Biogeosciences Discuss., referee comment RC2
https://doi.org/10.5194/bg-2021-250-RC2, 2021

Comment on bg-2021-250

Anonymous Referee #2

Referee comment on "Estimating dry biomass and plant nitrogen concentration in pre-Alpine grasslands with low-cost UAS-borne multispectral data – a comparison of sensors, algorithms, and predictor sets" by Anne Schucknecht et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-250-RC2, 2021

This manuscript provided by Schucknecht et al. presents a study of estimating pre-Alpine grassland aboveground dry biomass and plant nitrogen concentration using Unmanned Aerial Systems with two low-cost multispectral sensors. This study tested three statistical models including linear models, random forest, and gradient boosting machines to predict dry biomass and plant nitrogen concentration from UAS multispectral imagery. Three science questions on (1) whether spectral information of UAS sensors is enough for mapping dry matter and plant nitrogen, (2) the need for machine learning hyper-parameter tunning, (3) and model performance with different sensors, statistical models, and inputs were addressed. Results show that the two UAS multispectral data sets can achieve moderate performance to quantify grassland dry biomass and plant nitrogen concentration. Specifically, the best performance of quantifying dry biomass came from the combination of random forest, all predictors, and REM sensor data. The best model for plant nitrogen concentration was achieved by using random forest, all predictors, and SEQ sensor data. Considering the rapid development of UAS remote sensing for mapping vegetation traits, this study is necessary and interesting. The manuscript is well-structured and easy to follow. However, the current manuscript has several issues with experimental designs and the result interpretation. I suggest a major revision for the current format. Here are some comments that may be helpful to improve the manuscript.

Main issue:

1. The motivation for comparing these two UAS sensors is not clear. Are these two types of sensors are popularly used in UAS remote sensing studies? How the findings from the two sensor comparison are relevant to other studies and the UAS remote sensing community? Overall, SEQ and REM sensors are very similar. These two sensors have similar pixel resolution, similar wavelengths in green (550/560 nm), red (660/668 nm), and red edge (735/717 nm). Furthermore, the manuscript pointed out that SEQ performed well for predicting plant nitrogen concentration, while REM had a better performance for

predicting dry biomass. However, it is not clear why these two sensors had such different performances in the current manuscript. The analysis and explanation for sensor performance on dry biomass and nitrogen predictions need to be strengthened.

In Table 2, you labeled 790nm as near infrared. However, we usually refer to 700-800nm as red edge, while wavelengths beyond 800nm as near infrared. From the soil-vegetation radiative transfer modeling view, red edge wavelengths are vital for vegetation chlorophyll content and nitrogen content retrieval. The near infrared is more sensitive to the vegetation canopy structure such as leaf area index and total biomass. From my interpretation, SEQ has two red edge bands and could potentially get better results for nitrogen concentration retrieval, but not dry biomass as lacking information in near infrared. Meanwhile, REM has information on near infrared which is good for biomass retrieval.

2. The motivation for selecting Gradient Boosting Machines and Random Forest is also not clear. Why not other more popular machine learning or statistical approaches, such as partial least-squares regression, LASSO, Ridge, or Neural Networks?

The purpose of applying machine learning algorithms is not only to achieve good model predictive performance. Many machine learning algorithms like random forest can help to identify the relative importance of each feature input. This feature importance analysis is very necessary to understand the relationship between feature inputs and the predicted variables. However, such analysis is missing in this study. I strongly recommend further feature importance analysis to identify scientific linkage among input variables and the predicted variable to strengthen the manuscript result interpretation.

3. The UAS multispectral data were collected from one single flight in each site. How robustness of these results across different growth stages and dates is uncertain?

4. Machine learning parameter tunning is a very necessary and common step to implement model training. However, this manuscript highlights the hyper-parameter tunning as one major research question. The innovations of this study need to be strengthened.

Minor issues:

- There are many abbreviations in Figure 2. The caption should add explanations of these abbreviations for readers.
- The reflectance values in Figure 4 look quite different from the two sensors. Do you have ground reflectance collection to validate your reflectance?
- The manuscript mentioned that mountain regions have frequent cloud occurrences to argue the weakness of Copernicus satellite missions. However, UAS data collection under cloudy environment also has data quality issues. The manuscript may need to discuss such potential issues and mitigation strategies.
- Most parts of the manuscript used nitrogen concentration. However, Figure 6 used nitrogen content in the (c) and (d) subplots.
- The same issue of nitrogen concentration on Figure 7.
- Figure 8 (d) has clear shadows. The reflectance from these shadows needs to be either corrected to real surface reflectance to quantify vegetation traits or simply removed. I don’t think the current estimates for areas in tree shadows are right.
- The figure panel design of Figure 8 is strange. We normally put RGB into the first subplot. You have paired maps for DM and N. These paired subplots could be in one row.