Gerard, France; Hooftman, Danny; van Langevelde, Frank; Veenendaal, Elmar; White, Steven M.; Lloyd, Jon. 2017. **MODIS VCF should not be used to detect discontinuities in tree cover due to binning bias. A comment on Hanan et al. (2014) and Staver and Hansen (2015)**. *Global Ecology and Biogeography*, 26 (7). 854-859. 10.1111/geb.12592

Contact CEH NORA team at
noraceh@ceh.ac.uk

Title: MODIS VCF should not be used to detect discontinuities in tree cover due to binning bias – a comment on Hanan *et al.* (2014) and Staver and Hansen (2015)

Authors: France Gerard[1], Danny Hooftman[1,2], Frank van Langevelde[3], Elmar Veenendaal[4], Steven M. White[1,5], Jon Lloyd[6,7].

[1]Centre for Ecology and Hydrology Wallingford, Crowmarsh Gifford, OX10 8BB, UK, ffg@ceh.ac.uk
[2]Lactuca: Environmental Data Analyses and Modelling, Speenkruid 100, 1112NC, Diemen, The Netherlands
[3]Resource Ecology group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[4]Plant Ecology and Nature Conservation, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[5]Wolfson Centre for Mathematical Biology, Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford, Oxfordshire OX2 6GG, UK
[6]Department of Life Sciences, Faculty of Natural Sciences, Imperial College (Silwood Park), Buckhurst Road, Ascot, SL5 7PY, UK
[7]Centre for Tropical Environment and Sustainability Sciences (TESS) and College of Marine and Environmental Sciences, James Cook University, Cairns, Qld, Australia

Corresponding author: France Gerard
Number words Abstract: 195
Number words main body of text: 2047
Number of references: 22

ABSTRACT

In their recent paper, Staver and Hansen (*Global Ecology and Biogeography*, 2015, 24, 985-987) refute the case made by Hanan *et al.* (*Global Ecology and Biogeography*, 2014, 23, 259–263) that the use of classification and regression trees (CARTs) to predict tree cover from remotely sensed imagery (MODIS VCF) inherently introduces biases, thus making the resulting tree cover unsuitable for showing alternative stable states through tree cover frequency distribution analyses. We here provide a new and equally fundamental argument why the published frequency distributions should not be used for such purposes. We show that the practice of pre-average binning of tree cover values used to derive cover values to train the CART model will also introduce errors in the frequency distributions of the final product. We demonstrate that the frequency minima found at tree covers 8 % to 18 %; 33 % to 45 %; and 55 % to 75 % can be attributed to numerical biases introduced when training samples are derived from landscapes containing asymmetric tree cover distributions and/or a tree cover gradient. So it is highly likely that the CART, used to produce MODIS VCF, delivers tree cover frequency distributions that do not reflect the real world situation.

Main body of text:

The MODIS VCF tree cover product of Hansen *et al.* (2002b) provides worldwide estimates of percentage tree cover derived from MODIS data. Discontinuities in tree cover frequency distributions derived from this data have been used to support the hypothesis that the observed distribution of forest, savanna and grassland vegetation in the tropical and boreal regions of the world represent alternative stable states for equivalent environmental conditions (Hirota *et al.*, 2011; Staver *et al.*, 2011; Favier *et al.*, 2012; Murphy & Bowman, 2012; Scheffer *et al.*, 2012; Xu *et al.*, 2016). But recently Hanan *et al.* (2014, 2015) suggested that the adopted classification and regression trees (CART) approach, used to produce the MODIS VCF tree cover estimates, introduced a systematic bias which makes the MODIS VCF product inappropriate for the analysis of % tree cover frequency distributions. This point was countered by Staver and Hansen (2015) arguing (i) that the approach taken by Hanan *et al.* (2014), using simulated EO data and pseudo satellite metrics to demonstrate that an artificial bias is generated by the CART approach, does not reflect the complexity and variability of landscapes and vegetation across the globe and (ii) that the CART model used by Hanan *et al.* (2014) was highly pruned with very few nodes (nodes = 9) compared to that used for the MODIS VCF product (nodes = 109) resulting in a less smooth gradient of % tree cover values. In this analysis we show that the VCF tree cover product likely also contains systematic bias which is introduced in the pre-processing of calibration data used to train the CART model. Such errors would exacerbate the problem of aggregation in the CART predictions already demonstrated by Hanan *et al.* (2014).

The VCF product is the result of a continuously developing method which has been documented in a succession of publications (DeFries *et al.*, 1999; Defries *et al.*, 2000; Hansen *et al.*, 2002a; Hansen *et al.*, 2002b; Hansen *et al.*, 2003; Hansen *et al.*, 2005). The most recent version is referred to as MOD44B collection 5, it is available at a 250 m resolution and supersedes the previous collections, including the 500 m MOD44B collection 3. Collection 5 was used in the most recent study of Xu *et al.* (2016), while collection 3 was used in the majority of the publications about discontinuities in tree-cover distributions (Hirota *et al.*, 2011; Staver *et al.*, 2011; Favier *et al.*, 2012; Murphy & Bowman, 2012; Scheffer *et al.*, 2012).

The approach used to create the VCF product relies on two critical components: (i) the creation of training samples and (ii) the implementation of a CART model that derives % tree cover from a collection of MODIS-based metrics. The design (number of nodes, the choice of regression variables from the pool of MODIS-based metrics, and the regressions) of the CART model is determined by the training samples.

In other words, during training, the design of the CART is tailored to best reproduce the % tree cover values of the training samples. The method for deriving the samples, that are used to train the CART model, has remained the same between collections (DiMiceli *et al.*, 2011; Townshend *et al.*, 2011) and so the argument below applies to all VCF versions, including the most recent MOD44B collection 5.

Summarizing from Hansen *et al.* (2002b) (also in VCF User Guide), training samples are created as follows: 30 m Landsat Thematic Mapper image pixels are assigned one of four discrete % tree cover classes (0 %, 25 %, 50 %, 80 %) with their boundaries defined as (0-10 %, 11-40 %, 41-60 %, 61-100 %) respectively. These classified 30 m Thematic Mapper images are then re-gridded to produce 500 m (or 250 m) training pixels, matching the size of the MODIS pixels. The resulting % tree cover values represent the average of the 30 m pixel class values contained within each 500 m pixel. In short, the training values represent weighted averages of the discrete values 0 %, 25 %, 50 % and 80 %, with the weights defined by the number of 30 m pixels found within the 500 m pixels having one of the respective four values. We here evaluated the impact of deriving a % tree cover gradient in this manner, which we will henceforth refer to as 'averaging with pre-average binning', by first (i) considering a range of one-dimensional cover distributions through a semi-analytical experiment and then (ii) evaluating the effect of spatial auto-correlation through a two-dimensional Monte Carlo simulation experiment.

The semi-analytical experiment involved simulations using the Beta distributions (Taboga, 2012). The Beta distribution enabled us to represent a wide variety of possible continuous tree cover distributions found within a training pixel (i.e. 500 m pixel) by varying the shape parameters α and β of the distribution (see Fig. S1.1) and the % tree cover range (i.e. 0-100 %, 0-80 %, 20-100 %) (see Appendix S1 in Supporting Information for detail). Pre-average binning into bin-classes 0.5 % ($\sim$ 0 %), 25 %, 50 % and 80 % is achieved through a subdivision of the distribution $\boldsymbol{B}(\alpha, \beta)$ into four pieces, using three break points $z_1 = 0.1$; $z_2 = 0.4$ and $z_3 = 0.6$ and calculating the inferred mean $\left\langle \hat{X} \right\rangle$:

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} \frac{z_1}{2} \mathbf{B}(\alpha,\beta) dx + \int_{z_1}^{z_2} \left(z_1 + \frac{z_2-z_1}{2}\right) \mathbf{B}(\alpha,\beta)\, dx + \int_{z_2}^{z_3} \left(z_2 + \frac{z_3-z_2}{2}\right) + \int_{z_3}^{1} \left(z_3 + \frac{1-z_3}{2}\right) \mathbf{B}(\alpha,\beta) dx}{\int_0^1 \mathbf{B}(\alpha,\beta) dx} \quad (1)$$

where $\mathbf{B}(\alpha, \beta) = x^{(\alpha-1)}(1 - x)^{(\beta-1)}$ and α and β are the shape parameters ($\gtrless 0$); the denominator $\int_0^1 \boldsymbol{B}(\alpha, \beta) dx$ can be regarded as a normalising factor. Full details of the derivation of Eqn 1 can be found in Appendix S1 in Supporting Information. Testing for any bias was undertaken through a comparison of the inferred mean $\left\langle \hat{X} \right\rangle$ with the true mean $\bar{x}$. Specifically, we tested for many random

values of α and β (> 0.5 and ≤ 6) postulating that if there was no bias introduced through binning, the inferred mean would match the true mean.

Results from this semi-analytical experiment clearly demonstrate that binning % tree covers prior to averaging must introduce biases in the CART training samples when the original tree cover distribution within a tile is asymmetric. It is only when the distribution of % tree cover is perfectly symmetric that no error in the binned averages will occur (see Eqn A5 in Appendix S1 in Supporting Information). The magnitude of the bias and its location across the % tree cover axis is defined by the shape of the distribution and the % tree cover range within the tile, but a consistent tendency is for an overestimation in averages near the 20 % to 30 % and 65 % cover values and underestimation in averages near the 10 % to 15 % and the 40 % to 50 % cover values (Fig. 1). This pattern, although small, matches the discrepancy observed in the validation data for Africa shown in Fig. 1(b) of Staver and Hansen (2015).

Monte Carlo simulations were designed to deliver theoretical two-dimensional landscapes of % tree cover, in the form of a 1000 x 1000 grid of values in a similar fashion to the above approach, but in this case allowing for spatial auto-correlation effects to be examined. From these, training pixels % tree cover values were derived from averaging 20 x 20 grid cells (e.g. ~30 m Landsat pixels) into larger sized tiles (e.g. 500 m MODIS pixels). Tree covers were either binned or kept un-binned prior to the averaging into tiles. Again, we reasoned that if there was no bias introduced through binning, the variation in % tree cover over the resulting training pixels would be similar without a skew towards specific values of % tree cover. We focused on the effects of spatial auto-correlation, which is inherent in most landscapes (Gomez-Sanz *et al.*, 2014) and in real tree cover gradients, by modelling and comparing (i) fully random Uniform and Beta distributions of tree cover, (ii) a sharp boundary landscape, in which half of the landscape was assigned a 100 % tree cover and the remaining half a 0 % tree cover, (iii) a patchy landscape of auto-correlated Beta distributions across the 0 % to 100 % tree cover range, (iv) a linear gradient of tree cover decreasing from 100 % to 0 % tree cover along the landscape x-axis, and (v) a complex gradient representing Beta distributions across the 0 % to 100 % tree cover range. More detail about the Monte Carlo simulations with example landscapes is given in Appendix S2 in Supporting Information.

Noting first that, when the two-dimensional simulations were set up to produce the same random non-spatially auto-correlated landscapes as in the semi-analytical experiment, the Monte Carlo approach gives almost identical results as the application of Eqn 1 (Fig. S2.4). It further seems that there is the introduction of a second strong bias effect in the case of landscapes with a tree cover gradient (Fig. 2). This result is considered in more detail in Appendix S2 and shows that in such a situation pre-average

binning causes the training pixel values to inevitably converge towards the bin-class averages. In other words, in gradually changing landscapes, the binning procedure considerably lowers the variation in % tree cover. This results in a dominant proportion of training pixels with bin-class average values of 0 %, 25 %, 50 % and 80 %. Hence, binning tree cover values before averaging will inevitably underrepresent spatial variation and gradual transitions in tree cover.

Having been trained by biased samples, the CART model must necessarily propagate these biases to the resulting % tree cover map. However because of (i) the complexity of real landscapes, (ii) the large number of nodes in the model, (iii) the introduction of a stepwise regression at each node to deliver a more contiguous range of values (Hansen *et al.*, 2002b), and (iv) the variability in the uncertainties of the predicted tree cover, this bias will not always be immediately apparent. Of the published data we know of, the clearest evidence of a bias can be observed in (Hansen *et al.*, 2005) and explains the 38 %-45 % discontinuity shown in Fig.1(b) of Staver and Hansen (2015). Hansen *et al.* (2005) produced, applying the VCF approach, 500 m % tree cover maps for a region in Zambia using 500 m training samples that did not undergo pre-average binning at the 30 m Thematic Mapper pixel scale. Comparison with independent validation data show that in this case, the CART variants being tested were able to reproduce the continuous tree cover gradient observed in the validation data (Fig. 3(a)). But when the same validation data was used to evaluate the MODIS VCF collection 3 created using training samples that had undergone pre-average binning - Fig. 3(b) and also Fig.1(b) in Staver and Hansen (2015) - a clear reduction in values near the 10 % and 40 % cover range is revealed, indicating a bias. Similarly, the scattergram in Sexton *et al.* (2013) that compares the MODIS VCF collection 5 with validation data for 3 sites in America and one in Costa Rica suggests there may be a slight frequency minimum around the 60 % tree cover range, matching the conditions shown in Fig. 1 (g), (h) and (i). However without a statistical test for unimodality (Hartigan & Hartigan, 1985) the data in this case is inconclusive.

The bias will not always be readily noticeable. Moreover, the bias will not exist over areas where the corresponding training samples were derived from areas where there was no gradient and/or the tree cover distributions were symmetrical. As an example, the validation VCF scattergrams for the taiga-tundra transition zone, shown in Montesano *et al.* (2009) for the MODIS VCF collection 4, reveal no clear clustering of tree covers around the bin-class averages.

We have demonstrated here that through the combined semi-analytical and two-dimensional Monte Carlo simulations and through data published by Hansen *et al.* (2005) that tree cover frequency minima and maxima could be caused by a bias in the samples used to train the CART model. This bias will be present

in all MODIS VCF collections, in areas where the majority of the corresponding training samples were derived from landscapes that have an asymmetric tree cover distribution, contain a tree cover gradient, or a combination of both. This effect is likely to be further enhanced by the CART through the inherent aggregation of predicted values around nodal means. In support of Hanan *et al.* (2014), we argue that the MODIS VCF tree cover product should not be used to detect discontinuities in tree cover as, although landscapes vary across the globe, a majority will contain local tree cover distributions that are asymmetric and some will contain tree cover gradients.

ACKNOWLEDGEMENTS

REFERENCES

DeFries, R.S., Townshend, J.R.G. & Hansen, M.C. (1999) Continuous fields of vegetation characteristics at the global scale at 1-km resolution. *Journal of Geophysical Research-Atmospheres*, **104**, 16911-16923.

Defries, R.S., Hansen, M.C., Townshend, J.R.G., Janetos, A.C. & Loveland, T.R. (2000) A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, **6**, 247-254.

DiMiceli, C.M., Carroll, M.L., Sohlberg, R.A., Huang, C., Hansen, M.C. & Townshend, J.R.G. (2011) Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover. In: (ed. U.O. Maryland), College Park, MD, USA.

Favier, C., Aleman, J., Bremond, L., Dubois, M.A., Freycon, V. & Yangakola, J.-M. (2012) Abrupt shifts in African savanna tree cover along a climatic gradient. *Global Ecology and Biogeography*, **21**, 787-797.

Gomez-Sanz, V., Bunce, R.G.H. & Elena-Rossello, R. (2014) Landscape assessmenst and monitoring. *Forest Landscapes and Global Change* (ed. by J. Azevedo, A.H. Perera and M.A. Pinto), pp. 199-. Springer Science + Business Media, New York.

Hanan, N.P., Tredennick, A.T., Prihodko, L., Bucini, G. & Dohn, J. (2014) Analysis of stable states in global savannas: is the CART pulling the horse? *Global Ecology and Biogeography*, **23**, 259-263.

Hanan, N.P., Tredennick, A.T., Prihodko, L., Bucini, G. & Dohn, J. (2015) Analysis of stable states in global savannas – a response to Staver and Hansen. *Global Ecology and Biogeography*, **24**, 988-989.

Hansen, M.C., Townshend, J.R.G., Defries, R.S. & Carroll, M. (2005) Estimation of tree cover using MODIS data at global, continental and regional/local scales. *International Journal of Remote Sensing*, **26**, 4359-4380.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Marufu, L. & Sohlberg, R. (2002a) Development of a MODIS tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, **83**, 320-335.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., Dimiceli, C. & Carroll, M. (2002b) Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, **83**, 303-319.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C. & Sohlberg, R.A. (2003) Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. *Earth Interactions*, **7**

Hartigan, J.A. & Hartigan, P.M. (1985) The DIP test of unimodality. *Annals of Statistics*, **13**, 70-84.

Hirota, M., Holmgren, M., Van Nes, E.H. & Scheffer, M. (2011) Global Resilience of Tropical Forest and Savanna to Critical Transitions. *Science*, **334**, 232-235.

Montesano, P.M., Nelson, R., Sun, G., Margolis, H., Kerber, A. & Ranson, K.J. (2009) MODIS tree cover validation for the circumpolar taiga-tundra transition zone. *Remote Sensing of Environment*, **113**, 2130-2141.

Murphy, B.P. & Bowman, D.M.J.S. (2012) What controls the distribution of tropical forest and savanna? *Ecology Letters*, **15**, 748-758.

Scheffer, M., Hirota, M., Holmgren, M., Van Nes, E.H. & Chapin, F.S. (2012) Thresholds for boreal biome transitions. *Proceedings of the National Academy of Sciences*, **109**, 21384-21389.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C. & Townshend, J.R. (2013) Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, **6**, 427-448.

Staver, A.C. & Hansen, M.C. (2015) Analysis of stable states in global savannas: is the CART pulling the horse? – a comment. *Global Ecology and Biogeography*, 1-3.

Staver, A.C., Archibald, S. & Levin, S.A. (2011) The Global Extent and Determinants of Savanna and Forest as Alternative Biome States. *Science*, **334**, 230-232.

Taboga, M. (2012) *Lectures on probability theory and mathematical statistics*. CreateSpace Independent Pub.

Townshend, J.R.G., Hansen, M.C., Carroll, M.L., DiMiceli, C.M., Huang, C. & Sohlberg, R.A. (2011) User Guide for the MODIS Vegetation Continuous Fields product Collection 5 version 1. In, p. 12. University of Maryland, MD, USA.

Xu, C., Hantson, S., Holmgren, M., van Nes, E.H., Staal, A. & Scheffer, M. (2016) Remotely sensed canopy height reveals three pantropical ecosystem states. *Ecology*, n/a-n/a.

SUPPORTING INFORMATION

Appendix S1. Semi-analytical simulations
Appendix S2. Monte Carlo simulations
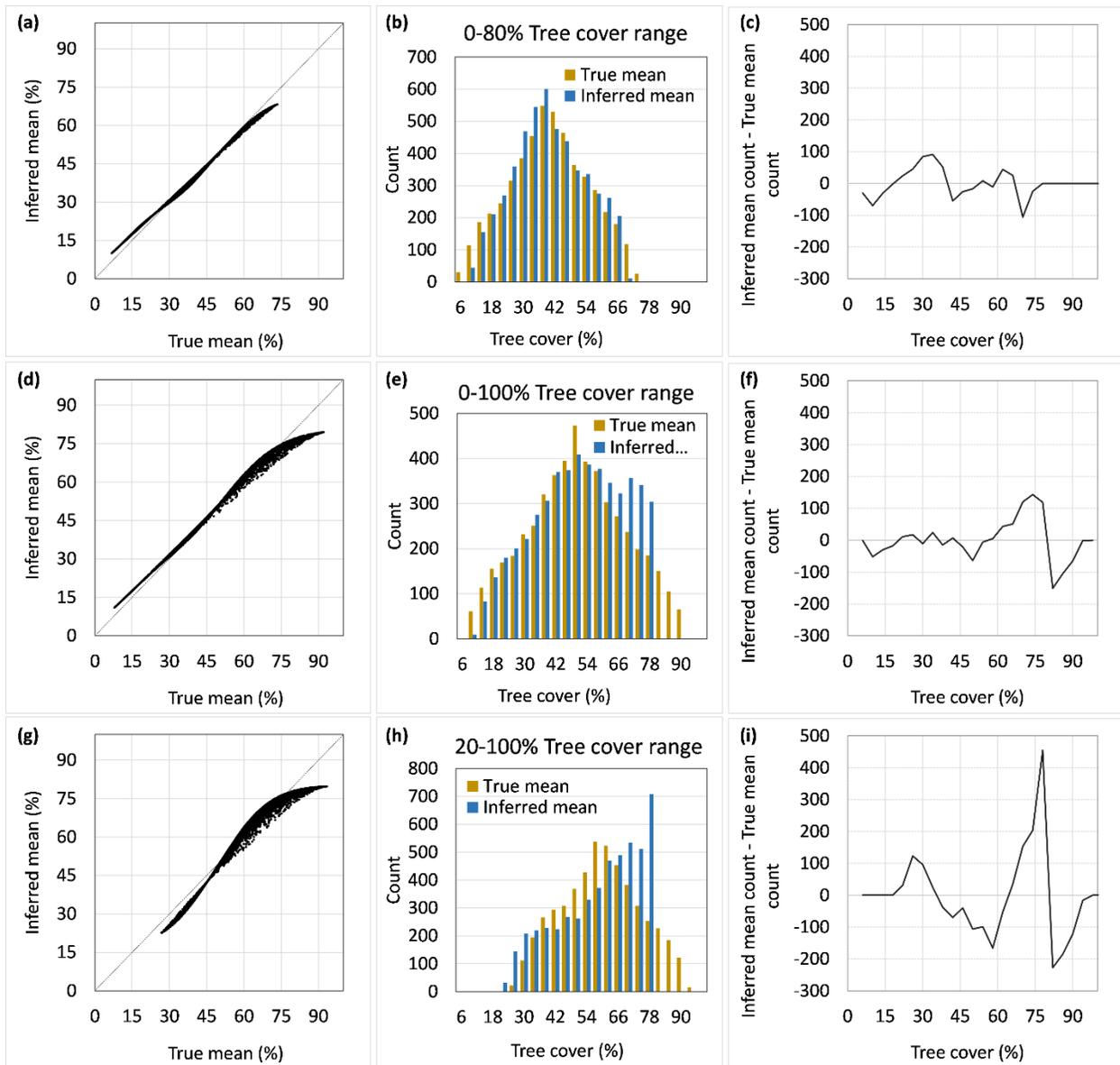
Figure 1. (a), (d) and (g) Scatter plots of training sample values (e.g. 500 m pixel) derived from averaging across smaller units (e.g. 30 m pixels) with (Inferred mean) and without (True mean) pre-average binning; (b), (e) and (h): the histogram counts of the sample values; (c), (f) and (i) the difference in histogram counts. The inputs are 1000 simulated Beta distributions made to fit a tree cover range of 0 % - 80 % (a, b and c); 0% - 100 % (d, e and f) and 20 % - 100 % (g, h and i) respectively, with the Beta shape variables α and β varying between 0.5 and 6.
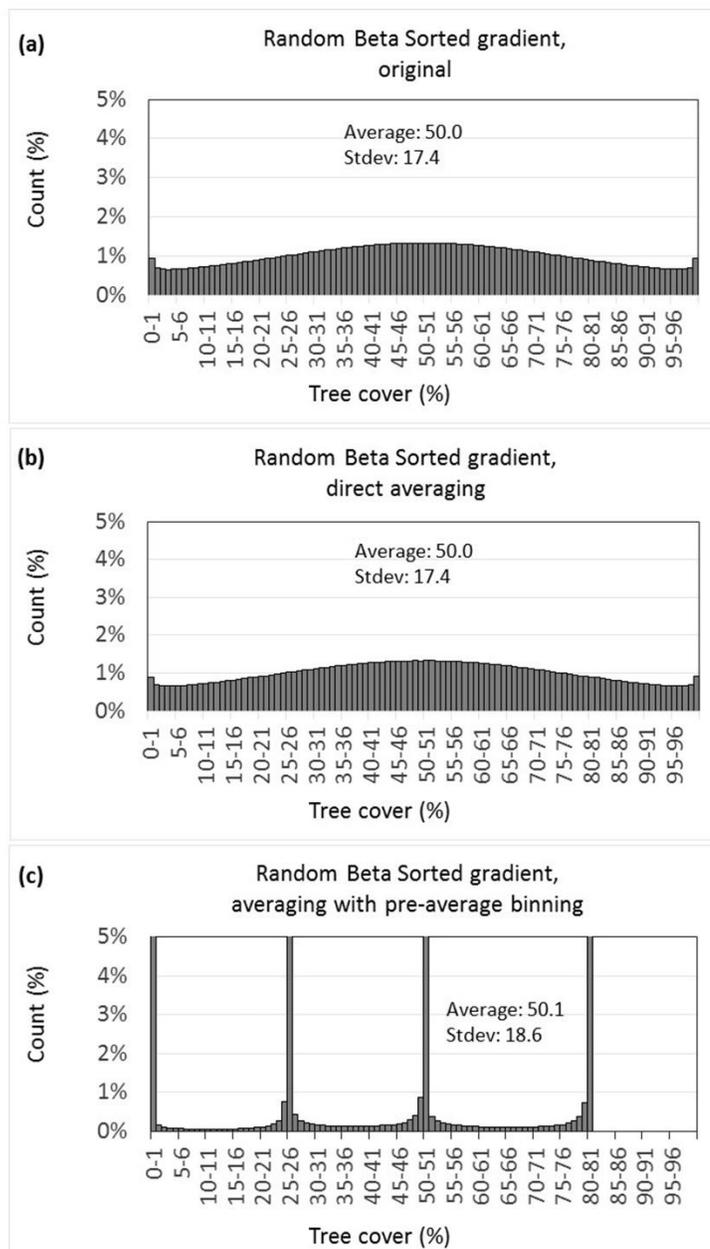
Figure 2. (a) Histograms of the original tree cover values; and the training samples values (e.g. 500 m pixel) acquired through averaging (b) without pre-average binning or (c) with pre-average binning. The original values are from Monte Carlo simulated theoretical 2 dimensional landscapes showing a complex gradient defined using Beta distributions (See Appendix S2 for more detail).
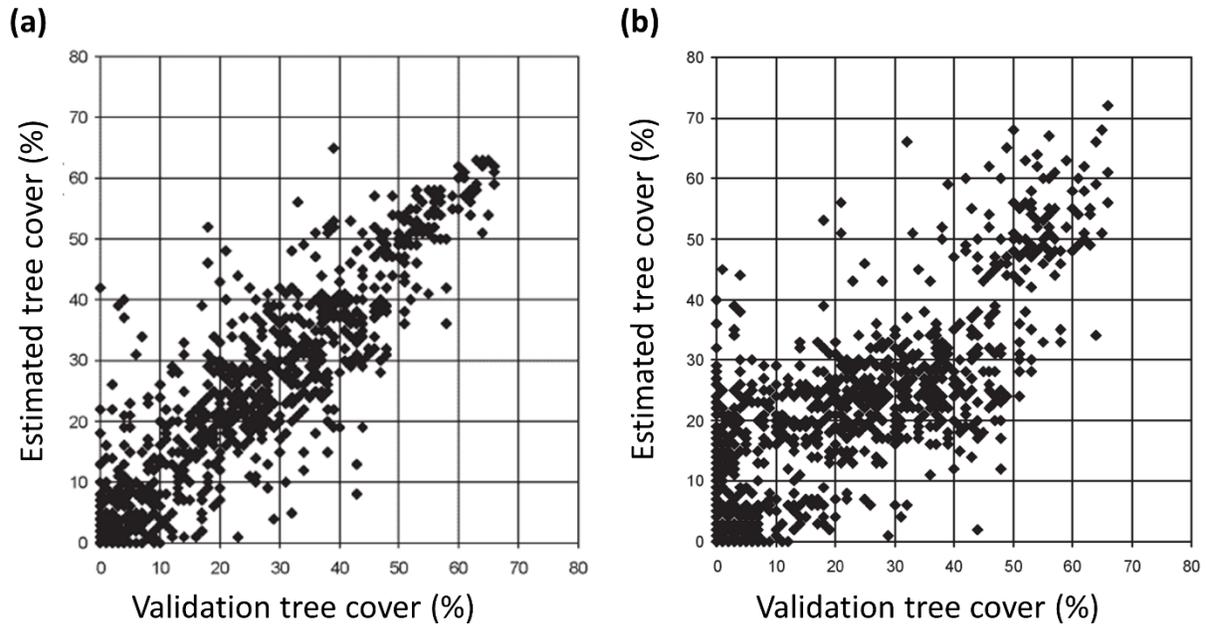
Figure 3 (a) Fig. 7(d) of Hansen *et al.* (2005) showing the validation results for tree cover estimates predicted by one of the four CART variants that were tested for Zambia; all four variants used training sample values representing averages without pre-average binning, which resulted in very similar validation plots, one of which is shown here. (b) Fig. 12 of Hansen *et al.* (2005) showing the validation results for the tree covers of the MODIS VCF collection 3 product using the same Zambia reference data as in (a); here a CART was using training sample values representing averages with pre-average binning.

APPENDIX S1

SEMI-ANALYTICAL SIMULATIONS
For our simulation we utilise the *Beta Distribution.* Examples of the Beta distribution with the shape variables α and β varying between 0.5 and 6 are given in Fig. S1.1.
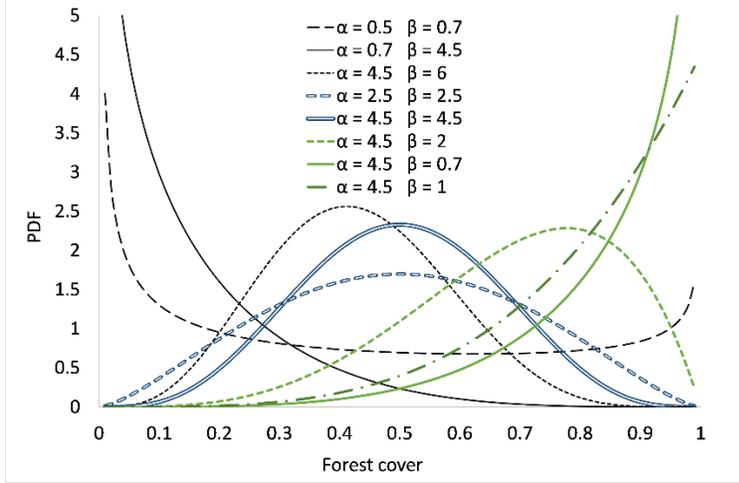
Figure S1.1   Examples of the Beta distribution for α and β varying between 0.5 and 6

We can evaluate the impact of binning by subdividing the continuous Beta distribution $B(\alpha, \beta)$ and comparing the inferred mean $\langle \hat{x} \rangle$ with the true mean $\bar{x}$. First taking the simple case of subdividing the continuous distribution into two, we calculate the inferred mean $\langle \hat{X} \rangle$:

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} \frac{z_1}{2} B(\alpha, \beta)\, dx + \int_{z_1}^1 \left(z_1 + \frac{1-z_1}{2}\right) B(\alpha, \beta)\, dx}{\int_0^1 B(\alpha, \beta)\, dx} \tag{A1}$$

which can also be written as

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} \frac{z_1}{2} B(\alpha, \beta)\, dx + \int_{z_1}^1 \frac{z_1}{2} B(\alpha, \beta) dx + 0.5 \int_{z_1}^1 B(\alpha, \beta)\, dx}{\int_0^1 B(\alpha, \beta)\, dx} \tag{A2}$$

As is evident from the symmetries in Fig. S1.1

$$Pr[(B\!:\!\alpha, \beta) \leq x] = 1 - Pr[(B\!:\!\alpha, \beta) \leq (1-x)] \tag{A3}$$

from which it then follows that

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} \frac{z_1}{2} B(\alpha, \beta)\, dx + \int_{z_1}^1 \frac{z_1}{2} B(\alpha, \beta) dx + 0.5 \int_0^{(1-z_1)} B(\beta, \alpha)\, dx}{\int_0^1 B(\alpha, \beta)\, dx} \tag{A4}$$

which reduces to (and using a slightly altered notation for convenience)

$$\langle \hat{x} \rangle = 0.5z_1 + 0.5 \frac{B[(1-z_1), \beta, \alpha]}{B(1, \alpha, \beta)} \tag{A5}$$

Equation (A7) shows that unless $z_1 = 0.5$ and $\beta = \alpha$ (i.e. the distribution is symmetrical) then $\langle \hat{x} \rangle \neq \bar{x}$.

It is relatively straight forward to extend this analysis to the case of three breaks (to create four bins as in Hansen *et al.* (2002b)):

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} \frac{z_1}{2} B(\alpha, \beta) \, dx + \int_{z_1}^{z_2} \left(z_1 + \frac{z_2 - z_1}{2} \right) B(\alpha, \beta) \, dx + \int_{z_2}^{z_3} \left(z_2 + \frac{z_3 - z_2}{2} \right) B(\alpha, \beta) dx + \int_{z_3}^{1} \left(z_3 + \frac{1 - z_3}{2} \right) B(\alpha, \beta) dx}{\int_0^1 B(\alpha, \beta) \, dx} \tag{A6}$$

which then becomes

$$\langle \hat{x} \rangle = \frac{\int_0^{z_1} 0.5z_1 B(\alpha, \beta) \, dx + \int_{z_1}^{z_2} (0.5z_1 + 0.5z_2) B(\alpha, \beta) dx + \int_{z_2}^{z_3} (0.5z_2 + 0.5z_3) B(\alpha, \beta) dx + \int_{z_3}^{1} (0.5z_3 + 0.5) B(\alpha, \beta) dx}{\int_0^1 B(\alpha, \beta) dx} \tag{A8}$$

expressed alternatively as,

$$\langle \hat{x} \rangle = \frac{\int_0^{z_2} z_1 B(\alpha, \beta) \, dx + \int_{z_1}^{z_3} z_2 B(\alpha, \beta) dx + \int_{z_2}^{1} z_3 B(\alpha, \beta) dx + \int_{z_3}^{1} B(\alpha, \beta) dx}{2 \int_0^1 B(\alpha, \beta) \, dx} \tag{A9}$$

which in terms of incomplete beta functions we express as

$$\langle \hat{x} \rangle = \frac{z_1 B(z_2, \alpha, \beta) + z_2 [B(z_3, \alpha, \beta) - B(z_1, \alpha, \beta)] + z_3 B[(1-z_2), \beta, \alpha] + B[(1-z_3), \beta, \alpha]}{2B(1, \alpha, \beta)} \tag{A10}$$

With the incomplete Beta distribution easily solved numerically, using for example the pbeta function in R and introducing the same break points as Hansen *et al.* (2002b) (i.e. 0.1 for bins 0-10% and 10-40%; 0.4 for bin 40-60% and 0.6 for bin 60-100%), we can now investigate how $\langle \hat{x} \rangle$ varies typically with $\bar{x}$. The simulations of 1000 random values of α and β (with α and β varying between 0.5 and 6) reveal that, near the outer breakpoints (0.1 and 0.6), the inferred mean $\langle \hat{x} \rangle$ values tend to be similar to the true mean $\bar{x}$ (Fig. 1). Near the 0.6 break point however, many $\langle \hat{x} \rangle$ values are lower than the true mean $\bar{x}$, which matches the discrepancy observed in the validation data for Africa. This becomes even more apparent when we compare the simulated histograms of the inferred and true means (Fig. 1). The R code that was used to implement the simulations has been stored in the Zenodo repository and can be downloaded from the following link http://doi.org/10.5281/zenodo.240916 .

APPENDIX S2

MONTE CARLO SIMULATIONS
We used theoretical landscapes of tree cover, to evaluate the variation of cover in a map created from averaging over units (cells) into larger sized tiles. We specifically looked at the effects of spatial auto-correlation created by a tree cover gradient. The theoretical landscapes consisted of a 2 dimensional grid of a million cells (1000 x 1000), each cell representing "virtual true" tree covers between 0% and 100%. As in Hansen *et al.* (2002b), 20 x 20 cells where spatially merged into tiles resulting in a total of 2,500 tiles (i.e. in Hansen *et al.* (2002b) cells represent 30 m TM pixels and tiles represent 500 m Modis pixels). The % tree cover for the tiles was calculated using (1) the Hansen et al (2002b) method, which we refer to as 'averaging with pre-average binning' and (2) the 'direct averaging' method. Averaging with pre-average binning involves a binning of the cells' based on their "virtual true" cover value prior to the merging. Each cell is assigned to one of 4 bin-classes: 0-10%; 10-40%; 40-60% and 60-100% and as in (Hansen *et al.*, 2002b) is then given the representative value of the bin-class it belonged to, i.e., 0%, 25%, 50% or 80%. The % tree cover of the individual tiles is the average of these allotted values. Direct averaging calculates the tile values by averaging the original "virtual true" cover values of the cells. The comparison between 'averaging with pre-average binning' and 'direct averaging' is done through histograms showing the distribution of the respective tile cover values (Note that we do not consider spatial distributions but only statistical distributions).

We carried out simulations for six landscape types, implementing the two approaches 'averaging with pre-average binning' and 'direct averaging', and producing their respective histograms (i.e. 2 x 25,000 per land scape type) with histogram bins set to 1% cover intervals. The ensuing histograms were averaged to produce a single histogram per approach and landscape type by taking the average of each bin. Landscape types 1 and 2 contains no spatial autocorrelation, while 3, 4, 5 and 6 do. Renditions of landscape types 1, 2, 4 and 6 were generated through Monte Carlos simulations. Landscape types 3 and 5 was represented through a single deterministic simulation.

- Landscape type 1, the 'Random Uniform Tree Cover', represents a fully random distribution of forest cover values. Each of the 25,000 Monte Carlo runs fills up the 1000 x 1000 grid with values drawn randomly from a uniform distribution of cover values ranging between 0% and 100%.
- Landscape type 2, the 'Random Beta Tree Cover', represents a fully random Beta distribution of forest cover values and is a 2-dimensional rendition of our semi-analytical simulations (see supplement 1). Each of the 25,000 Monte Carlo runs fills up the 1000 x 1000 grid with values drawn randomly from a Beta distribution of cover values ranging between 0% and 100%. The Beta-distribution parameters α and β are chosen randomly from the range between 0.5 and 6. An example is shown in Fig. S2.1(a).
- Landscape type 3 is the 'Sharp Boundary Landscape'. Here the 1000 x 1000 grid is split into two to show a single sharp boundary running parallel with the y-axis of the grid. Cells left of the boundary are assigned a 100 % tree cover, cells right of the boundary a 0 % tree cover.
- Landscape type 4, the 'Auto-correlated Beta Tree Cover Distribution' is a patchy distribution of auto-correlated Beta distributions. Here the 1000 x 1000 grid is subdivided into 2500 tiles of 20 x 20 grid cells. For each tile an α and β value is chosen at random from the range between 0.5 and 6. The tree cover for each cell within this tile is then drawn from the Beta distribution defined by the α and β pair. The resulting landscape shows tree cover values that are auto-correlated within each tile, but randomly different from tile to tile, producing a patchy tree cover landscape (Fig. S2.1(b))
- Landscape type 5, the 'Linear Tree Cover Gradient' has a linear gradient of decreasing forest cover from 100% to 0% along the x-axis of the 1000 x 1000 grid.

Landscape type 6, the ' Random Beta Sorted Tree Cover' is a 'noisy' gradient that represents a Beta distribution from 100% to 0% tree cover, matching the distributions of our semi-analytical simulation (see supplement 1). The 25,000 Monte Carlo runs produced a range of gradients along the x-axis of the 1000 x

1000 grid. The gradients were defined by the Beta-distribution parameters α and *β* chosen randomly from the range between 0.5 and 6. These min/max values for α and *β* were chosen to represent realistic landscape gradients. An example is shown in Fig. S2.1(c). The Matlab code (v7.14.0.739) for this complex gradient has been stored in the Zenodo repository and can be downloaded from the following link http://doi.org/10.5281/zenodo.240916 .

RESULTS

Without spatial auto-correlation: 'Random Uniform Tree Cover' and 'Random Beta Tree Cover' –
Figs. S2.2 and S2.3
For the 'Random Uniform Tree Cover' scenario, the results from both averaging methods are near-identical. As expected, the 'Random Beta Tree Cover' scenario shows the same bias as seen in the results from the semi-analytical simulations with a 0-100 % tree cover range (Fig. S2.4). 'Averaging with pre-average binning' results in a larger number of tiles with values greater than 60 % and fewer tiles with values between 40 % and 55 %.

With spatial auto-correlation: 'Sharp Boundary Landscape'
The result from the 'Sharp Boundary Landscape' simulation, displays no difference between both averaging methods, with the resulting histograms showing two peaks at the minimum (0 %) and maximum (80 %) tree cover values (not shown).

With spatial auto-correlation: 'Auto-correlated Beta Tree Cover Distribution' – Fig. S2.3.
The patchy landscape represented through the 'Auto-correlated Beta Tree Cover Distribution' shows near identical results as the 'Random Beta Tree Cover' simulation. Hence, no further bias is introduced when the auto-correlation is confined within the tiles.

With spatial auto-correlation: 'Linear Tree Cover Gradient' (not shown) and 'Random Beta Sorted Tree
Cover' Gradient – Fig. 2.
The addition of gradients does not alter the outcome for the whole landscape average and standard deviation values, which remain near-identical between the 'averaging with pre-average binning' and 'direct averaging'. However, for both the simple 'Linear Tree Cover Gradient' and the complex 'Random Sorted Beta Tree Cover' gradient the histogram of the resulting tile values skews substantially towards the bin-class values (0, 25, 50 & 80) when 'averaging with pre-average binning' is implemented. Mechanistically the skew is caused by auto-correlation among cells within tiles that is inherently in gradients. As a result, after binning, tiles would often contain cells with identical values, which decreases the variation among tiles and the range in tile values considerably.
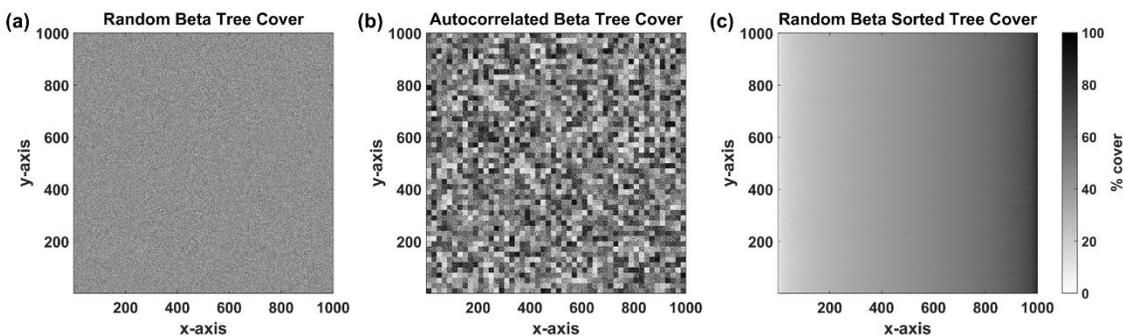
Figure S2.1 Examples of the landscapes generated by the Monte Carlo simulations: (a) Linear Tree Cover Gradient; (b) Auto-correlated Beta Tree Cover; and (c) Beta Tree Cover Gradient.
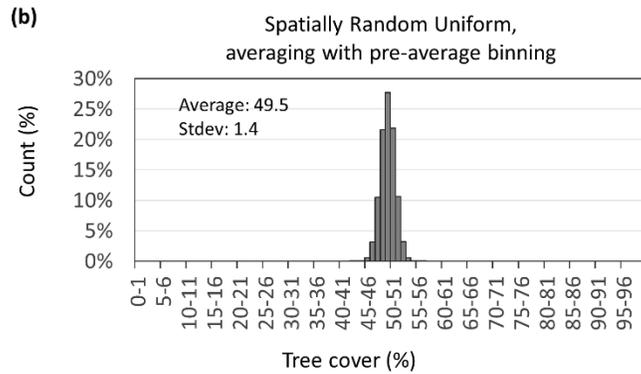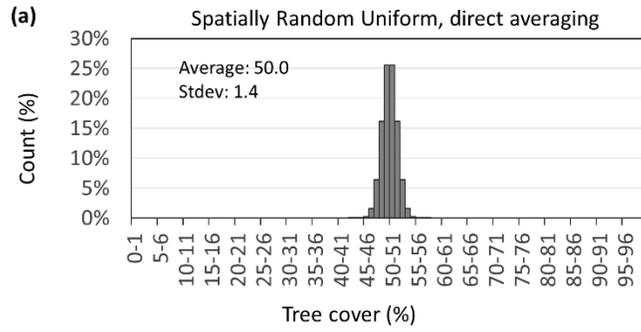
Figure S2.2 Histogram of calculated tile values from theoretical landscapes with no spatial auto-correlation: 'Random Uniform Tree Cover': (a) tile values resulting from 'direct averaging'; (b) tile values resulting from 'averaging with pre-average binning'.
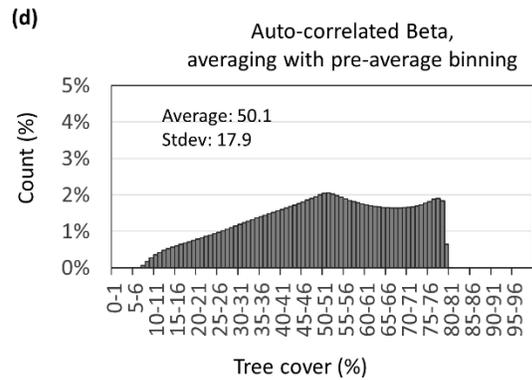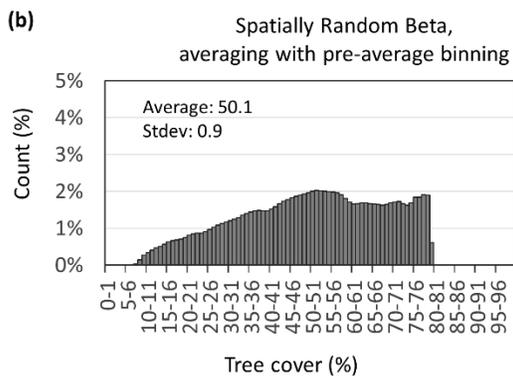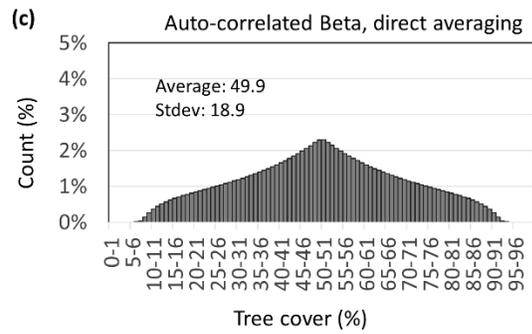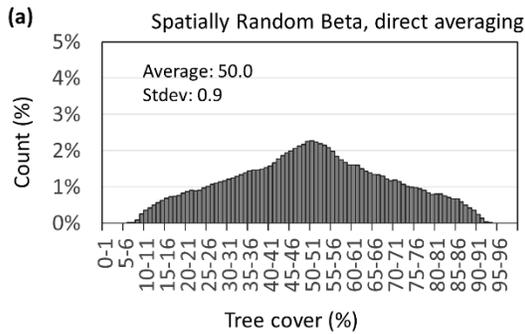
Figure S2.3 Histogram of the calculated tile values from theoretical landscapes with no spatial auto-correlation: 'Random Beta Tree Cover ' (a and b) and spatial auto-correlation: 'Auto-correlated Beta Tree Cover' patchy distribution (c and d). Tile values resulting from 'direct averaging' (a and c); and from 'averaging with pre-average binning' (b and d).
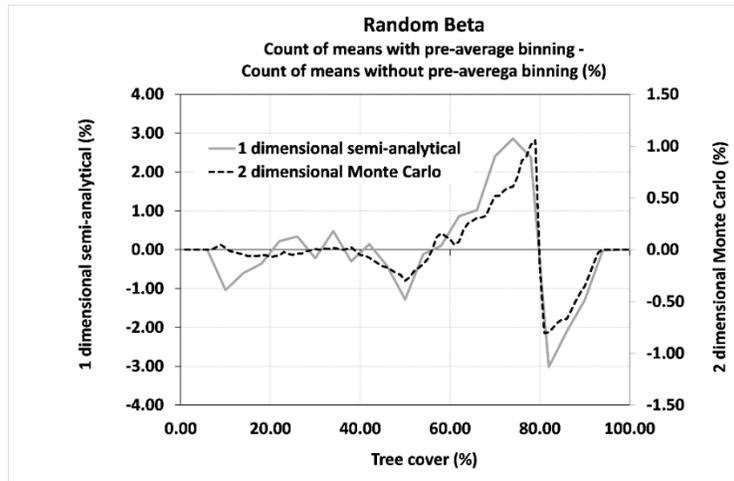
Figure S2.4 A comparison of the histograms count of means obtained through one-dimensional semi-analytical simulations and two-dimensional spatially random Monte Carlo simulations of Beta distributions. The plot shows the difference in histogram counts of means between averaging with and without pre-average binning.