Mean height and variability of height derived from Lidar data and Landsat images relationship

Cristina Pascual[1], Warren Cohen[2], Antonio García-Abril[1], Lara A. Arroyo[3], Rubén Valbuena[1], Susana Martí-Fernández[1] and José Antonio Manzanera[1]

[1]Technical University of Madrid (UPM), E.T.S.I Montes,
Ciudad Universitaria s.n., 28040 Madrid, Spain. c.pascual@upm.es
[2]Forestry Sciences Laboratory, Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97311, USA
[3]Centre for Remote Sensing and Spatial Information Science, School of Geography, Planning and Architecture, University of Queensland, Brisbane, QLD 4072, Australia.

Abstract

The mean height and standard deviation of the height of the forest canopy, derived from lidar data show to be important variables to summarize forest structure. However lidar data has a limited spatial extent and very high economic cost. Landsat data provide useful structural information in the horizontal plane and have easy access. The integration of both data sources is an interesting goal for sustainable forest management. Different spectral indices (NDVI and Tasseled Cap) were obtained from 3 Landsat scenes (March 2000, June 2001 and September 2001). In addition, mean and standard deviation of lidar height werecalculated in 30x30m blocks. Correlation and forward stepwise regression analysis was applied between these two variables sets. Best correlation coefficients are achieved among mean lidar height versus NDVI and wetness for the three dates (range between 0.65 to -0.73). Others authors indicate that wetness is one of the best spectral indices to characterize forest structure. Best regression models include NDVI and wetness of June and September as dependent variables (adjusted r2: 0.55 – 0.62). These results show that lidar data can be useful for training Landsat to map forest structure but it should be interesting to optimize this approach.

Keywords: Lidar, Landsat, mean height, Forest structure

1. Introduction

Canopy structure can be defined as the organization in space and time, including the position, extent, quantity, type and connectivity of the aboveground components of vegetation (Parker, 1995; Lefsky et al., 1999). Structure includes vertical (e.g. number of tree layers, understory vegetation) and horizontal features (e.g. spatial pattern of trees, gaps) as well as species richness (Maltamo et al., 2005).

Mean height and standard deviation of height derived from lidar data, have shown to be variables that synthesise forest structure of the canopy. Zimble et al. (2003) used lidar-derived tree height variances to distinguish between single-story and multi-story classes. Lefsky et al. (2005a) pointed that mean height and height variability derived from lidar data are strongly related to canopy indices related to stand structure. These authors consider that these variables represent the same kind of enhancement of lidar data that the tasseled cap indices represent for optical remote sensing. Pascual et al. (2008) found that mean, median and standard deviation of height derived from lidar were useful for distinguishing among horizontally heterogeneous forest structure types.

Small footprint airborne laser scanners provide detailed information on vertical distribution of forest canopy structure (*Hyyppa* et al., 2008), but over a limited spatial extent with a high economic cost. Landsat data provide useful structural information in the horizontal plane and are much more accessible (Cohen & Spies, 1992). Therefore the integration of optical remotely sensed imagery and lidar data provides improved opportunities to fully characterize forest canopy attributes and dynamics (Wulder et al., 2007).

Hudak et al. (2002) developed spatial extrapolation of lidar data over Landsat images. The combination of lidar derived metrics and optical images has been also developed (Chen et al., 2004; Lefsky et al., 2005b). In addition, two coincident lidar transects, representing 1997 and 2002 forest conditions in boreal forest of Canada, were compared using image segments generated from Landsat ETM+ imagery (Wulder et al., 2007).

Given the relation between mean and standard deviation of height derived from lidar and forest structure, the objective of the present work is to evaluate the relationship between summaries derived from lidar and the spectral information of the Landsat satellite. Final aim of this work is to establish whether Landsat data can be used to predict lidar forest canopy height (mean and standard deviation).

2. Methods

2.1 Study are

A 127.10 ha (1293 x 983 m) area on the western slopes of the Fuenfría Valley (40º 45´N, 4º 5´ W) in central Spain was selected as the study area. The Fuenfría Valley is located in the northwest portion of the Madrid region (Figure 1). The predominant forest is Scots pine (*Pinus sylvestris,* L) with abundant shrubs (*Cytisus scoparious* (L.) Link., *C. oromediterraneus* Rivas Mart. et al., *Genistaflorida* L.*)* in some areas.

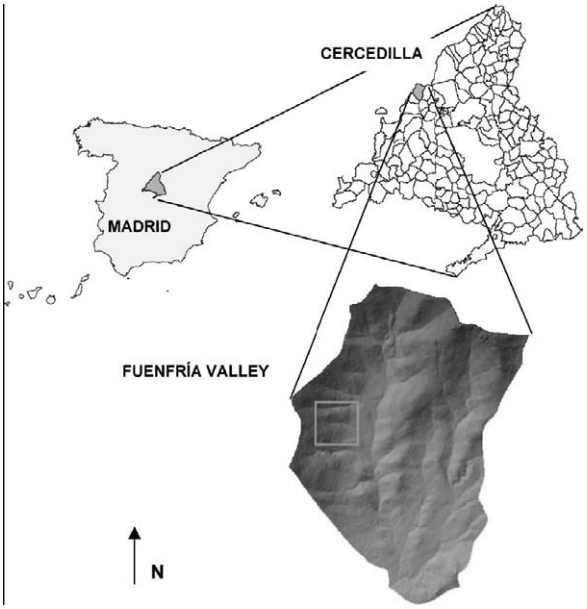

Figure 1: Study site. Fuenfría Valley in the village of Cercedilla, northwest of Madrid (Spain).

There are small pastures in the lowest part of the hillside. In the north sector of the study site there is an extensive rocky area. The site has a mean annual temperature of 9.4ºC and

precipitation averages 1180 mm per year. Elevations range between 1310 and 1790 m above sea level, with slopes between 20% and 45%. The general aspect of the study site is east.

2.2. Lidar data

A small-footprint lidar dataset was acquired by Toposys GmbH over the study area in August, 2002. The Toposys II lidar system recorded first and last returns with a footprint diameter of 0.95 m. Average point density was 5 points m^{-2}. The raw data (x, y, z-coordinates) was processed into two digital elevation models by TopoSys using the company's proprietary software. The digital surface model (DSM) was processed using the first pulse reflections and the digital terrain model (DTM) was constructed using the last returns. Filtering algorithms were used to identify canopy and ground surface returns for an output pixel resolution of 1 m horizontal and 0.1 m vertical resolution. According to Toposys calculations, the DSM and DTM, horizontal positional accuracy was 0.5 m and vertical accuracy was 0.15 m.

To obtain a digital canopy height model (DCHM), the DTM was subtracted from the DSM. Both the DTM and DCHM were validated before use by means of land surveying with total station and ground-based tree height measurements. The vertical accuracies, (Root Mean Square Error, RMSEs) obtained for the DTM in open areas and the DCHM under forest canopy were 0.30 m and 1.3 m, respectively (Pascual, 2006). These accuracies were acceptable for this study, and were in agreement with previous studies. For example, Clark et al. (2004) reported RMSEs for DTMs ranging from 0.06 to 0.61 m and for DCHMs ranging from 0.23 m to 2.41 m in tropical landscapes.

2.2. Image data and preprocessing

In this study we used three Landsat ETM+ images from scene path/row (201/32) that correspond to three different dates (March 15th, 2000, June 6th, 2001 and September 10th, 2001). Landsat images were georeferenced and radiometrically calibrated.

June and September's Landsat images were co-registered at the Alcalá University's Geography Department, using digital highways maps of Madrid region (E 1:50.000). RMSE was under 30 m (1 pixel), projection system was UTM (Datum Europeo 1950) and pixel of 30 m. We carried out a validation of the image co-registration in the study area using a series of easily recognisable points.

In the Landsat image of March 15th, a subset area of 30X30 km was orthorectified. Control points were selected, taking as reference September's georeferenced image. The source of altitudinal information was a 20-m pixel DTM of Madrid region. We used 38 control points homogenously spread out over the subset image. RMSE was 11.49 m (0.4 pixels). The COST absolute radiometric correction model of Chavez (1996) was applied to each image to convert digital counts to reflectance.

2.3. Lidar DCHM summaries (mean and standard deviation) and spectral indices.

The DCHM lidar (1 m pixel) was degraded to 30 m cell blocks providing a 30m-grid of 32 rows and 41 columns. Mean and standard deviation of the 900 lidar height values contained inside each 30 x 30 m block were calculated. Two new images 30 m pixel of mean and standard deviation of lidar height values were obtained.

Regarding optical images, NDVI and Tasseled Cap (TCAP) (brightness, greenness and wetness)

for March, June and September Landsat images were calculated.

2.4. Sampling design and statistical analysis

We create a mask of the forest Pinus sylvestris L. by unsuppervised classification of September image to exclude bare soil, rocks, pasture and shrubs for subsequent analysis (Figure 2). In addition, a systematic sampling design was performed to reduce the spatial autocorrelation inherent to the remote sensing imagery. Sampling procedure was estimated based on semivariograms calculations. We selected pixels at 130 – 150 m of distance each others (one of every 4 or 5 pixels). This distance was obtained through semivariograms calculation. Semivariograms 30 m lag (h) of the lidar DCHM mean height and Wetness tasseled component were calculated using the free distribution software Variowin 2.2. (Pannatier, 1996). Mean lidar height was selected for being the most interesting dependent variable and the TCAP Wetness component for being related to forest structure (Cohen & Spies, 1992). Variogram lag (h) was 30 m. The semivariance tends to stability at 130 -150 m. Two samples were obtained one for statistical model building and the other for an independent model validation.

Pearson's correlation among Landsat spectral indices and lidar statistical descriptors was performed. Furthermore, forward step regression analysis (p enter = 0.05; p remove = 0.05) between both variable set were also carried out. All statistical analysis were conducted with STATISTICA 6.1 software. Before proceeding to regression analysis, the normality of the dependent and independent variables was verified and transformed whenever was needed.

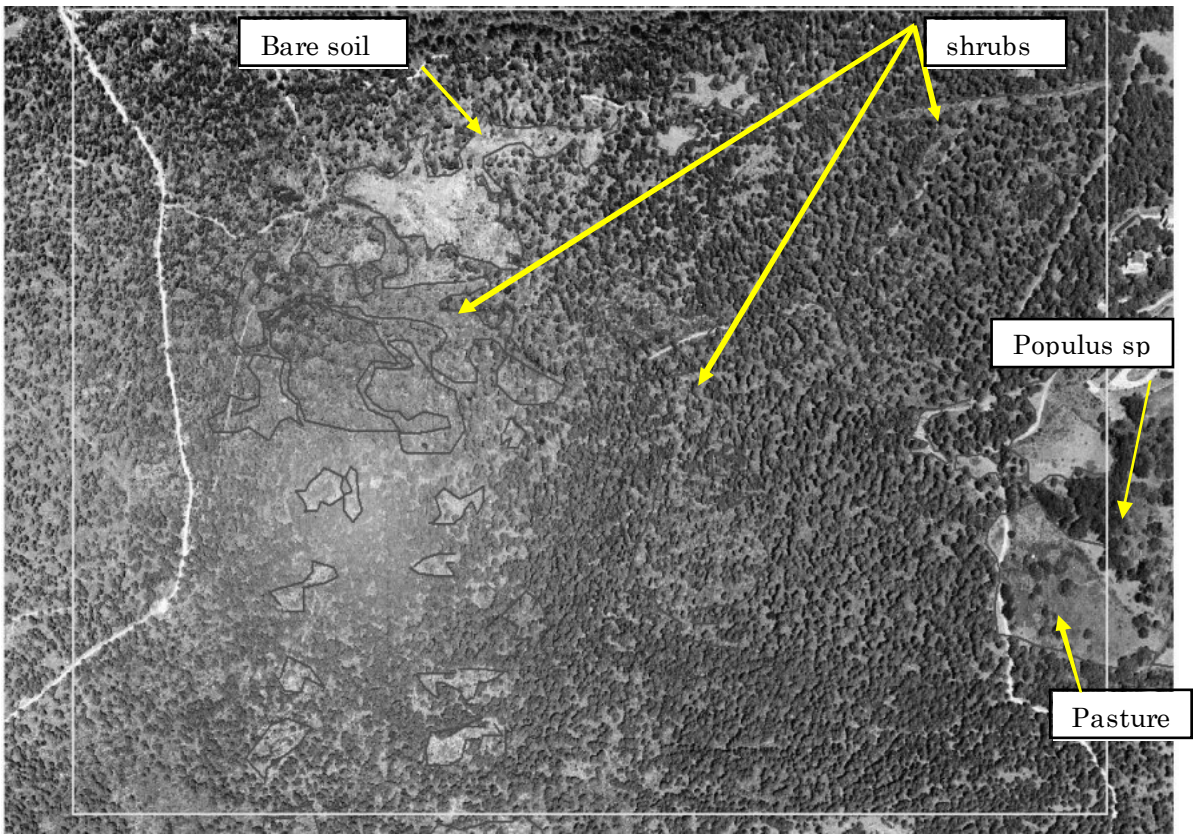

Figure 2. 0.5m pixel digital orthophotography of the study area (yellow frame). Different cobertures (pasture, bare soil, shrubs and *Populus* sp) were digitalised and labelled.

3. Results and Discussion

Mean lidar height and standard deviation of lidar height provided two images (Figure 3) with a gradient from black to white that represents the spatial variation of canopy height.

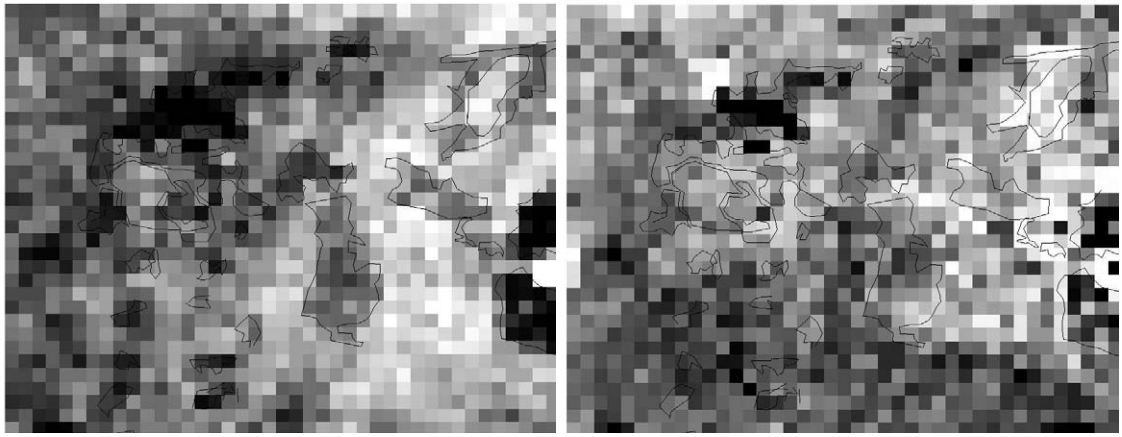

Figure 3. Mean lidar height image (30m pixel)(left) and Standard deviation of lidar height (right). Black to white gradient represent growing height values. Vectorial digitized covertures are included.

Correlations among NDVI indexes (Figure 4) and the squared mean lidar height ($\sqrt{hmean}$) indicated a moderately strong relationship among these variables ($r = 0.65$, $r = 0.70$ y $r = 68$; $p = 0.05$; $n = 47$ for March, June and September respectively). Standard deviation of lidar height (sd_30) demonstrated a scarce degree of relation with all NDVI indices for the three dates (Table 1).

Table 1: Pearson´s correlation between lidar derived metrics and spectral indices (n = 47).

March 15th				
	NDVI	**Br**	**Gr**	**We**
$\sqrt{hmedia}$	0.65*	-0.50*	0.46*	0.64*
sd_30	0.20	-0.16	0.18	0.20
June 6th				
	NDVI	**1/Br**	**Gr**	**Log(-We)**
$\sqrt{hmedia}$	0.70*	0.65*	0.50*	-0.72*
sd_30	0.30*	0.13	0.26	-0.11
September 10th				
	NDVI	**1/Br**	**Gr**	**Log(-We)**
$\sqrt{hmedia}$	0.68*	0.59*	0.34*	-0.73*
sd_30	0.29*	0.15	0.17	-0.04

*significant correlations $p < 0.05$; Br, Gr and We are brightness, greenness and wetness Tasseled components derived from ETM+.

Lu et al. (2004) found strong correlations between NDVI and forest attributes derived from field measurements. Nevertheless, Hall et al. (1995) and Franklin et al. (1997) do not consider this spectral index especially appropriate for the study of the forest attributes because of the weak correlation that has shown with certain parameters of vegetation. Regarding this, Lu et al. (2004) indicate that conclusions to its application vary depending on the biophysical parameters and the characteristics of the study area.

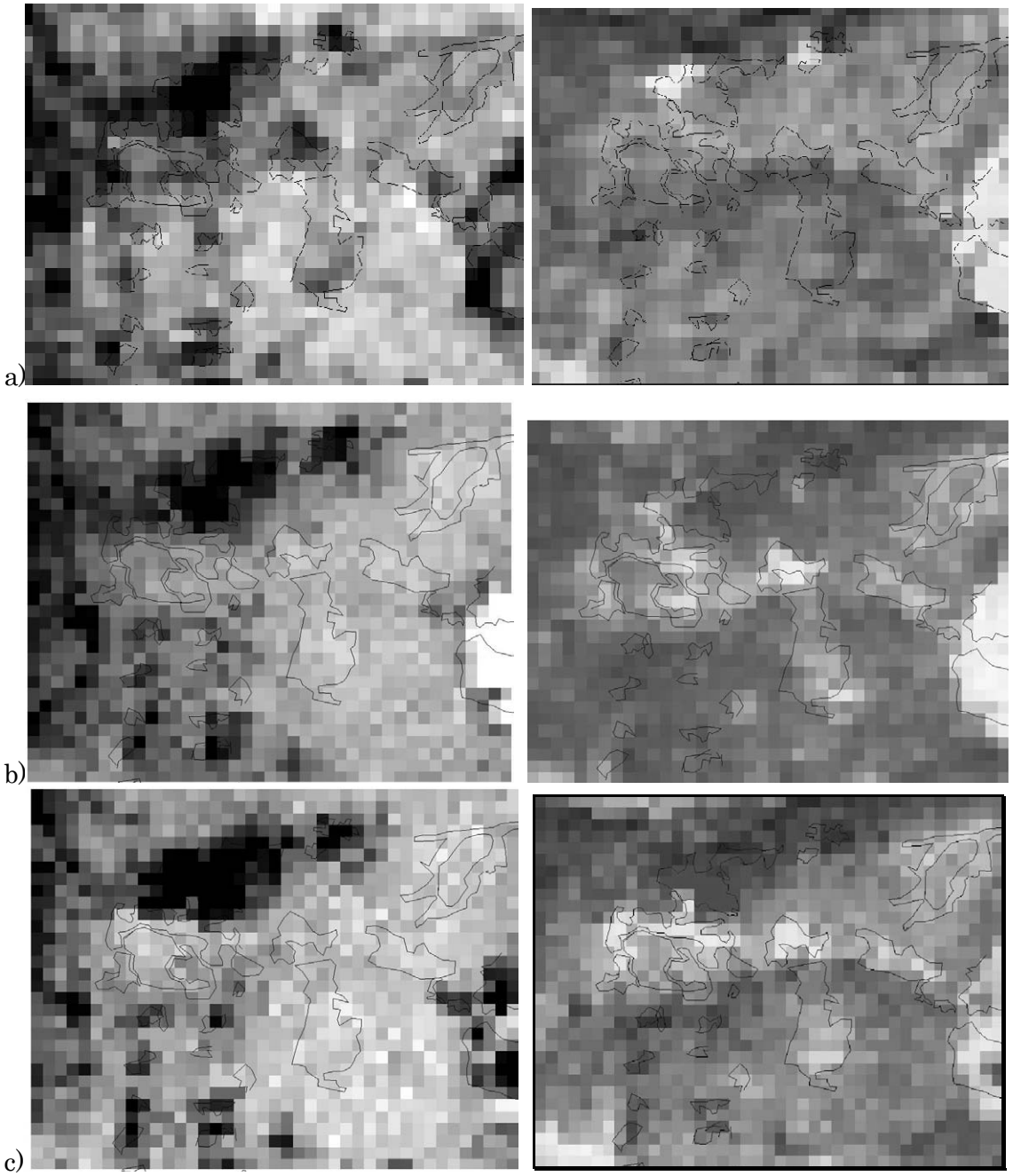

Figure 4. NDVI (left) and Colour composition of the TCAP components (right): Brightness in red channel; Greenness in green channel and Wetness in blue channel.g a) March 15th; b) June 6th and c) September 10th with the feature digitalized covertures (bare soil, pasture, shrubs)

Regarding TCAP transformation, the brightness and wetness of June and September (Figure 4) presented moderately strong correlations with the squared mean lidar height ($r = 0.65$, $r = -0.72$ y $r = 0.59$ $r = -0.73$; $p = 0.05$; $n = 47$ respectively) (Table 1). When considering Tasseled Cap components of each date separately, wetness presented higher correlations with the squared mean height. Other authors have also reported strong correlations between wetness component and multiple forest attributes measured in the field such as the dbh (diameter at breast height), crown diameter, mean height and basal area (Cohen & Spies 1992; Cohen et al. 1995). Wetness has been considered the most interesting spectral index to estimate forest structure of dense formations (Cohen & Spies 1992; Cohen et al. 1995; 2001). In addition, this component has

revealed as most significant when studying the temporal evolution of forests, such as mortality (Collins & Woodcok, 1994), harvesting and silvicultural activities (Wilson & Sader, 2002; Healey et al., 2005) or the evaluation of damages by plagues (Skakun, et al., 2003). Standard deviation of lidar height (sd_30) and Tasseled components revealed weak and not significant correlations (Table 1).

Regarding regression analysis (Table 2), the three models presented coefficients of determination ranging from 0.55 to 0.63. Standard deviation of height derived from lidar (SD_30) was excluded of regression analysis due to low Pearson´s correlation (Table 1).

None of the three models presented colinearity problems (i.e. linear relationship among the independent variables problems). The Variance Inflation Factor (VIF), as indicator of multicolinearity, did not present for any variable values close to 5 or 10. According to Montgomery, et al. (2002) those are the thresholds that question estimated regression coefficients by minimum squares.

Table 2: Forward stepwise regression models

Name	Models (forward stepwise regression)	r^2 adjusted	RMSE
Mod. NDVI	$\sqrt{hmean} = 1.137 - 0.0043 \cdot NDVI\_mar + 0.0085 \cdot NDVI\_jun$	0.55	4.07
Mod. TCAP	$\sqrt{hmean} = 3.970 + 0.133 \cdot Gr\_mar - 0.907 \cdot Log(-We\_sep)$	0.62	4.58
Mod. Mixed	$\sqrt{hmean} = 2.832 - 0.666 \cdot Log(-We\_sep) + 0.140 \cdot \sqrt{NDVI\_jun}$	0.59	4.32

A validation of regression analysis was performed using an independent sample of 54 pixels. Observed versus predicted values were represented in scatterplot graphs (Figure 5). All models showed a moderately strong adjustment (r = 0.73, p = 0.000; r = 0.72, p = 0.000 y r = 0.79, p = 0.000, n = 54 for Mod. NDVI, TCAP and MIXED respectively). Based on validation results best regression model were Mod. NDVI and Mod. MIXED.

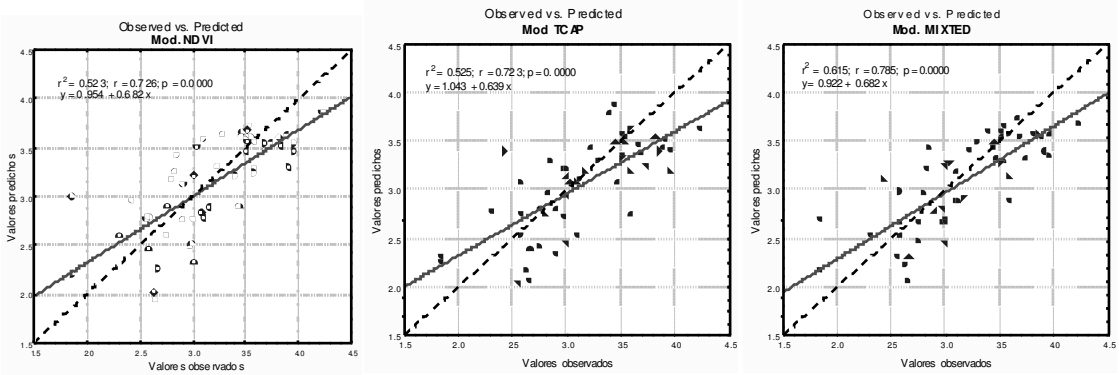

Figure 5. Scatterplots of independent (n=54) validation regression models (Observed vs. predicted). Left (Mod. NDVI); middle (Mod. TCAP) and right (Mod. MIXED).

Conclusions

Mean lidar height derived from lidar for the Scot pine forest of Cercedilla was estimated through the combination of spectral indices derived from Landsat images. Wetness TCAP component showed higher correlations with mean lidar height derived from lidar. Wetness

relationship with forest structure has been reported by different authors. Regression models were explicative, because of the relationships among variables. Nevertheless regression models presented a high variability (r2: 0.55 – 0.62) that diminishes its predictive capacity. These results show that lidar data can be useful for training Landsat to map mean height. Given the relationship among mean lidar height derived from lidar and the forest structure, landsat data can help to characterize forest structure. This approach should be analyzed in future research.

References

Chavez, P., 1996. Image-based atmospheric corrections—Revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025–1036.

Chen, X., Vierlinga, L., Rowell, E. and DeFelice, T., 2004,. Using lidar and effective LAI data to evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in a ponderosa pine forest. *Remote Sensing of Environment*, 91, 14–26.

Clark, M.L., Clark, D.B. and Roberts, D.A., 2004. Small-footprint lidar estimation of sub-canopy elevation and tree heigth in a tropical rain forest landscape. *Remote Sensing of Environment*, 91, 68 - 89.

Cohen, W. B. and Spies, T. A., 1992. Estimating structural attributes of Douglas-Fir western Hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, 41,1-17.

Cohen, W. B., Spies, T. A. and Fiorella, M., 1995. Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. Intern*ational Journal of Remote Sensing*, 16, 721–746.

Cohen, W. B., Maiersperger, T. K., Spies, T. A. and Oetter, D. R., 2001. Modeling forest cover attributes as continuous variables in a regional context with Thematic Mapper data. *International Journal of Remote Sensing*, 22, 2279–2310.

Collins, J. B. and Woodcock, C. E., 1994. Change detection using the Gramm–Schmidt transformation applied to mapping forest mortality. *Remote Sensing of Environment*, 50, 267–279.

Franklin, S. E., Lavigne, M. B., Deuling, M. J., Wulder, M. A. and Hunt Jr., E. R., 1997. Estimation of forest leaf area index using remote-sensing and GIS data for modeling net primary production. *International Journal of Remote Sensing*, 18, 3459–3471.

Hall, F. G, Shimabukuro, Y. E. and Huemmrich, K. F., 1995. Remote sensing of biophysical structure using mixture decomposition and geometric reflectance models. *Ecology Applications*, 5, 993-1013.

Healey, S. P., Cohen, W. B., Zhiqiang, Y. and Krankina, O. N., 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment*, 97, 301-310.

Hudak, A. T., Lefsky, M. A., Cohen, W. B. and Berterretche, M., 2002. Integration of LiDAR and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82, 397–416.

Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, S. and Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing,* 29, 1339-1366.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G, Spies T. A. and Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-Fir Western Hemlock forests. *Remote Sensing of Environment*, 70, 339-361

Lefsky, M. A., Hudak, A. T., Cohen, W. B. and Acker, S. A.,. 2005a. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of*

Environment, 95, 517–531.

Lefsky, M. A., Turner, D. P., Guzy, M. and Cohen, W. B. ,2005b. Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modelled forest productivity. *Remote Sensing of Environment*, 95, 549-558.

Lu, D., Mausel, P., Brondízio, E. and Moran, E., 2004. Relationship between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management,* 198: 149-167.

Maltamo, M., Packalen, P., Yu, X., Eerikainen, K., Hyyppa, J., Pitkanen, J., 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, 216, 41–50.

Montgomery, D., Peck, E. and Vining, G., 2002. Introducción al Análisis de Regresión Lineal. Compañía Editorial Continental. México.

Pannatier, Y. (1996). Variowin: Software for Spatial Data Analysis in 2D. Springer-Verlag. New York.

Parker, G. G.,1995. Structure and microclimate of forest canopies. In Forest Canopies—A Review of Research on a Biological Frontier (M. Lowman & N. Nadkarni, Eds.), Academic, San Diego, pp. 73–106.

Pascual, C., 2006. Análisis de la estructura forestal mediante teledetección: LiDAR (Light Detection And Ranging) e imágenes de satélite, Ph.D. thesis. Technical University of Madrid (UPM), Madrid.

Pascual, C., García-Abril, A., García-Montero, L.G, Martín-Fernández, S., Cohen, W.B.,2008. Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous Pinus sylvestris stands. *Forest Ecology and Management,* 255, 3677-3685.

Skakun, R. S., Wulder, M. A. and Franklin, S. E., 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sensing of Environment*, 86, 433-443.

Wilson, E. H. and Sader, S. A, 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80, 385– 396.

Wulder, M.A., Hana, T., White, J.C., Sweda, T. and Tsuzuki, H., 2007. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*, 110, 123-137

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. and Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, 87, 171 - 182.