Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-110-RC2, 2018

Biogeosciences Discussions

Interactive comment on "On estimating Gross Primary Productivity of Mediterranean grasslands under different fertilization regimes using vegetation indices and hyperspectral reflectance" *by* Sofia Cerasoli et al.

Anonymous Referee #2

Received and published: 6 June 2018

This is a well-designed and well-executed experimental field study investigating the skill of spectral reflectance indicators of temporal and treatment variability in grassland primary productivity. It builds incrementally on prior works in a well-trodden area. The methods of data analysis might be improved with alternative graphical displays and statistical analyses that would provide more insights, as suggested below. Some interpretations and conclusions draw distinctions regarding the relative skill of different reflectance metrics without sufficient quantitative support, as explained below. In addition, the paper could do more to advance functional models of GPP beyond the

data-intensive, highly empirical sort that is explored. Lastly, the paper's sentence level writing could be improved.

The study design introduces nutrient treatments which influence species composition GPP to some degree but the successive analyses discard this structure when testing the skill of reflectance metrics as predictors of GPP. GPP varies over time, across replicates and across treatments but all of this variability is pooled when testing the reflectance metrics. The paper might provide more insights by structuring the analysis to test treatment and temporal variability separately. Perhaps a repeated measures ANOVA could help, for example, by testing for significant effects of date, treatment, and one reflectance metric on variability in GPP. This would be akin to the linear mixed effect analysis of GPP and PAIgr.

GPP and PAIgr both vary over time and PAIgr varies across treatments (GPP appears to as well but apparently the statistical testing does not support this). These patterns are displayed well with Figs 2 and 3 but what is missing is display of a scatter of GPP versus PAIgr, and also of GPP versus (selected) reflectance metrics. These relationships should be shown, with symbols that differentiate the treatments and display individual replicates. The relationship that emerges (slope) would offer insights about the effective light use efficiency per unit green leaf area. The term 'effective' here refers to the combination of a maximal LUE with any limitations by water, light, or nutrients. This is the sort of parameterization that would be needed in a functional model. In fact, it would be interesting to test if any of the reflectance metrics have skill in predicting variability in GPP / PAIgr, thus capturing patterns in LUE rather than just green plant area.

Table 5 shows the skill of various multivariate linear models that include a suite of reflectance metrics selected to represent those available from different observing system types. This is a highly empirical approach to analysis and does not seem particularly useful in my opinion. The results are likely to be very heavily tuned to the specific dataset on hand and is not likely to be generalizable beyond the current study. For example, the Hyp-B step one selection includes a simple, linear model involving 13 unique bands. Biophysical or ecological functional models tend to use one or two metrics to represent structural (PAIgr) and functional (LUE) attributes of an ecosystem's capacity for primary production. This paper's approach throws every possible indicator and combination at the variability in the data and thus lends little practical insight into the theory with very limited capacity for transferability. A more thoughtful approach grounded in theory and practice would be more useful.

The study's test of linear models includes VIs and bands, but not band ratios. Given that the approach is highly empirical in nature, there does not seem to be a good reason to omit band ratios or other simple mathematical combinations of bands (e.g. unique normalized difference ratios). Testing a wider range of combinations could be warranted to see if any other indicators happen to rise to the top in terms of predictive skill.

The paper's interpretations and conclusions suggest that bands are better than VIs as predictors of GPP but this is not reasonably supported by the quantitative results. Table 5 shows a small, marginal, and questionable increase in adjusted R2 for Hyp-B compared to Hyp-VIs, and a decrease in adjusted R2 for S2-B compared to S2-VIs. In any case, the differences in explanatory power over all of these cases is less than 0.0247 R2, or 2.5% of the variability in GPP, indicating that all are equally good at predicting GPP. For L8, a case might be made, however the band metric has many more variables thrown at the problem (6 bands compared to just NDVI) and when these other bands are included in a step two selection, the NDVI model with bands had high skill. Surely bands and VIs are equally skillful for the other observing system types. Corresponding edits need to be made to section 4.2.

One of the advantages of VIs is that they normalize for a wide range of background reflectance, sun-sensor geometry, and atmospheric effects in ways that direct bands do not. This point seems to be lost on the authors and is important for developing indicators that can be transferred to remote sensing (space or airplane) over large

areas and across large gradients in surface and atmospheric conditions. Discussion about this should be included in the paper.

Akaike or Bayesian Information Criteria need to be adopted to evaluate the relative skill of the selected linear models, penalizing models that select more variables.

It is worth noting that soil moisture is essentially equal on all four reflectance observation dates, while temperature increased steadily from the first to the last observation date. Correspondingly, the statement on P12, L19 that suggests that the Hyp model represents changes in canopy water content might need to be revised. Canopy water content was not observed and soil water content did not differ over the four sampling dates. It is possible that canopy water content differed substantially from soil water content over this time series but that has not been established with quantitative, direct observations.

It is unfortunate that the study did not include an additional observation period in the mid to late June as PAIgr continued to decline.

The introduction is very well written and cited. One paper that might be useful to add to the framing and discussion is that of Asner et al. 2004 in PNAS ("Drought stress and carbon uptake in an Amazon forest . . .").

Page 5, L12: "All nutrients were added at . . ." seems to suggest N for nitrogen, or is N for nutrients here?

Measurement of soil moisture at only 10 cm depth may not be adequate to represent the soil water content being experienced by the grassland plants. It would be best to also measure a deeper profile of moisture.

It is interesting that the nutrient treatments allegedly altered the functional composition of the grassland plots, however pre-treatment data are not presented and this would be essential to demonstrate that the compositional shifts were due to the treatments themselves. Unless it can be established with data, the corresponding statement (P9, L23) should be corrected to omit suggestion that the treatments caused the compositional differences.

It is surprising that treatment effect was significant for respiration but not for GPP considering that both have similar spreads and error bars. Double check results here.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-110, 2018.