

Summary of the research and my overall impression

Merrick and coauthors present a novel dataset of remotely sensed vegetation indices (VIs) (NDVI, EVI, NIRv, NIRvrad, FCVI) from an UAS in a tropical forest canopy in Panama. They explore both spatial and temporal variability between indices and highlight potential uses for these indices at those varying scales. Specifically, the authors explore temporal correlations between GPP and VIs over the course of a day, diurnal changes in the spatial variation between VIs, and dominant spatial scales for variability in VI signals.

The paper is generally well written and structured and provides exciting insights on how VIs relate to each other. Both the dataset and the comparison are novel and within the scope of BG. Additionally, such direct comparisons between VIs are highly valuable because they provide insight in a field saturated with different VIs as to which VIs are most applicable for certain questions and specific strengths and limitations of each. The data collection approach is largely appropriate for the study, however, the temporal resolution of measurements is a major limitation. Additionally, the authors make claims about their findings in relation to SIF measurements that are not sufficiently substantiated. These major concerns are outlined with more specifics below.

Overall, this is an interesting study that will be of interest to the scientific community but needs some revisions to clarify what their findings are vs. what their findings imply. Therefore, I recommend this paper be accepted with major revisions.

Major Concerns

- The methods section is quite dense and difficult to follow. This makes it challenging for the reader to connect measurement approach to the presented results. I recommend the authors present some sort of conceptual figure showing their measurement approach and processing. I think this will be highly beneficial, particularly for a study that explores spatial variability.
- Only one day of GPP data is available. This has led to two specific issues:
 - I am concerned about the validity of a single day's worth of GPP data. I feel as though the statistics used to partition GPP from NEE may be insufficient with only one day available. It's worth some discussion about the limitations of this approach at a minimum.
 - In section 3.1, the authors explore the diurnal trend in VI, PAR, and GPP data. They use this trend to draw conclusions over the utility of NIRvrad as a proxy for GPP. However, I do not believe one day of data is sufficient to draw such strong conclusions. Additionally, there is insufficient discussion over how potential physical (illumination, viewing direction, etc.) or environmental effects (drought, seasonality, etc.) may impact these conclusions and the limitations posed by one day of data. Finally, Figure 1 appears to show a higher correlation between GPP and PAR than between GPP and NIRvrad – therefore significantly undercutting the authors' main claim in this section – that NIRvrad is an appropriate proxy for GPP over short temporal scales. To me, this section would be better off as a discussion of how NIRvrad in fact does **not** sufficiently capture diurnal variability in GPP – and moreso reflects changes in PAR. I also

recommend the authors provide a bit of additional commentary on why the other VIs show low correlations with GPP data.

- The authors repeatedly draw the conclusion that presented VI data is suitable for separating out the physiological from the structural component of the SIF signal when SIF measurements are available. However, the authors are not presenting SIF data and therefore not substantiating this claim with sufficient results or appropriate citations. Specific comments are included in specific examples. I feel that much of the SIF discussion in fact takes away from the authors main conclusions and novelty of their other results as it focuses the discussion on what they aren't doing (normalizing SIF with VI data). In particular, the majority of the introduction focuses on SIF. I recommend the authors cut down on this discussion significantly and make it more clear what conclusions they are drawing from their results vs. potential directions for future work.

Specifics:

- Lines 16-18: The statement 'presented here for the first time' is a bit misleading since you are not presenting these VI's for the first time, you're presenting them at this specific field site for the first time. Additionally, this opening does not make it clear the scientific question or problem you are trying to address or appropriate background information.
- Line 38: Unoccupied might be a more appropriate term, as presumably the UAS was piloted (just not with someone on board)
- Line 57: 'SIF is mechanistically linked to photosynthesis of plants, and thereby, has also been shown to be more sensitive to changes in forest canopy function and structure than RIs' – this deserves a citation. I also don't think you can say it's more sensitive to changes in forest canopy *structure* (although function yes). See the following for comparisons between SIF and VI's (among others):
 - Cheng, R., Magney, T. S., Dutta, D., Bowling, D. R., Logan, B. A., Burns, S. P., Blanken, P. D., Grossmann, K., Lopez, S., Richardson, A. D., Stutz, J., & Frankenberg, C. (2020). Decomposing reflectance spectra to track gross primary production in a subalpine evergreen forest. *Biogeosciences*, 17(18), 4523–4544. <https://doi.org/10.5194/bg-17-4523-2020>
 - Magney, T. S., Bowling, D. R., Logan, B. A., Grossmann, K., Stutz, J., Blanken, P. D., Burns, S. P., Cheng, R., Garcia, M. A., Köhler, P., Lopez, S., Parazoo, N. C., Raczka, B., Schimel, D., & Frankenberg, C. (2019). Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), 11640–11645. <https://doi.org/10.1073/pnas.1900278116>
 - Pierrat, Z., Nehemy, M. F., Roy, A., Magney, T., Parazoo, C., Laroque, C., Pappas, C., Sonnentag, O., Bowling, D. R., Seibt, U., Ramirez, A., Helgason, W., Barr, A., & Stutz, J. (2021). Tower-based remote sensing reveals mechanisms behind a two-phased spring transition in a mixed-species boreal forest. *Journal of Geophysical Research: Biogeosciences*. <https://doi.org/10.1029/2020JG006191>

- Line 87-89: It's worth mentioning which ecosystem types because this is not true across all ecosystems/some types show much better performance than others. The citations you have all have ecosystem type information.
- Lines 99-101: Again it's worth mentioning ecosystem type here (ie: specifically tropical in your case) – this doesn't necessarily apply for all ecosystems/we don't have enough studies testing this across varied vegetation cover.
- Lines 111-113: This deserves a citation (or several).
- Line 124: The introduction deserves some final statement about the broader aims of this work. What ultimate goal this information provides.
- Line 146: there's a period . typo after 12 ms
- Line 160: As mentioned above there should be additional discussion on the limitations of only one day of data.
- Line 173: I believe the original citation for NDVI is:
 - Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Figure 1: There appears to be some sort of accidental grid to the side of panel d?
- Lines 236-238: 'Our results demonstrate that UAS-based data are suitable for normalizing SIF at high spatial resolution in addition to recording structural heterogeneity of a tropical forest' – your results don't really demonstrate this because you don't have SIF data. Maybe if you say they have 'the potential' however I still think this distracts here from the other findings.
- Line 239: 'Because NIRv and NIRvrad use NDVI, these results also indicate that including NIR reflectance or NIR radiance is the largest contributing factor to this variability' – This is built into the definitions of NIRv and NIRvrad so I would rephrase this to reflect that.
- Lines 250-251: rephrase for clarity to 'The low variability and high means at midday of NIRv, FCVI, and NIRVrad indicate that...'
- Line 266: 'strong peak' is a bit of an overstatement, it seems much more rounded to me
- Line 277: remove 'note how'
- Lines 286-297: This discussion of SIF is much better because it acknowledges the potential, but also notes that SIF measurements are not available. This however also deserves some citations
- Line 313: Remove 'for the first time' – it's confusing as you're not presenting new indices, you're presenting new data at this particular location
- Line 317: I do not believe you can draw this conclusion with one day of data (see my major concern above)
- Lines 334-337: SIF discussion here is distracting from your main points
- Lines 345-346: You do not show that these measurements can be used to separate the components of a SIF signal and you're also not really showing how to use it as an estimate of fPAR, APAR, or GPP. Also worth noting this is for a tropical ecosystem.