Thank you to reviewer #2 for your insightful review. In response to your concerns about applying our results across the tropics and discussing height thresholds, we have expanded our Discussion section to clarify the strengths and weaknesses of our approach in mapping uncertainty across the tropics, and added additional relevant information regarding plot heights. Reviewer comments are in italics, responses are in blue. Line numbers are from the original manuscript.

This manuscript undertakes an analysis of the accuracy of the MODIS Vegetation Continuous Fields product with particular reference to sparsely wooded ecosystems using tropical forest and savanna field inventory data.

The manuscript is well written and presented. The figures are of high quality and the analysis is clearly described. As the authors related the VCF has been subject to considerable analysis and discussion over some period of time. It could reasonable be postulated that all products from moderate resolution sensors struggle with accurate discrimination along the ecotone between forest and grassland due to non-linearilties in the reflectance and VIs applied to derive them, and due to the enormous variation in the morphology, architecture, density and clumping, and phenology of the overstory.

The authors have to work with a limited inventory data set - limited in geographical coverage, and limited in sampling the above mentioned variation. They quite reasonably focus on the issue of clumping of the tree fraction within pixels and undertake a nice analysis based on this. I do wonder why they did not explore the actual representativeness of the field inventory sites versus the VCF pixel resolution in the manner of Roman et al., 2009 and subsequent publications.

[M.O. Román, C.B. Schaaf, C.E. Woodcock, A.H. Strahler, X. Yang, R.H. Braswell, P.S. Curtis, K.J. Davis, D. Dragoni, M.L. Goulden The MODIS (collection V005) BRDF/albedo product: assessment of spatial representativeness over forested landscapes Remote Sens. Environ., 113 (2009), pp. 2476-2498, 10.1016/j.rse.2009.07.009]

This is not in any way disqualifying since the analysis here is coherent and valid in itself.

We chose to use ‘clumping’ and ‘overlap’ parameters to scale the field sites up to MODIS VCF pixel sizes because it allows us to simulate a huge variety of tree cover distribution, with enormous potential variation between the site and the pixel. We find that this is a good way to fully consider the very different potential structural compositions of savannas, particularly those within ecotones. The approach of Román et la. (2009) is an interesting technique that could be used to constrain uncertainty further by measuring the coefficient of variation between the site and pixel, and would be something to consider in future research.

However, I think that there are several issues that the authors need to address more fully especially in their Results and Discussion.

1. The manuscript proposes a correction to VCF for savannas and forests and combined. The authors state in the Results (lines 224-229) that VCF estimates forest tree cover well and greatly under-estimates savanna tree cover. However, based on Figure 1. the performance of VCF at the forest sites looks pretty terrible based on the sites being in a pretty clear blob with very wide variation between Trobit and VCF. This makes me wonder about the representativeness issue for these Trobit sites and VCF pixels and whether clumping issues *are the only thing happening here. In any case, Figure 1 does not suggest that VCF is doing well in forest but is this a sampling issue?*

On a site-by-site basis it is true that there are many forest points that do not appear to be accurately measured by MODIS VCF as seen by their distance to the 1:1 line. The main difference we want to highlight is that MODIS VCF behaves very differently in forests versus savannas: where in savannas it consistently overestimates, in forests it behaves much less consistently, both over- and under-estimating cover. The lack of significant tree cover gain or loss for forests, after correction, across all combinations of clumping/overlap (Fig. 3) further demonstrates this. We have edited the text in section 3.1 of the Results to clarify that MODIS VCF does not perform 'well' in the TROBIT ecotone forests, but it does measure cover without the systematic offset observed in the savannas, which is our focus in this paper.

We have also inserted text in the Methods section to clarify that the forest and savanna sites in TROBIT were selected because they represent the forest-savanna ecotone, and so while MODIS VCF appears to have some issues with TROBIT forests, this does not necessarily hold true for other forest types. Further work looking at the product's effectiveness across different forest types may yield greater insight into how MODIS VCF performs in case of denser forest types.

2. I am concerned about proposing an overall correction to VCF based on such a limited sampling of the ecotone between grassland and forest. I wonder if it would be better to more clearly identify in the written text, the kind of savanna that is sampled by the Trobit.

The figure is designed to show potential areas of VCF mis-estimates rather than a correction. Note that we only show difference and significance maps - not an overall VCF correction map. This is important to help inform future targeted VCF assessments. We have emphasised (within the description of Fig. 2) how our post-correction maps are meant to be used: as indicators for where the product may be more or less reliable.

Figure A1 shows the distribution of sites. Although the sites appear to sample the gradient between the amazon and the cerrado reasonably well (however there is enormous variation within cerrado from wooded to very open short sparse shrubland), they do not sample the variation in savanna structure in Africa and Australia very well. The African sampling is confined to the tree cover gradient in West Africa passing from the Guinean Savana to the Sahel, whilst the sampling in Australia is confined to areas around the small tropical rainforest area and surrounding Einasleigh Uplands. in northern Queensland. There is enormous variation in structure, morphology, and phenology across African and Australian savannas. I would find the study more compelling if it: 1) paid more attention to the actual species, structure , phenology and composition of the sites; 2) constrained the narrative to an analysis of VCF for these particular systems; and 3) suggested a correction approach that is relevant to these kinds of systems and maintains the focus on the issue of sparsly wooded systems. I believe that it is a stretch to propose an overal correction to VCF from this study simply becuse of the available inventory data are limited in coverage of geographical and vegetation diversity.

We chose to do a tropic wide analysis as it is the ecological modelling studies carried out at these scales that often use VCF without considering VCF's underlying uncertainties (see references in the introduction, and discussion in the original m/s). The main purpose of our paper is to highlight to the global vegetation modelling community that VCF may have significant and systematic errors even when you take a broad view, and to encourage caution in the use of VCF. The reviewer and our introduction provide references where local-scale VCF assessments have been conducted, but none have translated their results demonstrating tropics-wide spatial implications.

However, the reviewer is rightly concerned that our analysis could be construed as a usable "correction map" of tropics-wide VCF. We placed several safeguards against this in the original m/s:

1. We only show difference maps, and no definitive "correction map"
2. Figure 2 maps are for all four clumping/overlap scenarios, and although these show qualitatively similar general patterns, there are also significant quantitative differences between them.
3. Disagreement between these maps are highlighted by the "Significance" map in Fig. 2
4. Much of the discussion is dedicated to constraining these uncertainties between scenarios further.

In addition, in the revised m/s, we make a number of changes and additions to make clear that we are presenting "areas of concern" when using VCF and not definitive corrected VCF maps:

1. We removed any reference to "corrected VCF" when discussing tropics-wide maps.
2. In the discussion, we included a map (Figure r2.1 - to be included as Fig A6) showing where to prioritise future in-situ reference data collection to help constrain our uncertainty maps further. We also included the following text in the discussion:

   "Finally, using a limited number of field plots will create additional uncertainty when the calibration is applied across the tropics. To identify where additional field plots would reduce this uncertainty we combined our uncertainty maps (Fig. 2) with distance from TROBIT plots (See Fig A6 in appendix). "

3. Additionally, as the reviewer points out, vegetation structure, morphology and phenology are all variables that could make a difference. Yet, although inter- and intra-continental differences in these variables between TROBIT plots are already quite large (Torello-Raventos et al. (2013) and Veenendaal et al. 2015), all continents show similar patterns across our clumping/overlap scenarios for savanna, i.e. a smaller underestimation or slight overestimation of tree cover at small and large covers and a substantial underestimation at intermediate covers (Fig. r2.2). To acknowledge the impact of plot representativeness (within the highly variable global tropical forest-savanna ecotone) on our analysis we have added the following text in our Discussion:

> "Using a limited number of field plots will create additional uncertainty when the calibration is applied globally across this highly variable tropical forest-savanna ecotone. The map in Fig. A6 combines our uncertainty maps (Fig. 2) with distance from TROBIT plots, and highlights Southeast Asia, Central America, and Mexico as areas where additional in-situ tree cover observations would help constrain uncertainties. Field data from the northwestern region of South America, the southeast of the African continent, and Madagascar would also help. Finally, factors such as topography, soil type and moisture content, phenology and cloud cover, and landscape heterogeneity, can affect the accuracy of remotely-sensed products, including MODIS

VCF. Data, characterising these at plot level, would help identify potential confounding factors affecting MODIS VCF performance and so help constrain uncertainties further."

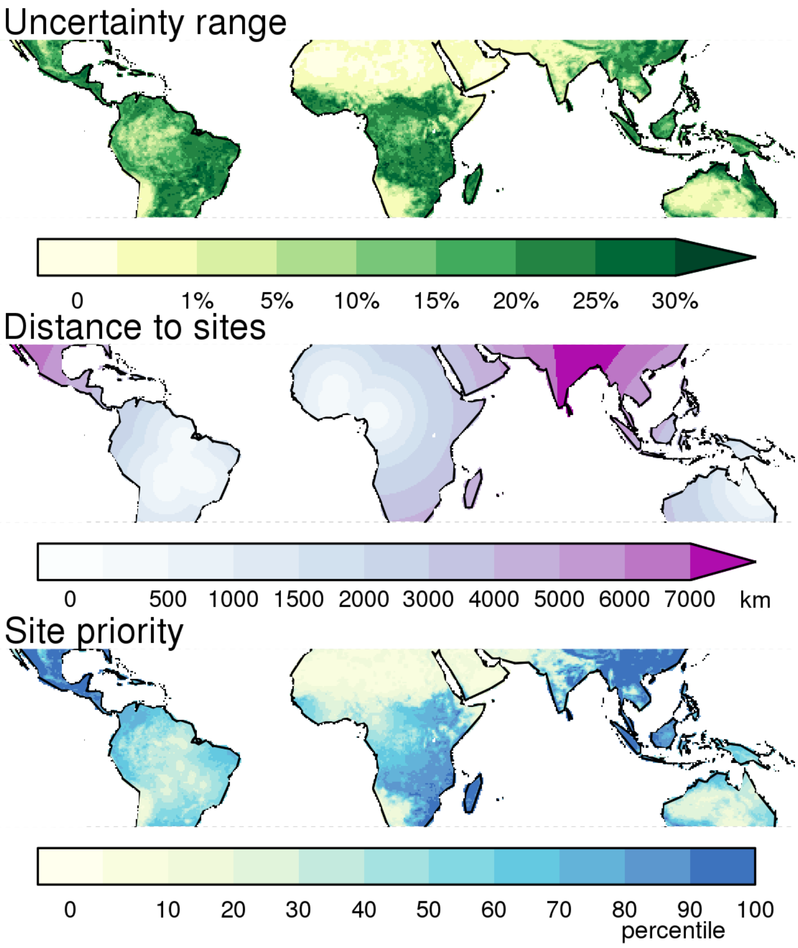

Fig r2.1. (Top) uncertainty range of potential VCF mismatch, calculated as maximum 90% percentile minus minimum 10% quantile over the four scenarios in Fig. 2. (Middle) geographic distance to closest TROBIT site. (Bottom) based on top and middle maps, priority areas for additional data collection to further constrain map uncertainties of Fig. 2.

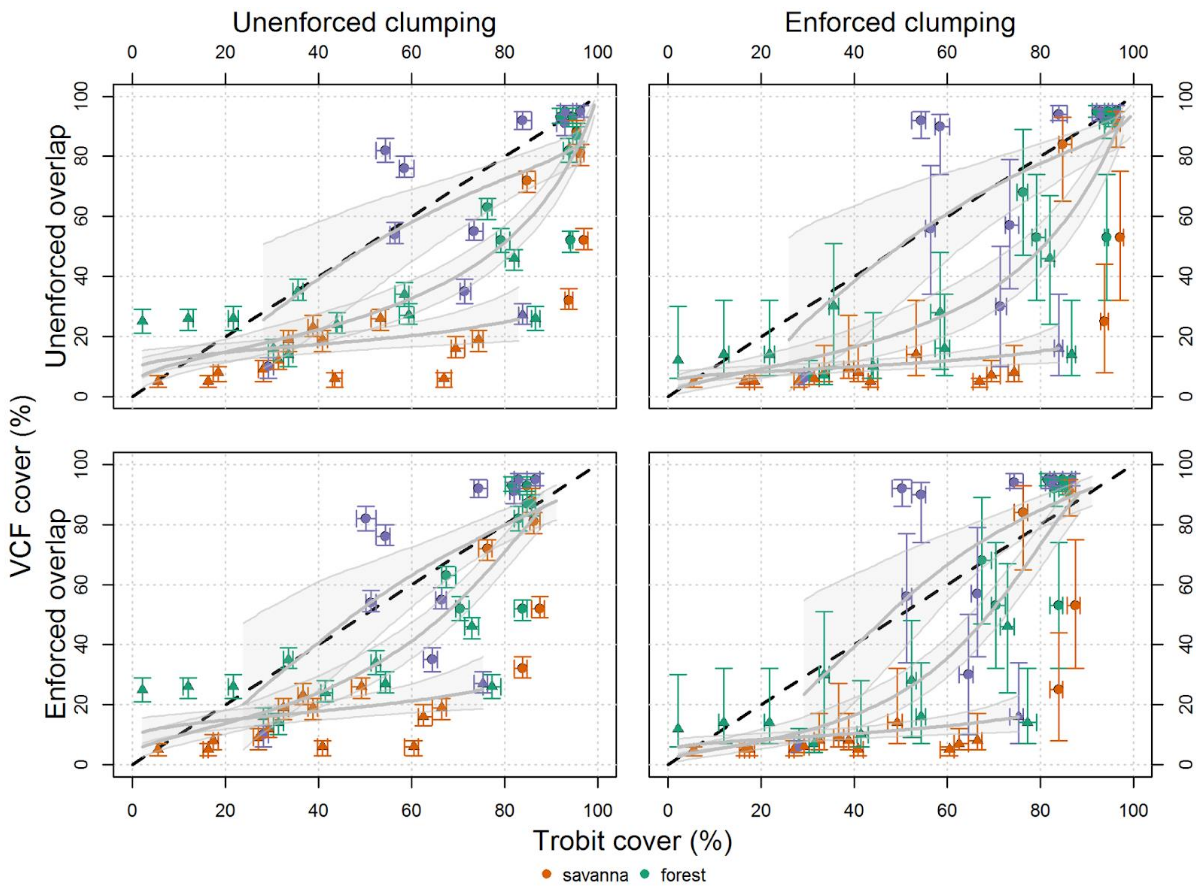

Fig r2.2: TROBIT vs VCF as per Fig 1 in the m/s, but with points coloured by continent: South America in green, Africa in red, Australia in purple. Triangles are savanna, circles are forests. Greyed regression polygons match the coloured forest, savanna and overall regression cures in Fig. 1.

3. A key issue with VCF and this analysis is the thorny one of the definition of tree cover. Really with VCF and any other remote sensing products it should be about detection of woody cover (of any sort) from the canopy reflectance which is distinguished from the background soil and understory wherein the sensitivity is constrained by pixel resolution and the discriminatory capacity of available reflectance bands. I think that the manuscript gets a bit bogged down in this thorny issue since VCF has a certain definition which is apparently based on height. This is problematice in many ways since although related, height does not have a one to one relationship with canopy extent. However this MODIS product has been around for a long time with clear definition (of height). This rather means that analysis using these Trobit sites needs to provide information on the "tree height" distributions at these sites. This once again returns to representativeness, and maybe explains why the forest data are so scattered for Trobit versis VCF in Figure 1 . So again I want more information about what the vegetation is at these Trobit sites. As a result, the Discussion is a bit convoluted. The section between lines 301 and 319 therefore is rather confusing and muddled to read. A whole lot of speculation about where trees are > 5 m or not is included. To make statements about this one needs evidence of the tree height distribution at the Trobit sites and a matching analysis. I treat VCF as attempting to estimate woody cover. I don't care if it is getting at trees > 5 m or less.

So this is kind of a furphy. The results in Figure 1 for savannas are pretty convincing about underestimation of woody cover which is probably unsurprising given the VCF method, but readily corrected based on your analysis at least for these particular savanna systems. If the Trobit data are limited then a whole lot of speculation about tree height really is not relevant and just clouds the Discussion. It damages the message which is best from the savanna analysis. Hence I think that clearer and more simplified findings and discussion points are required.

We agree with the reviewer that for RS-based products, such as VCF and MCD12Q1, that are derived from optical sensor data (i.e. surface reflectance), implementing a height threshold will be difficult. However, this height threshold is clearly stated in the VCF's definition and, similarly to the % cover thresholds in the IGBP land cover class definitions, is taken at face value by field-based scientists and modellers. As a matter of fact, model outputs are often adapted to match this 5 m threshold (e.g. Kelley et al., 2013; Lasslop et al., 2018). It is therefore important we discuss height. And while reviewer #2 is correct in saying that we cannot conclusively prove that height is a factor at play with MODIS VCF's accuracy with the limited data at hand, we find it important to highlight the importance of caution when assuming the 5 m limit.

However, we also agree that there was room for improvement in the manner in which we tackled this thorny issue and have improved our text in the Results and Discussion part of the manuscript. We have now provided the TROBIT plot heights and cover type by Torello-Raventos et al. (2013) in Table A1. Using this height information, we have also shown there is no relationship between TROBIT height and the difference between VCF and TROBIT covers (Fig r2.3 - to be included at Fig. A7 in the revised m/s). While this does not definitively disprove the existence of an assumed height threshold, the fact that there is no reduction in error as the upper stratum exceeds 5m warrants further investigation into the 5m limit.

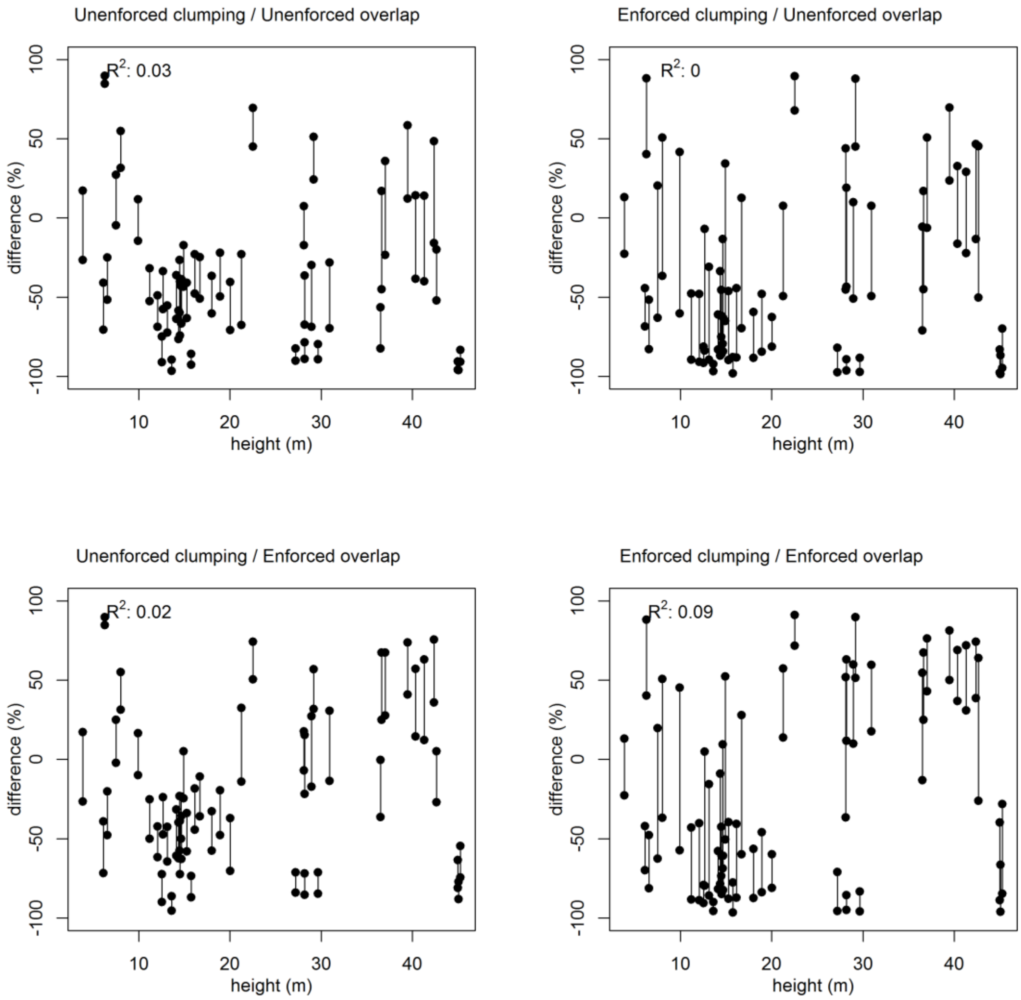

Fig r.2.3: TROBIT upper stratum height vs difference between VCF and TROBIT % cover over our four clumping and overlap scenarios. Upper and low bars represent 10-90 percentiles of uncertainty range based on convolution of VCF and TROBIT covers uncertainties from Fig. 1.

In summary, the study is well written, the methods are good, and the presentation is excellent. It is an interesting and useful study. However the authors need to address the following major issues before it can be considered for publication.

1. Limit the scope to developing correction to VCF that can be applied in the systems where Trobit inventory is available and representative.

2. Address the issue of representativeness of the Trobit sites, describe them much more fully, provide information of tree heights or tree height distributions or information from the literature on typical structure for these types of savannas and forests.

3. Focus the correction of the savannas for which your analysis is more convincing and explain the variation in the forest tree cover and why you think that VCF is good based on Figure 1.

4. Simplify the issues around definitions and detection of "trees". If you have no height data you can't really comment on this. Make sure that the thrust of the Results and Discussion is clear and clean and does not jump around between phrases about height definition when the imagery is seeing tree canopies not heights. If you can better describe the height distributions and representativeness of the sites, then discussion of the height definition can be useful.

In conclusion, these are the steps we have taken to address your concerns:

1. We will continue to map our findings across the tropics as this scale is one that best informs work done using VCF in ecological modelling, but we have made our limitations very clear and have emphasised that our findings are a guide to where there is likely more or less uncertainty, and not a correction within itself.
2. Additional information regarding TROBIT plots have been added to the paper, describing the types of plots selected for the project on the whole, with relevant plot-specific details added to table A1.
3. We have clarified that VCF use in savannas is more concerning than in forests because of a persistent systematic offset, and better described how VCF performs in forests in the Results section.
4. The Discussions section has been re-worked to better explain that the assumed height threshold in VCF has not been reflected in our results, and why it is an important issue to discuss.

References:

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B. and Willis, K. O.: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10(5), 3313–3340, doi:https://doi.org/10.5194/bg-10-3313-2013, 2013.

Lasslop, G., Moeller, T., D'Onofrio, D., Hantson, S. and Kloster, S.: Tropical climate–vegetation–fire relationships: multivariate evaluation of the land surface model JSBACH, Biogeosciences, 15(19), 5969–5989, doi:https://doi.org/10.5194/bg-15-5969-2018, 2018.

Torello-Raventos, M., Feldpausch, T., Veenendaal, E., Schrodt, F., Saiz, G., Domingues, T., Djagbletey, G., Ford, A., Kemp, J., Marimon, B., Marimon-Junior, B. H., Lenza, E., A Ratter, J., Maracahipes, L., Sasaki, D., Sonké, B., Zapfack, L., Taedoumg, H., Daniel, V. and Lloyd, J.: On the delineation of tropical vegetation types with an emphasis on forest/savanna transitions, Plant Ecology & Diversity, 6(1), 101–137, doi:10.1080/17550874.2012.762812, 2013.

Veenendaal, E. M., Torello-Raventos, M., Feldpausch, T. R., Domingues, T. F., Gerard, F., Schrodt, F., Saiz, G., Quesada, C. A., Djagbletey, G., Ford, A., Kemp, J., Marimon, B. S., Marimon-Junior, B. H., Lenza, E., Ratter, J. A., Maracahipes, L., Sasaki, D., Sonké, B., Zapfack, L., Villarroel, D., Schwarz, M., Yoko Ishida, F., Gilpin, M., Nardoto, G. B., Affum-Baffoe, K., Arroyo, L., Bloomfield, K., Ceca, G., Compaore, H., Davies, K., Diallo, A., Fyllas, N. M., Gignoux, J., Hien, F., Johnson, M., Mougin, E., Hiernaux, P., Killeen, T., Metcalfe, D., Miranda, H. S., Steininger, M., Sykora, K., Bird, M. I., Grace, J., Lewis, S., Phillips, O. L. and Lloyd, J.: Structural, physiognomic and above-ground biomass variation in savanna–forest transition zones on three continents – how different are co-occurring savanna and forest formations?, Biogeosciences, 12(10), 2927–2951, doi:10.5194/bg-12-2927-2015, 2015.