Response to referee #1 comment

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-RC1, 2021

We thank referee #1 for the constructive comments. Please find below, how we want to address the raised issues (referee comments in italic) in a revised version of the manuscript.

Main issue 1 (regarding single date acquisition of UAS and field data)

The study is entirely based on a single field campaign, therefore during one specific stage of grasslands phenological development (presumably early season). This is an important limitation which is not discussed. I would recommend to more explicitly highlight this and discuss the implications.

We are aware of the limitation of a single field campaign. The study presented in the manuscript tests the general approach of grassland trait estimation with an application of immediate grassland trait mapping. To increase the general validity or the results of a mono-temporal campaign, we selected several sites differing in species composition, current growth height and nutrient status, which allowed us to compile a dataset of variable grassland traits. We will include this aspect in the material and method section and add a paragraph in the discussion chapter to address the applicability of the study and requirements for other phenological stages (multi-temporal applications).

Main issue 2 (regarding limitations of UAS)

The study in the introduction suggests that UAS can bring significant advantages as compared to satellite or airborne data in mountain regions. However, I would argue they have also significant limitations, especially when low-cost sensors are used. By looking at the study results, I have the impression (perhaps wrong) that issues related to data acquisition and quality (e.g. long acquisition time, calibration and incident radiation measurement) might play an important role in explaining the relatively poor model performances. However, this is not mentioned or discussed in a clear way (there are some points, but not really a section discussing the challenges in UAS data acquisition and quality). I would find valuable to see some more discussions on the issues that might be related to UAS use (even better if supported by some analyses).

We agree with the referee that studies utilizing UAS data have their own challenges and drawbacks. Therefore, we will include a section in the discussion chapter that will discuss the challenges with UAS and low-cost sensors in general and the limitations of our study in particular (including the point raised in issue 1). Anyhow, we assume that the relative differences between sensors and predictor sets etc. are still valid and honest outcome of this study, even though there could be noise due to the UAS data acquisition.

Minor issues

Introduction:

There is a bit of mix between Alpine/pre Alpine etc, while most statements are valid for both. Perhaps if the area falls within the Alpine space (geographic region) there is no need to specify Pre. Alpine is sufficient and more details about the sites are given in the methods.

We prefer to keep the distinction between Alpine and pre-Alpine to highlight that our study covers the hilly Alpine foreland and not the high-elevation pastures, which would be associated with Alpine. Anyhow, we will adapt some sentences to make this clearer.

Perhaps 'often long' is unnecessary

"Often long" will be removed in the revised manuscript.

The sentence is unclear

Unfortunately, we are not sure to which sentence this comment refers. Could you please specify it?

86-89 This might be truth for specific cases, but it is important to keep in mind the limitations of sensors technologies onboards UAS.

Thank you for the comment. We will include a sentence on the limitations of UAS sensors.

Here and elsewhere it is mentioned canopy height data were not 'available' without explanation. I would suggest avoiding that, as this is explained in the methods. Otherwise, a short justification should be added here too.

We will follow the suggestion of the referee and adapt the respective sentences accordingly.

Materials and methods

There is no mentioning of the phenological stage of vegetation during the field campaign. This is an important factor.

We will add the information on the phenological stage by checking the photos of the field campaign.

9.50 to 16.30 is a quite long interval with expected variations in solar angle and, in mountain regions, shadows and possibly cloudiness. This could be quite a relevant factor affecting the data acquisition.

We agree with the referee and will include this issue in the discussion chapter (under the new section about challenges of UAS data and their acquisition).

Is there any indication of the geolocation accuracy?

The accuracy of the GNSS is provided in line 180ff: The exact coordinates of the GCPs` centres were obtained with a Global Navigation Satellite System (GNSS) receiver (Viva GNSS GS 10, Leica Geosystems AG, Switzerland) run in static mode for 10 minutes which resulted in an accuracy of 0.3 cm in horizontal direction and 0.5 cm in vertical direction in post-processing mode (Datasheet of Leica Viva GNSS GS10 receiver, 2020).

An area of 3x3 pixels seem very small considering geolocation errors. Assuming the plot should be somehow representative of a wider area, would not be more prudent to have a larger window?

We selected the 3x3 pixel window to approximately cover the area of the subplot (0.25 m x 0.25 m). Due to the high accuracy (0.003 m) of the GNSS measurements in relation to the window size of ~ 0.25 m, we expect just minor location errors. Another aspect is that although we aimed to sample homogenous plots, we realized in the field (and by further data analysis) that there were small-scale variations due to different plant species. These small-scale variations would be neglected by a larger window size.

Results

It is somewhat surprising the NIR does not follow DM or height, as this should be rather straightforward (unless for very small range). Is there any factor related to the acquisition that might be causing this issue?

In Figure 4 of the manuscript we just plotted the spectral profiles of a few selected samples (corresponding to min and max values of DM and N as well as the ones that have DM or N concentration values that approximately correspond to the 25th, 50th and 75th percentile). In Figure 4, the NIR reflectance does follow DM, except for the maximum value. We now created additional figures (Figure 1 below) to show the relationship between the reflectance of the NIR band and DM (and canopy height). In fact, there is a positive relationship between NIR and DM, but this relationship

is not the strongest. Therefore, the appearance of Figure 4 depends very much on the selection of the samples.

Potential reasons for the rather week relationship could be the different species compositions of the subplots, differences during the acquisition (time during the day, clouds), as well as the radiometric correction of the multispectral sensors, which can be challenging and a source of uncertainty (Olsson et al., 2021). We will include the new plots and address the issue in the discussion of the revised manuscript version.

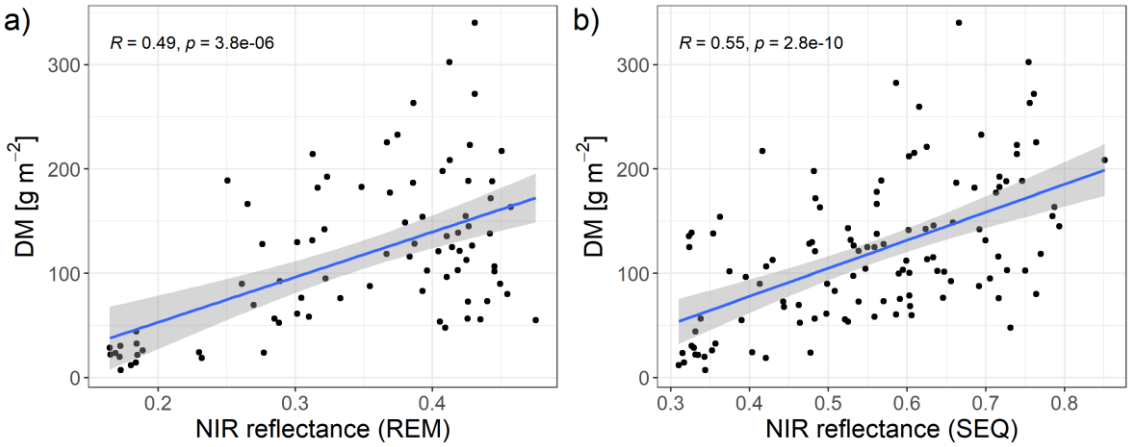

Figure 1. Scatterplots of NIR reflectance vs. DM for the REM sensor (a) and SEQ sensor (b) with linear model fit (Spearman correlation coefficient and p-value indicated in the plot). Note: there are less data points for REM as there were no flights with this sensor at the Eschenlohe site.

The doubt of a strong influence of acquisition factors is also supported by the very poor performance of regression as compared to machine learning and the improved performances on DM including VI in the validation. Also the important role of ground canopy height may suggest that as this variable is clearly not affected by the UAS data acquisition. I would suggest to run some tests and eventually add some considerations in the discussions.

Thanks for raising this issues. As we conceive the issue, in our data the traits are not strongly correlated to spectral signals and VI in bivariate manner. However, it does not only seem to be because of the acquisition error, but other confounding factors involved as well. We will do more diagnosis and discuss the arguments of the reviewer on this topic in the discussion and add relevant literature as evidence.

The word 'notably' is very often repeated. Sometime is a bit redundant

We will adapt the text accordingly.

Figure 7. It is a bit strange to see many points along a line (i.e. same N content) in Fig.7b. Is it correct?

We have checked the figure and it is correctly reflecting the model outcome. The points along a line means that in this model set-up, GBM does not differentiate most of the samples in predicting N concentration. This is because the domain of the input data values of the test site are new to the algorithm, makes the algorithm to guess it is constant. We will discuss this in the discussion and add potential strategies how to overcome this (i.e. using synthetic sampling and generative modelling).