Supporting Information to the paper

'Remote sensing of beta diversity: evidence from plant communities in a semi-natural system'

Applied Vegetation Science

Hoffmann, S., Schmitt, T.M., Chiarucci, A., Irl, S.D.H., Rocchini, D., Vetaas, O.R., Tanase, M.A., Mermoz, S., Bouvet, A., & Beierkuhnlein, C.

Corresponding author: Samuel Hoffmann (samuel.hoffmann@uni-bayreuth.de)

Appendix 2. Description of Lidar based metrics derived for La Palma, Canary Islands

Prepared by CESBIO in the frame of the of ECOPOTENTIAL project

Mihai A. Tanase[1,2], Stéphane Mermoz[2], Alexandre Bouvet[2] and Thuy Le Toan[2]

[1] University of Alcala, Madrid, Spain

[2] Center for the Study of the Biosphere from Space (CESBIO), Toulouse, France

Use of this document

1. This report is based on unpublished data and has not been peer-reviewed.
2. This report may be copied for distribution within the ECOPTENTIAL consortium.
3. Distribution to a wider audience should be with permission of the authors.
4. Latest draft: March 15th 2017

For further details regarding processing, modelling and cal/val activities please contact the CESBIO team (mihai@tma.ro, mermozs@cesbio.cnes.fr, Alexandre.Bouvet@cesbio.cnes.fr).

Version control

Version 1	Uploaded on the ECOPTENTIAL repository (March, 15th 2017)

I. Input data

1. Remote sensing data

Point cloud tiles (214, centered at 28.7N 17.87W) from Airborne Laser Scanning (ALS) were downloaded from the Spanish National Geographic Institute (IGN, in its Spanish acronym) in compressed LAZ format (Fig. S1).

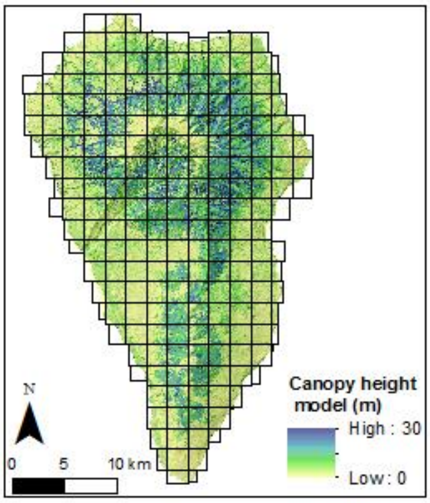

Fig. S1. Extent of LAS tiles (black squares) over La Palma island.

2. In situ data

Relevant *in situ* data were not available. By relevant we refer to datasets allowing for calibration and validation of lidar derived forest EVs (i.e., above ground biomass, canopy height, canopy cover, LAI).

II. Lidar processing

The Lidar data were examined for extent, point density, consistency, overlapping areas or gaps, and the accuracy of the existing classification (Fig. S2). Several issues were noticed: i) the presence of data gaps, ii) the presence of points over sea iii) noise (in shadowed areas or steep slopes), iv) overlapping flight lines and, v) inaccurate point classification (i.e., for the ground class). These issues were addressed as follows:

i) data gaps: no other source of ALS data was available. Therefore, two small areas (roughly 200m by 400m) centered at 28.5N/17.85W and 28.618N/17.844W present gaps in the derived vegetation metrics, DEM, and DSM layers. The gap size does not allow for accurate interpolation as evident when examining interpolated DEM and DSM layers. Therefore, these two areas were masked out in all layers.

ii) points over sea were eliminated using a combination of rules on height and RGB values (from orthophotos), using las2las, an open source tool within the Lastools software package;

iii) noise: eliminated using a similar rule approach as for points over sea;

iv) still present in the dataset as the point clouds did not provide information on overlapping (as was the case Sierra Nevada and Sierra de Baza).

v) general point classification: the point classification was fairly accurate for building and vegetation classes. However, the classification does not properly differentiate between vegetation and ground points over steep slopes which resulted in flattening out most of the ridges (Fig. S3). Therefore, ground points were reclassified using the open source Multiscale Curvature Classification (MCC) algorithm (Evans and Hudak 2007). The algorithm was designed for forested environments on rough surfaces and was proved to produce the highest success rates at identifying ground and non-ground returns for similar datasets (Montealegre et al. 2015). Depending on tile, different *scale* and *curvature* parameters were used. Their values were determined by iterative tests starting from the default ones. The iterative testing showed that using a scale parameter of 2 and a curvature parameter of 0.3 provides for the least confusion between ground and vegetation points for most tiles. However, for 22 tiles[1] located on the volcano`s eastern slopes the scale and curvature

[1] 214-3182, 216-3180/3182/3184, 218-3180/3182/3184/3192/3194, 220-3184, 222-3166/3190, 224-3186/3188/3190, 226-3184/3186/3188/3190, 228-3184/3186/3188.

parameters providing better classification results for the ground points were 1.5 and respectively 0.4. These parameters allowed for ridges to maintain their natural shape. A combination of high slopes, dense vegetation, low point density, and flight parameters have resulted in a low number of shots hitting the ground. Ground points are therefore sparse and their interpolation inadequate when spatial resolutions below 10 m are used. As a result, the area covered by the 22 tile presents artefacts in the high resolution (5 m) DEM. Although different tools were used to classify ground point (Fusion, MCC, Lastools) the artefacts were not completely resolved. Therefore, it is advised not to use the 5m DEM on the eastern slopes of the volcano.

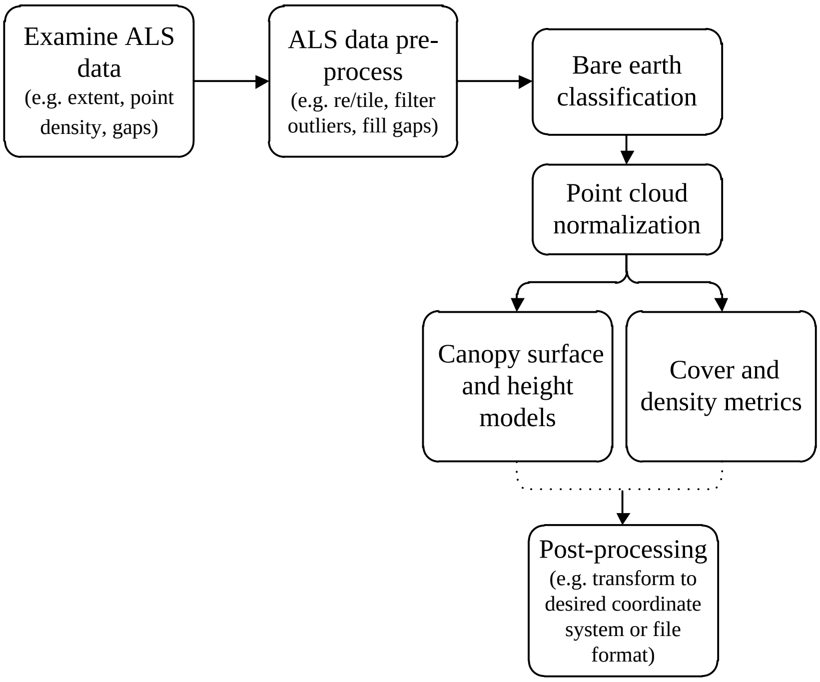

Fig. S2. Flowchart of Airborne Laser Scanning (ALS) data processing and ALS-based metrics generation

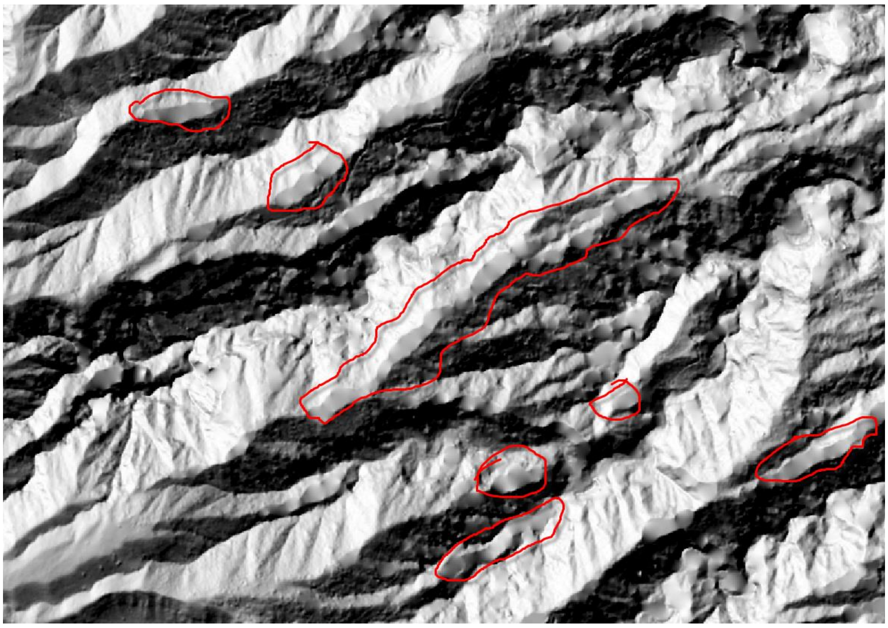

Fig. S3. Flattened hill tops (circled in red) in the shaded DEM. The DEM was obtained using the original classification available in the downloaded data.

The digital elevation and surface models (DEM and DSMs) were interpolated in ArcGIS© at various spatial resolutions using the MCC classified ground points. The ten-meter spatial resolution DEM was used to compute the height above ground (i.e. normalized height) when generating ALS-based layers characterizing forest structure (e.g., canopy height, cover). ALS-based layers where derived using the original (IGN) classification of the point cloud data to take advantage of buildings classification which was fairly accurate. The normalized point cloud data were used to produce generic Lidar-based metrics (canopy closure, canopy density, percentiles, etc.). Such Lidar metrics represent proxies of forest structural characteristics and can be used with parametric or non-parametric models to derive above ground forest biomass or other structural characteristics of interest.

The lidar-based layers (i.e., metrics) were post-processed (masked) to eliminate areas with data gaps. Vector files showing LA Palma island limits were used to eliminate interpolated areas around the island shore (caused by the irregular pattern of available tiles). The vector file also contained the location of two areas (about 8 ha each) where data gaps in the point cloud were detected.

III. Output data

The full extent of the La Palma island was processed using Fusion Area Processor (AP), a suite of tools/scripts developed for large area processing. Fusion is an open software developed and maintained by the USDA Forest Service. The advantage of using AP (as compared to process individual tiles and mosaic them) is the seamless surfaces (no artifacts along border tiles) obtained through automatic buffering of the tiles. AP also offers support for multi-core processing which decreases the processing time. The metrics uploaded to ECOPTENTIAL repository are described below. Further metrics ('Furthermetrics.txt' on request) were generated and are described separately ('Gridmetrics.pdf' on request). In addition, metrics for the following strata ('Stratametrics.txt' on request) were also computed: 0-0.5m, 0.5-2m, 2-5m, 5-10m,10-15m,15-20m,20-25m, and >25 m. Space limitation on the FTP repository impedes the storage of all these metrics. On request, these layers can be provided through other means.

Files naming: AreaName_LayerAcronymTresholds_SpatialResolution**.**extension

DEM: digital elevation models produced from Lidar data.

The LAS files were re-classified using the MCC algorithm of Evans and Hudak (2007) into 'ground' and 'unassigned' points. Returns classified as 'ground' were interpolated to a raster surface at various spatial resolutions.

LAS Dataset to Raster (ArcGIS©) settings:

- Triangulation: Natural Neighbor
- Point Thinning type: Windows Size
- Point Selection Method: Closest to mean

Slope: slope in degree based on the above DEMs.

DSM: digital surface model based on the same Lidar dataset as the DEM. First returns were interpolated to raster surface at various spatial resolutions. The above-mentioned LAS Dataset to Raster (ArGIS©) settings were used for interpolation.

Chm_average (canopy height model): average of normalized elevations (i.e., height above ground) within the resolution cell. Produced through Fusion Area Processor from Lidar point cloud data using as a reference surface the 10m spatial resolution DEM.

In areas with slopes above 50 degrees the CHM values may exceed 60m (the maximum height of *Pinus canariensis*) due to overhanging areas on steep slopes. Masking areas with high slopes may improve accuracy (tree type vegetation is not likely to occur on such slopes). Please notice that tree heights are generally below 30m for all species. Pine trees above 30m height may occur only exceptionally. Therefore, CHM values over 30m are likely to be related to trees over steep slopes or overhanging areas and should be set to 30m and respectively masked out. A masking layer may be derived by combining information on slope, elevation (over 2000m only shrub vegetation occurs) and Return Proportion (RP, see below) for different strata (e.g., LaPalma_RP_StrataOver25m_20m).

The layer was generated for the entire island. Over developed areas, layer values correspond to infrastructure and should be masked out using an appropriate layer and/or NDVI values.

TFC (tree fraction cover, %): first returns over 2 m height divided by the total number of first returns within the resolution cell. Produced through Fusion AP from the Lidar point cloud data using as a reference surface the 10m spatial resolution DEM.

The layer was generated for the entire island. Over developed or agricultural areas, layer values correspond to infrastructure and respectively crops and should be masked out using an appropriate layer and/or NDVI values.

VF (vegetation fraction): number of all returns over 0.5m height divided by the number of all returns within the resolution cell. Produced through Fusion AP from Lidar point cloud data using as a reference surface the 10m spatial resolution DEM. VF might be useful to asses cover for all vegetation types (including shrubs).

The layer was generated for the entire island. Over developed areas, layer values correspond to infrastructure and should be masked out using an appropriate layer and/or NDVI values.

RP: Return proportion: proportion of returns from different strata with respect to the total number of returns in the resolution cell. Useful to appraise vegetation vertical structure (i.e., cover, number of strata etc.). For percentage values, multiply with 100.

The layer was generated for the entire island. Over developed areas, layer values correspond to infrastructure and should be masked out using an appropriate layer and/or NDVI values.

LAIe: effective Leaf Area Index (LAIe): computed based on the gap probability (P) as: **LAIe=-ln(P)**

The gap probability (P) is computed as the ratio of ground returns to the total number of returns (Fieber et al. 2014). LAIe term is used to address the lack of correction for clumping effects and the presence of woody elements (Fieber et al. 2014, Fieber et al. 2016).

The LAIe was set to no data (i.e., -9999) for pixels were no ground returns were recorded and the total number of returns was above zero. Such pixels correspond to very dense vegetation (mostly on the eastern slopes of the volcano) were shots did not reach the ground. A 30m LAIe was also produced to reduce the presence of such pixels. Artefacts may be present in the LAIe layer particularly for the 22 tiles were the classification of the ground returns was problematic.

Notes:

- All data sets in EPSG 4083.
- No data areas are designated -9999 (in *.tif files).
- Datasets covering the full extent of La Palma Island were generated at various spatial resolution as indicated in the file name.

- The 5m resolution DEM was produced to accommodate other needs (e.g. hydrological modelling). Due to the low point cloud density, such versions of the DEM may contain gaps. Please check and mask accordingly before use.
- Twenty-two tiles (see above) might present artifacts in some of the produced Lidar metrics. These artifacts are related to the steep topography and the characteristics of the ALS flight.

References

Evans, J.S., & Hudak, A.T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. IEEE Transactions on Geoscience and Remote Sensing, 45, 1029-1038

Fieber et al. (2014). Effective LAI and CHP of a Single Tree From Small-Footprint Full-Waveform LiDAR. IEEE Geoscience and Remote Sensing Letters, 11, 1634-1638.

Fieber et al. (2016). CHP Toolkit: Case Study of LAIe Sensitivity to Discontinuity of Canopy Cover in Fruit Plantations, IEEE Transactions of Geoscience and Remote Sensing, in press.

Montealegre, A.L., Lamelas, M.T., & Riva, J. (2015). Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. Remote Sensing, 7, 8631-8654.

Morsdorf et al. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction, Remote Sensing of Environment, 104, 50-61.