Dear Editor, please find attached our revised manuscript. We have dealt with all the reviewers' comments and your recommendations and hope this improved version has addressed all the outstanding issues.

We have highlighted the limitations of the global extrapolation further by referring to 'calibration' instead of 'correction' throughout the text. We have also replaced Figure 2 with a new figure (Figure 5) that combines Figure 2 and Figure A7 and now shows (i) the post-calibration change in tree cover that is statistically significant across all four scenarios, (ii) the uncertainty range of the calibration, (iii) priority regions for field surveying. The accompanying text in the results and parts of the discussion further stress the uncertainty of our calibration. A detailed outline of the changes in response to the reviewers' comments is below.

We have split the methods section into four sub-sections: 2.1 EO Products and Field data; 2.2 Converting In-Situ Canopy Area Index to MODIS VCF / Landsat TCC percent tree cover; 2.3 Calculating Uncertainty Under Different Overlap-Clumping Scenarios; 2.4 Mapping MODIS VCF Uncertainty Across The Tropics.

We have included appendix figures A2 and A3 (adapted to show the approach used for Landsat TCC) as Figures 1 and 2 in the main text. Figure 2 now also includes an example of the effects of unenforced and enforced clumping on 30 m x 30 m Landsat TCC pixels.

<u>Reviewer1:</u>
Thank you to reviewer #1 for your further review. In response to your concerns about ensuring the reader is aware of how the reference data were collected and in particular the fact that the CAI is derived from allometric relationships, we have split the methods section into 4 sub-section and included the paragraph describing reference data collection in what is now Section 2.2 . With respect to including the Landsat TCC in our analysis, we are now presenting results of the TROBIT reference data - Landsat TCC tree cover comparison Please refer to the supplementary pdf for a comprehensive response to your review.

For PDF:
"...the key steps of reference data collection should still be summarized and disclosed in this paper. It is critical to let the readers beware that the CAI reference used in the research was not directly measured in the field, but was calculated based on allometric equations."

In what is now Section 2.2 we describe how the reference data is collected and clearly highlight the fact that CAI is based on allometric equations: *"CAI is defined as the sum of the projected areas of individual tree crowns divided by the ground area. In the TROBIT project (Torello-Raventos et al. (2013) and Veenendal et al. (2015)), plot-wide CAI is made up of the sum of the upper-stratum, mid-stratum, and subordinate-stratum crown areas. Membership to a stratum is determined by the tree's dbh (upper-stratum: dbh > 10 cm, mid-stratum: 2.5 cm < dbh < 10 cm, and subordinate-stratum: dbh < 2.5 cm, height > 1.5 m). About 50 trees per stratum per plot were measured to derive plot-specific allometric relations between stem diameter and crown area (supplement B of Torello Raventos et al., 2013). These were then applied to the whole plot to establish plot-level CAI. For the allometric relationships, tree crowns were treated as circles and the individual tree projected crown area was determined using the average of crown radii measured along the four cardinal points (i.e. from the centre of the stem to the distance furthest from the stem). "*

"...with regards to the suggestion of including the Landsat VCF in additional analysis. I have to rather insist on the comment...such an evaluation can still provide useful insights on how big a role the scale mismatch between the MODIS sensor and field plots plays in the data comparison. It can also provide useful insights on whether or not the biases in MODIS VCF are inherited by the later generation of remote sensing products…"

We have incorporated the additional Landsat TCC product evaluation into our work. Taking advantage of a recent version4 release which provides a 2005 and 2010 layer, we compared the average of 2005 and 2010 tree cover with the TROBIT reference data.

The text in methods was updated as follows:

"We used the 2005 and 2010 30m Landsat TCC version 4 product (https://lcluc.umd.edu/metadata/global-30m-landsat-tree-canopy-version-4), and worked with the 2005 and 2010 average values. The product was downloaded manually from https://e4ftl01.cr.usgs.gov/MEASURES/GFCC30TC.003/. "

"For Landsat TCC, where the Landsat TCC pixels (30m x30m) are smaller than the TROBIT field sites, we calculated a TCC percent tree cover to match the TROBIT field site size by summing the percent tree cover within the TCC pixel part found inside the TROBIT field site and then dividing the sum by the TROBIT site area. As TROBIT site orientation was not recorded, we randomized the angle between the

TROBIT site and TCC pixel grid for each of the1000 samples when generating the probability distribution. "Enforced clumping" was performed as per MODIS VCF (Fig. 2)with the direction of clumping randomized. "

Figure r1, incorporated into the new Figure 2 illustrates the clumping procedure for Landsat TCC.

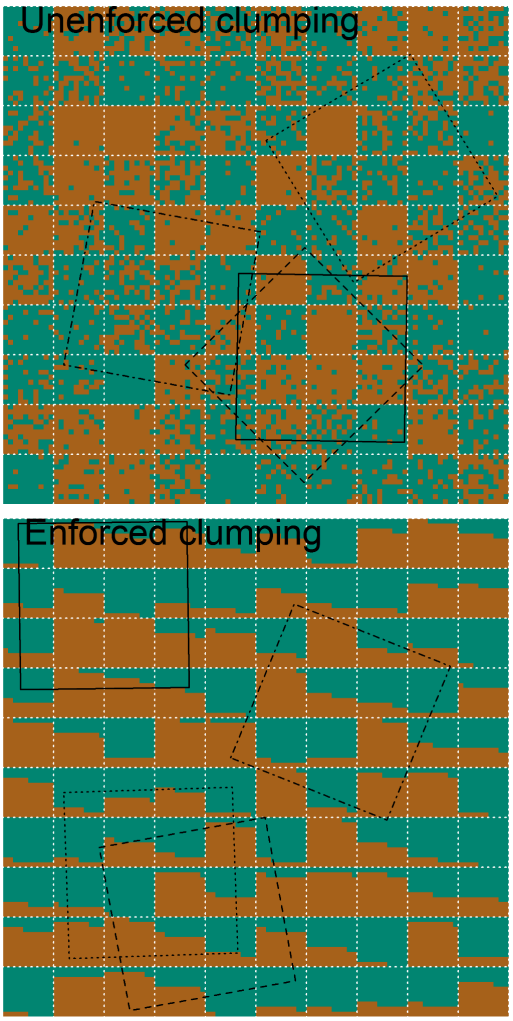

Figure r1: Example of the effects of unenforced and enforced clumping on 30 m x 30 m Landsat TCC pixels with a mix of tree covers (green) and non-tree cover (brown). White dotted lines are TCC pixel boundaries. Clumping all the cover to one side of the pixel (right bottom) affects the average canopy cover value of a 100 m x 100 m-sized TROBIT site (black boxes).

We found that using MODIS VCF to train the model used to estimate Landsat TCC tree cover propagates the MODIS VCF biases. We have highlighted this is in the text:

"Similar patterns can be observed with Landsat TCC (black line, Fig. 4). There is a significant underestimation of tree cover in the lower cover ranges up to 59% when there is enforced overlap, and up to 82% when overlap is not enforced. In savanna sites (orange line, Fig. 4) the underestimation (at 95% confidence) is significant and consistent for covers below 75 - 80 % (without enforced overlap) or below 52 - 60% (with enforced overlap). In forest sites (green line, Fig. 4) there is no systematic difference."

"While MODIS VCF is a powerful and accessible tool to map tree cover, our field data-based calibrations indicate that the latest MODIS VCF collection 6 is missing a lot of woody cover, even when uncertainty introduced by site canopy overlap and clumping within the MODIS VCF pixel are accounted for. The Landsat TCC product, which may be viewed as an alternative with a higher spatial resolution, behaves in a similar manner."

"Similar results for MODIS VCF and Landsat TCC also suggest that by training TCC with VCF tree cover these biases have been propagated into the finer-scale product. Models calibrated using MODIS VCF (Brandt et al., 2017; Lasslop et al., 2020; Burton et al., 2019; Kelley et al., 2019, 2021) also risk inheriting these biases and should therefore be validated using other sources of data. "

We also highlight and incorporate in our introduction and discussions the only other Landsat TCC evaluation carried out by Montesano et al., (2016):

Introduction:

"Likewise, validation of the finer-scale TCC product has been limited to its penultimate version and to the taiga-tundra circumpolar region (Montesano et al., 2016)."

"Similarly to MODIS VCF (Montesano et al., 2009), Montesano et al., (2016) revealed an overestimation of the taiga-tundra low tree covers in the finer-scale Landsat TCC, suggesting that using VCF as training has propagated these overestimations into the higher resolution product."

Discussions:

"According to its definition, MODIS VCF only maps trees that are 5 m or taller (Hansen et al. 2003), while the TROBIT CAI includes all trees with a minimum dbh of 2.5 cm, as well as trees with a height exceeding 1.5 m when dbh < 2.5 cm. This could explain our observed underestimation in the lower tree cover ranges for both MODIS VCF and Landsat TCC. In fact Montesano et al. (2016) showed an improved match between Landsat TCC and their lidar-derived tree cover reference data when reducing the height threshold from 5 m to 2 m. However, because of how our field

reference CAI is derived, we were not able to conclusively link the 5 m threshold to our observed underestimation."

"To ensure the appropriate use of both products, we suggest that where field data are available, the products should be calibrated for use in the target region. However, calibrating on a large scale using field data as a reference presents several challenges. Firstly, different in-situ measurement techniques tend to measure different types of tree cover (e.g. Fiala et al., 2006; Korhonen et al., 2006; Rautiainen et al., 2005) and each will require a specific conversion method to enable direct comparison with MODIS VCF or Landsat TCC. For example, Montesano et al. (2016) 's comparison did not acknowledge VCF's and thus TCC's 'within canopy gaps,' which may explain their observed underestimation in covers above 80%. "

Reviewer2:

Thank you to reviewer #2 for your further review. In response to your concerns about using a limited number of sites to represent a very varied tropical savanna-forest ecotone, we have further edited the text to highlight the limitations of the global extrapolation as follows:

We have altered the text in the introduction to: *"In this study, we evaluate MODIS VCF Collection 6 in tropical savannas and forest areas by comparing VCF's tree cover percentage to corresponding field data."*
"We then, for MODIS VCF, characterise the observed bias in woody covers across both savanna and forest ecosystems and apply our calibration across the tropics to highlight the regions most likely affected by these inaccuracies. We finish by discussing the implications the uncovered biases may have on tropical vegetation and terrestrial biogeochemical modelling."

We also refer to 'calibration' instead of 'correction' throughout the text. We have also replaced Figure 2 with a new figure (Figure 5) that combines Figure 2 and Figure A7 and now shows (i) the post-calibration change in tree cover that is statistically significant across all four scenarios, (ii) the uncertainty range of the calibration, (iii) priority regions for field surveying. The accompanying text in the results and parts of the discussion further stress the uncertainty of our calibration.