Dear Editor and Reviewer #2

This paper developed a new annual 250-m grided AGB dataset from 2000 to 2019 over the Qinghai-Tibetan Plateau by using ground in-situ measurements, multi-years UAV images and MODIS datasets. UAV images were used as the bridge to overcome scale mismatches between ground samples and coarse MODIS satellite pixel scales. Many efforts have been devoted on UAV observation works as well as field sample works at large region scales. In general, I think this is a good paper and is within the scope of ESSD. I have a few comments for authors' consideration.

Response: We appreciate your insightful comments on our paper. The comments provided have been extremely helpful to us. We have revised the manuscript in response to your comments and carefully proofread the manuscript to minimize typographical, grammatical, and bibliography errors. The point-to-point responses to your comments are listed below in **blue**.

Point 1. In table 1, the acquisition times of UAV sampling and field sampling of AGB in 2019 were mainly in the growth season and therefore the UAV estimations of AGB also in this season. However, the MODIS pixel level vegetation indices were composited by MVC method which can reflect the best grown condition in the whole year. Aboveground biomass may still be available for several months after sampling. The temporal mis-matches between field work and MODIS composites may lead to estimation errors.

Response: Thank you for your comments. As you pointed out, the UAV photos and measured AGB samples we acquired in 2019 were mainly from July to August.

- In constructing/validating the AGB estimation model at the pixel scale, the MODIS NDVI, EVI, and kNDVI indices closest to the sampling time were chosen to minimize the difference between sampling and satellite acquisition times.

  As you suggested, we added an explanation in section 2.4.2 (lines 162-166):

  "After that, the corresponding vegetation indices closest to the time of the UAV sampling were extracted to construct/validate a pixel-scale AGB estimation model. In addition, the kNDVI index was calculated to overcome the NDVI saturation issue based on the equation kNDVI= TANH ($NDVI^2$) (Camps-Valls et al., 2021)."

  We also explained it in section 2.5.4 (lines 232-234):

  "The MODIS vegetation indices and other spatial metrics corresponding to each GRID or RECTANGLE mode were then extracted using the ArcGIS software. Here, the MODIS

NDVI, EVI, and kNDVI indices closest to the sampling time were chosen to minimize the time difference between sampling and satellite overpass."

- The annual maximum vegetation indices were used as the input parameters for the AGB inversion of the entire QTP region. As in previous studies, we used MVC method to obtain the maximum value of annual vegetation index (NDVI, EVI and kNDVI), and assumed that AGB corresponding to this period also reached the maximum value in the growing season. We explained this in Section 2.4.2 (lines 165-166).

  "The annual maximum vegetation indices were calculated by the maximum value composition (MVC) algorithm to estimate the spatial AGB distribution of QTP from 2000 to 2019 (Holben, 1986; Wang et al., 2021; Gao et al., 2020)."

- Although the vegetation index closest to the sampling time was selected when constructing the pixel scale estimation model, there was still a time difference between the ground samples and MODIS indices, which would lead to estimation error. We pointed out this limitation in section 4.4 (lines 451-453)

  "Although the MODIS index closest to the sampling time was chosen for the construction/validation of the AGB estimation model, there was still a time difference between the measured samples and the MODIS indices, which might lead to estimation errors."

Point 2. In section 2.3.2, the BELT flying mode were used for three GRID routes and four ground sampling quadrats were sampled in the BELT routes. However, in Figure 2 and Section 2.5, how were the BELT images at 2-m height used was not introduced. It seems only 20-m UAV images were used for development of UAV estimation model.

Response: Thank you for your comments. As you mentioned, we only used 20m high UAV photos when constructing the AGB estimation model at the quadrat scale.

- However, in fieldwork, the BELT flight mode served as a bridge between traditional and UAV sampling at 20m. Compared with GRID mode, the size of BELT was relatively small (40 m × 40 m), and the flight altitude and speed were set to 2m and 1m/s, respectively. It ensured the staff had enough time to place a sampling frame directly below the aircraft lens to capture it in the 2m and 20 m UAV photos. We explained this in section 2.3.1 (lines 129-133).

  "The BELT mode is similar to GRID, but is designed to obtain near-ground UAV image data with higher resolution (Figure 3b). It can be combined with the traditional sampling method to ensure the consistency of UAV images with the ground samples (Figure 3d). Typically, the BELT size is set to 40 m × 40 m, and the flying height and speed are set to 2 m and 1m/s to ensure that field crews have enough time to place sampling frames under the UAV waypoints. As with the GRID mode, 16 UAV images can be captured in a single flight."

- Although our previous research confirmed that 2m photographs could be used to model sample-scale grassland AGB, the reasons why 20m high UAV photographs were chosen for the quadrat-scale model in this study were as follows: Firstly, the spatial coverage of a 20m-high UAV photo (26 m×35 m) is much wider than a single 2m-high UAV photo, making it easier to match to the MODIS pixel scale. We have added an explanation in section 2.5.2 (lines 210-216) :

  "Since the spatial coverage of a 20m-high UAV photo (26 m×35 m) is much wider than a single 2m-high UAV photo, making it easier to match to the MODIS pixel scale. Hence, the 20m-high UAV photos containing the sample frames were chosen for constructing the quadrat-scale AGB estimation model. A total of 906 pairs of quadrat-scale UAV-field AGB observation data were collected, with good spatial representativeness (Figure 1 a, red dots). The observed AGB values ranged from 0 to 450 g/m^2, with mean and median values of 59.75 g/m^2 and 33.04 g/m^2, respectively, most of which were less than 100 g/m^2 (Figure 5a). The cropped 20-meter-high UAV image indices and the measured AGB values were used as the independent and dependent variables to build the RF model (Figure 2).
  "

- Secondly, using the 20m-high UAV photo containing the sample frames to construct the quadrat-scale model ensures an exact match of spatial scales between the independent and dependent variables. Furthermore, it facilitates the estimation of the AGB of an entire 20m-high UAV photo. We explained this in section 4.1 (lines 350-357)

  "Spatial scale matching of dependent and independent variables was achieved in estimating AGB values at different scales. First, at the quadrat scale, the independent variables were all derived from cropped 20-meter-high UAV images corresponding to the ground samples (Figure 3e). Then, the 20-meter-high UAV image was cropped into ~2000 quadrat-sized patches to ensure consistency with the quadrat-scale model, and the average of these patches was used as the final AGB at the photo scale. Finally, by averaging the AGB of 16 or 12 UAV photos within the MODIS pixel, the AGB value matching the MODIS pixel scale was calculated (Figure A1). With these three steps, we successfully upscaled the measured AGB from the traditional quadrat scale (0.5 m×0.5 m) to the photo scale (26 m×35 m) and MODIS pixel scale (250 m×250 m).

- In addition, we discussed the quadrat-scale AGB estimation models based on the 2-meter and 20-meter UAV photos in section 4.3 (lines 384-388):

  "At the quadrat scale, consistent with our previous study, we further confirmed that the UAV RGB images could be used to estimate grassland AGB (Zhang et al., 2022a; Zhang et al., 2018). Similar to the 2-meter-high UAV image, the indices from the 20-meter-high UAV image could be used to estimate the grassland AGB at the quadrat scale (R^2=0.73, RMSE=44.23 g/m^2, Figure 6a). Compared with the 2-meter-high UAV image, the 20-meter-high UAV image is more suitable for matching the MODIS pixel due to its wider spatial coverage (26 m ×35 m)."

Point 3. For one MODIS pixel, it seems more than 16 UAV images at 20-m height are needed to cover the whole pixel. I'm not quite understanding the GRID, RECTANGLE and BELT flight modes. Does it mean the UAV only take pictures in the waypoints and there are gaps among those pictures? The authors can explain more about how it works as traditionally we will make overlaps among pictures.

Response: Thank you for your comments. As you mentioned, all three flight modes (GRID, RECTANGLE, and BELT) are designed to shoot only at the pre-set waypoints.

- Compared to the commonly used MOSAIC flight mode (which requires a guaranteed overlap rate between photos to obtain a full view of an area), our designs are more in line with the traditional eco-sampling concept, which can better balance the spatial representation and accessibility of samples for efficient sample collection. An explanation of the three flight modes was added in Section 2.3.1 (lines 126-135):

"GRID, RECTANGLE, and BELT are the most commonly used flight modes in the FragMap software. GRID and RECTANGLE modes have 16 and 12 waypoints for capturing UAV images within a MODIS pixel range (Figure A1). Their flying height and speed are set to 20 m and 3m/s, respectively. The spatial coverage of a 20-meter-high UAV photo is about 26 m × 35 m. The BELT mode is similar to GRID, but is designed to obtain near-ground UAV image data with higher resolution (Figure 3b). It can be combined with the traditional sampling method to ensure the consistency of UAV images with the ground samples (Figure 3d). Typically, the BELT size is set to 40 m × 40 m, and the flying height and speed are set to 2 m and 1 m/s to ensure that field crews have enough time to place sampling frames under the UAV waypoints. As with the GRID mode, 16 UAV images can be captured in a single flight. Compared with the MOSAIC flight mode (which requires a guaranteed overlap rate between photos to obtain a full view of an area), our design is more in line with the traditional ecological sampling concept. It allows for a better balance of spatial representation and accessibility of samples, resulting in efficient sample collection."

- Thus, for the plot size of the MODIS pixel (250 m × 250 m), it typically takes 40 minutes to complete sampling using MOSAIC mode, while it takes only 10 minutes using GRID and RECTANGLE flight modes. Our flight modes significantly reduce the sampling time and offer the possibility of obtaining more samples matching the MODIS pixel size.

- In addition, to further clarify the impact of the number of UAV samples on the AGB estimation at the MODIS pixel scale, we conducted a comparative analysis of the AGB estimation results using 1 photo to 16 photos in a step-by-step incremental manner. As shown in Figure 7a, when the number of UAV photos increased to 4, the growth rate of the correlation coefficient slowed down and tended to be stable. It indicated that although the 16 photos could not cover the entire MODIS, they were sufficiently spatially representative. The relevant results for this part were presented in Section 3.3 (lines 294-300):

“ Moreover, the correlation between NDVI and UAV-estimated AGB increased with the number of UAV photos. It increased rapidly as the number increased from 1 to 4 (from 0.74 to 0.86), then slowed down and stabilized (from 0.87 to 0.88). In addition, we compared the scatter plots and fitting lines between NDVI and different AGB estimation methods (Figure 7b-f). The results showed a weak linear relationship between the traditionally measured AGB and NDVI, with an R^2 of 0.29. With the UAV sampling method, the linear relationship was greatly improved and increased with the number of photographs. The fit coefficient R^2 increased from 0.54 to 0.78, much higher than the traditional sampling method (Figure 7).

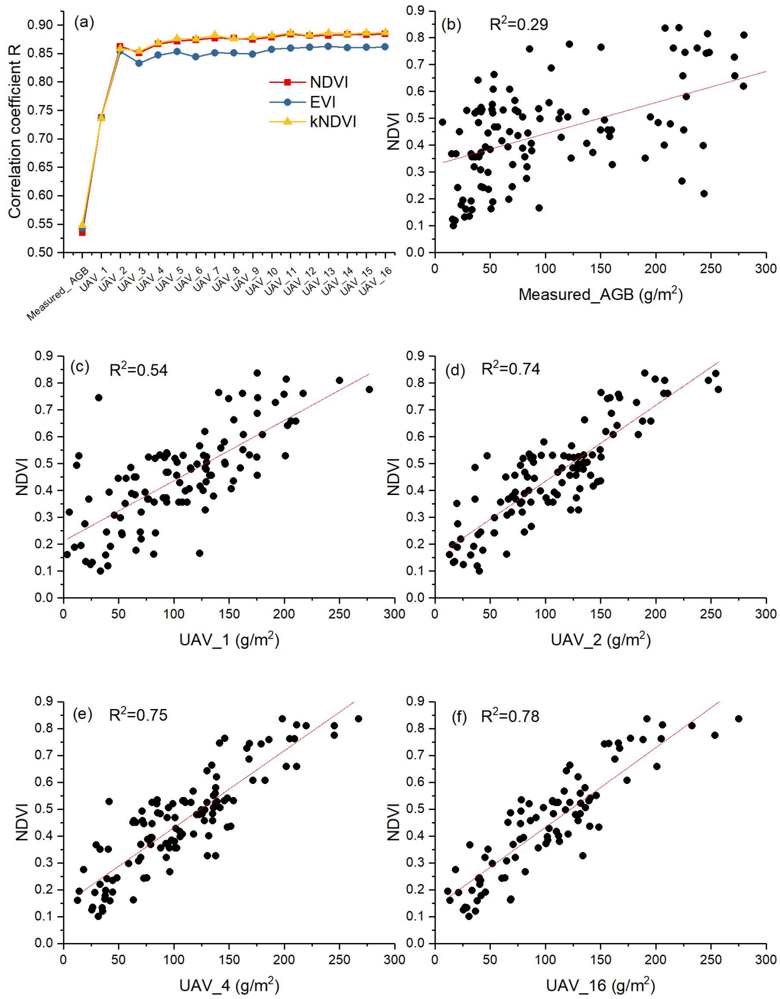

Figure 7. Correlation between MODIS vegetation indices and different AGB estimation methods (a); scatter plots of NDVI with different AGB estimation methods (b-f). UAV_x, x represents the number of UAV photos used to estimate the average AGB at the MODIS pixel scale. Here, x ranges from 1 to 16.

”

- Admittedly, neither GRID nor RECTANGLE flight mode captures the entire MODIS pixel. We discussed this limitation in Section 4.4 and proposed a solution for the next step (lines 461-463):

  “Moreover, although the UAV images in GRID or RECTANGLE mode could cover most areas of a MODIS pixel, full pixel coverage was still not achieved. Therefore, we will gradually upscale to MODIS pixels by combining UAVs with Sentinel-2 or Landsat images.”

Point 4. Page 18, line 324, “The reason was that the GIRD mode could obtain 16 photos in the MODIS pixel at a time, while the RECTANGLE mode could only take 12 photos”. Figure A1(b) should be cited to explain the RECTANGLE model.

Response: Thank you for your comments. As you suggested, we modified it as (lines 364-365):

“The reason is that GIRD mode can take 16 pictures within a MODIS pixel, while RECTANGLE mode only takes 12 pictures (Figure A1).”

Point 5. Page 22, line 407, “AGB_2000.tif represents this TIFF file describing the alpine grassland AGB condition of QTP in 2005” should be in 2000.

Response: Thank you for your comments. As you suggested, we modified it as (lines 468-469):

“For example, AGB_2000.tif represents this TIFF file describing the alpine grassland AGB condition of QTP in 2000.”