Responses to Reviewer #2's Comments

Overall Assessment:

This manuscript presents a comprehensive comparison of long-term vegetation trends and interannual variability (IAV) across the Asia-Oceania region using NDVI, VOD, and VISIT-simulated GPP. The study provides valuable insights into how different observation principles capture divergent aspects of ecosystem dynamics. The manuscript is generally well-written and logically structured. I recommend a minor revision to address a few methodological caveats and expand the discussion, which will further strengthen the robustness and impact of the study.

Response:

We thank Reviewer #2 for the positive assessment and constructive suggestions. We have addressed the main methodological points by examining temporal-window sensitivity, validating VISIT GPP against long-term eddy-covariance-based GPP, and revisiting the interpretation of the VISIT driver-sensitivity experiments. We will also expand the discussion of VOD band dependence and model structural uncertainty in the revised manuscript.

Specific Comments:

1. Influence of VOD Microwave Bands on Discrepancies:

The authors utilized the VODCA CXKu dataset, which integrates C-, X-, and Ku-bands. As is well known in microwave remote sensing, signal penetration depth depends heavily on canopy structure and the specific microwave frequency. The C, X, and Ku bands have relatively short wavelengths, meaning their penetration into dense canopies is limited compared to lower frequencies. Therefore, the discrepancy observed between NDVI and VOD, or between VOD and GPP—particularly in high-biomass regions like tropical Southeast Asia—might be partially driven by the limited penetration depth of these specific bands, which may capture signals mostly from the upper canopy rather than the entire woody biomass. I suggest discussing how these band-specific limitations might translate to the observed discrepancies. Furthermore, it would be highly beneficial to discuss whether employing an L-band VOD

product (e.g., VODCA-L, or similar products from SMOS/SMAP), which has deeper canopy penetration, might yield different trend and IAV alignments.

Response:

We thank the reviewer for highlighting the importance of microwave-frequency dependence in interpreting the discrepancies involving VOD. We agree that the frequency-dependent sensitivity of VODCA CXKu should be considered more explicitly, particularly in densely vegetated regions.

The C-, X-, and Ku-band observations contributing to VODCA CXKu generally have shallower canopy penetration than longer-wavelength L-band observations. As described by Zotta et al. (2024), VODCA CXKu is primarily sensitive to vegetation water-content dynamics in the upper canopy, whereas L-band observations are theoretically more sensitive to a larger fraction of the canopy, including woody components such as branches and trunks. Therefore, particularly in high-biomass forests such as those in tropical Southeast Asia, CXKu VOD should not be interpreted as a direct measure of total woody biomass. Differences in the vegetation components represented by CXKu VOD, canopy greenness represented by NDVI, and carbon uptake represented by GPP may therefore contribute to the observed discrepancies. However, the present analysis does not allow us to isolate the contribution of microwave frequency from other sources of disagreement.

We also agree that comparison with L-band VOD would be valuable for evaluating the dependence of our results on microwave frequency. Because of its deeper canopy penetration and greater relative sensitivity to woody components, L-band VOD may show different trend and IAV relationships with NDVI and GPP from those obtained using CXKu VOD. However, closer agreement is not necessarily expected, because vegetation water content and woody structure, canopy greenness, and ecosystem carbon uptake need not vary synchronously. The

VODCA L product described by Zotta et al. (2024) covers 2010–2021, so an equivalent comparison would require a substantially shorter common analysis period than the 2000–2021 period used in this study.

In the revised manuscript, we will expand the Discussion to explain how the frequency-dependent sensitivity and limited canopy penetration of the C-, X-, and Ku-band observations may affect the interpretation of the VOD-related discrepancies, particularly in dense forests. We will also clarify that our present conclusions regarding VOD apply specifically to VODCA CXKu and should not be automatically generalized to L-band VOD. Comparison with L-band products will be discussed as an important direction for evaluating the robustness of the observed trend and IAV relationships across microwave frequencies.

Reference:

Zotta, R.-M., Moesinger, L., van der Schalie, R., Vreugdenhil, M., Preimesberger, W., Frederikse, T., de Jeu, R., and Dorigo, W.: VODCA v2: multi-sensor, multi-frequency vegetation optical depth data for long-term canopy dynamics and biomass monitoring, Earth Syst. Sci. Data, 16, 4573–4617, https://doi.org/10.5194/essd-16-4573-2024, 2024.

2. GPP Validation Using Flux Tower Data:

The study relies on GPP derived from the process-based terrestrial ecosystem model VISIT. Given the abundance of eddy covariance flux towers that have accumulated long-term GPP data across the Asia-Oceania region (e.g., AsiaFlux, OzFlux), it is now highly feasible to verify the trend and IAV of the VISIT-simulated GPP against in-situ observations. I strongly encourage the authors to perform such a validation. This would clarify whether the regional discrepancies observed between satellite indicators and GPP (e.g., in Siberia and Southeast Asia) are genuine physiological divergences or artifacts stemming from the model's structural uncertainties.

Response:

We thank the reviewer for this constructive suggestion. We agree that validation against eddy-covariance-based GPP is important for assessing whether discrepancies involving VISIT GPP may partly reflect model uncertainty. Following this suggestion, we conducted a site-scale comparison of VISIT GPP with eddy-covariance-based GPP from long-term Flux Shuttle records across Asia–Oceania.

We evaluated both interannual variability and long-term trends. Interannual variability 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. Sites were selected based on the availability of sufficiently long and complete annual GPP records overlapping 2000–2021.

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 this comparison as demonstrating that VISIT consistently reproduces tower-scale GPP variability. Rather, the results indicate that model-related uncertainty must also be considered when interpreting discrepancies between VISIT GPP and the satellite indicators.

We also examined the geographical and ecosystem coverage of the validation because the reviewer specifically highlighted Siberia and Southeast Asia. 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 Flux Shuttle 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 reference for a tropical evergreen broadleaf forest ecosystem, 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.

The availability and completeness of daytime- and nighttime-partitioned GPP differ among sites. We therefore adopt a hierarchical site-selection approach rather than requiring both DT_VUT and NT_VUT to be available at 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 examined so that uncertainty related to NEE partitioning is represented in the validation.

The current IAV comparisons and site locations are summarized in Fig. R1. 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.

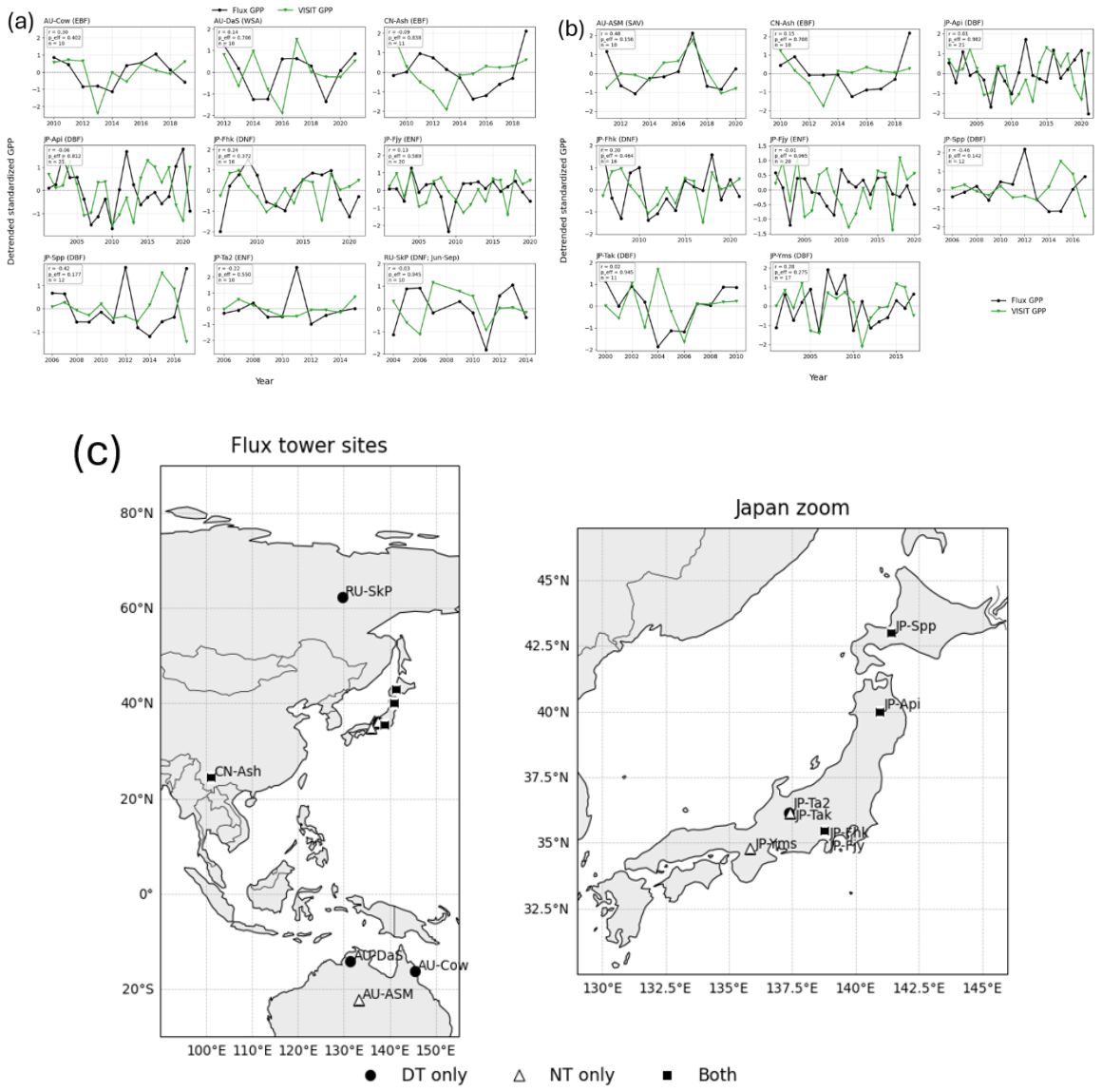


Figure R1. Site-scale comparison of VISIT GPP with eddy-covariance-based GPP from Flux Shuttle sites across Asia–Oceania.

(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, distinguishing the availability of daytime- and nighttime-partitioned GPP. 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 differences in spatial representativeness may contribute to VISIT–Flux disagreement. In addition, tower-derived GPP is not directly observed but is estimated by partitioning measured NEE, introducing uncertainty associated with the partitioning method. The flux comparison therefore cannot uniquely attribute disagreement to VISIT 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.

We will incorporate the finalized site-level IAV and long-term trend comparisons into the revised manuscript and Supplement. We will also revise the Discussion and Conclusions to clarify that discrepancies involving VISIT GPP may reflect a combination of differences in ecological sensitivity, observational and spatial-representativeness uncertainty, NEE-partitioning uncertainty, and VISIT model uncertainty, whose relative contributions cannot be fully separated in the present study.

3. Growing Season (GS) Definition Mismatch (L125):

The authors define the growing season (GS) using specific Ta and NDVI thresholds. While the rationale behind applying this GS mask to NDVI is sound, the authors chose to use annual averages for GPP without applying the same GS mask. I wonder to what extent the observed discrepancies between GS-averaged NDVI and annual-averaged GPP are simply artifacts of this temporal mismatch. I recommend conducting a sensitivity test where the exact same GS mask is applied to the GPP data prior to the NDVI vs. GPP comparison. This ensures that observations from the exact same temporal window are compared, isolating true functional divergence from methodological inconsistencies.

Response:

We thank the reviewer for pointing out this potential source of inconsistency. We agree that applying the same temporal window to NDVI and GPP is important for assessing whether differences in temporal aggregation contribute to the observed NDVI–GPP discrepancies.

Following the reviewer's suggestion, we applied the same GS definition to monthly SNU NDVI and VISIT GPP before annual aggregation and repeated the IAV comparison. The results changed only slightly: the mean signed NDVI–GPP correlation across Asia–Oceania was 0.374 with the original annual-GPP treatment and 0.366 with the common-GS treatment. In addition, 96.38% of pixels retained the same correlation-significance classification. These results indicate that the difference between GS-averaged NDVI and annual-averaged GPP has only a limited influence on the overall NDVI–GPP IAV agreement across the study domain. For completeness, the corresponding spatial comparison is reproduced here as Fig. R2.

A plausible explanation for this limited change is that VISIT may already simulate relatively low GPP during months excluded by the GS definition, particularly in cold regions. Because the GS definition is held constant across years for each grid cell, excluding months with consistently low GPP would be expected mainly to rescale the annual GPP series rather than substantially alter its interannual pattern. Such an approximately constant rescaling would have little effect on the correlations calculated from the standardized and detrended series. In grid cells where all months satisfy the GS criteria, the two aggregation methods are identical. These considerations may help explain the small overall change, although the magnitude and interannual variability of GPP outside the defined GS have not been quantified here.

Importantly, the limited sensitivity to temporal aggregation does not imply that the remaining NDVI–GPP discrepancies necessarily represent physiological divergence. Differences in the vegetation properties represented by NDVI and GPP, observational uncertainty, and uncertainty in the VISIT simulations may also contribute. This sensitivity test specifically evaluates the effect of using a common temporal window on the IAV comparison and does not, by itself, establish the robustness of the long-term trend comparison to temporal aggregation.

We will include this sensitivity analysis in the Supplement, summarize its main finding in the revised manuscript, and revise the Discussion accordingly. We will also clarify that 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. Thus, this sensitivity analysis evaluates the effect of applying the same temporal window to NDVI and GPP, rather than the effect of interannual changes in growing-season timing.

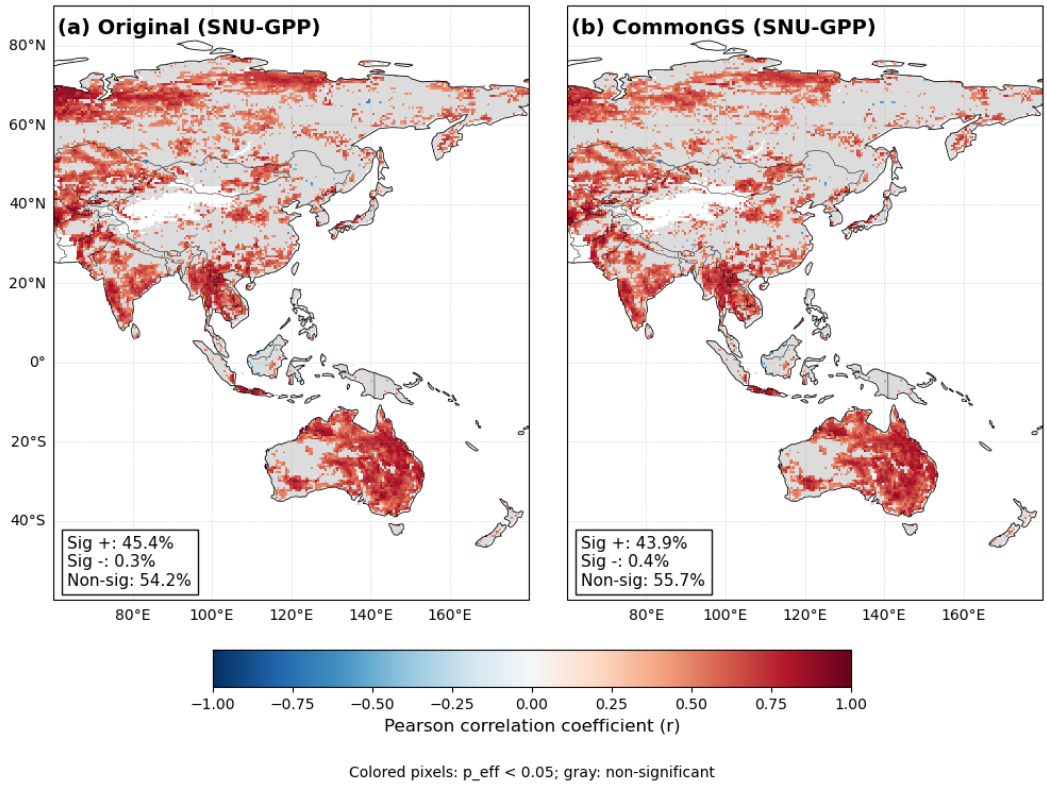


Figure R2. Sensitivity of NDVI–GPP interannual variability agreement to temporal aggregation.

Spatial distributions of the signed Pearson correlation coefficient between detrended SNU NDVI and VISIT GPP for (a) the original treatment using growing-season NDVI and annual GPP and (b) the common-GS treatment in which the same growing-season definition was applied to both NDVI and GPP. Statistical significance was evaluated using an effective sample

size accounting for lag-1 temporal autocorrelation. The mean signed correlations across Asia–Oceania were 0.374 and 0.366 for panels (a) and (b), respectively.

4. Model-Dependent Attribution Analysis (L170):

The attribution analysis based on the VISIT model's sensitivity experiments (varying drivers like precipitation, temperature, vapor pressure, etc., one at a time) is methodologically sound. However, the authors should explicitly acknowledge that the "importance" of a specific driver (e.g., temperature dominance in Siberia or precipitation in Australia) is heavily dependent on the internal mathematical formulation and parameterization of the VISIT model. A factor identified as highly sensitive in this analysis reflects the model's internal structural logic, which does not necessarily perfectly mirror its importance in reality. Employing a different ecosystem model could potentially yield a completely different hierarchy of factor importance. This caveat regarding model structural uncertainty should be explicitly discussed in Section 4.3.

Response:

We thank the reviewer for highlighting this important limitation. We agree that the inferred importance of individual drivers depends on the structure and parameterization of VISIT.

The sensitivity experiments quantify how GPP responds to prescribed forcing changes within VISIT and therefore provide information on the model’s internal sensitivity to individual drivers. However, they do not independently establish the relative importance of these drivers in real ecosystems. The magnitude and ordering of the simulated responses depend on the model formulation, parameterization, and representation of vegetation and environmental constraints. A different ecosystem model could therefore produce a different hierarchy of driver importance under the same forcing conditions.

We will explicitly discuss this limitation in Sect. 4.3 and clarify that the inferred driver effects should be interpreted as model-dependent sensitivities rather than as definitive real-world attribution. We will also revise the corresponding statements in the Results and Discussion so that regional driver rankings are clearly described as responses simulated within VISIT. Comparisons with other ecosystem models and independent observational constraints would be needed to evaluate how robust these inferred driver relationships are beyond the VISIT framework.