BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../ R. VALBUENA ET AL.

Canonical Correlation Analysis for Interpreting Airborne Laser Scanning Metrics along the Lorenz Curve of Tree Size Inequality

RUBÉN VALBUENA[1], PETTERI PACKALEN[2], TIMO TOKOLA[2] AND MATTI MALTAMO[2]

[1] *European Forest Institute HQ. Yliopistokatu 6. 80100 Joensuu, Finland*

[2] *University of Eastern Finland. Faculty of Forest Sciences. PO Box 111 Joensuu, Finland*

Abstract

The objective of this study was to explore the explanatory capacity of airborne laser scanning (ALS) metrics with regard to tree size inequality properties from the forest. With this purpose, we selected the analysis of the Lorenz curve as a method for determining complexity in forest structure. The Lorenz curve is a representation of the relations of relative dominance among trees in the forest. Therefore, it presents a detailed description of the balance between overstory and understory, providing with valuable information on the degree of inequality among tree sizes in the forest. The methodology chosen was a canonical correlation analysis (CCA) of ALS metrics against regular quantiles along the Lorenz curve. Results demonstrated that most explanatory power can be yielded from indices of concentration of return heights, such as the L-coefficient of variation (i.e., Gini coefficient). This is highly relevant as it demonstrates the Lorenz curves from tree sizes and return heights to be closely related. Moreover, the study of separate canonical components allowed us to observe the correlation of certain metrics with each segment of the curve, detailing the effects that can be observed in ALS surveys in relation with tree stocking balance relations in multilayered forests. The first CCA component was more related to the dominant canopy, and therefore it influences the ALS surveys in a greater extent. This dominant layer is mainly described by canopy cover metrics, and thus it depends mainly on the forest stand relative density. The second CCA component was more related to the development of the understory, which influences the total amount of returns observed and the skewness of their heights. Future research studying the Lorenz curve from ALS surveys could provide forest inventories with important relations on forest structural characteristics.

Key words: airborne laser scanning, forest inventory, tree size inequality, Lorenz curve, canonical correlation analysis

Introduction

The Lorenz curve $M(x_r)$ is a cumulative distribution function (CDF) of a relative probability density distribution (PDF). This PDF is a rescaled density ratio of two distributions, and therefore a relative PDF (Handcock et al. 1999). In the case of forestry, these two distributions are the distribution of diameters at breast height (DBH) and its basal area-weighted counterpart (Gove and Patil 1997). A basal area-weighted PDF relates intrinsically to its original distribution, as it is defined by the quadratic relation between the diameter and the area of a circle (e.g., Gove 2003). The Lorenz curve therefore expresses the dominance of each tree in relation to its relative contribution to the total basal area and stem density. Lexerød and Eid (2006) and Valbuena et al. (2013a) employed this property to study tree size inequality and subsequently classify forest structural types.

The Gini coefficient (*GC*; Gini 1912, Glasser 1962) is a statistical descriptor directly related to the Lorenz curve. The *GC* is the ratio between the second and first L-moments, and it is therefore often referred to as L-coefficient of variation (*L.CV*; Hosking 1990). It is therefore a second order descriptor of concentration, i.e. relative dispersion. Previous forestry research work has concentrated mainly on the advantages of using (Knox et al. 1989, Lexerød and Eid 2006, Duduman 2011). However, more detailed analysis of the Lorenz curve and the L-moments has an interest in the sense of obtaining a scale invariant comparison of structural properties of the forest. For this reason, the attention has very recently been turned to studying tree size inequality by L-moments, especially with regard to their relations with airborne laser scanning (ALS) surveys (Ozdemir and Donoghue 2013, Valbuena et al. 2013b). The role of ALS in studying the complexity of forest structure is based on the capacity of

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

ALS to partially penetrate the dominant canopy, providing information about the understory (Bollandsås et al. 2008).

Figure 1 summarizes the relation between the Lorenz curve and the *GC*. In the Lorenz plot, the diagonal line represents a situation of complete equality, in which the relative difference between the PDF and the weighted PDF is zero, and hence *GC*=0 (Weiner and Solbrig 1984). The Lorenz curve has this property as it is multiplicatively scale invariant, in other words, two PDFs present the same Lorenz curve if they differ by a multiplicative constant (Handcock et al. 1999). Therefore, the *GC* provides a measure of variability in tree diameters which is invariant across development classes, and the value *GC*=0 is observed at any forest stand with all trees of equal DBH, regardless of their size. Furthermore, Valbuena et al. (2012) pointed out the importance in forestry of comparing Lorenz curves against the line rendered by a theoretical uniform PDF, which asymptotically obtains the middle value of *GC*=0.5. This line may therefore be used as a reliable discriminator between even and uneven-sized forest areas (Valbuena et al. 2013a). Finally, the highest dispersion in a PDF is given by a maximally bimodal distribution, which asymptotically obtains the maximum value of *GC*=1. In forestry this theoretical condition would be represented by a situation of one sole big tree accounting for most basal area, accompanied for an infinite number of very small seedlings accounting for most stem density (Staudhammer and LeMay 2001).

In this study, we explored the explanatory capacity of ALS remote sensing for describing the tree size inequality properties of the forest. Valbuena et al. (2013b) analyzed the relation of ALS metrics to specific indicators, which are descriptors about the amplitude and symmetry of the Lorenz curve. A natural step forward was to focus on a more profound analysis of the relations observed along the whole curve. The analysis of the entire Lorenz curve was chosen in order to observe the effect that relations of relative dominance among trees has in the return cloud rendered from an ALS survey. The aim of this research is to acquire a more profound knowledge of the effects that ALS metrics have at different segments of the Lorenz curve.

Materials and Methods

This study site was 800 ha of boreal forests surveyed with an ALTM Gemini sensor (Optech, Canada) at the municipality of Kiihtelysvaara in the province of North Karelia (Finland; approx. lat.: 62°31' N; lon.: 30°10' E; 130 –150 m above sea level). Dominant
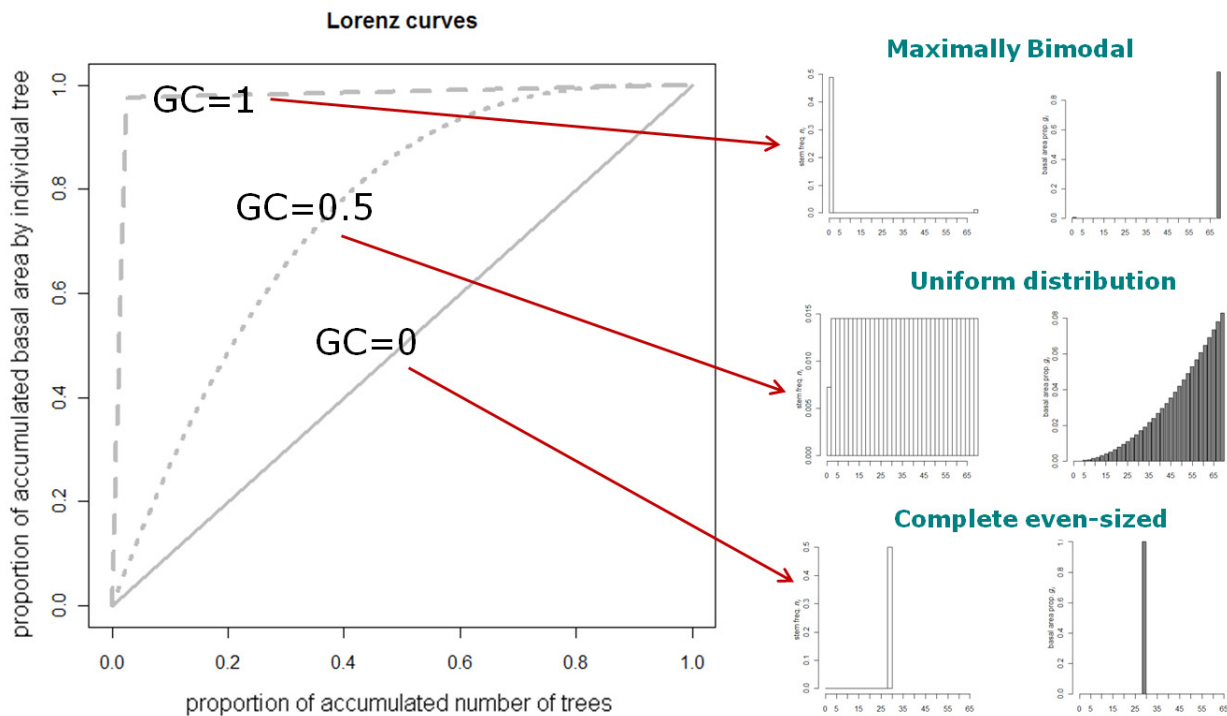

Figure 1. Lorenz curves of theoretical tree diameter distributions (white bars) and their corresponding basal area-weighted distributions (dark bars)

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

tree species is Scots pine (*Pinus sylvestris* L.), accounting for 72% of total standing volume, with a minor proportion of Norway spruce (*Picea abies* [L.] Karst.) and deciduous trees. The field sample consisted of 79 squared plots with sides sizing 20, 25 or 30 m, varying in relation to stand density. Plot location was determined subjectively, with the intention to include the range of variability in the area. Every tree (i) with DBH larger than 5 cm or height larger than 4 meters was measured and recorded. Basal areas were calculated for single stems (g_i, m^2) and ranked (r) according to decreasing DBH. Following Valbuena et al. (2012), sample Lorenz curves were obtained at plot-level as the cumulative proportions (p_i) of the total basal area in relation to the cumulative proportions of stem density x_r ($p_j^\# = n_j^\#/n$) accounted from each ranked tree. The sample bias-corrected estimator of *GC* developed by Glasser (1962) was employed for calculating the relative tree basal area differences observed within each forest plot.

We applied high density ALS data (11.9 pulses· m^{-2}). The ALS returns were processed using Terrascan software (Terrasolid, Finland), and a digital terrain model (DTM) was produced from returns classified as ground. Return heights above ground level were obtained by subtracting the DTM altitude underneath each individual ALS return. A set of ALS metrics was generated with FUSION software (version 3.1, USDA Forest Service), using 1 m as canopy cover threshold (McGaughey 2012). The ALS metrics detailed on Table 1 were computed from the height-above-ground distributions, following the state-of-the-art in ALS remote sensing (Næsset 2002). They included the abovementioned L-moments and their ratios (Hosking 1990), and therefore the Gini coefficient was, in this sense, also considered for the distribution of ALS return heights above ground.

Statistical analyses and modelling were carried out in R environment (version 2.15; R Development Core Team 2011). Least absolute shrinkage and selection operator (LASSO) method for variable selection (Tibshirani 1996) was implemented using package *GLMnet* (version 1.9-3; Friedman et al. 2010). LASSO is a predictor shrinkage method, which is a special case (α=1) of a penalized least squares method called elastic net. It minimizes the sum of squared residuals subject to the sum of the absolute values of the estimated coefficients ($\Sigma|\beta_j|$), that is to say, the L1-norm of the coefficients (t). LASSO was in this case used as a variable selection method since some of the coefficients are shrunken to zero, and hence discarded, by constraining the latter sum by a threshold $\Sigma|\beta_j| \le \pm$, also called the lasso parameter. The optimization algorithm, the least angle regression (LARS) (Efron et al. 2003),

Table 1. Summary of ALS-derived predictors

Lidar metrics		Symbol	
Moment statistics			
Central tendency		*Mean; Mode*	
Dispersion		*Var; SD; CV; MAD.P50; MAD.mode; CRR*	
Skewness/Kurtosis		*Skew; Kurt*	
Order statistics		*Max; Min; P50*	
Height quantiles		*Pi for i = 1,10,20,...,80,90,99*	
L-moments		*Li for i = 1,2,3,4*	
L-ratios		*L.CV; L.skew; L.kurt*	
Canopy cover			
		All returns	First returns
Total count		*Count.total*	*Count.total.f*
Count/Percent. > 1 m		*Count; Cov*	*Count.f; Cov.f*
Count/Percent.	> mean	*Count.mean; Cov.mean*	*Cov.mean.f*
Count/Percent.	> mode	*Count.mode; Cov.mode*	*Cov.mode.f*

SD: standard deviation: CV: coefficient of variation; MAD: Median absolute deviation above median (P50) and mode; CRR: Canopy relief ratio (see McGaughey, 2012)

was used to compute the entire path of LASSO solutions by stagewise additive fitting. The optimal solution was selected by choosing the elastic net coefficient (λ) which provided the smallest mean squared error in leave-one-out cross validation (LOOCV) (Zou and Hastie 2005). LASSO was carried out using the *GC* as response variable, in order to limit the number of predictors involved in the canonical correlation analysis (CCA) to only those truly related to the Lorenz curve of tree size inequality.

Regular intervals along the entire Lorenz curve were considered in order to observe the effects of ALS metrics at different portions of them. The sample Lorenz curves computed at plot-level were divided at regular quantiles, therefore obtaining a multivariate response $y=\{M(.05), M(.10), ... , M(.90), M(.95)\}$. CCA was carried out between this response and the predictor dataset resulted from the LASSO selection. Being in this case y a multivariate response, CCA was selected with the purpose of observing relations at different portions of the Lorenz curve. In addition, CCA is used as the basis for transforming the feature space for k-MSN imputation, a method widely employed for ALS estimation in forest inventories (Maltamo and Packalen 2014). A closer look to the individual canonical components, therefore, provides detail on which ALS metrics are most influential in the estimation of indicators related to the Lorenz curve. As the dispersion of the variables considered may differ significantly, we chose the variance-normalized version of CCA

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

weighted in relation to others (Stage and Crookston 2007). F-statistic significance test was applied to determine which canonical vectors were relevant to analyze.

Results

The variable reduction strategy carried by LASSO resulted in a reduced predictor dataset with 25 ALS metrics (Figure 2). We chose selecting the λ which provided the minimum LOOCV error (left vertical line), as opposed to choosing the error contained within one standard error of the minimum λ (right vertical line). We considered that prior variable reduction would otherwise have resulted in an excessively reduced predictor dataset for the actual purpose of CCA.

The significance F-test was positive for two canonical vectors in the CCA. In Table 2, the top 25th percentile of the absolute coefficient values are denoted in bold, therefore signifying those variables most involved in the canonical correlation. The first CCA component ($r = 0.97$, $p\text{-}value < 0.001$) explained variance predominantly in the left part of the Lorenz curve (lower quantiles), whereas the second CCA component ($r = 0.90$, $p\text{-}value < 0.001$) did in the right part of the

Lorenz curve (higher quantiles). Note that as coefficients have been computed over z-standarized variables, they can be directly compared with the purpose of analyzing their explanatory power. See Table 1 for the meaning of the predictor variables.

Discussion

In forestry, the meaning of the Lorenz curve discrete quantiles considered in this study $y = \{M(.05), M(.10), \dots, M(.90), M(.95)\}$, is grounded on the relative dominance of individual trees within the forest (Weiner and Solbrig 1984). Each of these $M(x_r)$ represents the proportion of cumulative basal area which corresponds to the proportion of cumulative stem density x_r accounted by the trees ranking $\geq r$ (Valbuena et al. 2013b). For the concave Lorenz curves hereby considered (from trees ranked according to decreasing DBH), the left part of the Lorenz curve represents the relative dominance of the upper layers. These approximately correspond to the $M(.05-.25)$ quantiles (Figure 3). On the other hand, the right part of the Lorenz curve represents the presence and development of understory and suppressed trees, approximately around the $M(.55-.95)$ quantiles. Thus, as the Lorenz curve represents relative tree dominance, each portion of the curve is related to the different layers that can be found in a multi-structured forest. The relative dominance of the upper strata is therefore represented in the right tail of the Lorenz curve, where-
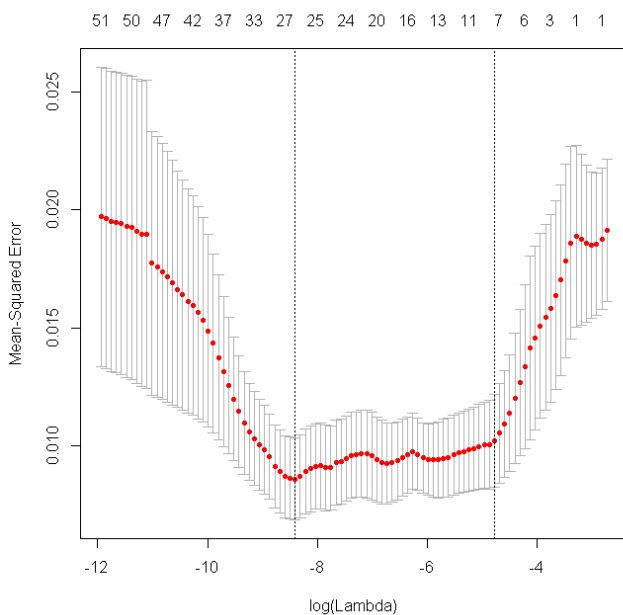

Figure 2. LASSO results. The mean squared error (MSE) curve obtained by LOOCV is shown in red colour, accompanied by corresponding bands extending one standard deviation. The x-axis shows the elastic net coefficient λ in logarithmic scale, whereas the top is annotated with the corresponding number of predictors selected. Hence, stagewise additive fitting sequences from right to left. Vertical lines are the optimization thresholds: minimum MSE (left) and MSE within one standard error of the minimum (right)
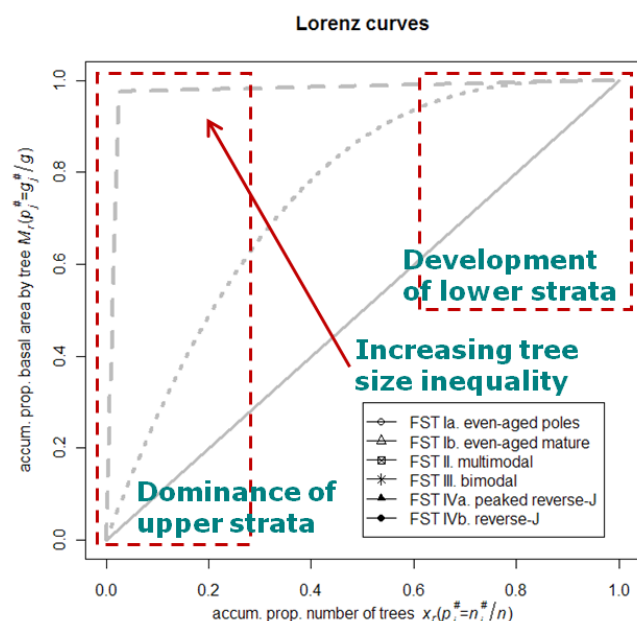

Figure 3. Relation of each portion of the Lorenz curve with the different components of forest structure

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

as the left tail describes the relative rarity of the lower strata.

The results obtained in the CCA analysis showed the interest of considering discrete portions along the Lorenz curve. They allowed us to attain a more profound understanding of the effect that ALS metrics have at different segments of the curve. It can be observed in Table 2a that each canonical component was roughly focused on either half of the curve. This is denoted by larger absolute coefficient values (numbers in bold) obtained for either tail of the curve at each CCA component. The $M(.05-.25)$ quantiles obtained higher values in the first CCA component. Hence, the predictors showing higher coefficient values in this component in Table 2b are more related to the upper strata. On the other hand, the $M(.55-.95)$ quantiles were mainly represented in the second CCA, and its corresponding ALS metrics are thus more related to the degree of development in the lower strata.

The middle part of the Lorenz curve, in the region of $M(.20-.50)$, is the area that distances itself the most from the diagonal. It is therefore the segment most related to the GC of DBH inequality, as the GC equals to the area comprised between the Lorenz curve and the diagonal (Gini 1912). This middle part was represented in both canonical components, showing some of the highest absolute values and therefore most influencing the relation with ALS metrics. Consequently, those predictors having large absolute values on both components are those with most explanatory

Table 2. Canonical correlation analysis (CCA) components for the Lorenz curve $M(x_r)$. See Table 1 for the meaning of the predictor variables

(a)	CCA components	
Response variable	CCA1	CCA2
$M(.5)$	-1.78	1.53
$M(.10)$	**3.29**	-3.98
$M(.15)$	**-2.54**	-0.75
$M(.20)$	-0.31	1.88
$M(.25)$	1.68	1.37
$M(.30)$	-0.69	1.66
$M(.35)$	1.41	-4.08
$M(.40)$	**-3.28**	4.00
$M(.45)$	-0.05	**-4.14**
$M(.50)$	**2.72**	0.44
$M(.55)$	-1.56	**4.43**
$M(.60)$	-1.09	1.44
$M(.65)$	1.41	-4.79
$M(.70)$	-0.49	0.27
$M(.75)$	0.31	-0.52
$M(.80)$	0.68	3.31
$M(.85)$	-0.86	**-4.85**
$M(.90)$	0.45	2.88
$M(.95)$	-0.39	-0.37

(b)	CCA components	
Predictor variable	CCA1	CCA2
Mode	-0.01	0.10
Var	0.08	-1.2
CV	**5.08**	**-9.97**
P01	0.44	-0.14
P05	-1.15	0.05
P20	0.14	0.61
P40	-0.23	0.72
Max	-0.27	2.26
L3	1.00	**-3.47**
L.CV	**-7.56**	**14.05**
L.kurt	0.25	0.61
MAD.P50	0.15	0.98
MAD.mode	0.27	**-2.88**
CRR	-0.40	2.22
Count.total	0.21	2.26
Count.total.f	0.30	-0.75
Count	1.23	**2.58**
Count.f	**-1.45**	**-3.98**
Cov	**-1.70**	-1.40
Cov.mean	-0.05	0.21
Cov.mode	**2.05**	0.93
Cov.mean.f	1.08	1.62
Cov.mode.f	**-1.65**	0.37
Cov.mode / Count.total.f	-0.32	-1.87

power with regards of tree size inequality. Results demonstrated that such ALS metrics were actually indices of concentration, i.e. dispersion of return heights relative to their average. One such ALS metric was the coefficient of variation (CV), which is the ratio between the standard deviation and the mean, second and first moment respectively. Most importantly, the ALS metric having the highest absolute coefficient values was the L-coefficient of variation ($L.CV$), which is also the ratio between the second and first L-moments. It is noteworthy to mention that $L.CV$ is actually the GC of ALS return heights (Hosking 1990). This demonstrates the Lorenz curves from tree sizes and return heights to be closely related. There is therefore a chance for finding scale invariant relations between the Lorenz curve of tree diameter PDFs and ALS return height PDFs. The upper layers of the canopy have relation of dominance with the understory both in terms of the number of seedlings growing underneath and the number of ALS returns reaching it (Bollandsås et al. 2008). The relation of dominance between overstory and understory is similar in both cases, and therefore the Lorenz curve and the L-moments are reliable means of studying these relations of balance among vertical strata in multilayered forests from ALS remote sensing (Ozdemir and Donoghue 2013, Valbuena et al. 2013a).

The upper strata were mainly represented in the first CCA component, and therefore ALS metrics with higher absolute coefficients in that component are more related to the dominant canopy. It can be observed in Table 2b that canopy cover metrics were more related to properties of the relative dominance of the overstory. ALS metrics with highest absolute values for the first CCA were: the proportion of return heights above their mode ($Cov.mode$) and above 1 m (Cov). Valbuena et al. (2013b) also found canopy cover metrics to affect all the indicators related to the Lorenz curve. Moreover, the type of return also seemed to be an important characteristic in terms of relative dominance of the upper strata, since the proportion of first returns above the mode ($Cov.mode.f$) was amongst the most relevant ALS metrics in the first component. Also, the total number of first returns ($Count.f$) was significant at both components, which may indicate this metric to be related with relations of relative density affecting all strata. Canopy cover may therefore be one of the forest properties most affecting the left segment of the Lorenz curve, as it is directly linked with relative dominance.

The relations with the lower strata could be observed independently in the second CCA component, on the other hand. After those already mentioned, the most relevant metric in this component was the third

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

L-moment (*L3*). As this is a metric describing asymmetry (Hosking 1990), this result is consistent with previous research describing the importance of components of Lorenz asymmetry in defining the relative importance of the understory (Valbuena et al. 2013a, 2013b). Other important metrics in the second component found in this study were the median absolute deviation from the mode (*MAD.mode*) and the total return count above 1 m (*Count*). It is worth noting that variance in the response $M(x_r)$ is intrinsically higher at the right tail of the Lorenz curve and lower at the left tail (Handcock et al. 1999). Hence, using z-standarization for the response (Stage and Crookston 2007), succeeded in emphasizing the effects over the higher quantiles in the second CCA component, which otherwise may have remained concealed. As a result, the effect on the understory was revealed, and these ALS metrics present most explanatory power with regards to the degree of development in the understory.

Conclusions

The L-coefficient of variation of ALS return heights, which is the Gini coefficient, was the most relevant ALS metric in the canonical correlation analysis. Therefore, this paper demonstrates the Lorenz curves of inequality in tree sizes and ALS return heights to be closely related. Each tail of the Lorenz curve provides different information on the relations among trees in a forest: degree of dominance in the overstory and development in the understory. Canonical correlation analysis along the Lorenz curve therefore reveals the correlation of certain metrics with each segment of the curve, detailing the effects that can be observed in ALS surveys in relation with tree stocking balance relations in multilayered forests. Canopy cover metrics are related to the overstory in terms of stand relative density, whereas the understory is associated with asymmetry metrics.

References

Bollandsås, O.M. Hanssen, K.H. Marthiniussen, S. and Næsset, E. 2008. Measures of spatial forest structure derived from airborne laser data are associated with natural regeneration patterns in an uneven-aged spruce forest. *Forest Ecology and Management* 255(3–4): 953-961.

Duduman, G. 2011. A forest management planning tool to create highly diverse uneven-aged stands. *Forestry* 84: 301-314

Efron, B. Hastie, T. Johnstone, I. and Tibshirani, R. 2004. Least angle regression. *The Annals of Statistics* 32 (2): 407-499.

Friedman, J. Hastie, T. and Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33(1): 1-22.

Gini, C. 1912. Variabilità e mutabilità. In: Pizetti, E. and Salvemini, T. (Eds.) 1955) Reprinted in Memorie di metodologica statistica. Rome: Libreria Eredi Virgilio Veschi. English translation in *Metron* 63(1): 3-38

Glasser, G.J. 1962. Variance Formulas for the Mean Difference and Coefficient of Concentration. *Journal of the American Statistical Association* 57(299): 648-654.

Gove, J.H. 2003. A note on the relationship between the quadratic mean stand diameter and harmonic mean basal area under size-biased distribution theory. *Canadian Journal of Forest Research*. 33:1589–1590.

Gove, J.H. and Patil, G.P. 1998. Modeling the basal area-size distribution of forest stands: a compatible approach. *Forest Science*. 44(2): 285–297.

Handcock, M.S. Fienberg, S. and Lievesley, D. 1999. Relative Distribution Methods in the Social Sciences . Springer. Secaucus, NJ, USA

Hosking, J.R.M. 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 52: 105-124

Knox et al. 1989. Population dynamics in Loblolly Pine stands: changes in skewness and size inequality. *Ecology* 70: 1153-1167

Lexerød, N.L. and Eid, T. 2006. Assessing suitability for selective cutting using a stand level index. *Forest Ecology and Management* 237: 503-512

Maltamo, M. and Packalen, P. 2014. Species specific management inventory in Finland. In: Maltamo, M. Naesset, E. and Vauhkonen, J. (Eds.): Forestry applications of airborne laser scanning: Concepts and case studies. Managing Forest Ecosystems 27. Dordrecht*: Springer*.

McGaughey, R.J. 2012. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. Version 3.10. Pacific Northwest Research Station. Seattle, WA.: USDA Forest Service.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80: 88-99.

Ozdemir, I. and Donoghue, D.N.M. 2013. Modelling tree size diversity from airborne laser scanning using canopy height models with image texture measures. *Forest Ecology and Management* 295: 28-37.

R Development Core Team. 2011. R: A Language and the? Environment for Statistical Computing.

Stage, A.R. and Crookston, N.L. 2007. Partitioning error components for accuracy-assessment of near neighbor methods of imputation. *Forest Science* 53(1): 62-72.

Staudhammer, C.L. and LeMay, V.M. 2001. Introduction and evaluation of possible indices of stand structural diversity. *Canadian Journal of Forest Research* 31: 1105–1115.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 58 (1): 267-288.

Valbuena, R. Packalén, P. Martín-Fernández, S. and Maltamo, M. 2012. Diversity and equitability ordering profiles applied to study forest structure. *Forest Ecology and Management* 276: 185-195.

Valbuena, R. Packalén, P. Mehtätalo, L. García-Abril, A. and Maltamo, M. 2013a. Characterizing Forest Structural Types and Shelterwood Dynamics from Lorenz-based Indicators Predicted by Airborne Laser Scanning. *Canadian Journal of Forest Research* 43: 1063-1074.

Valbuena, R. Maltamo, M. Martín-Fernández, S. Packalén, P. Pascual, C. and Nabuurs, G.J. 2013b. Patterns

BALTIC FORESTRY

CANONICAL CORRELATION ANALYSIS FOR INTERPRETING AIRBORNE LASER /.../   R. VALBUENA ET AL.

of Covariance between Airborne Laser Scanning Metrics and Lorenz Curve Descriptors of Tree Size Inequality. *Canadian Journal of Remote Sensing* 39 (S1): S18-S31.

Weiner, J. and Solbrig, O. 1984. The meaning and measurement of size hierarchies in plant-populations. *Oecologia* 61: 334–336.

Zou, H. and Hastie, T. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society B* 67(2): 301-320.

КАНОНИЧЕСКИЙ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ДЛЯ ИНТЕРПРЕТАЦИИ МЕТРИК ВОЗДУШНОГО ЛАЗЕРНОГО СКАНИРОВАНИЯ ВДОЛЬ КРИВОЙ ЛОРЕНЦА НЕРАВНОМЕРНОСТИ РАЗМЕРОВ ДЕРЕВЬЕВ

Р. Вальбуена, П. Пакален, Т. Токола, Ма. Малтамо

Резюме

Целью данного исследования было изучить пояснительную способность метрик воздушного лазерного сканирования (ALS) в отношении свойства неравномерности размеров деревьев леса. Для этого в качестве метода для определения сложности в структуре леса.нами использован анализ кривой Лоренца. Кривая Лоренца является индикатором отношений относительного доминирования среди деревьев в лесу. Таким образом, она представляет собой подробное описание баланса между верхним и нижним ярусом насаждения, предоставляя ценную информацию о степени неравномерности размеров деревьев в лесу. Выбранная методология была основана на использовании канонического корреляционного анализа (*ККА*, ССА) метрик воздушного лазерного сканирования против регулярных квантилей вдоль кривой Лоренца. Результаты показали, что наибольшей пояснительной способностью обладают показатели концентрации высот отражения, такие как L-коэффициент вариации (т. е., коэффициент *Gini*). Это имеет большое значение, поскольку демонстрирует, что кривые Лоренца, зависящие от размеров деревьев и высоты отражения тесно связаны между собой. Кроме того, изучение отдельных канонических компонентов позволило наблюдать корреляцию определенных метрик с каждым сегментом кривой, детализирующим эффекты, которые могут наблюдаться при проведении инвентаризации плотности деревьев в многоярусных лесах с использованием воздушного лазерного сканирования. Первый компонент ССА был более связан с доминирующим пологом, и поэтому он в большей степени влияет на инвентаризацию с использованием воздушного лазерного сканирования. Доминирующий ярус в основном описывается метриками покрова и, таким образом, зависит преимущественно от относительной плотности древостоя. Второй компонент ССА был более связан с развитием подлеска, влияющего на общую сумму возвратов и асимметрию его высоты. Последующие исследования с использованием кривой Лоренца, построенной на данных воздушного лазерного сканирования, смогут обеспечить инвентаризацию лесов важными данными о структурных характеристиках насаждений.

Ключевые слова: воздушное лазерное сканирование, инвентаризация леса, неравномерность размеров деревьев, кривая Лоренца, канонический корреляционный анализ