REVISTA DE TELEDETECCIÓN
Spanish Association of Remote Sensing
(2016) 45, Special Issue, 15-25
ISSN 1133-0953
EISSN 1988-8740
http://dx.doi.org/10.4995/raet.2016.4029

INVITED PAPER

Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods

Valbuena, R.*, Maltamo, M., Packalen, P.

University of Eastern Finland, School of Forest Sciences, PO Box 111. Joensuu, Finland.

Abstract: The area-based method has become a widespread approach in airborne laser scanning (ALS), being mainly employed for the estimation of continuous variables describing forest attributes: biomass, volume, density, etc. However, to date, classification methods based on machine learning, which are fairly common in other remote sensing fields, such as land use / land cover classification using multispectral sensors, have been largely overseen in forestry applications of ALS. In this article, we wish to draw the attention on statistical methods predicting discrete responses, for supervised classification of ALS datasets. A wide spectrum of approaches are reviewed: discriminant analysis (DA) using various classifiers –maximum likelihood, minimum volume ellipsoid, naïve Bayes–, support vector machine (SVM), artificial neural networks (ANN), random forest (RF) and nearest neighbour (NN) methods. They are compared in the context of a classification of forest areas into development classes (DC) used in practical silvicultural management in Finland, using their low-density national ALS dataset. We observed that RF and NN had the most balanced error matrices, with cross-validated predictions which were mainly unbiased for all DCs. Although overall accuracies were higher for SVM and ANN, their results were very dissimilar across DCs, and they can therefore be only advantageous if certain DCs are targeted. DA methods underperformed in comparison to other alternatives, and were only advantageous for the detection of seedling stands. These results show that, besides the well demonstrated capacity of ALS for quantifying forest stocks, there is a great deal of potential for predicting categorical variables in general, and forest types in particular. In conclusion, we consider that the presented methodology shall also be adapted to the type of forest classes that can be relevant to Mediterranean ecosystems, opening a range of possibilities for future research, in which ALS may show great predictive potential.

Key words: airborne laser scanning, discriminant analysis, maximum likelihood, minimum volume ellipsoid, naïve Bayes, support vector machine, artificial neural networks, random forest, nearest neighbour.

Clasificación de etapas de desarrollo forestal a partir de datos de vuelos lidar nacionales de baja densidad: comparación de métodos de aprendizaje automático

Resumen: Los métodos de estimación por áreas son ya habituales para el uso de escaneo láser aerotransportado (ALS) en la predicción de atributos forestales descritos por variables continuas: biomasa, volumen, densidad, etc. No obstante, apenas se ha prestado atención a los métodos de clasificación por aprendizaje automático, que por otro lado son frecuentes en diversos campos de la teledetección, como la identificación de coberturas del suelo por sensores multiespectrales. En este artículo hemos centrado la atención en métodos estadísticos destinados a predecir variables discretas, para obtener clasificaciones supervisadas de datos ALS. Varios métodos han sido revisados: clasificadores de análisis discriminante (DA) –máxima verosimilitud, elipsoide de volumen mínimo y Bayesiano ingenuo–, máquinas de vector soporte (SVM), redes neuronales artificiales (ANN), selvas aleatorias (RF), y vecino más próximo (NN). La

* *Corresponding author: rubenval@uef.fi*

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

comparación se ha realizado en el contexto de una clasificación en las etapas de desarrollo (ED) utilizadas en la gestión forestal de los bosques de Finlandia, utilizando para ello datos de baja densidad de su vuelo nacional. RF y NN produjeron las matrices de error más equilibradas, con predicciones por validación cruzada no sesgadas para todas las EDs. Aunque SVM y ANN mostraron las mayores precisiones, obtuvieron resultados muy dispares entre las distintas EDs, siendo óptimas sólo para algunas concretas. DA obtuvo los peores resultados, y sólo se observó que pudieran ser beneficiosas en la predicción de rodales en la etapa de diseminado. Nuestros resultados muestran que, además de la confirmada capacidad del ALS para predecir variables de cuantificación de las masas forestales, también existe potencial en la clasificación de variables categóricas en general, y tipologías forestales en particular. En conclusión, consideramos que estas metodologías se pueden también adaptar a las clases naturales de edad y tipos estructurales que sean relevantes en ecosistemas mediterráneos, abriendo un abanico de posibilidades para investigación en ALS, con mucho potencial.

Palabras clave: escaneo láser aerotransportado, análisis discriminante, máxima verosimilitud, elipsoide de volumen mínimo, Bayesiano ingenuo, máquinas de vector soporte, redes neuronales artificiales, selvas aleatorias, vecino más próximo, clases naturales de edad.

1. Introduction

Supervised machine learning methods build a classification model from training data (Hastie *et al.,* 2009). In the context of remote sensing, and airborne laser scanning (ALS) sensors in particular, forest types (silvicultural development classes, in this case) are linked to field plots using metrics describing the distribution of ALS returns backscattered from that same sampled area (Næsset, 2002; Maltamo *et al.,* 2006). These metrics are then used as independent variables (*X*), and the classification model is implemented for making predictions throughout the scanned area (McInerney *et al.,* 2010; McRoberts, 2012). There is a large assortment of statistical learning methods, which are often denominated non-parametric since, although flexible, the meaning of their parameters is usually difficult to grasp (Rogan *et al.,* 2008). For this reason, the simplest methods are preferred for generalizing clear phenomena, while more complex ones have the advantage of being flexible enough for describing intricate relations. The latter case may be found if the targeted development classes are only to be described as a complex combination of different ALS metrics. The goal is to find general but flexible rules, not too prone to outliers. This is to be found as a trade-off between having a too simplistic method lacking predictive power, while avoiding over-fitting to the sample by representing noisy data.

1.1. Discriminant analysis: classification based on maximum likelihood or minimum volume ellipsoid

In the context of ALS estimation, discriminant analysis has been scarcely employed for stratifying the forest areas into areas of homogeneous species composition (Holmgren *et al.,* 2008; van Aardt *et al.,* 2008; Kim *et al.,* 2009; Ørka *et al.,* 2009; Zhang *et al.,* 2011) or structural types (Maltamo *et al.,* 2014; Torresan *et al.,* 2014). These approaches have focused mainly on classification by linear discriminant analysis, while approaches to delimit quadratic boundaries –i.e. hyperplanes– among classes (Geerling *et al.,* 2007) have been largely overseen. One of the most widespread methods in other fields of remote sensing is the maximum likelihood (ML) classifier (Foody *et al.,* 1992; Blackard and Dean, 1999), which uses a probability model to discriminate the targeted classes. In ML, multivariate Gaussian kernels for each class to be separated are fitted using the covariance structure of the training data. A hard classification is carried out under the criteria of assigning the class with maximum posterior probability density in the feature space (*X*). Another alternative is to employ the minimum volume ellipsoid (MVE) estimator, which searches the hyperplanes using Rousseeuw's (1986) ellipsoids. For each class, MVE consists of determining the minimal hypervolume ellipsoid containing more than half of the training plots belonging to that class.

1.2. Naïve Bayes estimator

The Naïve Bayes (NB) classifier is a collection of independent classifiers based on Bayesian probability. The assumption is that the relation with the response classes is independent for every predictor. Therefore, instead of searching for the ML hyperplanes, a separate classifier is deducted for each predictor variable. Otherwise, NB is very similar to ML, in the sense that it is based on adjusting Gaussian kernels to the training data. As ML, the final hard classification in NB is based on selecting the classes with higher posterior probability. Only few studies have tested the ability of NB classification using ALS data (García-Gutiérrez *et al.,* 2010).

1.3. Support vector machine

In support vector machine (SVM) the hyperplane separating classes is defined by structural risk minimization. That is to say, the cost function is defined as a combination of maximizing the distance between the hyperplane and the training samples, and minimizing the error of samples falling into another class's hypervolume. As in ML, the hyperplanes are computed based in class posterior probabilities and quadratic optimization. Although SVM is very common for classification in multispectral remote sensing (Heikkinen *et al.,* 2011), studies including ALS are rather scarce (Dalponte *et al.,* 2008; García *et al.,* 2011).

1.4. Artificial neural networks

Artificial neural networks (ANN) are a set of mathematical models, in this case multinomial log-linear, defining the distribution of the training data in both X and Y. In an ANN, predictors are the input of the network, whereas the output is the response variable. Hence, dependencies among variables are depicted from the structure of the network. These relations are decomposed into neurons, which are interconnected by weights which are iteratively adapted (Atkinson and Tatnall, 1997). In computing predictions, there is a search for matches with patterns observed for each class in the training data, by converging into a minimum mean squared error. The ANN method is therefore well-suited for identifying complex patterns, although it has also been criticized for being a "black box" solution, as the resulting networks are very difficult to interpret (Rogan *et al.,* 2008). Although the use of ANN for classification is common in the remote sensing literature (Blackard and Dean, 1999), to our knowledge, its application with ALS sensors has so far been confined to prediction of continuous variables (Niska *et al*., 2010).

1.5. Random Forest

Random forest (RF) is becoming increasingly popular in ALS remote sensing (Falkowski *et al.,* 2009; Yu *et al.,* 2011; Reese *et al.,* 2014; Valbuena *et al.,* 2014). RF consists of bootstrapping –i.e. applying random sampling with replacement– the training data and computing a classification tree with each bootstrap sample. A classification tree consists of generating a stepwise classification by partitioning the explained variance recursively, in a succession of binary splits of predictor thresholds determined under the criterion of residual sum of squares minimization (Martinuzzi *et al.,* 2009). The outcome is an ensemble: a collection of classification models that can be combined by leveraging their collective strengths, finally using the majority value for the classification (Hastie *et al.,* 2009). However, while being potentially powerful in the accuracy of the final classification, the collection of classification trees is overall useless when the primary interest is on the relations underlying those predictions (Torresan *et al.,* 2014). In the context of ALS, RF classification has been used to discriminate species (Ørka *et al.,* 2012), vegetation classes (Reese *et al.,* 2014), or successional stages (Falkowski *et al.,* 2009).

1.6. Nearest neighbour imputation methods

Nearest neighbour (NN) methods consist of carrying out a classification based on computing statistical distance metrics to reference sample plots in the feature space X (Kilkki and Päivinen, 1987; Tomppo, 1993). Advantages of nearest neighbour methods include simplicity and capacity for modelling complex covariance structures (e.g., Valbuena *et al.,* 2014). Although the use of nearest neighbour-based methods is widespread in ALS estimation of continuous variables (Maltamo *et al.,* 2006; Hudak *et al.,* 2008; McInerney *et al.,* 2010; Vauhkonen *et al.,* 2010), only few studies

have considered its employment for forest area classification (Dalponte *et al.,* 2008; Falkowski *et al.,* 2010; Korpela *et al.,* 2010; Torresan *et al.,* 2014).

2. Objectives

In this study, we aim at obtaining a classification of forest areas into silvicultural development stages used in practical forest management in Finland. We carried out a comparison of supervised classification methods using predictors derived from an area-based method for ALS processing, as an initial screening for selecting the most suitable method to the targeted classes and their relations to the area-based ALS metrics. The objective of the particular piece of research presented in this article was therefore to select the supervised classification method most suitable for this task. We were interested not only on the overall accuracy of classification for all the silvicultural development classes, but also on determining whether each machine learning method was better suited for certain development classes than the others, especially in detecting if any method may be biased toward underestimating or overestimating any specific class.

3. Material and methods

3.1. Study area and forest development classes

This research was conducted in Finland, focusing in a study area located in the region of North Karelia, surrounding Joensuu and extending to Outokumpu in the West and Pyhäselkä to the South. The total extent of the area was 252,000 ha, approximately 200,000 ha of which are covered by forest, the remaining (~20%) being lakes, urban and agricultural land. Most of these forests are stand-wise managed, and silvicultural development classes (DC) are employed in many decision-making stages in forest planning. These DCs are (codes denoting each class derive from their original names in Finnish language):

- Seedling stands (*T1*). Planted seedlings with average height lower than 1.3 m.
- Sapling stands (*T2*). Their heights are higher than 1.3 m, and diameters at breast height DBH <8 cm.
- Young thinning stands (*02*). Forest with poles of DBH=8-16 cm and dominant heights around 7-9 m.
- Advanced thinning stands (*03*). Trees mainly having DBH>16 cm.
- Mature stands (*04*). These stands have quadratic mean diameter QMD=18-25 cm.
- Shelterwood stands (05). Fairly uncommon in silvicultural systems dominated by rotation management, they are used in Boreal forests for natural regeneration of spruce.
- Regeneration stands (*Y1*). They are multi-layered forests that can be derived as a successional stage of many of the above DCs. While containing trees at a mature overstory, which serves as shelter, it will also have a spruce-dominated understory of seedlings and saplings.
- Seed-tree stands (*S0*). Used for natural regeneration of pine and birch, they are stands where few trees have been left after harvesting to provide with seeds for the next generation.

Field data consisted of 679 randomly positioned field circular plots sizing 255 m^{-2}, all measured during the summer 2013. These DCs were identified from the Finnish Forest Centre's (Suomen Metsäkeskus, SMK) stand database for the initial random allocation of plot positions. The DC for which each sample plot belonged to was nevertheless reassessed in the field by the expert field crew, assuring that no field plot was included if there was a discrepancy between the database and the actual DC observed in the forest. High-grade global navigation satellite systems (GNSS) were employed to determine plot centre positions, as required for their combination with the ALS datasets (Valbuena, 2014). GNSS receiver was a Trimble GeoXH with an external antenna elevated 5 m from the ground, and differentially corrected at post-processing stage using Trimble VRS network. We refer the reader to Valbuena *et al.* (2016)

for more details about the field data acquisition campaign.

3.2. ALS data and processing

Data was downloaded from the National Land Survey of Finland (NLS, 2013). The survey was appointed to Blom Kartta Oy (Finland). Laser data were acquired during May 2012 with an ALS60 system from Leica Geosystems (Switzerland). Flying at a height of 2300 m above ground rendered an average density of 0.91 points per squared-meter. Software FUSION (version 3.1, USDA Forest Service) was used to compute the ALS predictors (McGaughey, 2012), which were L-moments and their ratios, and canopy cover metrics (Valbuena *et al.,* 2015) from the ALS returns backscattered from the position of the field plots.

3.3. Statistical methods

Different machine learning methods for supervised classification were tested in R statistical environment (version 3.1.0; R Development Core Team, 2014). Quadratic discriminant analysis was implemented with package *MASS* (Venables and Ripley, 2002), for ML (*mle* method) and MVE (*mve* method) classification. Package *e1071* (Meyer *et al.,* 2014a) was employed for NB classification using function *naiveBayes*, and also for computing a SVM *C*-classification using *svm* function. For ANN implementation, we computed a feed-forward neural network with a single hidden layer, which was carried out with *nnet* package (Ripley, 1996). Liaw and Wiener's (2002) package *randomForest* was used for the implementation of the RF algorithm. NN imputation was carried out using package *class* (Venables and Ripley, 2002). The accuracy of these methods was assessed by leave-one-out cross validation. That is, each sample plot was employed to validate the prediction of separate models fitted after eliminating that given plot from the training dataset. Cross-validated contingency matrices were therefore elaborated for detailed accuracy assessment, using the *CrossTable* function included in package *gmodels* (Warnes, 2013) for inferring their statistical significance. Bias was assessed as the discrepancy between producer and user's accuracies for each forest structural types. The degree of misclassification was evaluated by the final overall accuracy and Cohen's (1960) kappa coefficient (κ), as implemented in package *vcd* (Meyer *et al.,* 2014b).

4. Results

4.1. Maximum likelihood classification

The ML classification reached an overall accuracy of 64.6%, obtaining a coefficient of agreement κ=0.57 (Table 1). Overall, it can be affirmed that there was a slight overestimation of multi-layered forests (*05* and *Y1*) and underestimation of even-sized (*02*, *03* and *04*) ones. The ML method was actually the one showing a best prediction for shelterwood plots (*05*), being able to detect 48.4% of these areas. Also, 26.53% seed-tree areas (*S0*) were mistakenly classified as regeneration stands (*Y1*). Cross-validated predictions showed that 24.2% of sapling (*T2*) stands were classified as being seedling (*T1*) areas, therefore having a strong bias toward *T1*.

4.2. Minimum volume ellipsoid

The overall accuracy for the MVE classification was as low as of 58.2%, showing a coefficient of agreement of only κ=0.50 (Table 2), which were the lowest of all methods. 34.4% of sampling

Table 1. Maximum likelihood estimator: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	35	52						6	37.6%
T2	1	151	4					4	94.4%
02		8	76	22	1	1			71.0%
03			21	86	11	1	1		71.7%
04				15	43	9		1	63.2%
05			6	11	22	15	6		25.0%
Y1			7	7	3	5	4	13	10.3%
S0		1	2				3	25	80.6%
Producer's accuracy	97.2%	71.6%	65.5%	61.0%	53.8%	48.4%	28.6%	51.0%	679

stands (*T2*) were classified as being young forests (*02*). Shelterwood (*05*) areas were also overestimated, mainly due to misclassifications from even-sized (*03* and *04*) forests. As in ML, there was an overestimation of regeneration (*Y1*) areas, mainly due to 22.4% of seed-tree plots (*S0*) being wrongly classified as *Y1*.

4.3. Naïve Bayes estimator

For the NB classification, overall accuracy was 67.2% and the coefficient of agreement κ=0.60 (Table 3). While not providing the best results, the cross-validated contingency table showed that the final classification was one of the least unbiased of all the comparisons. However, as in ML the problem of overestimating seedling stands (*T1*) persisted, as 23.6% of sapling (*T2*) stands were mistakenly classified as being *T1*.

4.4. Support vector machine

The SVM classification reached the best general results of all tested methods, as overall accuracy was 75.0% and the coefficient of agreement was κ=0.68 (Table 4). However, details in Table 4 show that classes *T1* and *Y1* were virtually neglected. On the other hand, SVM was seemingly well-suited for seed-tree areas (*S0*), being able to correctly detect 77.6% of them.

4.5. Artificial neural networks

Results of the ANN classification were also quite satisfactory, with an overall accuracy of 71.7% and κ=0.64 (Table 5). As in SVM, there was an overestimation of sapling (*T2*) plots, with 80.6% of seedling (*T1*) areas being wrongly classified as *T2*, reversing the tendency showed by ML, MVE and NB. Also like SVM, results were especially erroneous for multi-layered DCs (*05* and *Y1*).

Table 2. Minimum volume ellipsoid estimator: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	24	23						4	47.1%
T2	12	112						4	87.5%
02		73	77	13	1	3	1	4	44.8%
03			30	93	17	1		1	65.5%
04			2	16	49	10	6	1	58.3%
05		4	2	16	13	14	3	2	25.9%
Y1			5	3		3	4	11	15.4%
S0								22	100.0%
Producer's accuracy	66.6%	52.8%	66.4%	66.0%	61.3%	45.2%	28.6%	44.9%	679

Table 3. Naïve Bayes estimator: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	34	50						5	38.2%
T2	2	152	3					5	93.8%
02		9	85	18	1	4	1		72.0%
03			28	103	23	8	1		63.2%
04				12	48	9		1	68.6%
05				5	8	6	5	4	21.4%
Y1				2		4	1	7	7.1%
S0		1		1			6	27	77.1%
Producer's accuracy	94.4%	71.7%	73.3%	73.0%	60.0%	19.4%	7.1%	55.1%	679

Table 4. Support vector machine: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	0								0.0%
T2	36	210	9					10	79.2%
02		2	83	13		4	1		80.6%
03			24	121	25	5	3		68.0%
04				3	52	15		1	73.2%
05				3	3	5	2		38.5%
Y1						1	0		0.0%
S0				1		1	8	38	79.2%
Producer's accuracy	0.0%	99.1%	71.6%	85.8%	65.0%	16.1%	0.0%	77.6%	679

4.6. Random forest classification

The RF results showed the best trade-off between exactness and precision. While being roughly unbiased for all classes, the overall accuracy and coefficient of agreement reached 72.6% and κ=0.66 respectively (Table 6). Although there was a slight overestimation of advanced thinning areas (*03*), mainly due to misclassification of other similarly even-sized areas (*02* and *04*), Table 6 shows no major flaws in this classification toward a certain DC. Compared to other methods, the RF algorithm had the best results in identifying regeneration forests (*Y1*), correctly detecting 50.0% of the sample plots.

4.7. Nearest neighbour imputation

Results for NN were fairly similar to those obtained by RF, as the contingency matrix showed unbiased classification for all DCs (Table 7). Accuracy measures were however lower, with an overall accuracy of 65.4% and a coefficient of agreement κ=0.57, comparable to those obtained by methods based on discriminant analysis (see 4.1 and 4.2). However, NN imputation showed to be especially better suited for detecting multi-layered forests than any other methods, as it was capable to obtain some of the best results for *05* and *Y1* simultaneously.

5. Discussion

Results showed some critical differences among methods. Table 8 includes a summary of methods' performance, for comparison. Overall accuracies ranged 58.2-75.0%. Some degree of convergence and similarities between some groups of methods could, however, be also discussed. Most of the errors observed occurred between DCs that were, regarding the actual chronosequence of

Table 5. Artificial neural networks: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	7	6						2	46.7%
T2	29	205	5					8	83.0%
02		1	82	12		5		3	79.6%
03			29	105	23	6	7	4	60.3%
04				17	55	16	1	1	62.5%
05				5	2	2	1		20.0%
Y1				1			1		50.0%
S0				1		2	4	31	81.6%
Producer's accuracy	19.4%	96.7%	70.7%	74.5%	68.8%	6.5%	7.1%	63.3%	679

Table 6. Random forest: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	17	14						2	51.9%
T2	19	194	7					8	85.1%
02		5	85	16		4	1		76.6%
03			23	105	23	5	3		66.0%
04				15	51	11	2	1	63.8%
05			1	3	5	8	1	1	42.1%
Y1				2	1	3	7	4	41.2%
S0		3						33	91.7%
Producer's accuracy	47.2%	89.8%	73.3%	74.5%	63.8%	25.8%	50.0%	67.3%	679

Table 7. Nearest neighbour imputation: contingency table.

	Observed								
Predicted	*T1*	*T2*	*02*	*03*	*04*	*05*	*Y1*	*S0*	User's accuracy
T1	9	19						3	29.0%
T2	26	186	9					9	80.9%
02		5	69	23	2	4	1	1	65.7%
03			32	87	22	5	5	1	57.2%
04			1	18	48	7	1	1	63.2%
05			3	7	7	14	1		43.8%
Y1			1	5		1	6	3	37.5%
S0		2	1	1	1			31	86.1%
Producer's accuracy	25.7%	87.7%	59.5%	61.7%	60.0%	45.2%	42.9%	63.3%	679

successional stages, close to one another (i.e. cells close to diagonal in confusion matrices, Tables 1-7). Such errors are less important in terms of the applicability of the resulting maps in forest management. For example, mistakenly classifying a young forest (*02*) as an advance thinning (*03*) would have a lesser impact in terms of erroneous forest management decision-making than a misclassification as a shelterwood area (*05*). In that same sense, confusion between seed-tree plots (*S0*) and seedling (*T1*) or even sapling (*T2*) areas may also be of lesser importance (i.e. cells situated in the corners opposite to diagonal in confusion matrices, Tables 1-7). Final Cohen's (1960) kappa (κ) values would have been higher if a matrix of weights would have been applied taking these into account (Meyer *et al.,* 2014b). However, there was no objective way to implement this, and we therefore decided to skip such practice, to allow comparability with similar studies.

Table 8. Comparison of methods.

Method	overall accuracy (%)	kappa coefficient (κ)
Maximum likelihood (ML)	64.6	0.57
Minimum volume ellipsoid (MVE)	58.2	0.50
Naïve Bayes (NB)	67.2	0.60
Support vector machine (SVM)	75.0	0.68
Artificial neural networks (ANN)	71.7	0.64
Random forest (RF)	72.6	0.66
Nearest neighbour (NN)	65.4	0.57

Most approaches showed a tendency to overestimate even-sized (*02*, *03* and *04*) forests and underestimate multi-layered (*05* and *Y1*) methods (most critically SVM and ANN, but also NB and RF in a lesser extent). These biases may be due to the low density nature of the ALS dataset employed, and the impossibly for the machine learning algorithm to discriminate forest areas where intermediate returns may be backscattered from diverse tree strata (multi-layered), of simply from bare trunks in even-sized forests. This tendency was reverted by DA methods (ML and MVE), which principally overestimated multi-layered forests (*05* and *Y1*). Regarding the bias of the final classification, NN obtained the most balanced contingency matrix (Table 7), showing proportions for all DCs fairly equal to the observed ones.

Detection of seedling areas (*T1*) was challenging for most algorithms. The case was especially critical for SVM, which totally neglected this class, probably due to the overlapping of the sapling areas (*T2*) in the ALS metric values used in the training sample. Methods based on probabilities, ML and NB, were therefore those which performed best for this specific task. They could therefore be suitable for the cases where identifying this specific DC may be critical for the final use of the classification maps.

Regarding the overall accuracy of the methods tested, we notice that approaches could be grouped into three types of outcomes. First, the results obtained by the MVE-based classification were insufficient for most practical purposes, and therefore that approach is not recommended. Some other methods obtained fairly satisfactory outcomes: ML, NB and NN. In that sense, NN showed a balanced confusion matrix with an overall accuracy that could suffice for many purposes in forest management. NN has been appreciated as a simple method that can be really efficient, if sufficient and representative training data is available (McInerney *et al.,* 2010; McRoberts, 2012; Valbuena *et al.,* 2014). Rest of methods, SVM, RF and ANN, obtained results that outperformed the others'. However, the above-mentioned biases observed can render them unreliable for many purposes, as it can be a sign of over-fitting to the sample. This could easily be the case for SVM (Dalponte *et al.,* 2008) and ANN (Rogan *et al.,* 2008). The RF method may provide the best trade-off between accurate classification not too biased toward certain DCs. This may be the reason why it has been chosen by many authors for classifying ALS datasets (Falkowski *et al.,* 2009; Ørka *et al.,* 2012; Reese *et al.,* 2014 Torresan *et al.,* 2014).

Each method showed different advantages and some serious pitfalls to pay attention when evaluating the different methods. The choice of method can therefore depend on the assets showed by each of them. Tables 1-7 show the accuracy of supervised machine learning methods (Hastie *et al.,* 2009) for building classification

models from training data. Although it requires the use of field data, the advantage is that classifications are adapted to the pilot areas, and the method itself presents no trouble to be carried out elsewhere, as classifications are always calibrated to local field plots. Another advantage of employing supervised classifications is that the categories considered can be tailored to the forest management needs and the particularities of each study area. Therefore, the presented methodology shall also be adapted to the type of forest classes that can be relevant to Mediterranean ecosystems, opening a range of possibilities for future research, in which ALS may show great predictive potential.

6. Conclusions

We observed that RF and NN had the most balanced error matrices, with cross-validated predictions mainly unbiased for all DCs. Although overall accuracies were higher for SVM and ANN, their results were very dissimilar across DCs, and they can therefore be only advantageous if certain DCs are targeted. DA methods underperformed in comparison to other alternatives, and were only beneficial for the exact case of detecting seedling (*T1*) stands. These results demonstrate that, besides of the well demonstrated capacity of ALS for quantifying forest stocks, there is a great deal of potential for predicting categorical variables in general, and forest types in particular. Overall RF classification was probably the best approach in terms of a trade-off between obtaining an accurate classification unbiased across all the DCs.

Acknowledgements

This research was funded by SMK. Special thanks to Juho Heikkilä and Jussi Lappalainen (SMK), Heli Laaksonen (NLS), and Aki Suvanto (Blom Kartta Oy) for their support at different stages of this study, including the delivery of data for this research. We are grateful for anonymous reviewers' contribution to improve the quality of the final manuscript.

References

Atkinson, P.M., Tatnall, A.R.L. 1997. Introduction: Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), 699-709. http://dx.doi.org/10.1080/014311697218700

Blackard, J.A., Dean, D.J. 1999. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24(3), 131-151. http://dx.doi.org/10.1016/s0168-1699(99)00046-0

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. http://dx.doi.org/10.1177/001316446002000104

Dalponte, M. Bruzzone, L., Gianelle, D. 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions in Geosciences and Remote Sensing,* 46(5), 1416-1427. http://dx.doi.org/10.1109/tgrs.2008.916480

Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E., Hudak, A.T. 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment*, 113(5), 946-956. http://dx.doi.org/10.1016/j.rse.2009.01.003

Falkowski, M.J., Hudak, A.T., Crookston, N.L., Gessler, P.E., Uebler, E.H., Smith, A.M.S. 2010. Landscape-scale parameterization of a tree-level forest growth model: A *k*-nearest neighbor imputation approach incorporating LiDAR data. *Canadian Journal of Forest Research*, 40(2), 184-199. http://dx.doi.org/10.1139/x09-183

Foody, G.M., Campbell, N.A., Trodd, N.M., Wood, T.F. 1992. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric Engineering & Remote Sensing*, 58(9), 1335-1341,

García, M., Riaño, D., Chuvieco, E., Salas, J., Danson, F.M. 2011. Multispectral and LIDAR Data Fusion for Fuel Type Mapping Using Support Vector Machine and Decision Rules. *Remote Sensing of Environment*, 115(6), 1369-1379. http://dx.doi.org/10.1016/j.rse.2011.01.017

García-Gutiérrez, J., Martínez-Álvarez, F., Riquelme, J.C. 2010. Using Remote Data Mining on LIDAR and Imagery Fusion Data to Develop Land Cover Maps. *Trends in Applied Intelligent Systems. Lecture Notes in Computer Science*, 6096, 378-387. http://dx.doi.org/10.1007/978-3-642-13022-9_38

Geerling, G.W., Labrador-Garcia, M., Clevers, J.G.P.W., Ragas, A.M.J., Smits, A.J.M. 2007. Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. *International Journal of Remote Sensing*, 28(19), 4263-4284. http://dx.doi.org/10.1080/01431160701241720

Hastie, T., Tibshirani, R., Friedman, J.H. 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer

Heikkinen, V., Korpela, I., Tokola, T., Honkavaara, E. Parkkinen, J. 2011. An SVM Classification of Tree Species Radiometric Signatures Based on the Leica ADS40 Sensor. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4539-4551. http://dx.doi.org/10.1109/tgrs.2011.2141143

Holmgren, J., Persson, Å., Södermann, U. 2008. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote* Sensing, 29(5), 1537-1552. http://dx.doi.org/10.1080/01431160701736471

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment,* 112(5), 2232-2245. http://dx.doi.org/10.1016/j.rse.2007.10.009

Kilkki, P., Päivinen, R. 1987. *Reference sample plots to combine field measurements and satellite data in forest inventories*. University of Helsinki, Department of Forest Mensuration and Management, Research Notes, 19, 209-215.

Kim, S., McGaughey, R.J., Andersen, H.E., Schreuder, G. 2009. Tree Species Differentiation Using Intensity Data Derived From Leaf-on and Leaf-Off Airborne Laser Scanner Data. *Remote Sensing of Environment*, 113(8), 1575-1586. http://dx.doi.org/10.1016/j.rse.2009.03.017

Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T., Hyyppä, J. 2010. Tree species classification in airborne LiDAR data: influence of stand and tree factors, intensity normalization and sensor type. *Silva Fennica*, 44(2), 319-339. http://dx.doi.org/10.14214/sf.156

Liaw, A., Wiener, M. 2002. Classification and Regression by randomForest. *R News*, 2(3), 18-22.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., Kangas, J. 2006. Non-parametric estimation of stem volume using laser scanning, aerial photography and stand register data. *Canadian Journal of Forest Research*, 36(2), 426-436. http://dx.doi.org/10.1139/x05-246

Maltamo, M., Kallio, E., Bollandsås, O.M., Næsset, E., Gobakken, T., Pesonen, A. 2014 Assessing dead wood by airborne laser scanning. In: Maltamo M., Næsset E., Vauhkonen J. (Eds.) *Forestry Applications of Airborne Laser Scanning. Concepts and Case Studies. Managing Forest Ecosystems,* 27. Dordrecht: Springer. pp. 375-395. http://dx.doi.org/10.1007/978-94-017-8663-8_19

Martinuzzi, S., Vierling, L.A. Gould, W.A. Falkowski, M.J. Evans, J.S. Hudak, A.T., Vierling, K.T. 2009. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, 113(12), 2533-2546. http://dx.doi.org/10.1016/j.rse.2009.07.002

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(1), 88-99. http://dx.doi.org/10.1016/s0034-4257(01)00290-5

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.

McInerney, D.O., Suarez-Minguez, J., Valbuena, R., Nieuwenhuis, M. 2010. Forest canopy height retrieval using LiDAR data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. *Forestry,* 83(2), 195-206. http://dx.doi.org/10.1093/forestry/cpq001

McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *Forest Ecology and Management*, 272, 3-12. http://dx.doi.org/10.1016/j.foreco.2011.06.039

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F. 2014a. *e1071: Misc Functions of the Department of Statistics, TU Wien*. R package version 1.6-4.

Meyer, D., Zeileis, A., Hornik, K. 2014b. *vcd: Visualizing Categorical Data*. R package version 1.3-2

Niska, H., Skön, J.P., Packalén, P., Tokola, T., Maltamo, M., Kolehmainen, M. 2010. Neural networks for the prediction of species specific plot volumes using airborne laser scanning and aerial photographs. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1076-1085. http://dx.doi.org/10.1109/tgrs.2009.2029864

NLS - National Land Survey of Finland 2013. *Laser scanning data*. Available at http://maanmittauslaitos.fi. (accessed on September, 2013).

Ørka, H.O., Næsset, E., Bollandsås, O.M. 2009. Classifying Species of Individual Trees by Intensity and Structure Features Derived From Airborne Laser Scanner Data. *Remote Sensing of Environment*, 113(6), 1163-1174. http://dx.doi.org/10.1016/j.rse.2009.02.002

Ørka, H.O., Gobakken, T., Næsset, E., Ene, L., Lien, V. 2012. Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Canadian Journal of Remote Sensing*, 38(2), 125-138. http://dx.doi.org/10.5589/m12-021

Reese, H., Nyström, M., Nordkvist, K., Olsson, H. 2014. Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation. *International Journal of Applied Earth Observation and Geoinformation*, 27, 81-90. http://dx.doi.org/10.1016/j.jag.2013.05.003

Ripley, B.D. 1996. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., Roberts, D. 2008. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sensing of Environment*, 112(5), 2272-2283. http://dx.doi.org/10.1016/j.rse.2007.10.004

Rousseeuw, P.J. 1986. Multivariate estimation with high breakdown point. In: Grossman, W., Pflug, G., Vincze, I., Wertz, W. (eds.) *Mathematical Statistics and Applications,* 283-297. Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-009-5438-0_20

Tomppo, E. 1993. National Forest Inventory of Finland. In: Nyyssönen, A. (ed.). *Proceedings of FAO/ECE meeting of experts on global forest resources assessment, Kotka, Finland*. The Finnish Forest Research Institute, Research Papers, 469, 204-205.

Torresan, C., Strunk, J., Zald, H.S.J, Zhiqiang, Y., Cohen, W.B. 2014. Comparing statistical techniques to classify the structure of mountain forest stands using CHM-derived metrics in Trento province (Italy). *European Journal of Remote Sensing*, 47, 75-94. http://dx.doi.org/10.5721/eujrs20144706

Valbuena, R. 2014. Integrating Airborne Laser Scanning with Data from Global Navigation Satellite Systems and Optical Sensors. In: Maltamo M., Næsset E. and Vauhkonen J. (Eds.) *Forestry Applications of Airborne Laser Scanning. Concepts and Case Studies. Managing Forest Ecosystems,* 27. Dordrecht: Springer. pp. 63-88. http://dx.doi.org/10.1007/978-94-017-8663-8_4

Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., Maltamo, M. 2014. Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves. *ISPRS Journal of Photogrammetry and Remote Sensing, 95*, 23-33. http://dx.doi.org/10.1016/j.isprsjprs.2014.06.002

Valbuena, R., Mehtätalo, L., Maltamo, M., Packalen, P. 2015. *A rule-based method using L-moments for direct detection of shade-intolerant regeneration in Boreal forests from airborne laser scanning*. Silvilaser 2015. La Grande Motte. France. pp. 47-49. Available at: http://www.isprs-geospatialweek2015.org/files/Proceedings_Silvilaser2015.pdf (accessed on January, 2016).

Valbuena, R., Maltamo, M., Packalen, P. 2016. Classification of multi-layered forest development classes from low-density national airborne lidar datasets. *Forestry*, 89. http://dx.doi.org/10.1093/forestry/cpw010

Van Aardt, J.A.N., Wynne, R.H., Scrivani, J.A. 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. *Photogrammetric Engineering and Remote Sensing*, 74(8), 1033-1044. http://dx.doi.org/10.14358/pers.74.8.1033

Vauhkonen, J., Korpela, I., Maltamo, M., Tokola, T. 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, 114(6), 1263-1276. http://dx.doi.org/10.1016/j.rse.2010.01.016

Venables, W.N., Ripley, B.D. 2002. *Modern Applied Statistics with S-Plus.* New York: Springer.

Warnes, G.R. 2013. *gmodels: Various R programming tools for model fitting.* R package version 2.15.4.1.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R. 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 28-37. http://dx.doi.org/10.1016/j.isprsjprs.2010.08.003

Zhang, Z., Liu, X., Peterson, J., Wendy, W. 2011. Cool temperate rainforest and adjacent forests classification using airborne LiDAR data. *Area*, 43(4), 438-448. http://dx.doi.org/10.1111/j.1475-4762.2011.01035.x