Using a Multistructural Object-Based LiDAR Approach to Estimate Vascular Plant Richness in Mediterranean Forests With Complex Structure

Javier Lopatin, Mauricio Galleguillos, Fabian E. Fassnacht, Andrés Ceballos, and Jaime Hernández

Abstract—A multistructural object-based LiDAR approach to predict plant richness in complex structure forests is presented. A normalized LiDAR point cloud was split into four height ranges: 1) high canopies (points above 16 m); 2) middle-high canopies (8–16 m); 3) middle-low canopies (2–8 m); and 4) low canopies (0–2 m). A digital canopy model (DCM) was obtained from the full normalized LiDAR point cloud, and four pseudo-DCMs (pDCMs) were obtained from the split point clouds. We applied a multiresolution segmentation algorithm to the DCM and the four pDCMs to obtain crown objects. A partial least squares path model (PLS-PM) algorithm was applied to predict total vascular plant richness using object-based image analysis (OBIA) variables, derived from the delineated crown objects, and topographic variables, derived from a digital terrain model. Results showed that the object-based model was able to predict the total richness with an r^2 of 0.64 and a root-mean-square error of four species. Topographic variables showed to be more important than the OBIA variables to predict richness. Furthermore, high-medium canopies (8–16 m) showed the biggest correlation with the total plant richness within the structural segments of the forest.

Index Terms—Bootstrapping, LiDAR, object-based analysis, partial least squares path model (PLS-PM), vascular plant richness.

I. Introduction

BIODIVERSITY is an essential element of the Earth system from which all humans benefit directly or indirectly [1]. The current and future states of biodiversity have acquired increasingly more scientific and political relevance. This is based on the greater knowledge about the adverse effects that the decrease in biodiversity may produce on ecosystem services on which human well-being depends [2], [3].

J. Lopatin, A. Ceballos, and J. Hernández are with the Laboratory of Geomatics and Landscape Ecology, Faculty of Forest and Nature Conservation, University of Chile, 11315 Santiago, Chile (e-mail: javierlopatin@gmail.com; aceballos@ug.uchile.cl; jhernand@uchile.cl).

M. Galleguillos is with the Department of Environmental Sciences, School of Agronomic Sciences, University of Chile, 11315 Santiago, Chile (e-mail: mgalleguillos@renare.uchile.cl).

F. E. Fassnacht is with the Institute of Geography and Geoecology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany (e-mail: fabian.fassnacht@kit.edu).

Theoretical and empirical studies have suggested that local biodiversity is strongly influenced, in a positively correlated fashion, by the environmental heterogeneity [4]. More complex environments can host a greater number of ecological niches, which, in turn, can be colonized and inhabited by a greater number of species [5], [6]. Different remote sensors allow capturing spatial heterogeneity in two ways: spatial variability or heterogeneity (horizontal variability) and topographical–structural variability or heterogeneity (vertical variability) [6].

Some studies have used topographical information, i.e., microrelief and their spatial variability [7], [8], along with predictors describing the configuration and variability of the vertical structure of the aboveground vegetation [6] to estimate environmental heterogeneity. Topographical and structural information captured by LiDAR sensors may therefore be suitable to provide biodiversity-related variables in different ecosystems [9], particularly in forests with high vertical complexity [6]. Such data are expected to have large potential for biodiversity characterization [10]. For example, the suitability of a combination of LiDAR and hyperspectral data to estimate vascular plant richness in deciduous Mediterranean forests was demonstrated in [11]. A link between imaging spectroscopy and LiDAR with floristic composition and forest structure was presented by Higgins *et al.* [12], and the influences of terrain and vegetation structures on local avian species richness in two mixed-conifer forests were examined by Vogeler *et al.* [13].

Object-based image analysis (OBIA) is a segmentation/classification and feature extraction technique that has been used among other applications for environmental monitoring requirements, conservation goals, spatial planning enforcement, or ecosystem-oriented natural resources management [14]. The OBIA approach is able to group homogeneous pixels into meaningful objects based not only on their radiometric values but also on multiple attributes such as texture, context, and shape. This approach was found to produce more reliable and less noisy results when applied to high-resolution images [15], [16]. A large set of studies combined LiDAR data with OBIA technique for segmentation or feature extraction purposes. For example, a two-stage approach for characterizing the structure of *Pinus sylvestris* stands in forests of central Spain was presented in [17]. The urban landscape structure at the parcel level was also investigated by Zho and Troy [18] who applied an object-oriented approach on high-resolution digital aerial imagery and LIDAR data for the Baltimore area.
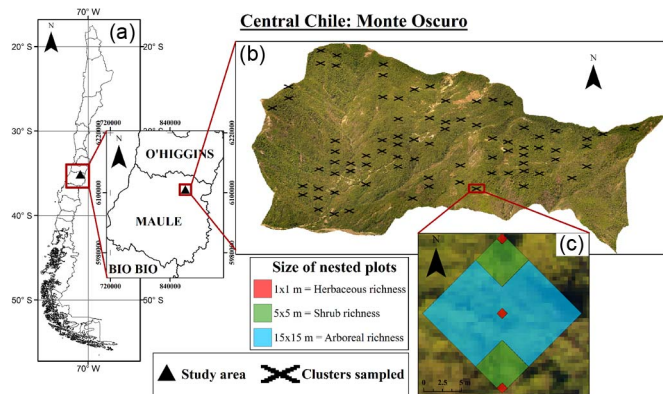
Fig. 1. (a) Location of the study area. (b) Distribution of field-measured richness values. (c) Detail of a sampling unit with nested plots.
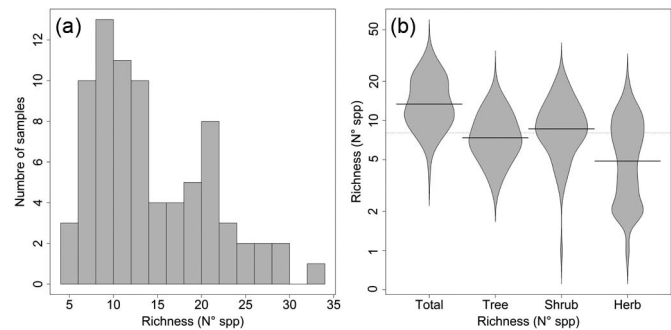

Fig. 2. (a) Histogram of the richness values. (b) Beanplots of the richness distribution. N° spp is the number of vascular plants species found in a (sub-) plot. Beanplots show the distribution of N° spp values of all plots of the corresponding richness class. Black horizontal lines indicate the median of the same data.

The objective of this study was to develop a method to predict local vascular plant richness in a secondary mixed deciduous Mediterranean forest. The underlying idea was to use only structural and topographical attributes derived from LiDAR in a multistructural object-oriented approach.

II. STUDY SITE AND DATA SET

A. Study Area

The study area, Monte Oscuro, is located in the Andes foothills of central Chile in the Maule Region (35°07"00" S, 70°55"30" W) [see Fig. 1(a)]. This area is associated with the Sub-Mediterranean Temperate bioclimatic zone, with a mean annual precipitation of 1000 mm, which is mainly concentrated between April and October. The site has an extent of 1295 ha, a mean altitude of 1075 m above sea level, and mostly southern aspect. Monte Oscuro is covered by secondary deciduous forest dominated by *Nothofagus obliqua* (Mirb.) Oerst., mixed with sclerophyll species.

B. Ground Data

We used the vascular plant richness as a response variable due to the clear dominance of this taxonomic group in Mediterranean forests and also because it is essential for the trophic network and ecosystem functioning [19]. A botanic survey was performed between January 2013 and January 2014. A 200 m × 200 m regular grid was applied to locate the square nested plots [20]. Steep zones (< 45%) and sites too close to trails (20 m) were not considered in the sampling [see Fig. 1(b)] to facilitate operational endeavor and avoid the effect of borders and bare soil in the plots. Finally, 80 square nested plots were located in the study site.

In each sampling unit, species within three height layers were registered and determined: trees (height of 2 m or more), shrubs (less than 2 m of height), and herbs (non-woody plants). Each plot was composed of six nested subplots. The largest one, in which only trees were registered, occupied 225 m² (15 m × 15 m), and it had its vertices aligned to the four cardinal points (N, E, S, and W). Shrubs were identified in two subplots of 25 m² (5 m × 5 m). Herbs were registered in three sub-subplots of 1 m² [see Fig. 1(c)].

Following the suggestions in [21], we considered the nested subplots to be a representative sample of the total plot area. Therefore, for each 225-m² plot, the total richness of vascular species was obtained by simply summing up all species encountered in the plots and subplots of the three height layers. Species occurring in more than one height layer (e.g., tree and shrub layers) were only counted once. The histogram of the richness values shows a bimodal distribution [see Fig. 2(a)]. This data distribution can be attributed to the fact that the area is considered as a second growth natural forest under some areas with silvicultural intervention, resulting in a high number of relatively young stands that were established after harvesting activities [22]. In Fig. 2(b), a distribution of the sample richness is illustrated.

C. LiDAR Data

A discrete pulse LiDAR from the Harrier 54/G4 Dual System (manufactured by Trimble Industries and provided by Digimapas Chile Ltd., Santiago, Chile) was used to survey the study site on March 2011. Characteristics of the Harrier 54/G4 Dual System include scanning frequency of 100 Hz, pulse rate of 100 kHz, scan angle [field of view (FOV)] of ±22.5°, laser beam divergence (instantaneous FOV) of 0.5 mrad, and an intensity of returns from a 1550-nm laser, obtaining an average point cloud density of 4.64 points/m² and a footprint of 29 cm. This instrument was also used in other studies, for example, in [22].

III. METHODS

A. Multistructural Height-Based Threshold

The LiDAR point cloud was classified into ground and nonground points for bare-earth extraction according to Briese [23]. Normalization of the LiDAR point cloud was performed to transform the altitude data units from meters above sea level to meters above the ground. Then, the normalized point cloud was classified into four classes of canopy heights (see Table I). These classes are based on the methods proposed by Etienne and Prado [24], who showed that they provide a representation of the vegetation structure of dominant species in the South-center of Chile.

TABLE I
HEIGHT-BASED THRESHOLD CLASSIFICATION

Height based threshold	Altitudinal ranges
High Canopies (HC)	Over 16 m.
Middle-High Canopies (MHC)	8 − 16 m.
Middle-Low Canopies (MLC)	2 − 8 m.
Low Canopies (LC)	0 − 2 m.

B. DEMs

The point cloud classified into ground and nonground points was interpolated to obtain a digital terrain model (DTM) and a digital canopy model (DCM). Furthermore, interpolation of the four segmented normalized point clouds was performed to obtain four pseudo-DCMs (pDCMs). All the obtained digital elevation models (DEMs) had 1 m^2 spatial resolution.

C. Object-Based Analysis

An object-oriented segmentation approach was used to delineate crown objects (polygons) of small size (less than 400 m^2), based on the DCM and the four derivate pDCMs. For the segmentation of the crown objects, we took into account the scale, the homogeneity, the shape, and the color of the DCM and the pDCMs. The scale is related to the minimum size required to identify a particular object, which depends on the resolution of the images [17]. Homogeneity is described by a mutually exclusive interaction between color and shape. Color refers, in this case, to the height of the objects, whereas shape is divided into two equally exclusive properties, i.e., smoothness and compactness, which define the boundaries of the polygons (crown objects) and their transition to others, respectively [14].

We applied a multiresolution segmentation algorithm, in eCognition software, to the DCM and the four pDCMs (scale parameter = 5, shape criterion = 0.3, and compactness criterion = 0.5). Results, which can be understood as crown objects, can be compound by a part of a crown, a single crown, or a group of crowns.

The crown objects derived from the DCM cover the full height ranges of the data set (LC to HC). These polygons were subsequently classified into the height classes given in Table I. No height-based classification was performed to the crown objects derived from the pDCMs (see Fig. 3).

D. Model Parameterization

Several topographical and OBIA predictor variables were obtained for the regression analysis. The topographic variables used are the generalized surface, the maximum heights, and the DEM, estimated in SAGA GIS(r) software. For each variable, the minima, mean, and maxima values within the plots were estimated. These variables were selected based on the findings of Ceballos *et al.* [11], who demonstrated that these topographic variables had high correlation with the total vascular richness in the same study area.

The object-based predictor variables were calculated by extracting: 1) the total number of crown-object polygons associated to each richness sampling plot; 2) their coverage (m^2/ha);
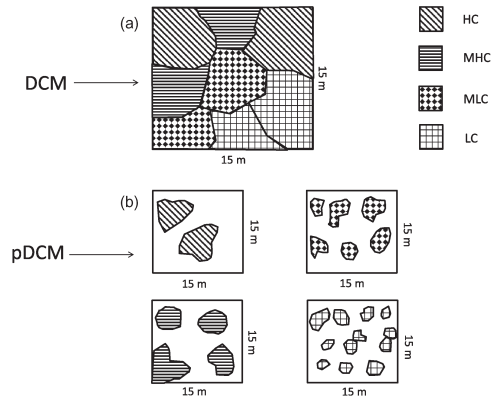

Fig. 3. Crown-object segmentation. (a) Crown objects were delineated from the DCM and classified according to the classes given in Table I. (b) Crown objects were delineated from the pDCMs.

TABLE II
SUMMARY OF THE OBIA PREDICTOR VARIABLES

Source of the variables	Number of variables
Classified DCM crown-objects	4+4+4+4
High Canopies (HC) crown-objects	4
Middle-High Canopies (MHC) crown-objects	4
Middle-Low Canopies crown-objects	4
Low Canopies crown-objects	4
Total	32

3) the standard deviation; and 4) the average of their areas. In the case of the crown objects derived from the DCM, a total of 16 variables were obtained (i.e., 4 for each altitudinal range given in Table I; compare also Fig. 3), and in the case of each pDCM, 4 variables were obtained. Therefore, a total of 32 predictor variables were derived from OBIA (see Table II).

E. Statistical Analysis

We used the partial least squares path model (PLS-PM) algorithm to predict total vascular plant richness. PLS-PM works as a fusion of regression analysis, principal components analysis, and path analysis [25]. This combination seems especially interesting in a remote sensing context, where often highly correlated predictors are applied and data reduction methods were found to improve model performances [26].

PLS-PM is a nonparametric method designed for prediction analysis and the development of theoretical models [27], which can be graphically represented by a path diagram that shows how the various elements (variables) are related to each other [25]. This algorithm was performed using the R-Project package "plspm."

PLS-PM has a series of model validation techniques that work on different levels (i.e., loading, weights, and r^2) [25], [27], with an overall validation of the model obtained with a pseudo-goodness-of-fit (GoF) proposed by Tenenhaus *et al.* [25]. In addition, a bootstrapping validation technique with 1000 boots was performed to assess the quality of the model. In each bootstrap iteration, we used the samples that were not drawn in the bootstrap (on average, 36.8% of the total

TABLE III
SELECTED VARIABLES

Latent Variable	Selected variables
DCM	Middle crown-objects, the Middle cover, and the Low cover.
HC	Number, Cover, and Standard Deviation of the crown-objects size.
MHC	Number and Cover of the crown-objects.
MLC	Number of crown-objects.
LC	Number and Cover of the crown-objects.

number of samples) as holdout sample for an independent validation [26].

For the selection of the predictor variables in latent variables, the heuristic "10 cases per indicator" provided by Tenenhaus *et al.* [25] to maximize the efficiency of PLS-PM was used. This statement is a recommendation for robustness in PLS-PMs, where the authors made a series of empirical tests to provide this conclusion. The predictor variables with highest loading and communality were selected in each latent variable.

F. Richness Prediction

Species richness, defined as the number of species associated with a known area [21], has proven to be a relevant indicator to determine the biodiversity of a given area, as well as a valuable element to be used by decision makers at different levels [28] to establish spatially explicit conservation strategies for biodiversity [29].

PLS-PM uses linear approaches to predict the total forest vascular plant richness, as described in (1). The prediction formula was standardized to obtain positive weights and to eliminate the intercept terms [27] as follows:

$$R = \left\{ \left[\sum_{k=1}^{n} \beta_{jk} \left(\frac{\left(\sum_{j=1}^{N} (w_{ij})(y_j) \right) - \overline{x}_j}{STD_j} \right)_k * STD_R \right] + \overline{x}_R \right\} + \varepsilon_{jk}$$

(1)

where R is the richness in the number of species, k is the flux of R in each predictor latent variable (LV), i corresponds to every predictor variable of LV_j, STD_j is the standard deviation of LV_j, $\overline{x}_j$ is the mean value of LV_j, STD_R is the standard deviation of the ground data richness, ε_{jk} is the mean of the ground data richness, and ε_{jk} is the associated error. The error of the model was estimated using root-mean-square error (RMSE).

IV. RESULTS AND DISCUSSION

A. Variable Selection

Following the heuristic "10 cases per indicator" provided by Tenenhaus *et al.* [25] to maximize the efficiency of PLS-PM, a maximum of 8 predictor variables were selected for each latent variable (because 80 observations were obtained for the modeling). The OBIA predictors selected for each LV are shown in Table III. Finally, a total of 19 predictor variables were
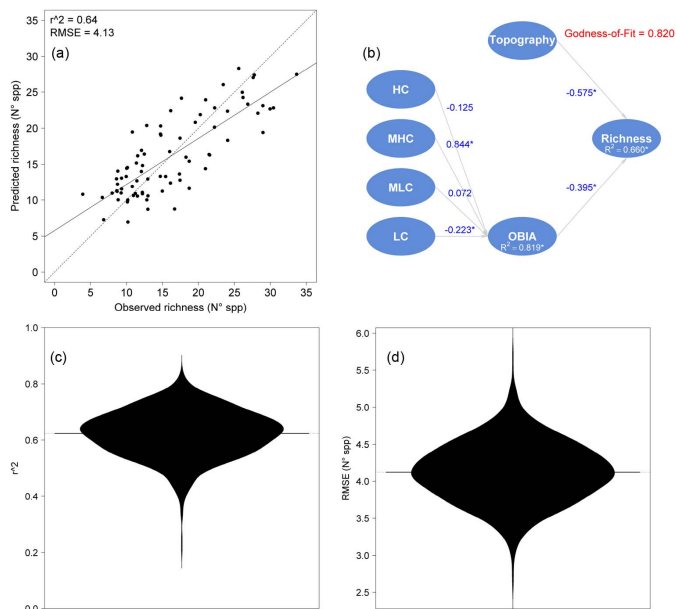

Fig. 4. Model adjustment of the data set and path diagram. (a) Scatterplot between observed and predicted richness values. (b) Path diagram. (c) and (d) r^2 and RMSE distribution obtained when applying the bootstrapping validation. ∗ Indicates that the path coefficient is significant with 95% confidence.

used for the modeling (8 are topographical, as proposed in [11], and 11 are OBIA based).

B. Model Validation

The loadings r and communalities were highly significant (< 0.9) in all cases, demonstrating that the predictor variables were strongly represented by the latent variables (LV's).

Furthermore, the redundancy of the richness was high (about 0.7), indicating that the LV's had a highly significant predictive ability.

The path diagrams of the prediction models of richness, with the path coefficients (β) between the predictor variables and the LV's, are shown in Fig. 4(b). The LV with the highest coefficient was topography, indicating that, at local scale, the (micro-) topography and the edaphic conditions provide more accurate predictive information for the estimation of plant richness than the structural vegetation information. It is possible that, with a smaller scale study, this relationship could change.

Middle-high canopies and the low canopies were the best forest structure segments to predict total richness, accounting for 84% and 20% of the effects of the vegetation in the total forest richness, respectively. On the contrary, the worst forest structure segments were the middle-low canopies and the high canopies, accounting for only 7% and 12% of the effects of the vegetation on the total forest richness, respectively.

The model showed a significant predictive capacity, with a GoF of 0.82, an r^2 of 0.64, and an RMSE of 4.13 species. The model showed a slightly improved ability to predict total plant richness compared with an earlier study, which reported an adjusted r^2 of 0.57 and an RMSE of 5.05, using LiDAR and hyperspectral data [11]. In [13], total species richness was estimated with an adjusted r^2 ranging from 0.05 to 0.42, with LiDAR and Landsat imagery. Finally, multispectral information

for predicting species richness in tropical forest with an r^2 of 0.64 was applied by [30].

The model showed a slight tendency to overestimate species richness in areas with low species diversity (under 15 species) and to produce a slight underestimation of species richness in zones of high field-measured species diversity (over 15 species). This is consistent with the bimodal distribution of the richness data set [see Fig. 2(a)], where the mean of the two modes (the valley between the two local maxima) is in the 15-species threshold value. The high number of young stands makes the left local maximum of the distribution higher than the right local maxima, producing the mitigation effect of the model.

V. Conclusion

An empirical PLS-PM based on object-based LiDAR data has been developed to predict plant vascular richness in a highly complex secondary Mediterranean forest in central Chile. Results have demonstrated that the inclusion of object-based information provided accurate results. Furthermore, we found that, in our study site, the effects of the (micro-) topography and edaphic conditions are more important than the vegetation structural features at a local scale. Moreover, middle-high canopies are the most useful variables to predict the total forest richness. Finally, LiDAR data showed to be a reliable source of information for the use of object-based approaches in forest ecosystems.

References

[1] J. E. Duffy, "Why biodiversity is important to the functioning of real-world ecosystems," *Front. Ecol. Environ.*, vol. 7, no. 8, pp. 437–444, Oct. 2009.

[2] P. Balvanera *et al.*, "Quantifying the evidence for biodiversity effects on ecosystem functioning and services," *Ecol. Lett.*, vol. 9, no. 10, pp. 1146–1156, Oct. 2006.

[3] S. R. Carpenter, E. M. Bennett, and G. D. Peterson, "Scenarios for ecosystem services: An overview," *Ecol. Soc.*, vol. 11, no. 1, pp. 29, 2006.

[4] S. Díaz, J. Fargione, F. S. Chapin, and D. Tilman, "Biodiversity loss threatens human well-being," *PLoS. Biol.*, vol. 4, no. 8, pp. 1300–1305, Aug. 2006.

[5] A. Dufour, F. Gadallah, H. H. Wagner, A. Guisan, and A. Butler, "Plant species richness and environmental heterogeneity in a mountain landscape: Effects of variability and spatial configuration," *Ecography*, vol. 29, no. 4, pp. 573–584, Aug. 2006.

[6] K. M. Bergen *et al.*, "Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for LiDAR and radar spaceborne missions," *J. Geophys. Res.*, vol. 114, no. G2, pp. G00E066-1–G00E06-13, Jun. 2009.

[7] K. J. Gaston, "Global patterns in biodiversity," *Nature*, vol. 405, no. 6783, pp. 220–227, May 2000.

[8] A. Pacini, S. Mazzoleni, C. Battisti, and C. Ricotta, "More rich means more diverse: Extending the environmental heterogeneity hypothesis to taxonomic diversity," *Ecol. Indicators*, vol. 9, no. 6, pp. 1271–1274, Nov. 2009.

[9] J. Dauber *et al.*, "Landscape structure as an indicator of biodiversity: Matrix effects on species richness," *Agric. Ecosyst. Environ.*, vol. 98, no. 1–3, pp. 321–329, Sep. 2003.

[10] B. Leutner *et al.*, "Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing," *Remote Sens.*, vol 4, no. 9, pp. 2818–2845, Sep. 2012.

[11] A. Ceballos, J. Hernández, P. Corvalán, and M. Galleguillos, "Combining LiDAR and hyperspectral remote sensing to estimate vascular plant richness of deciduous Mediterranean Forests of Central Chile," *Remote Sens.*, to be published.

[12] M. Higgins *et al.*, "Linking imaging spectroscopy and LiDAR with floristic composition and forest structure in Panama," *Remote Sens. Environ.*, vol. 154, pp. 358–367, Nov. 2014.

[13] J. Vogeler *et al.*, "Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests," *Remote Sens. Environ.*, vol. 147, pp. 13–22, May 2014.

[14] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[15] M. Machala and L. Zejdová, "Forest mapping through object-based image analysis of multispectral and LiDAR aerial data," *Eur. J. Remote Sens.*, vol. 47, pp. 117–131, 2014.

[16] B. Desclée, P. Bogaert, and P. Defourny, "Forest change detection by statistical object-based method," *Remote Sens Environ.*, vol. 102, no. 1/2, pp. 1–11, May 2006.

[17] C. Pascual *et al.*, "Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous Pinus sylvestris stands," *Forest Ecol. Manage.*, vol. 255, no. 11, pp. 3677–3685, Jun. 2008.

[18] W. Zho and A. Troy, "An object-oriented approach for analysing and characterizing urban landscape at the parcel level," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3119–3135, Jun. 2008.

[19] M. W. Palmer, P. G. Earls, B. W. Hoagland, P. S. White, and T. Wohlgemuth, "Quantitative tools for perfecting species lists," *Environmetrics*, vol. 13, no. 2, pp. 121–137, Mar. 2002.

[20] T. J. Stohlgren, *Measuring Plant Diversity: Lessons From the Field*. New York, NY, USA: Oxford University Press Inc., 2007.

[21] A. E. Magurran, *Ecological Diversity and Its Measurement*. Princeton, NJ, USA: Princeton Univ. Press, 1988.

[22] F. E. Fassnacht *et al.*, "Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass," *Remote Sens. Environ.* vol. 154, pp. 102–114, Nov. 2014.

[23] C. Briese, "Extraction of digital terrain models," in *Airborn and terrestrial Laser Scanning*, G. Vosselman and H. Maas, Eds. New York, NY, USA: Taylor & Francis, 2010, pp. 135–167.

[24] M. Etienne and C. Prado, "Descripción de la vegetación mediante la cartografía de ocupación de tierras," *Ciencias Agrícolas*, vol. 10. Santiago, Chile: Universidad de Chile, 1982.

[25] M. Tenenhaus, V. Esposito, Y.-M. Chatelin and C. Lauro, "PLS Path Modeling," *Comput. Stat. Data Anal.*, vol. 48, no. 1, pp. 159–205, Jan. 2005.

[26] F. E. Fassnacht *et al.*, "Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol 7, no. 6, 2547–2561, Jun. 2014.

[27] L. Lee, S. Petter, D. Fayard, and S. Robinson, "On the use of partial least squares path modeling in accounting research," *Int. J. Accounting Inf. Syst.*, vol. 12, no. 4, pp. 305–328, Dec. 2011.

[28] C. Bellard, C. Bertelsmeier, P. Leadley, W. Thuiller, and F. Courchamp, "Impacts of climate change on the future of biodiversity," *Ecol. Lett.*, vol. 15, no. 4, pp. 365–377, Apr. 2012.

[29] W. Turner *et al.*, "Remote sensing for biodiversity science and conservation," *Trends Ecol. Evol.*, vol. 18, no. 6, pp. 306–314, Jun. 2003.

[30] J. L. Hernández-Stefanoni *et al.*, "Modeling α- and β-diversity in a tropical forest from remotely sensed and spatial data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 19, pp. 359–368, Oct. 2012.