective sample size based on the lag-1 autocorrelation of the two detrended annual series. The arbitrary $r^2 < 0.3$ threshold will therefore be removed from the agreement classification.

For long-term trends, we retain Sen's slope and use the Hamed–Rao modified Mann-Kendall test to account for serial correlation. We will also replace the original three-category trend-agreement maps with significance-aware categories that distinguish: both indicators significantly increasing, both significantly decreasing, significant trends with opposite signs, mixed-significance cases, and cases where neither trend is significant. Cross-indicator comparisons will focus on trend direction and significance rather than directly comparing native-unit trend magnitudes; where magnitude comparisons are needed, standardized or relative measures will be used. In addition, we examined how the revised significance-aware trend classification changed when the temporal window was matched across indicators (Fig. R3).

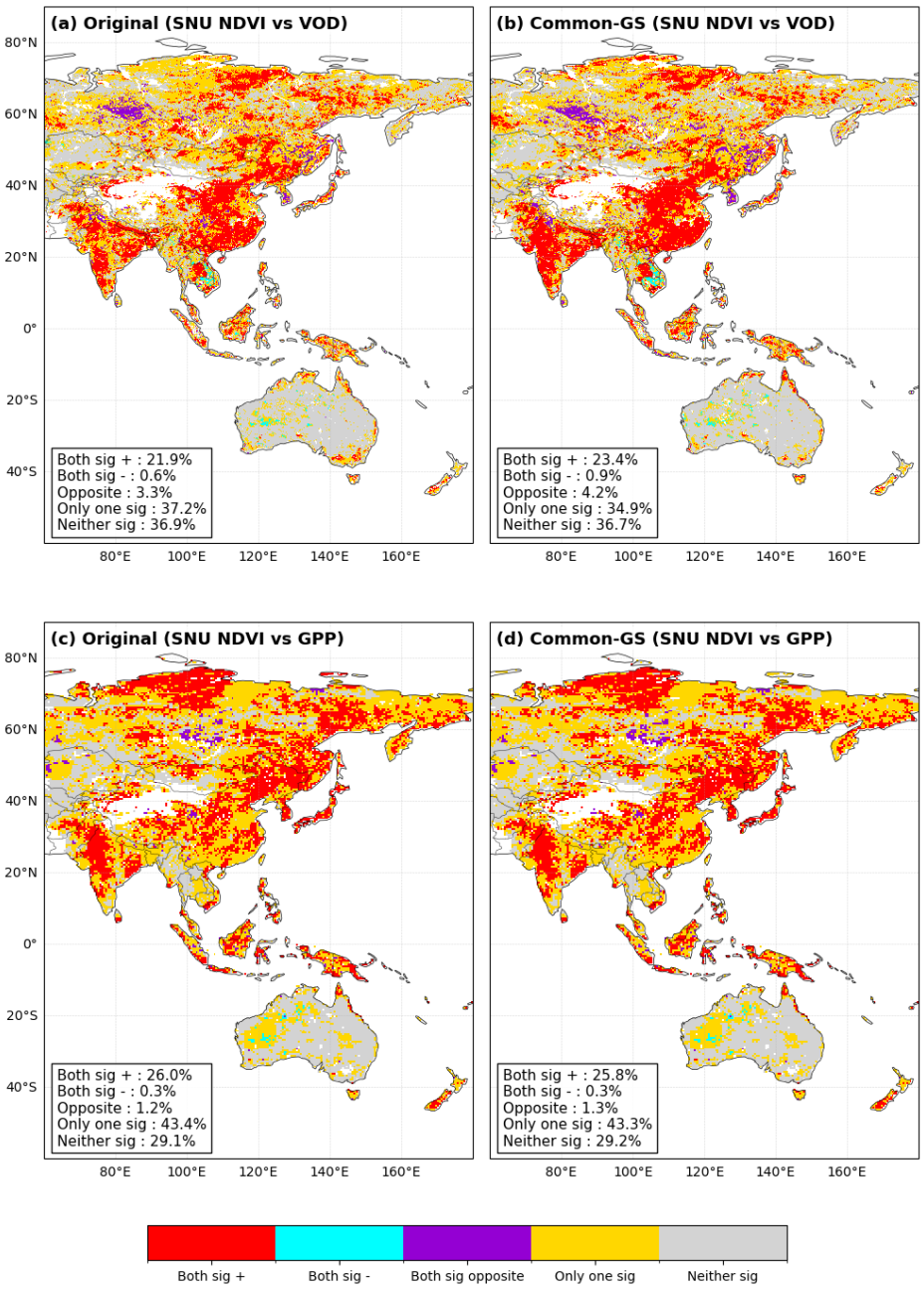


Figure R3. Sensitivity of the revised significance-aware trend classifications to temporal-window matching during 2000-2021.

Trend classifications are shown for SNU NDVI versus VOD using (a) the original temporal treatment and (b) the common-GS treatment, and for SNU NDVI versus VISIT GPP using (c) the original temporal treatment and (d) the common-GS treatment. Trend direction was determined using Sen's slope, and statistical significance was evaluated using the Hamed–Rao modified Mann-Kendall test to account for serial correlation. Pixels were classified as both significantly increasing, both significantly decreasing, significantly trending in opposite directions, significant in only one indicator, or non-significant in both indicators. Percentages indicate the cosine-latitude area-weighted fraction of the valid analysis domain assigned to each category.

We also agree that the large number of pixel-wise tests raises a multiple-testing issue. In the revised manuscript, we will retain the uncorrected results for transparency and apply a Benjamini–Hochberg false discovery rate (FDR) correction to the pixel-wise IAV-correlation and trend-significance tests as a sensitivity analysis. We will use this analysis to assess whether the major spatial conclusions remain robust after accounting for multiple testing.

Because the fixed GS definition will also be recalculated using the 2000–2021 study period as described in our response to Major Comment 1, the corresponding GS-dependent trend classifications will be updated in the revised manuscript.

3

The conclusions about modeled GPP and CO2 attribution are not adequately supported. Only one configuration of one ecosystem model is evaluated, so the results should not be generalized to model-based GPP. More information is needed on the VISIT configuration, spin-up, vegetation and land-use representation, parameterization, historical climate forcing, and so on. The sensitivity experiments are also unclear. The Methods define ALLCON as the simulation in which all climate drivers vary, whereas Fig. 7 additionally includes a CLIMATE component described as the total climate contribution. Moreover, Fig. 7 displays detrended anomalies, which cannot demonstrate that CO2 is the primary driver of a long-term trend. The authors need to present undetrended component trends, define a consistent factorial or difference-based decomposition including interaction terms, and compare the inferred contributions quantitatively. Validation against independent GPP constraints or comparison with a model ensemble would also be needed to determine whether the discrepancies are specific to VISIT or robust across GPP estimates.

Response:

We agree that the original description and interpretation of the VISIT sensitivity experiments were unclear and, in part, inconsistent. We also agree that results obtained from a single configuration of one ecosystem model cannot be generalized to model-based GPP as a whole, and that the driver-sensitivity experiments should not be interpreted as an independent attribution of real-world ecosystem responses.

After re-examining the experiment setup, we found that the definition of ALLCON in the submitted manuscript was incorrect. ALLCON denotes the baseline experiment in which the climate drivers and atmospheric CO_2 are held at their 1900 values. In the PRE, TEMP, VAPOR, and CLOUD experiments, only the corresponding climate driver varies through time, while the other climate drivers and atmospheric CO_2 remain fixed. In the CO_2 experiment, atmospheric CO_2 varies through time while climate is held fixed. CLIMATE represents the experiment in which the climate drivers vary jointly while atmospheric CO_2 remains fixed. Land use changes over time in ALLCON and all sensitivity experiments, following the same prescribed land-use history. We will revise the Methods so that the forcing varied in each experiment, the variables held fixed, and the reference condition are described explicitly and consistently.

We also re-examined the regional GPP responses without detrending, as suggested by the reviewer. For each sensitivity experiment, we calculated the undetrended difference from the ALLCON baseline and quantified its long-term trend using Sen's slope, with statistical significance evaluated using the Hamed–Rao modified Mann–Kendall test. This approach allows the long-term response to each prescribed forcing experiment to be evaluated relative to the same reference simulation, rather than inferring long-term contributions from detrended anomalies.

The new analysis changes our interpretation of the driver-sensitivity results. The CO_2-minus-ALLCON response showed a significant positive trend in all six regions examined, whereas the CLIMATE-minus-ALLCON response was significantly positive in Siberia and East Asia and was not significant in the other regions. However, the number of regions showing a statistically significant response does not by itself establish the relative magnitude or dominance of individual forcings. Moreover, the available experiments do not constitute a complete factorial decomposition. We will therefore remove the statement that rising CO_2 is uniformly the primary driver of long-term GPP increases. Instead, the results will be described as sensitivities of VISIT-simulated GPP to the prescribed forcing experiments. The climate-related responses also differed substantially among regions, indicating that the simulated response to climatic forcing is spatially heterogeneous within VISIT. The undetrended regional responses are summarized in Fig. R4.

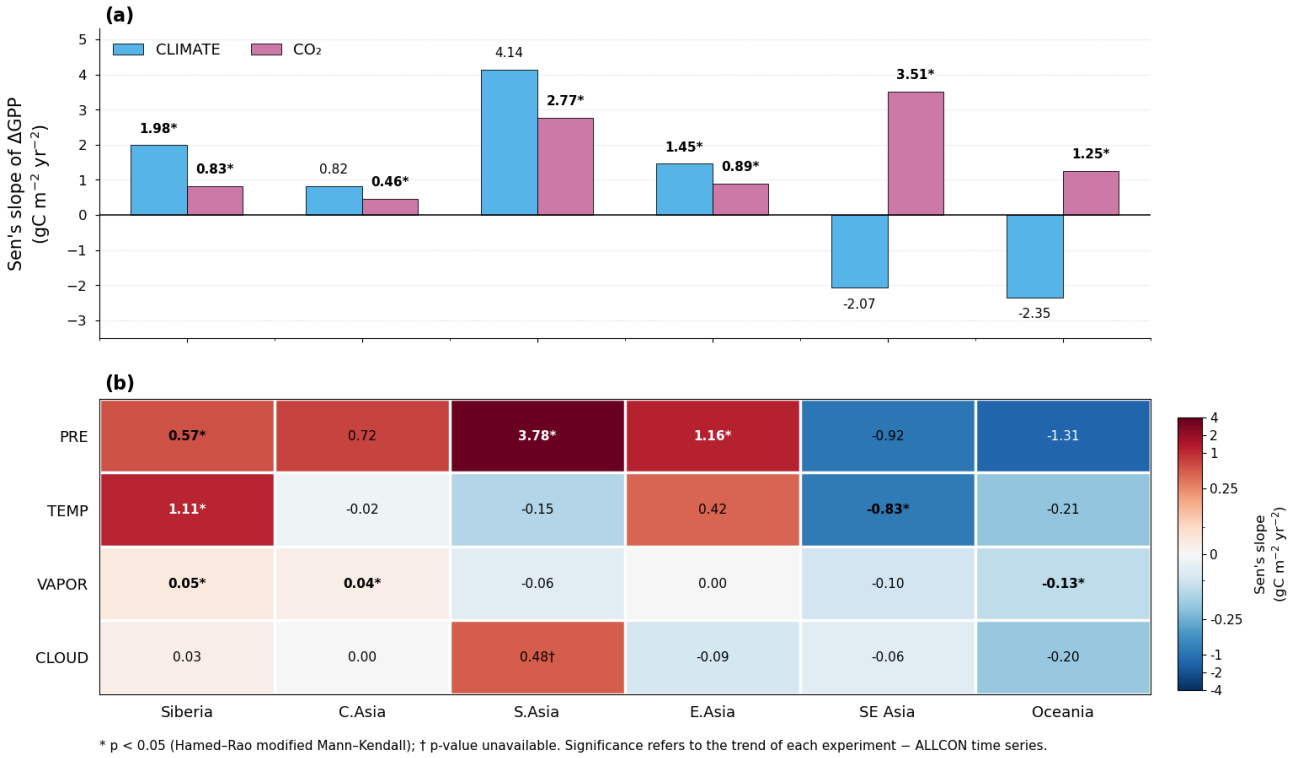


Figure R4. Long-term trends in regional VISIT GPP responses from the driver-sensitivity experiments during 2000–2021.

Sen's slopes were calculated from the undetrended difference between each sensitivity experiment and the ALLCON baseline. (a) Regional trends for the CLIMATE and CO_2 experiments. (b) Regional trends for the individual climate-driver experiments (PRE, TEMP, VAPOR, and CLOUD). Statistical significance was evaluated using the Hamed–Rao modified Mann–Kendall test. These results are interpreted as sensitivities within the VISIT model framework rather than as a complete causal attribution because the available experiments do not allow interaction effects among drivers to be uniquely separated.

Because the available experiments do not form a complete factorial ensemble, interactions among drivers cannot be uniquely separated. We will therefore avoid describing the analysis as a complete causal attribution. We will also expand the Methods to describe the VISIT configuration, spin-up, vegetation and land-use representation, parameterization, and forcing data more explicitly. In the Discussion, we will clarify that both the magnitude and the inferred hierarchy of the simulated responses depend on the structure and parameterization of VISIT and could differ in another ecosystem model.

In addition, following the reviewer's suggestion, we conducted a site-scale validation of VISIT GPP against eddy-covariance-based GPP from long-term Flux Shuttle records across Asia–Oceania as an independent

observational constraint. We evaluated both interannual variability and long-term trends. IAV was assessed using standardized and linearly detrended annual GPP series, with correlation significance adjusted for lag-1 temporal autocorrelation. Long-term trends were evaluated using Sen's slope and the Hamed–Rao modified Mann–Kendall test.

The site-scale comparison showed substantial variability among sites and generally limited agreement in interannual variability between VISIT and tower-derived GPP. Long-term trend agreement was also site dependent, with some sites showing the same trend direction and others differing in sign and/or magnitude. We therefore do not regard the flux comparison as demonstrating that VISIT consistently reproduces observed tower-scale GPP variability. Rather, the results indicate that uncertainty in modeled GPP must be considered when interpreting discrepancies between VISIT GPP and the satellite indicators.

We also examined the geographical and ecosystem coverage of the validation, particularly for the regions highlighted by the reviewer. For northern Eurasia, we identified a Russian site with a sufficiently long and usable GPP record and will include it in the revised validation. For Southeast Asia, we examined the available candidate records but did not identify a site with a sufficiently complete long-term GPP record that satisfied our selection criteria. We therefore cannot directly evaluate VISIT performance in Southeast Asia using the available long-term flux records. As an additional tropical evergreen broadleaf forest reference, we retain AU-Cow in northern Australia, for which a sufficiently long daytime-partitioned (DT_VUT) GPP record is available. AU-Cow is not treated as a geographical substitute for Southeast Asia, but it provides an additional evaluation in a tropical evergreen broadleaf forest ecosystem.

The availability and completeness of daytime- and nighttime-partitioned GPP differ among sites. We are therefore adopting a hierarchical site-selection approach rather than requiring both DT_VUT and NT_VUT to be available for every site. We prioritize sites for which both partitioning estimates provide sufficiently long and complete records, while retaining a site when only one partitioning estimate satisfies the data-availability criterion. The partitioning method used for each site will be documented explicitly. Where both estimates are available, their consistency will also be evaluated so that uncertainty related to NEE partitioning is represented in the validation.

The current IAV comparisons and site locations are summarized in Fig. R5. The site set will be updated in the revised manuscript to reflect the selection described above. Site-specific information, including the observation period, partitioning method, Flux and VISIT Sen's slopes, and their statistical significance, will be provided in the Supplement so that the long-term trend comparison can be evaluated explicitly.

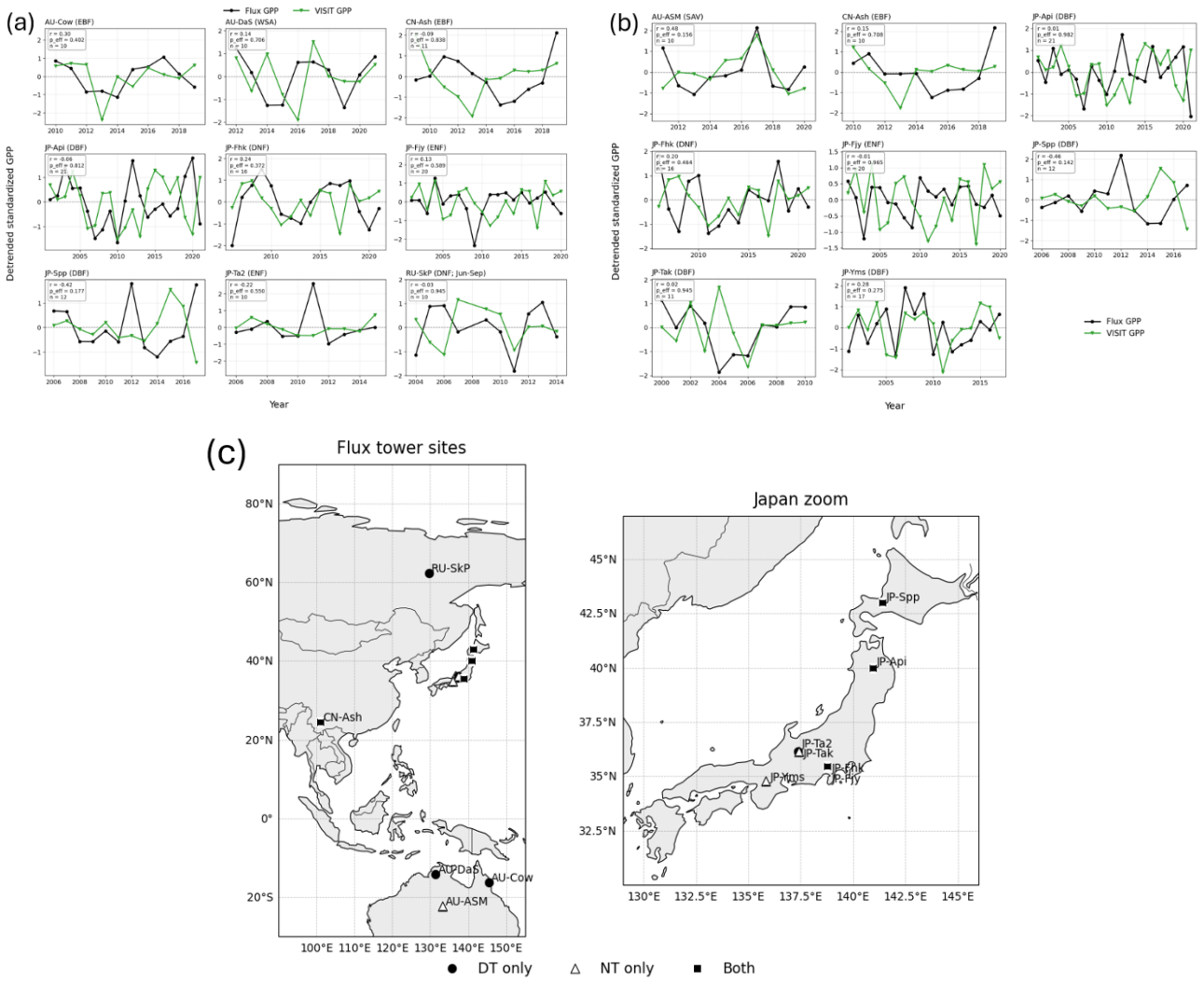


Figure R5. Site-scale comparison of VISIT GPP with eddy-covariance-based GPP from Flux Shuttle sites across Asia–Oceania.

(a, b) Standardized and linearly detrended annual GPP time series from VISIT and flux-tower estimates obtained using daytime (DT_VUT) and nighttime (NT_VUT) partitioning, respectively. Site-level Pearson correlation coefficients and p-values adjusted for lag-1 temporal autocorrelation are shown for the available long-term records. (c) Spatial distribution of the sites included in the current validation. Site-specific long-term trend estimates and their statistical significance will be provided in the Supplement, and the site set will be updated in the revised manuscript following the selection criteria described above.

This validation has important limitations. The footprint represented by an eddy-covariance tower is much smaller than the 0.5° VISIT grid, so spatial representativeness may contribute to differences between tower-derived and modeled GPP. In addition, flux-tower GPP is not directly observed but is estimated by partitioning measured NEE, introducing uncertainty associated with the partitioning method. The comparison therefore

cannot uniquely attribute VISIT–Flux disagreement to model structure. We will document the Flux Shuttle data source, site-selection and annual-data criteria, and the partitioning treatment used in the validation so that the data-quality screening is transparent.

Taken together, the revised sensitivity analysis and the independent flux comparison lead us to a more cautious interpretation of the GPP-related results. We will no longer interpret disagreement between satellite indicators and VISIT GPP primarily as physiological divergence, nor will we generalize the inferred driver hierarchy beyond the VISIT framework. We will revise the Discussion and Conclusions to state that the observed discrepancies may reflect a combination of differences in ecological sensitivity, satellite-observation and processing uncertainty, spatial-representativeness effects, NEE-partitioning uncertainty, and VISIT model uncertainty, whose relative contributions cannot be fully separated in the present study.

4

The interpretation of indicator disagreement is too categorical. NDVI, VOD, and GPP measure different variables, so low correlation among them is partly expected and does not by itself demonstrate ecological decoupling, dataset error, or meaningful complementary information. The manuscript repeatedly attributes regional differences to NDVI saturation, VOD contamination by soil moisture and freeze-thaw processes, or missing processes in VISIT, but these explanations are not fully support the results.

Response:

We agree that our interpretation was too categorical. NDVI, VOD, and GPP represent different, although related, vegetation properties and processes, so some degree of disagreement is expected. Low correlation alone does not demonstrate ecological decoupling, errors in individual datasets, or that the indicators necessarily provide complementary ecological information.

In the revision, we will distinguish clearly between observed statistical disagreement and possible mechanisms that could contribute to it. NDVI saturation, soil-moisture and freeze–thaw effects on VOD, and processes not represented in VISIT will be discussed as possible explanations or limitations rather than as demonstrated causes of the regional patterns. Where relevant, we will cite supporting literature while making clear that these mechanisms have not been directly tested in this study.

The additional analyses support this more cautious interpretation. Matching temporal windows changes the degree of agreement, particularly for NDVI–VOD, demonstrating that processing choices contribute to the reported discrepancies. However, disagreement that persists after temporal matching cannot, by itself, be attributed to ecological differences. The limited agreement between VISIT and tower-derived GPP also indicates that model-related uncertainty must be considered, although this comparison does not identify the causes of the regional discrepancies.

We will revise the Abstract, Discussion, and Conclusions accordingly, including removing the categorical statement that the inconsistencies represent "scientifically meaningful signals rather than analytical artifacts." We will frame the study as quantifying indicator-dependent differences and their sensitivity to methodological choices, while acknowledging that the relative contributions of ecological differences, observational uncertainty, and model limitations remain unresolved.

Minor comments:

L15-18: It is up to the band selection. I think it is difficult to interpret Ku-, C-, or X-band VOD consistently. How do they represent the proxy of biomass?

Response:

We agree that the VODCA CXKu product, which merges VOD retrievals from C-, X-, and Ku-band observations, should not be described as a direct or uniform proxy for biomass. As explained by Zotta et al. (2024), already cited in our manuscript, this product is primarily sensitive to vegetation water-content dynamics in the upper canopy. Merging the retrievals does not imply that the three frequency bands have identical physical sensitivities. Vegetation water content depends on both the amount of vegetation and its hydration state; therefore, changes in CXKu VOD cannot be attributed uniquely to changes in biomass. The relationship also depends on microwave frequency, vegetation type, and canopy density, particularly because these bands have limited penetration into dense canopies.

We will revise the Abstract at L15–18 to describe the results as agreement or disagreement among NDVI, VOD, and model-simulated GPP, rather than as direct evidence of corresponding changes in greenness, biomass, and productivity. We will also clarify the interpretation of VODCA CXKu in the Methods and Discussion and avoid treating it as a consistent measure of total woody biomass across ecosystems.

L39: Please cite some references.

Response:

Thank you for pointing this out. We will add references supporting both the optical basis of NDVI and the limitation of interpreting greenness changes as changes in photosynthetic activity. Specifically, Tucker (1979) and Myneni et al. (1995) support the description of NDVI, while Walther et al. (2016) and Hu et al. (2022) demonstrate that greenness and photosynthetic activity can exhibit different seasonal and interannual dynamics. Myneni et al. (1995) is already cited elsewhere in the manuscript and will also be cited explicitly here.

To make the statement more specific and align it with the supporting evidence, we will replace the broad reference to "other aspects of ecosystem change" with a reference to photosynthetic activity. The revised text will read:

"However, NDVI primarily reflects canopy greenness and optical properties (Tucker, 1979; Myneni et al., 1995). Changes in vegetation greenness do not necessarily correspond to changes in photosynthetic activity, as demonstrated by differences in their seasonal and interannual dynamics (Walther et al., 2016; Hu et al., 2022)."

L44-45: Yes, but it is up to the purpose of the study. For instance, if we focus on the canopy greenness, then there is not much problem. I think "vegetation dynamics" is bit unclear here.

Response:

We agree that the suitability of NDVI depends on the purpose of the analysis. NDVI is an appropriate indicator for monitoring canopy greenness, and our intention was not to question its usefulness for that purpose. The limitation concerns interpreting greenness changes alone as evidence of changes in other ecosystem processes, such as photosynthetic activity.

We will therefore replace the broad term "vegetation dynamics" with the specific vegetation property or process being discussed. In coordination with the revision described in our response to the preceding comment, the relevant text will clarify that NDVI is well suited to monitoring canopy greenness, while changes in greenness do not necessarily track seasonal or interannual changes in photosynthetic activity.

L15-20: The abstract interprets agreement as coherent changes in greenness, biomass, and productivity. This should be moderated because VOD is not a direct biomass measurement and GPP is not independently observed here.

Response:

We agree and will moderate this interpretation in the Abstract. Agreement among the indicators should not be presented as direct evidence of coherent changes in greenness, biomass, and productivity, because VOD is not a direct biomass measurement and VISIT GPP is a model estimate rather than an independent observation.

We will therefore revise the Abstract to describe agreement among the indicators without inferring corresponding changes in biomass or ecosystem productivity, and we will explicitly identify GPP as simulated by VISIT. We will also review the corresponding statements in the Results, Discussion, and Conclusions to ensure consistency with the revised Abstract. In the Results, trends and agreement will be described in terms of the measured or simulated indicators, without directly equating VOD changes with biomass changes or treating VISIT-simulated GPP as observed productivity. The ecological implications and associated limitations will be addressed in the Discussion, and the Conclusions will reflect the same qualifications.

L58: Here, the authors clarified VOD as a proxy of water status, but it represents a biomass in this study.

Response:

Thank you for pointing out this inconsistency. We agree that vegetation water status and biomass should not be used interchangeably when interpreting VOD. As described by Zotta et al. (2024), already cited in our manuscript, VODCA CXKu is primarily sensitive to vegetation water-content dynamics in the upper canopy. These dynamics can reflect changes in both vegetation amount and hydration state, so VOD changes cannot be attributed uniquely to biomass changes.

We will revise the statement at L58 and use this interpretation consistently throughout the manuscript. Specifically, we will refer to vegetation water content rather than using “water status” and “biomass” interchangeably, and clarify that the relationship between VOD and biomass depends on microwave frequency, vegetation type, and canopy density.

L58-60: Do NDVI, VOD, and GPP each clearly represent structural, hydrological, and functional vegetation changes, respectively? This correspondence may be overly simplistic or misleading. Please revise the sentence.

Response:

We agree that the original one-to-one correspondence was overly simplistic. The indicators have different but overlapping sensitivities and cannot be assigned exclusively to structural, hydrological, and functional changes. We will revise the sentence to describe more precisely what each indicator represents in this study:

"NDVI reflects canopy greenness and optical properties, VODCA CXKu is sensitive to vegetation water content, particularly in the upper canopy, and VISIT-simulated GPP represents model-estimated carbon uptake through photosynthesis. Their trends and interannual variability may therefore differ, but these differences cannot be assigned directly to distinct categories of ecosystem change."

This wording is consistent with the revised interpretation of VOD described in our responses to the preceding comments and explicitly identifies GPP as a model estimate.

L63-64: I do not agree with the statement that only a few studies have investigated this topic. Please provide appropriate references to support this claim. More generally, some of introduction sentences are presented without citations. Relevant references should be added where needed.

Response:

We agree that the broad statement that only a few studies have investigated this topic was insufficiently supported. Previous studies have examined different relationships among vegetation greenness, photosynthetic activity, and microwave-derived vegetation properties. For example, Walther et al. (2016) compared the seasonal dynamics of EVI and solar-induced chlorophyll fluorescence in boreal evergreen forests, while Hu et al. (2022) investigated the relationship between interannual variations in LAI and GPP. Dou et al. (2023), already cited in our manuscript, evaluated the reliability of VOD for estimating decadal and interannual carbon dynamics, including the effects of temporal aggregation.

We will remove the unsupported statement and revise the Introduction to acknowledge these studies explicitly. Building on this work, we will define our specific focus as comparing long-term trends and interannual variability among NDVI, VODCA CXKu, and VISIT-simulated GPP across Asia–Oceania during 2000–2021, while assessing the sensitivity of these comparisons to temporal aggregation. We will present this as the scope and contribution of our study without implying that comparisons among vegetation indicators are generally lacking.

We will also review the Introduction for statements requiring additional support and add appropriate references, including those addressing the interpretation and limitations of NDVI and VOD discussed in our responses to the preceding comments.

L95-110: The authors select the long-term datasets which cover 1980s-present but only use 2000-2021. Could you explain it?

Response:

The analysis period of 2000–2021 was selected as the common temporal overlap across the full set of datasets evaluated in this study. The MODIS record begins in 2000, while the VODCA CXKu record used here ends in 2021. Although SNU NDVI and VISIT GPP are available for longer periods, we use the same period for the cross-indicator analyses so that long-term trends and interannual variability are compared consistently. MODIS NDVI will be retained as an independent robustness check, as described below.

We will clarify this rationale in Sect. 2.2 and distinguish the full temporal coverage of each dataset from the common period used in our analyses.

L103-105: What is harmonized C,X,Ku-band VOD? Also, VODCA provides its own GPP product. It is worth testing it too.

Response:

Thank you for these suggestions. First, "harmonized C/X/Ku-band VOD" refers to the VODCA CXKu product, which combines VOD retrievals from multiple sensors operating in the C-, X-, and Ku-bands. As described by Zotta et al. (2024), the retrievals are statistically rescaled using cumulative distribution function (CDF) matching, with AMSR-E X-band observations serving as the reference, and the scaled overlapping observations are subsequently merged using weighting. This processing reduces systematic differences among the input records but does not make the physical sensitivities of the different frequency bands identical. We will explain this in Sect. 2.2 and use the product name "VODCA CXKu" consistently.

Second, we agree that VODCA2GPP provides a useful additional comparison. VODCA2GPP uses information derived from the merged VODCA CXKu record together with ERA5-Land 2 m air temperature and is calibrated against FLUXNET2015 GPP estimates (Wild et al., 2022). It therefore provides a GPP estimate based on a different approach from VISIT, although it is not independent of the VOD information used in our study.

We will include VODCA2GPP as an additional GPP product-sensitivity analysis over the common period 2000–2020. After harmonizing the temporal aggregation and spatial grid, we will compare its trend and IAV relationships with NDVI and VOD against those obtained using VISIT GPP. The detailed results will be presented in the Supplement and summarized in the main text. Because VODCA2GPP uses VOD-based predictors, agreement between VOD and VODCA2GPP will not be treated as an independent validation of VOD.

L105-106: Please explain why the VISIT model was selected for this analysis. The choice appears somewhat selective, and the criteria for model selection are unclear.

Response:

VISIT was selected because we had access to an established model configuration that covers the study period and can be rerun under controlled one-driver perturbations. This allows us to compare simulated GPP with satellite-based indicators while examining GPP responses to individual meteorological drivers—precipitation, temperature, vapor pressure, and cloud cover—as well as atmospheric CO_2 within a consistent model framework. The ability to directly operate and modify this established VISIT setup was therefore important for the specific sensitivity experiments required in this study.

Coordinated model intercomparison frameworks such as TRENDY provide factorial simulations designed to separate the effects of major forcing categories such as atmospheric CO_2, climate, and land-use change. However, the standard experiments available to us do not provide the specific one-at-a-time meteorological perturbations required for the present analysis. Our VISIT setup allowed these experiments to be conducted directly.

We will clarify these model-selection criteria in Sect. 2.2. We acknowledge that using a single model configuration does not characterize uncertainty across ecosystem models. We will therefore limit our conclusions to the VISIT simulations examined here and explicitly discuss the model dependence of the inferred responses.

L125-127: It is not matched with study period.

Response:

We agree that the reference period used to define the GS should be consistent with the study period. We will therefore recalculate the fixed GS definition using the monthly mean air temperature and NDVI for 2000–2021, with the same thresholds as in the original analysis. The GS will be defined for each grid cell and calendar month and then applied consistently across all years.

We will update all analyses that depend on the fixed GS definition and revise the Methods and Results accordingly. This will ensure that the climatological reference period used to define the GS is consistent with the 2000–2021 period used for the trend and IAV analyses.

L125-138: I think this data processing strategy should be carefully improved. The current method is not consistent.

Response:

We agree that the original temporal aggregation was inconsistent across the indicators. As described in our response to Major Comment 1, we have conducted sensitivity analyses using matched temporal windows and alternative VOD aggregation methods. These analyses show that temporal aggregation affects the degree of agreement, particularly between NDVI and VOD.

In the revised analysis, we will use a common fixed GS definition for NDVI, VOD, and VISIT GPP as the primary temporal aggregation approach. As explained in our response to the preceding comment, we will recalculate this GS definition using monthly mean air temperature and NDVI for 2000–2021 and update the affected analyses. The original single-month VOD sampling, the three-month mean centered on the NDVI peak month, the maximum VOD within the GS, and other alternative temporal definitions will be retained as sensitivity analyses.

We will reorganize Sect. 2.3 to distinguish the common-GS approach used in the main comparisons from the alternative aggregation methods used in the sensitivity analyses. This will allow us to assess how strongly the reported agreement and disagreement depend on temporal processing choices.

L139: I think they are inconsistent.

Response:

We agree. The statement that the masking procedures ensured consistency across all analyses was not fully accurate because the original temporal aggregation differed among NDVI, VOD, and GPP. We will clarify that spatial resampling and land masking were harmonized across the comparisons, whereas temporal aggregation was indicator-specific in the original analysis.

In the revised analysis, temporal aggregation will also be harmonized by applying the common fixed GS definition based on 2000–2021 to all three indicators. Alternative temporal aggregation methods will be described separately as sensitivity analyses.

Figure. 1: Please select a single NDVI product for the main analysis.

Response:

We agree that the role of the two NDVI products should be clarified so that the cross-indicator comparison is not interpreted as an inter-product benchmarking exercise. In the revised analysis, SNU NDVI will be used as the primary NDVI product in Fig. 1 and in the main cross-indicator analyses. SNU provides the long-term NDVI record used in the GS-based analyses and is also the NDVI product used in the temporal-sensitivity analyses presented here.

MODIS NDVI will be retained as an independent robustness check to assess whether the main NDVI-based findings are strongly dependent on the choice of NDVI product. We will clearly distinguish these roles in the Methods and Results.

L219-221: Assessing the consistency of IAV or trends among different products is not the primary goal of this study. If the authors intend to assess consistency across products, the same analysis should also be conducted using multiple VOD and GPP products.

Response:

We agree that cross-product consistency is not the primary objective of this study. We will revise the relevant statements so that the main objective is the comparison among indicator types based on different observation and modeling principles. SNU NDVI will be used in the main analysis, while MODIS NDVI will be retained as a supplementary sensitivity check to assess whether the main NDVI-based findings depend strongly on the choice of NDVI product.

We will also include VODCA2GPP as an additional GPP product-sensitivity test, as described in our response to the comment on L103–105. However, because VODCA2GPP uses VOD-based predictors, its agreement with VOD will not be treated as an independent validation. These supplementary comparisons help assess product dependence but do not constitute a comprehensive comparison of multiple NDVI, VOD, and GPP products.

A directly comparable alternative VOD product covering the full 2000–2021 study period is not included in the present analysis. We will explicitly acknowledge this limitation and restrict our conclusions to the datasets examined here.

Section 3.1-3.4: Because NDVI, VOD, and GPP were processed using different temporal definitions, the reported differences may reflect processing choices rather than actual differences among the indicators.

Response:

We agree. The additional common-GS and VOD-window sensitivity analyses show that some of the differences reported in Sects. 3.1–3.4 are sensitive to temporal processing, particularly for NDVI–VOD IAV, whereas other differences persist after temporal matching.

We will therefore revise these Results sections to present the common-GS comparisons as the primary results, using the fixed GS definition recalculated for 2000–2021. The original indicator-specific processing and alternative VOD windows will be presented as sensitivity analyses. We will distinguish findings that change with temporal aggregation from those that remain consistent across the tested methods.

We will also clarify that disagreement remaining after temporal matching does not, by itself, demonstrate an intrinsic ecological difference, because observational uncertainties and model limitations may still contribute.