Evaluating the potential of Airborne Laser Scanning (ALS) remote sensing for developing indicators applied in sustainability impact assessments with ToSIA tool

Marc Caufapé Poca

Master Thesis

MSc in European Forestry – Erasmus Mundus

Evaluating the potential of Airborne Laser Scanning (ALS) remote sensing for developing indicators applied in sustainability impact assessments with ToSIA tool

Marc Caufapé Poca

Master Thesis

MSc in European Forestry – Erasmus Mundus

2013

Brief presentation of the host institution

EFICENT-OEF is a Regional Office of the European Forest Institute (EFI), an international organization established in 1993. The office was founded in Freiburg (Germany) in 2009, and it has two different geographical scopes. It focuses its research activities on the different forestry sectors in Central European countries, but is also open to cooperation with other countries outside this area.

An integral component of EFICENT is the Observatory of European Forests (OEF), which is located in Nancy (France). OEF, in contrast to other Regional Offices, has a pan-European perspective and supports EFI by establishing a policy-oriented forest information platform, both at the national and the international levels. OEF aims to promote research collaboration between institutions, and to foster the linkages between information and policy decision-making processes. The institution also caters for courses/workshops for researchers and information experts to stimulate expertise exchange on current and emerging topics. All these goals are pursued through three different research programs:

- Microeconomics: Forest investment analysis
- Macroeconomics: Forest sector modeling, including Criteria & Indicators for Sustainable Forest Management (SFM)
- Observatory: Communication and participatory research with civil society

While mostly focused on forest economics, all programs also take into consideration the social and ecological aspects of SFM.

In order to successfully operate at the pan-European level, OEF maintains a close interaction with various stakeholders from European and international organizations (e.g. FAO, UNECE, FOREST EUROPE and the European Commission), being involved in ongoing policy processes in forestry and adjacent environmental and socio-economic sectors. OEF also cooperates with several Universities and research institutions in Europe, and its core funding is provided by four different donors: the French Ministry of Agriculture and Fisheries, the Regional Council of Lorraine, the Urban Community of Greater Nancy, and the *Institut National de la Recherche Agronomique* (INRA).

SWOT analysis

One of the main strengths of my hosting institution is its broad presence all over Europe, and specifically for OEF, is its very close cooperation with local institutions (e.g. INRA, AgroParisTech University). This improves the efficiency of the institution when making use of its budget. One of the biggest opportunities nowadays is the establishment of new offices in non-European countries (i.e. Brazil, China and USA), and the possibility to start new collaboration projects. On the other hand, one of the major weaknesses is the inability to work in longer term funded research projects, since forestry projects can usually take up to 30 years. Finally, one of main threats is the economical crisis that is affecting the entire Euro zone, compromising long-term research projects and their funding.

Abstract

The concept of sustainability has evolved from a narrow focus on sustainable wood production to a much broader definition that includes many multi-dimensional aspects, such as environmental, social and economic values. The development of criteria and indicators together with new sustainability impact assessment tools has been an active process during the last decades. ToSIA is an entirely data-driven tool that evaluates these impacts based on indicators, processes and material flows of value chains. Some of these indicators can be estimated with Airborne Laser Scanning (ALS) systems, which are rapidly spreading as a predominant technique for forest resource assessments. A three-staged procedure in which sample plots were used to develop empirical relationships between various metrics computed from laser data, and tree characteristics measured in the field has been applied. Different linear models have been selected to predict several forest variables in different forest stands within a study area. Four different attributes (Total volumes of log and pulp, Gini coefficient, & Probability of existence of Coarse Woody debris) were finally selected in a classification process into three different categories: forest areas for timber extraction, forest areas with a relatively high biodiversity value, and forest areas with a high biodiversity value but also potentially attractive for their recreational use. Gini coefficient was the main clustering criterion between the classes that accurately represented the different processes originally included in the ToSIA value chain, finally linking both applications.

Context and state of the art

Sustainability has been recognized as a target of good management in forestry for centuries (Carlowitz 1713). However, the concept of sustainable forest management has evolved from a narrow focus on sustainable wood production to a much broader definition that includes many multi-dimensional aspects: environmental, social and economic (MCPFE 2002, Lindner et al. 2010, Rosén et al. 2012). The development of consistent criteria and indicators for sustainable forest management has been an active process during the last decades. This development has also been linked with the creation of different sustainability impact assessment methods, such as ToSIA software (Tscherning et al. 2008). ToSIA compares quantified impacts on sustainability of Forest Wood Chains (FWC) between a baseline and one or more scenarios. FWCs are defined as chains of production processes (e.g. harvesting-transport-industrial processing), which are linked with products (e.g. volume of saw logs). Sustainability is determined by analyzing the different indicators for all the production processes along the same FWC. Then, sustainability values are calculated as products of the relative indicator values (i.e. indicator value expressed per unit of material flow) multiplied by the material flow entering each process (Palosuo 2010). Calculated sustainability values are then aggregated for the complete FWC, or just only for segments of the same (Lindner et al. 2010). ToSIA is a flexible and entirely data-driven tool, with the possibility to include additional indicators (Lindner 2011). New indicators need to be defined following a certain format, as stated in a “Data Collection Protocol” (Berg, 2008) that can be developed in collaboration with a wide range of information sources (Lindner et al. 2010).

Airborne Laser Scanning (ALS) sensors are rapidly spreading as a predominant remote sensing technique for forest resource assessment. Relationships between different metrics that may be derived from ALS data and tree and canopy characteristics have been widely studied (Hudak et al. 2009, Næsset 2004). Næsset (2002) proposed a practical two-stage procedure in which sample plots were used to provide individual stand estimates of tree variables from laser scanner data. Georeferenced sample plots could, in the first stage, be used to develop empirical relationships between various metrics derived from laser data and tree characteristics measured in the field. Such relationships can be later used to provide the corresponding estimates for each one of the forest stands covered by the laser data; therefore, carrying out forest inventories at larger scales (Næsset, Bjerknes 2011). Furthermore, ALS sensors also present certain limitations when applied in the assessment of forest resources (e.g. species differentiation). However, different authors have demonstrated that by combining ALS data with other sources of remote sensing and geographical information is possible to overcome some of them (Andersen et al. 2011, McInerney et al. 2010, Packalen et al. 2009, Valbuena et al. 2013). For this reason, there is a good potential for developing harmonized protocols by means of ALS in the estimation of a good number of sustainability indicators, especially those related with the environmental conditions (Wulder et al. 2008).

The main objective of this study was the implementation of ALS sensors in the definition of new protocols required in the estimation of several environmental sustainability indicators, all of them included in the “Data Collection Protocol” of ToSIA tool. All indicator values will be finally used in the delineation of different management units, which will be later used as input data for ToSIA.

Materials and methods

<u>Materials</u>

Study area

An area of approximately 800ha in the municipality of Kiihtelysvaara (62°31'N 30°10'E) in the province of North Karelia (Finland) was selected. Only the forest and other wooded land was analyzed, following the definition of The Food and Agriculture Organization (FAO 2011). However, the study area was also partly covered by other land uses such as open grasslands and water bodies. Commercial active management is mainly pursued in the area since around 70% belongs to UPM-kymmene (Finland), one of the biggest forest companies in the world (Appendix I). The rest of the forest area belongs to small private forest owners. The main tree species were Norway spruce (*Picea abies* (L. Karst.)) and Scots pine (*Pinus sylvestris* (L.)), with a minor proportion of broadleaved species (mainly *Betula ssp*.).

Sample plot inventory

The ground-truth dataset acquired in this study consisted on 79 sample plots, stratified within the different forest stand types within the study area, and plot positions were subjectively determined to assure full range coverage for all the forest variables considered. 67 of the sample plots were located into forest land owned by UPM (*U*), and the remaining 12 in areas owned by small private owners (*P*) (Appendix I). Plots areas were between 400-900m^2, depending on the density of trees in each one of them. Plot size was therefore limited by assuring the time required for plot mensuration to extend no longer than one working day. All seedling and sapling stands were ignored in the distribution.

Field measurements were carried out between May and June 2010, and in each one of the sample plots all trees with a diameter at breast height over 5cm, or a height over 4m were callipered, measured and mapped. The taxonomical species of each one of the selected individuals was then identified and recorded with a numerical code, where 1=*Pinus sylvestris*; 2=*Picea abies*; and 3=Deciduous ssp. (mainly *Betula pendula*). For every tree *i* within plot, diameter at breast height (dbh_i; cm) was calculated as the average of two different perpendicular diameter measures, and tree top height (h_i; m) was measured with a Vertex hypsometer. Basal area-weighted proportions of each species were computed, and dominant species were defined as those whose presence was higher than 60%. Within the sample plot inventory, the diameter at 6 m height was also measured in order to obtain more accurate estimates for Total volume (*Tvol*), Volume of logs (*Tvol_log*), and Volume of pulp (*Tvol_pulp*) for each individual tree (Laasasenaho 1982). For pine and spruce, the minimum diameter and length for the logs were established in 16cm and 4m respectively, whereas for deciduous trees, these values were 18cm and 3.1m. For all species, the maximum length of the logs was 6.1m, which corresponds to the maximum size that sawmills can process in Finland.

Airborne Laser Scanning data

Laser data was acquired by an aircraft carrying an ALTM Gemini 125kHz multi-pulse mode laser scanning system produced by Optech, Canada. The data was acquired on the 26 June 2009, and the plane was flown approximately 720m above the ground. The scanning angle was 26º with an approximate overlap of 55%, so that every location in the study area was covered by at least two flight lines. The swath width was approximately 320 m, and the laser pulse density was estimated at 11.9 measurements/m^2. This configuration was used to maximize the probability for trees to receive ALS hits in both sides, reducing the occurrence of shadow areas in which tree foliage obstructs the path of the laser from reaching behind them.

ToSIA - Forest Wood Chain (FWC)

ALS systems are potentially able to provide a wide range of indicators related with the assessment of forest resources. That is the reason why the FWC in this study was almost entirely built with processes that belong to the “Forest resources management” module of the program (Lindner et al. 2010) (Figure 2).

The first process (ALS Scanning of the study area) is related with the entire study area and the ALS point cloud that covers it. This process has two different output products; each one was measured in area units depending on a differentiation made between Forest and Other Wooded Land area and the rest of land uses (Food and Agriculture Organization-FAO 2011). Since the main objective of this study was the analysis of forest ecosystems, the value chain was only developed for the product 10003000. The first of the created branches corresponded to all those areas with a present economic interest in this hypothetical study. Products 10003009, 10002996 & 10002989 were measured in terms of volume of wood resources (m^3/ha), including information like *Tvol*; *Tvol_log* & *Tvol_pulp* as indicators of each production process (Table 2). The second half of the FWC enclosed all areas with a natural value relatively higher, and/or with a certain social value. All products within this part of the value chain are area-related. The distinction between the areas within different processes was made according to the values for the predicted variables in this study (Appendix IV), and the aggregation into these products is specific for this case study. Products of the same ID, regardless if input or output, always have the same characteristics.

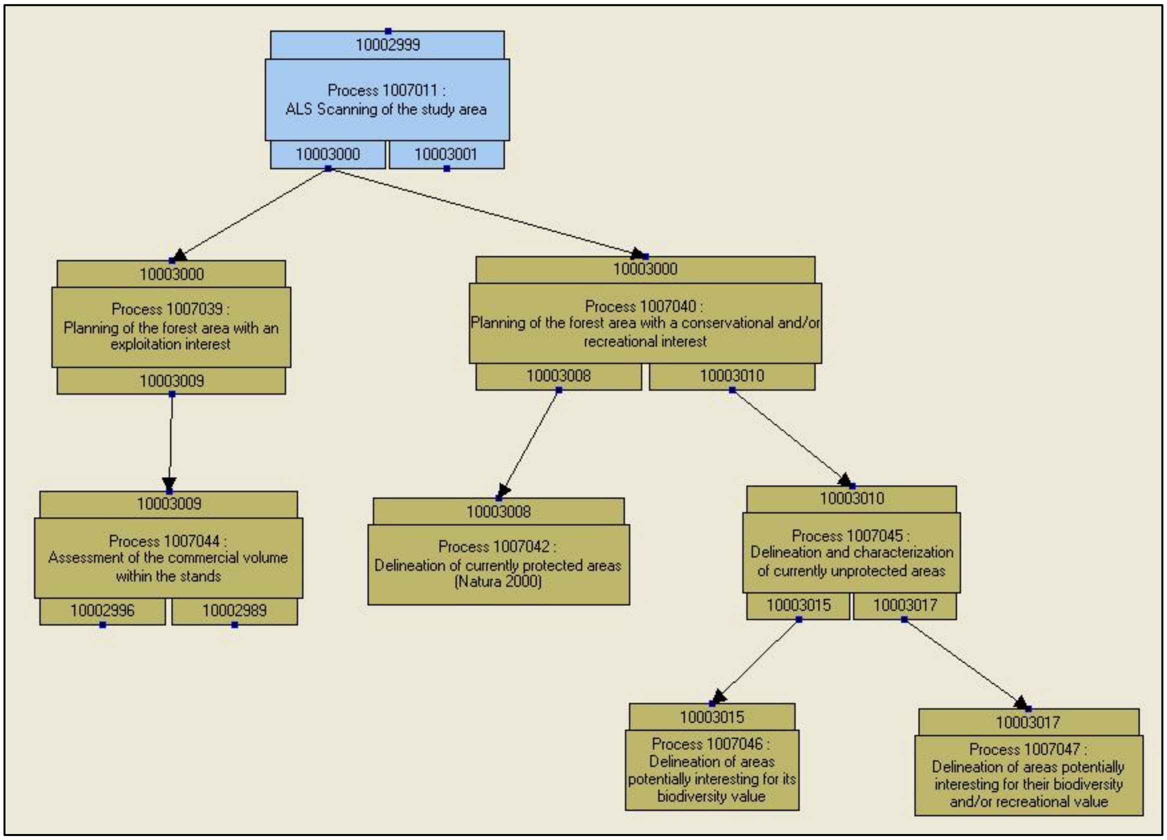


Figure 2. Forest Wood Chain created in ToSIA. Rectangles correspond to processes and their respective inputs and outputs are represented by numerical codes; arrows indicate the connexions between the different processes.

General methodology

The main methodology applied in this study is based on the practical two-stage procedure suggested by Næsset (2002), but with the inclusion of an additional third step (Figure 3).

Step I. Laser data processing

A Digital Terrain Model (DTM) was firstly interpolated using a Delaunay triangulation (Axelsson 2000), and all point elevations from the ALS data were scaled to above ground heights. Then, a large number of descriptive statistics was derived both at the plot and study area levels, being the same in both cases. For instance: (1) quantiles corresponding to 0, 10,..., 90 percentiles of the distribution; (2) maximum and minimum; (3) mean and mode; (4) standard deviation and coefficient of variation; etc. Metrics derived for the entire study area were stored in different grid (raster) files, and were later used in the implementation of the linear models. Grid cell size was established in 20m (400m^2), assuring a scale comparable to the size of the majority of the field sample plots. This first step was carried out using Fusion (McGaughey 2012).

Step II. Regression analysis

Suitable relationships were established between the different ALS descriptive statistics derived at the plot level and all the measurements obtained during the field inventory. The former acted as predictors (**X**), whereas the later were used as dependent variables (**Y**).

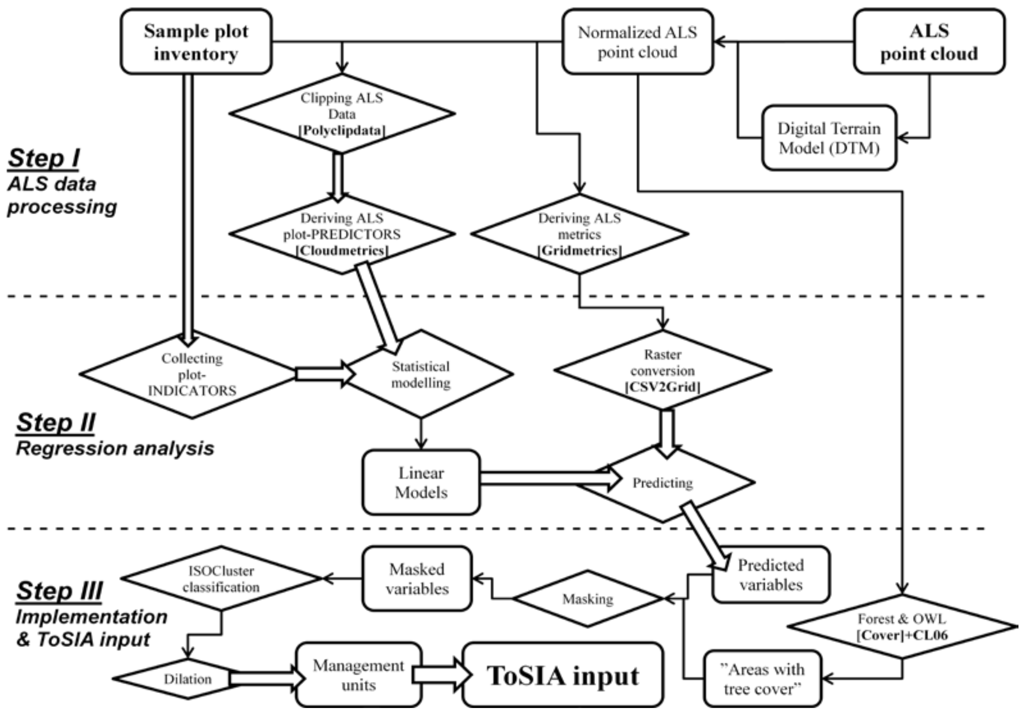


Figure 3. Flowchart of the procedure followed to obtain the polygon shape files used to calculate the different indicator values for the ToSIA FWC. Rectangles represent datasets and their subsets; processing steps are represented by rhombi; and the bold names between brackets correspond to the different FUSION commands used in each processing step.

Some of the dependent variables were directly measured in each sample plot, and then expanded to the hectare level: 1) Total Volume/ha (m^3/ha) (*Tvol*); 2) Total Volume of logs/ha (m^3/ha) (*Tvol_log*); and 3) Total Volume of pulp/ha (m^3/ha) (*Tvol_pulp*). On the other hand, some other parameters were derived from the previous ones:

- Gini coefficient *(GC):* measure of tree size inequality (Gini 1912, Weiner and Solbrig 1984). It quantifies the deviation from perfect equality, with a minimum value of zero when all trees are of equal size (volume in this study), increasing for more uneven-sized stands up to a maximum value of 1 (equation 1). Lexerød and Eid (2006) and Valbuena et al. (2012) demonstrated that *GC* is superior to all other indices when evaluating tree size equitability of forest stands. This coefficient presents the highest discriminator ability, capability of providing logical ranking of different distributions, and sensitivity to variation in sample size. In this study, it has been used as a measure of the structural complexity between the different forest stands.

(1)

Where,

Tvol – Total volume for each *i* and *j* tree, in m^3
n – Total number of trees inside each plot

- <u>Total aboveground biomass *(TAB)* & Total belowground biomass *(TBG)*</u>*:* different species-specific allometric equations were selected from the review made by Zianis et al. (2005). All selected equations have already been statistically tested in Finnish forest ecosystems (equations 2-7).

$$AB_{P.sylvestris} = 18.779 - 4.328 * dbh_i + 0.506 * dbh_i^2 \quad (2)$$

$$AB_{P.abies} = 19.018 - 4.806 * dbh_i + 0.565 * dbh_i^2 \quad (3)$$

$$AB_{B.pendula} = 0.00087 * dbh_i^{2.28639} \quad (4)$$

$$\mathrm{Log}(BG_{P.sylvestris}) = -1.89 + 2.74 * \log(dbh_i) \quad (5)$$

$$\mathrm{Log}(BG_{P.\ abies}) = -2.0274 + \log[h_i * (dbh_i^2)] * 0.8946 \quad (6)$$

$$BG_{B.pendula} = -3.887 + 1.3668 * \log(dbh_i^2 * h_i) \quad (7)$$

Where,

AB - total aboveground biomass of an individual tree, in kg
BG - total belowground biomass of an individual tree, in kg
dbh_i - diameter of the tree *i*, in cm
h_i – height of the tree *i*, in m

With these equations we calculated the above and belowground biomass for each single tree, so these values were later expanded at the hectare level, resulting in the indicators *TAB* and *TBG*, respectively.

In the end, eight different indicators were obtained in each sample plot: (1) Ownership type; (2) Dominant species; (3) Total volume/ha (m^3/ha); (4) Total volume of logs/ha (m^3/ha); (5) Total volume of pulp/ha (m^3/ha); (6) Gini coefficient (0-1); (7) Total aboveground biomass/ha (kg/ha); and (8) Total belowground biomass (kg/ha) (Table 1). Second indicator in the table (Dominant species) was calculated following an alternative methodology, based on the utilization of airborne multispectral imagery (see Particular cases - Proportions of species mixture).

Once both the predictors (**X**) and the indicators (**Y**) were gathered, the most suitable regression models for the prediction of plot dependent variables 3 to 8 were fitted. Model selection was based on the application of the Akaike Information Criteria (AIC) method (Akaike 1973). We performed a forward stepwise selection procedure of the predictor variables, iteratively finding the combination with the lowest AIC value. Models´ performance was assessed by leave-one-out cross-validation, from which relative the Root Mean Squared Errors (RMSE%) were evaluated from the prediction. Logarithmic transformations were applied in both sides of the equalities whenever heterocedasticity processes were detected (Table 2). The last two models in this table were not obtained following the above-mentioned procedure (see Particular cases - Probability of existence of standing/lying coarse woody debris). All procedures were carried out using R version 2.15.2 (The R Foundation for Statistical Computing 2012)

Step III. Implementation of the models & ToSIA input

Once all models were optimally adjusted they were extrapolated into the whole study area, and implemented for each one of the grid cells in which this was divided. The output of this process was imported into ArcMap10 (ESRI 2010) as five different raster (ascii) layers, one for each dependent variable. These raster layers were then masked according to the definition made by Food and Agriculture Organization-FAO (2011) for Forest and Other Wooded Land area, leaving out all other land uses (see Particular cases – Delineation of Forest and Other Wooded Land area).

Six out of the eight predicted variables or indicators were implemented into the FWC previously designed: Total Volume/ha (m^3/ha), Total Volume of logs/ha (m^3/ha), Total Volume of pulp/ha (m^3/ha), Gini coefficient (0-1), Probability of existence of downed coarse woody debris (1-7) and Probability of existence of standing coarse woody debris (1-7) (Table 2). The selection was made taking into account the created chain in itself, where a first differentiation was made between the economically interesting forest areas, and those with a higher natural relative and/or social value (Figure 2). The differentiation between the two branches of the FWC was based on a thresholding process of the predicted values for the selected indicators. An isocluster algorithm was applied, and the dataset was apportioned into homogeneous clusters according to the statistical distribution of the mentioned variables. After range-normalizing the variables (0-100), eight clusters were generated. The resulting class signatures were used as the input for an unsupervised classification method based on a Maximum Likelihood algorithm (Appendix III). The result of the classification process was then filtered with a Majority algorithm in order to render homogeneous class delineation by removing the high number of isolated pixels that appeared within the different classified areas. The eight different classes were then grouped into two new ones, which corresponded to the two different branches characterizing the FWC (Appendix IV). The discrimination between the two main groups presumed within the FWC was based on the value of the *GC*, *Tvol* & *Tpulp*. The group with an economical interest was then further developed by deriving from *Tvol* the commercial volume of timber per hectare, and by later differentiating it between *Tvol_log* and *Tvol_*pulp (Appendix V). On the other hand, the group enclosing all areas with a relatively high natural interest was developed by separating those areas currently under the protection of the Natura2000 scheme, from the ones without any protection. The protected area was directly delineated with the assistance of an additional source of geographical information (see Particular cases – Natura2000 protected areas), and then introduced into the FWC, without going through any previous classification process. Finally, the distinction between areas with a relatively high biodiversity value and those with not only a high natural value but also with some social interest, was based on the predicted values for *CDW* & *GC* (Appendix VI).

Eventually, the link between ALS-predicted sustainability indicators and ToSIA was accomplished with the creation and quantitative characterization of different polygon layers in ArcMap10 (ESRI 2010), each one representing a different process within the ToSIA FWC. This final step was not performed in the present study, but it has already been successfully tested in a previous Bachelor thesis (Korvenranta 2011).

Particular cases (modifications of the general methodology)

Due to several reasons, some of the input information and indicators for the ToSIA FWC were obtained without following the general methodology previously presented:

1. *Delineation of Forest and Other Wooded Land area*

As previously stated, only the forest land as defined by Food and Agriculture Organization-FAO (2011) was considered in the analysis. This definition encloses all land areas spanning more than 0.5ha, with trees higher than 5m, a canopy cover of at least 10%, and/or with trees able to reach these thresholds *in situ*. The first three criteria were directly applied to the ALS data: all laser hits with an aboveground height over 5m were filtered, and then their canopy cover was calculated with a threshold value of 10%. Once those areas were delineated in ArcMap10 (ESRI 2010), only those expanding more than 0.5ha were selected.

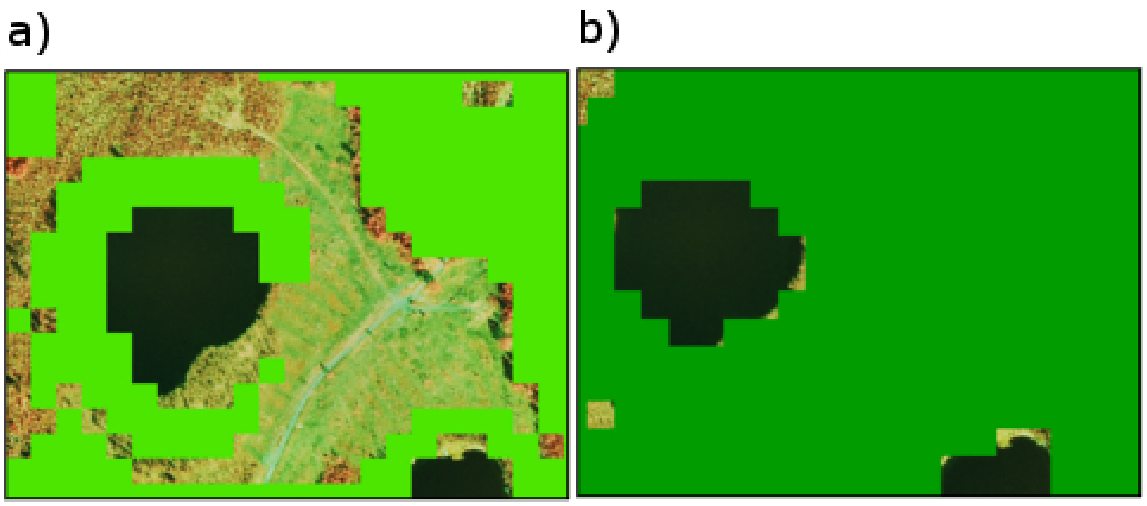


Figure 4. Specific zone within the study area that presented spatial differences between the areas delineated with ALS data alone (a), and the one derived from the CLC land use cover (b)

However, with the utilization of ALS data alone, it was not possible to consider all those areas that did not present the threshold values when the ALS flight was carried out, but that would potentially be able to reach them in the future (i.e. clear cut areas). In order to fulfill this last requirement in the definition, we decided to use an additional source of geographical information. One raster layer, corresponding to the CORINE Land Cover (CLC) classification from the year 2006, was obtained from the Finnish Environmental Administration (OIVA 2013). Only CLC categories including a tree cover over 10% and trees higher than 5m were selected, and the spatial difference between the area firstly delineated from the ALS data alone, and the one resulting from the reclassification of the CLC raster layer, was computed. Considering that the CLC layer used in this study was created three years before the acquisition of the ALS data (in 2009), we assumed that the difference corresponded to forest areas harvested during the period from 2006 to 2009. Eventually, all these differing areas were added to the area directly delineated with ALS data (Figure 4).

2. *Natura2000 protected areas*

One raster layer including all Natura2000 protected areas in Finland was obtained from the Finnish Environmental Administration (OIVA 2013). After a visual checking, we identified that one part of the protected site called Jukavaara-Särkilamminvaara fell inside our study region. This area is currently protected for its natural and landscape value at the national level, since it contains some endangered habitat types included in the Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora, so-called Habitats Directive. Some the endangered habitats that currently exist in this region are dystrophic ponds and lakes, ridges of forested habitats, and also wooded swamps. As stated before, this piece of land currently protected by the Natura2000 network, was directly subtracted and included in the ToSIA FWC, within the part of the chain developed from process 1007040.

3. *Probability of existence of standing/lying coarse woody debris (1-7)*

The usefulness of ALS data for inventorying Coarse Woody Debris (CWD) is based on the reflections detected from the gaps within partially closed canopy surfaces. However, CWD dynamics in managed forest areas is quite different as compared to unmanaged ones. In the latter, canopy gaps are only originated by natural mortality processes of the individuals, so a direct relationship can be established with the existence of deadwood material inside these gaps. On the other hand, in managed forest areas canopy gaps are not only the result of these kind of processes, but they can also be the result of silvicultural operations (e.g. thinning) (Packalén and Maltamo 2008). However, Pesonen et al. (2010) pointed out that after harvesting logging residues that form CWD may also be left in the stands, and thereby, the variation in canopy structure in managed forest areas may also indicate the existence of deadwood material. Following the advice made by Pesonen et al. (2009), who defined as challenging the construction of biologically functioning prediction models for the direct estimation of CWD volumes in commercial forest areas, we selected two different logistic models for the prediction of the probability of existence of lying and standing coarse woody debris in each grid cell, respectively (equations 8-9).

Due to a lack of field data, these two models were directly implemented into the whole study area without going through any regression analysis procedure.

$$\mathrm{DDW} = -1.2076 + 0.1249*(f\_h_{30}) \qquad (8)$$

Where: DDW - probability of existence of downed or lying coarse woody debris (1-14)

f_h_{30} - height at which the accumulation of first laser pulse heights is 30%

$$SDW = -4.0311 + 1.1525*(\overline{}) - 0.5433*(Ln(f\_p_0)) \quad (9)$$

Where: SDW - probability of existence of standing coarse woody debris (1-14)

f_h_{20} - height at which the accumulation of first laser pulse heights is 20%.

f_p_0 - proportion of first laser hits accumulating at the 0% height.

Finally, we calculated the probability of existence of Coarse Woody Debris (CWD) as the sum of the two probabilities in each grid cell. Therefore, the range of CWD was 1-14.

4. *Proportion of species mixture*

One of the main limitations when using ALS data alone in forest inventories is the differentiation between tree species. In order to overcome this limitation, we firstly took advantage of the species identification carried out during the field inventory for all measured trees. Secondly, we used another additional source of remote sensing information, this time one multispectral aerial image covering the entire study area.

Based on the species identification undertaken in the field, we calculated the species composition within each one of the sample plots, weighted by the basal area of each one of the individuals measured. The reason for weighting the species proportion by the basal area was based on the assumption that trees with a bigger diameter usually are characterized by bigger tree crowns, resulting in bigger reflectance responses when working with multispectral images. Once the mixture of species was calculated in each plot, a supervised classification procedure was implemented in the aerial image. This classification process was based on a maximum likelihood algorithm (Chuvieco 2002). During the training phase, five different land use classes were targeted: 1) Unforested; 2) Deciduous; 3) Pine; 4) Spruce and 5) Water bodies. Classes 1 and 5 were directly delineated based on a visual interpretation of the aerial image. On the other hand, for the other three classes (2 to 4) we used as ground-truth data all the values obtained for the different plot species composition. Then, the spectral signatures for the different categories were extracted, and these were later used in a soft classification process. Instead of classifying each pixel to one single class based on its highest probability value to belong to one specific category, we obtained the probabilities for each single pixel to belong to each individual category, independently. Only those raster files representing the probabilities of the grid cells to belong to classes 3, 2 and 4 were selected, in this order, to create a composite band RGB representation. The final band composition raster file represented the proportion of species mixture in each grid cell in which the study area was divided (Appendix VII).

Results

Table 1. Subset of the ten first sample plots with the values corresponding to the eight different indicators calculated for each one. Acronyms *Own*. and *Dom_sp* correspond to Ownership type and Dominant species, respectively. All the rest, correspond to the respective names given earlier to the different indicators.

Plot	Own.[a]	Dom_sp[b]	Tvol (m3/ha)	Tvol_log (m3/ha)	Tvol_pulp (m3/ha)	GC (0-1)	TAB (kg/ha)	TBG (kg/ha)
1	U	p	310.11	211.36	84.98	0.63	156143.23	50050.63
2	U	p	170.38	96.82	64.94	0.60	117846.20	42161.95
3	U	p	198.90	94.16	99.20	0.35	98638.66	34819.43
4	U	m	156.89	49.91	92.19	0.51	101647.95	22287.94
5	U	b	164.37	12.07	133.29	0.46	101787.44	13201.05
6	U	s	469.27	364.45	95.82	0.65	215107.88	45702.25
7	U	b	138.17	43.73	82.01	0.62	107745.78	21373.85
8	U	p	247.32	138.97	101.12	0.39	125560.50	45862.46
9	U	s	292.41	115.99	154.38	0.55	204007.12	36987.77
10	U	s	261.93	87.96	157.90	0.46	155193.18	26409.59

[a] U = UPM; [b] p = *Pinus sylvestris*; m = Mixed stands (no dominant sp.); b = *Betula sp*.; s = *Picea abies*

Table 2. Fitted models with the lowest AIC values for the different selected variables.

Dependent variable	Predictive model[a]	RMSE%
Tvol (m^3/ha)	-4.859 – 50.622**AAD* + 40.427**P70*	16.87
Log(Tvol_log) (m^3/ha)	-6.232 + 3.535**Max*	8.37
Log(Tvol_pulp) (m^3/ha)	9.278 – 2.523**Max* + 1.382**Mean*	6.88
GC (0-1)	-1.907 + 4.345**LCV* + 0.037**PRAb1-0.020*PRAbMd*	18.13
Log(TAB) (kg/ha)	9.849 – 0.894**Mean*	1.53
TBG (kg/ha)	-2986+2975.4**P70*	25.16
DDW (1-7)	-1.208+0.125**P30*	R2=0.22
SDW (1-7)	-4.0311 + 1.1525*(sqrt(F_H_{20})) – 0.5433*(Ln(F_P_0))	R2= 0.29

[a] *AAD* = Elevation Average Absolute Deviation; *P70* & *P30* = quantiles corresponding to the 70 & 30 percentiles of the pulse laser height distributions, respectively; *Max* = Maximum elevation; *Mean* = Mean elevation; *LCV* = L-moments Coefficient of Variation; *PRAb1* = Percentage of all returns above 1m aboveground height; *PRAbMd* = Percentage of all returns above the mode height; sqrt(F_H_{20}) = square root of the height at which the accumulation of first laser pulse heights is 20%; Ln(F_P_0) = natural logarithm of the proportion of first laser hits accumulating at the 0% height.

Eight different plot indicators were collected, either directly from the field measurements, or derived from the original data (Table 1). All plot indicators, except Ownership type (*Own.*) and Dominant species (*Dom_sp*), were regressed against the predictor variables derived from the distributions of the laser hits within each one of the plots.

The selection of the best linear models for the predictions was based on the lowest AIC values (Akaike 1973). Finally, logarithmic transformations were applied into the models when potential heterocedastic influences were detected in the predictions. The only models that did not require such transformations were the ones for the prediction of *Tvol_log*, *Tvol_pulp* and *TAB* (Table 2).

All selected models comprised a maximum of three predictor variables, which was settled as a threshold in order to avoid higher significant correlations between the different predictors. The lowest RMSE% values were obtained for *Tvol_log*, *Tvol_pulp* and *TAB* indicators, whereas they were much higher for *Tvol*, *GC* and *TBG*. The coefficient of determination (R^2) was found to be higher than 0.6 in almost all models, except for Total volume of pulp (0.40) and for Total belowground biomass (0.52) (Figure 5). The coefficient presented its highest values for *Tvol* and *Tvol_log*, 0.78 in both cases.

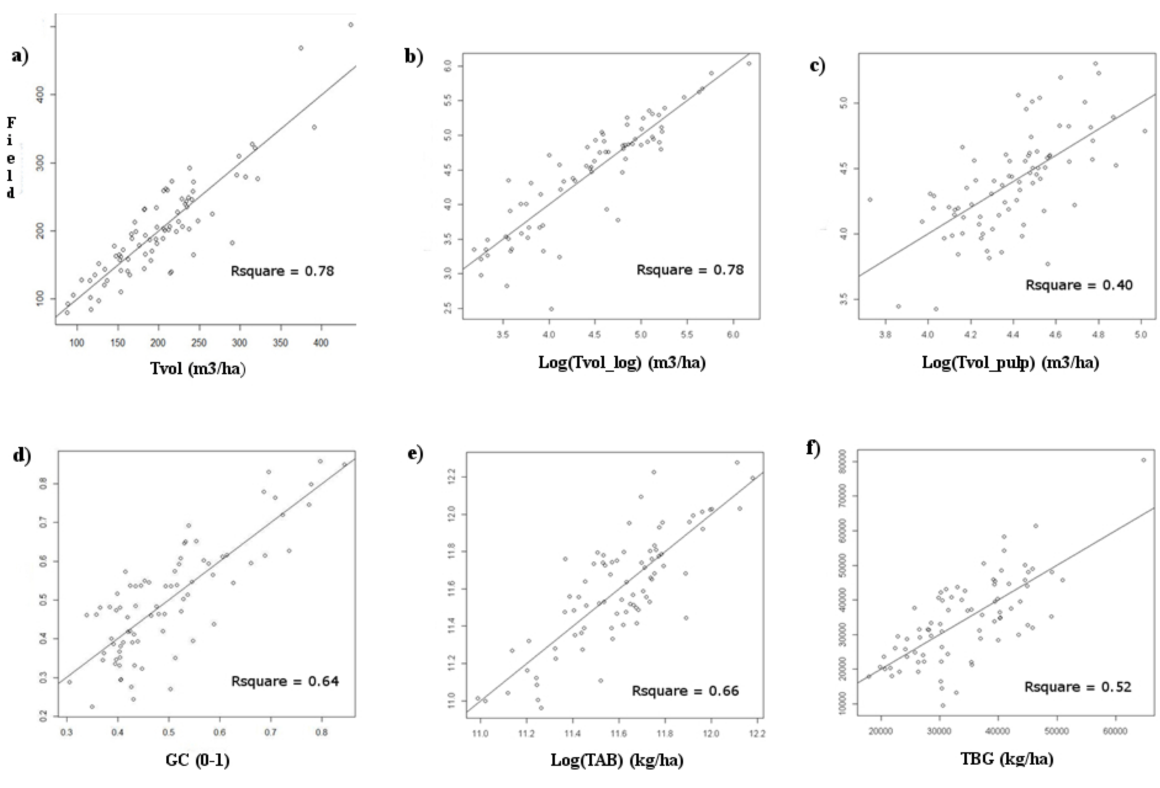


Figure 5. Scatter plots of the different predicted values for the dependent variables, versus the observed ones during the field sample inventory. a) Total volume (m^3/ha); b) Total volume of logs (m^3/ha); c) Total volume of pulp (m^3/ha); d) Gini coefficient (0-1); e) Total aboveground biomass (kg/ha); f) Total belowground biomass (kg/ha)

The next step was the extrapolation of the last six models throughout the entire study area, and the spatial representation of all predicted values for each dependent variable. The graphical representations for the distributions of the eight predicted dependent variables (Table 2) have been included in Appendix II.

Several clustering processes for *Tvol_log*, *Tvol_pulp*, *GC* & *CDW* were applied until we finally obtained the spectral signature of eight different classes, which were considered to accurately represent the different processes of the ToSIA FWC (Appendix III). The main clustering criterion was the *GC*, which divided the different areas into the two main groups presumed in the FWC (Figure 2): 1007039 where stands with low *GC* (<0.5) were selected for productive use; and 1007040 where the stands with high *GC* (>0.5) were selected for their biodiversity and/or recreational interest. However, an additional requirement for the inclusion of the forest areas into the first category was that the sum of *Tvol_log* & *Tvol_pulp* had to be at least 33% once both variables were normalized in their ranges. The method also succeeded in discriminating those areas more suitable foe saw-wood or pulp-wood production, achieving good class separability after disaggregating by *GC*. Figure 4 also illustrates how some of these classes would have been difficult to discriminate without using GC, as they overlap according to their *Tvol_log* and *Tvol_pulp*. On the other hand, the distinction within the areas with a relatively high natural interest was based on the values for *GC* and *CWD*. All areas with an interest only for biodiversity conservation presented a *GC* between 0.5-1 and CWD>7; whereas areas with an extra added social value were characterized by *GC* values between 0.5-0.75, and CWD<0.5.

The classification output is displayed in Appendix IV. Describing the different classes according to the order presented in the map legend: 1) In red, includes all the areas where the models did predict either negative values, or values really close to zero for all three predicted variables; 2) In yellow, corresponded to those areas where the sum of the predicted Total volume of logs (m^3/ha) and Total volume of pulp (m^3/ha) was lower than 33% of the variables normalized range. This category was also characterized by a *GC*<0.5 and a probability to find CWD lower than 7/14; 3) In pink, presented similar characteristics to the second one, except that *Tvol_pulp* was the most significant volume-related variable instead of *Tvol_log*; 4) In orange, included all those areas with an economically significant production of pulpwood, rather even-sized forest stands, and a high probability to find CWD (>0.5); 5) In brown, corresponded to all those areas with similar characteristics as the previous one, but with an economical interest due to their big amount of timber for saw logs production; 6) In light green, characterized by low values both for saw log and pulpwood yields, but with a Gini coefficient over 0.5 and a low probability to find CWD; 7&8) The darker greens, included all areas with a high production of either sawn logs or fibre, respectively, with a complex forest structure, and with a high probability of existence of deadwood material.

The result of the reclassification into two different classes is shown in Appendix V. Following the order for the classes previously specified, the aggregation into the two classes was carried out according to: 1) Other areas: which included categories 1, 2 & 3; 2) Harvesting: corresponding to the initial classes 4 & 5; 3) Biodiversity and/or recreation interest: categories 6, 7 & 8.

Finally, new categories 2 & 3 were further divided according to the different processes included in the FWC (Figure 2). For instance, category number 2 (Harvesting) was subdivided by differentiating initial categories 4 & 5, therefore, separating areas with a high timber yield for the production of saw logs from areas with a high content of pulpwood (Appendix VI). On the other hand, initial category 3 (Biodiversity and/or recreation interest) was subdivided into areas with or without recreational interest (Appendix VII).

Discussion & prospects for the future

Surprisingly, two of the lowest RMSE% values in Table 2 were obtained for the models predicting *Tvol_log* and *Tvol_pulp* metrics. In the sample plot inventory, not only the minimum tree diameter and minimum and maximum lengths for the logs criteria were applied to determine the two volumes for each individual tree. During the process, it was also preferred to optimize the economical benefit by subjectively varying the different lengths of the potential saw and pulp logs obtained from a certain tree. Since this subjective component will never be predictable with regression modelling processes, we found these two values rather unexpected. In contrast, the low value obtained for the R^2 in the model for *Tvol_pulp*, may indicate this phenomenon.

The study area analyzed by Pesonen et al. (2009) in the estimation of the probability of existence of coarse woody debris, presented highly similar conditions to the one analyzed in this study. The study area was located in a typically commercial forest managed area in the central part of Finland, and it was mostly owned by the same company, UPM. That is the reason why we decided to utilize the same two models (equations 8-9) in this study.

The combination of the four selected variables in the classification process (*Tvol_log*, *Tvol_pulp*, *GC* and *CWD*) was considered suitable to represent all the processes created in the ToSIA FWC. By establishing a limit of 33% out of the normalized ranges for the sum of the first two, we clearly indicated the economic interest of the different areas. Moreover, in this study we implemented the threshold at *GC*=0.5 defined as a border between even-sized and uneven-sized forests by Valbuena et al. (2012). The Gini coefficient has already been demonstrated to be a good indicator of the forest structure, which can be related with higher ecological values of forest stands. On the other hand, Edwards et al. (2010) characterized, for the Nordic European region, the public preferences to both the variation in tree size within tree stands and to the existence of harvesting residues. The perception of the first silvicultural attribute was defined with a bell-shaped curve, implying that very low or very high variations between trees were negatively perceived. Concerning the existence of deadwood material, the relation presented a negative value.

Despite some limitations (e.g. delineation of Forest and Other Wooded Land area), ALS sensors have demonstrated to have a great potential for estimating environmental sustainability indicators. For instance, even though only six out of the eight predicted forest variables were finally used in this study, the two biomass-related parameters could also be used to calculate new sustainability indicators (e.g. carbon stock).

This study has also demonstrated that is possible to overcome some of these limitations when combining ALS data with other sources of remote sensing information and/or geographical data. One clear example is the possibility of differentiating between species with the addition of multispectral aerial images into the analysis. This would allow for the prediction of new and more accurate indicator estimates, such as the segmentation of the value for the Total commercial volume into the different species (in this study corresponded to the sum of *Tvol_log*

& *Tvol_pulp*). This may be a highly interesting indicator in countries such as Finland, where commercial forestry and the forest sector in general, have an important role in the national economy. We are currently working on the improvement of the models by analyzing their sensitivity both species composition and ownership type. Due to time and space constraints in this manuscript, we decided not to include any of these results in the study. However, they will be an important part of the results and discussion in a future version of this report.

Finally, I would like to remark that there are numerous possibilities for the prediction of new indicators, representing each one a new single research challenge. For instance, once the vegetation cover type is characterized, it would be feasible to calculate the potential evapotranspiration of that specific land cover. Furthermore, Digital Terrain Models produced from ALS data are characterized for having a very high resolution. That can be really helpful when assisting decision-makers in avalanche or landslide risk management, since that would provide us with a better understanding of the dynamics behind these phenomena.

Acknowledgements

I would like to thank Dr. Ruben Valbuena and Dr. Diana Tuomasjukka, my two supervisors from the European Forest Institute-Headquarters office in Joensuu (Finland), for revising a preliminary version of this manuscript, and especially for supervising all my progress during my two internships at EFI. I also would like to thank Jean-Christophe Hervé from the *Institut National de l´Information Géographique et Forestière* in Nancy (France), for assisting me during my navigation through all the complicated statistical terms and concepts. My most sincere thanks to Dr. Petteri Packalén from the University of Eastern Finland, for providing the laser and inventory data and for answering all my questions.

Finally, I would also like to dedicate some words to all the staff at the Central European Regional Office of European Forest Institute - Observatory for European Forests (EFICENT-OEF), and especially to Dr. Patrice Harou for opening for me the doors of the office.

References

AKAIKE, H. 1973. *Information Theory and an Extension of the Maximum Likelihood Principle.Second International Simposium on Information Theory*. Akademiai Kiado, Budapest ed.

ANDERSEN, H. -E, et al. 2011. Using Multilevel Remote Sensing and Ground Data to Estimate Forest Biomass Resources in Remote Regions: A Case Study in the Boreal Forests of Interior Alaska. *Canadian Journal of Remote Sensing*, vol. 37, no. 6, pp. 596-611.

AXELSSON, P. 2000. DEM Generation from Laser Scanner Data using Adaptive TIN Models. *International Archives of Photogrammetry and Remote Sensing*, vol. 33, pp. 110-110-117.

BERG, S. 2008. *EFORWOOD Deliverable PD0.0.16: Manual for data collection for Regional and European cases*, Uppsala: Skogforsk.

CARLOWITZ, HCV. 1713. Sylvicultura oeconomica: Anweisung zur wilden Baum-Zucht, Leipzig: Braun.

CHUVIECO, Emilio. 2002. *Teledetección Ambiental :La Observación De La Tierra Desde El Espacio.* Madrid: Ariel. ISBN 8434480476.

EDWARDS, D., et al., 2010. *Public Preferences for Silvicultural Attributes of European Forests.* EFORWOOD project. Sustainability Impact Assessment of the Forestry-Wood Chain ed., Deliverable D2.3.3 ed. Scotland: Forest Research (FR), Northern Research Station.

ESRI. 2010. *ArcMap.* 10.0th ed. Redlands, California.

Food and Agriculture Organization - FAO. 2011. *State of World's Forests.* Rome, Italy.

GINI, C., 1912. Variabilita e mutabilita. Bologna.

HUDAK, A. T.; EVANS, J. S.and SMITH, A. M. S. 2009. LIDAR Utility for Natural Resource Managers. *Remote Sensing*, vol. 1, pp. 934-934-951.

KORVENRANTA, M. 2001. Metsähakkeen hyödyntämisen työllistävä vaikutus Kiteellä ToSIA-työkalulla ja paikkatietomenetelmillä arvioituna. *Itä-Suomen yliopisto; Historia- ja maantieteiden laitos*, 32.

LAASASENAHO, J. 1982. Taper Curve and Volume Functions for Pine, Spruce and Birch. *Communicationes Instituti Forestalis Fenniae*, vol. 108, no. 1, pp. 1-1-74.

LEXERØD, N. L. and Eid, T. 2006. An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, 222 (1), pp. 17-28.

LINDNER, M. et al. 2010. ToSIA - A Tool for Sustainability Impact Assessment of Forest-Wood-Chains. *Ecological Modelling*, 9/10, vol. 221, no. 18, pp. 2197-2205. ISSN 0304-3800.

MCGAUGHEY, R. J. 2012. *FUSION/LDV: Software for LIDAR Data Analysis and Visualization. Version 3.10.* U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, University of Washington. Available from:<http://forsys.cfr.washington.edu/fusion/fusionlatest.html>.

MCINERNEY, D. O., et al. 2010. Forest Canopy Height Retrieval using LiDAR Data, Medium-Resolution Satellite Imagery and kNN Estimation in Aberfoyle, Scotland. *Forestry*, vol. 83, no. 2, pp. 195-206.

MCPFE. 2012. Improved Pan-European Indicators of Sustainable Forest Management. In Ministerial Conference on the Protection of Forests in Europe, Vienna, Austria, pp. 6.

NÆSSET, E. 2004. Accuracy of Forest Inventory using Airborne Laser Scanning: Evaluating the First Nordic Full-Scale Operational Project. *Scandinavian Journal of Forest Research*, vol. 19, no. 12, pp. 554-554-557.

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), pp. 88-99.

NÆSSET, E.; and BJERKNES, K-O. 2011. Estimating Tree Heights and Number of Stems in Young Forest Stands using Airborne Laser Scanner Data. *Remote Sensing of Environment*, vol. 78, no. 3, pp. 328-328-340.

OIVA. 2013. *CORINE Land Cover 2006 (CLC2006).* . Ympäristö ja paikkatietopalvelu asiantuntijoille (Environment and Spatial Information Services) ed., Finnish Environmental Administration. 23/04/2013. Available from:<http://wwwp2.ymparisto.fi/scripts/oiva.asp>.

PACKALÉN, P.; and MALTAMO, M. 2008. Estimation of Species-Specific Diameter Distributions using Airborne Laser Scanning and Aerial Photographs. *Canadian Journal of Forest Research*, vol. 38, no. 7, pp. 1750-1760.

PACKALEN, P.; SUVANTO, A.and MALTAMO, M. 2009. A Two Stage Method to Estimate Species-Specific Growing Stock. *Photogrammetric Engineering and Remote Sensing*, vol. 75, no. 12, pp. 1451-1460.

PALOSUO, T., et al. 2010. Assigning results of the Tool for Sustainability Impact Assessment (ToSIA) to products of a forest-wood-chain. *Ecological Modelling*, 221 (22), pp. 15–25.

PESONEN, A., et al. 2009. Comparison of Field Sampling Methods for Assessing Coarse Woody Debris and use of Airborne Laser Scanning as Auxiliary Information. *Forest Ecology and Management*, vol. 257, no. 6, pp. 1532-1541.

PESONEN, A., MALTAMO, M. and KANGAS, A., 2010. The Comparison of Airborne Laser Scanning-Based Probability Layers as Auxiliary Information for Assessing Coarse Woody Debris. *International Journal of Remote Sensing*, vol. 31, no. 5, pp. 1245-1259.

ROSÉN, K., et al. 2012. Challenges in Implementing Sustainability Impact Assessment of Forest Wood Chains. *European Journal of Forest Research*, vol. 131, no. 1, pp. 1-5.

The R Foundation for Statistical Computing. 2012. *R Version 2.15.2.* , 26.10.2012.

TSCHERING, K., et al. 2008. Ex-ante Impact Assessments (IA) in the European Commission – an overview. In: *Sustainability impact assessment of land use changes* , Helming K., Pérez-Soba M., Tabbush P. (Ed.), pp. 17-33, Berlin: Springer.

VALBUENA, R., et al. 2012. Diversity and equitability ordering profiles applied to study forest structure. *Forest Ecology and Management* (1), pp. 185-195.

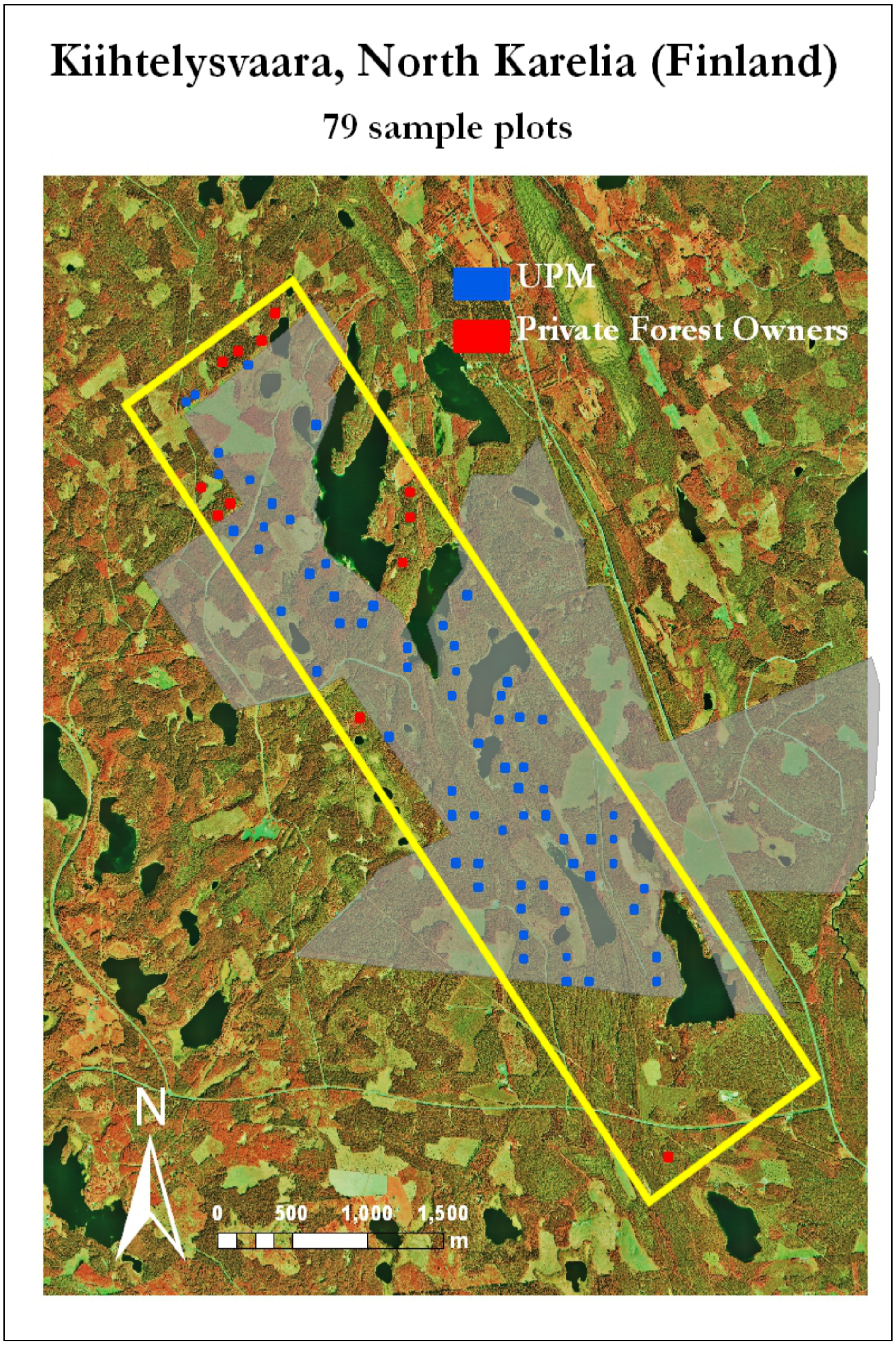
VALBUENA, R., et al. 2013. Within-species benefits of back-projecting laser scanner and multispectral sensors in monospecific Pinus Sylvestris forests. *European Journal of Remote Sensing*. 46, pp. 491-509.

WEINER, J.; and SOLBRIG, O. T. 1984. The Meaning and Measurement of Size Hierarchies in Plant Populations. *Oecologia*, vol. 61, no. 3, pp. 334-336.

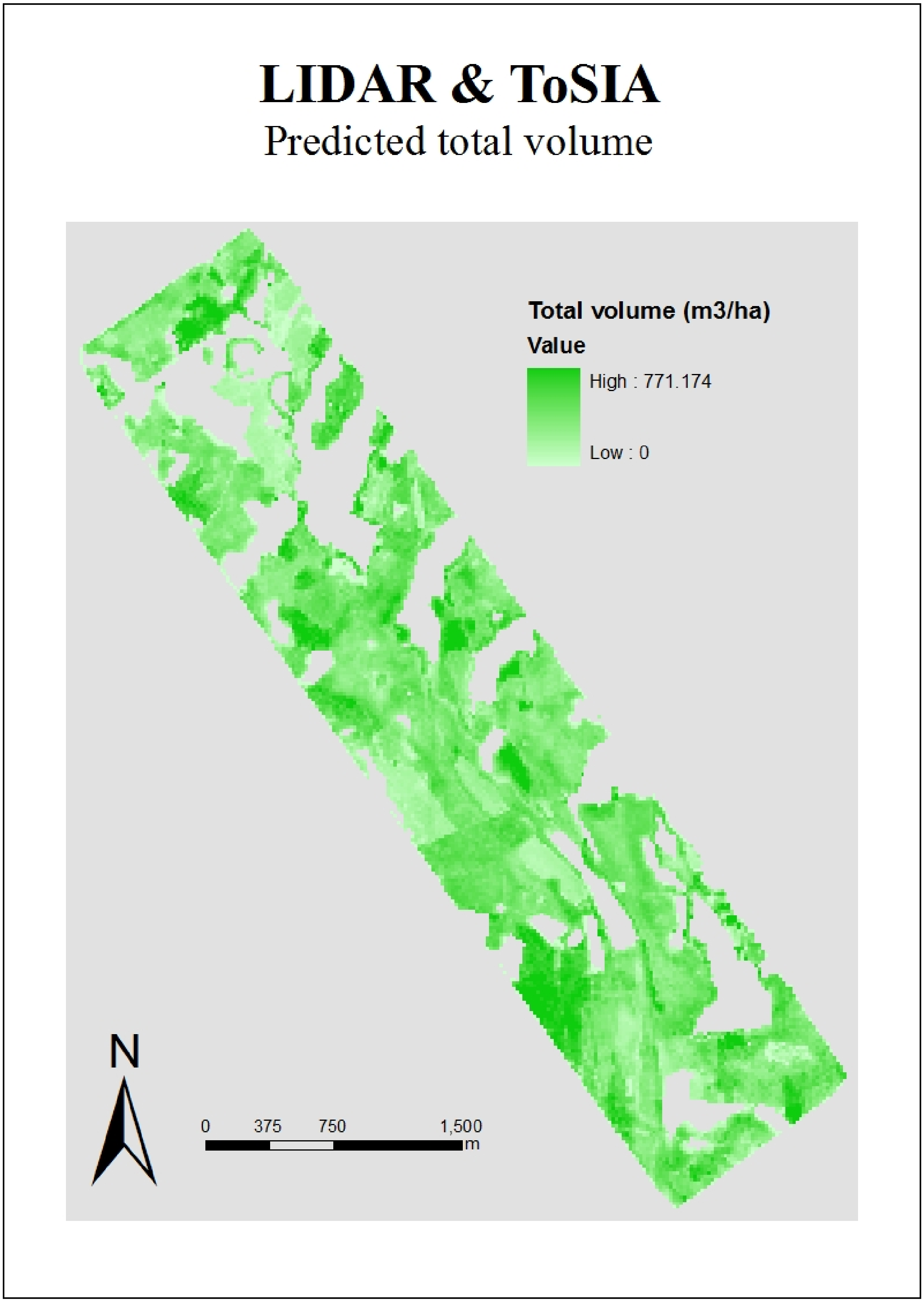
WULDER, M.A., et al. 2008. The role of LiDAR in sustainable forest management. *The Forestry Chronicle, 84 (13)*, pp. 807-826.

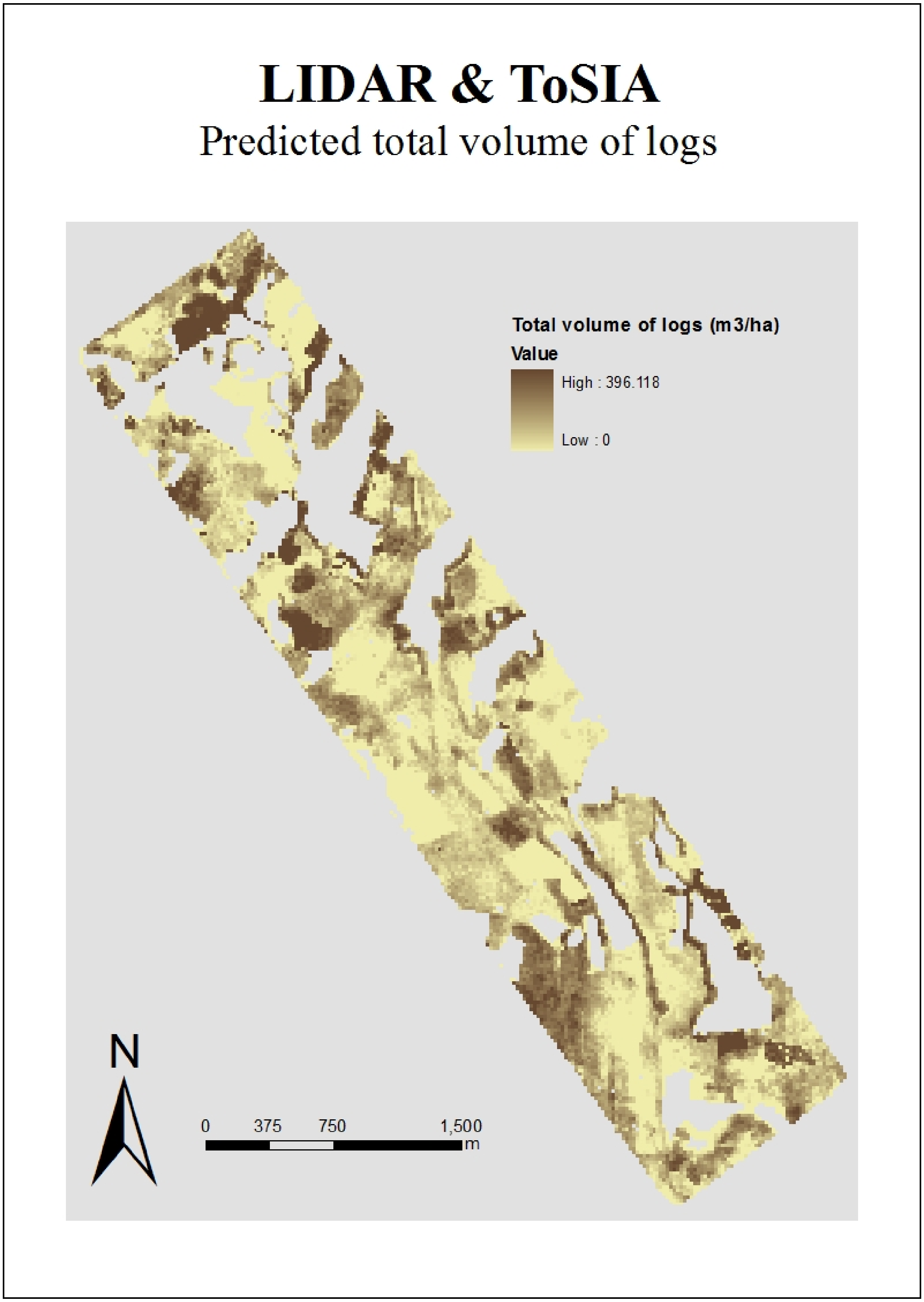
ZIANIS, D., et al. 2005. Biomass and Stem Volume Equations for Tree Species in Europe. *Silva Fennica*. Monographs 4.

Appendix I: Location of the 79 field sample plots within the study area (in yellow), and their differentiation according to ownership type. The area owned by UPM is shadowed in grey.



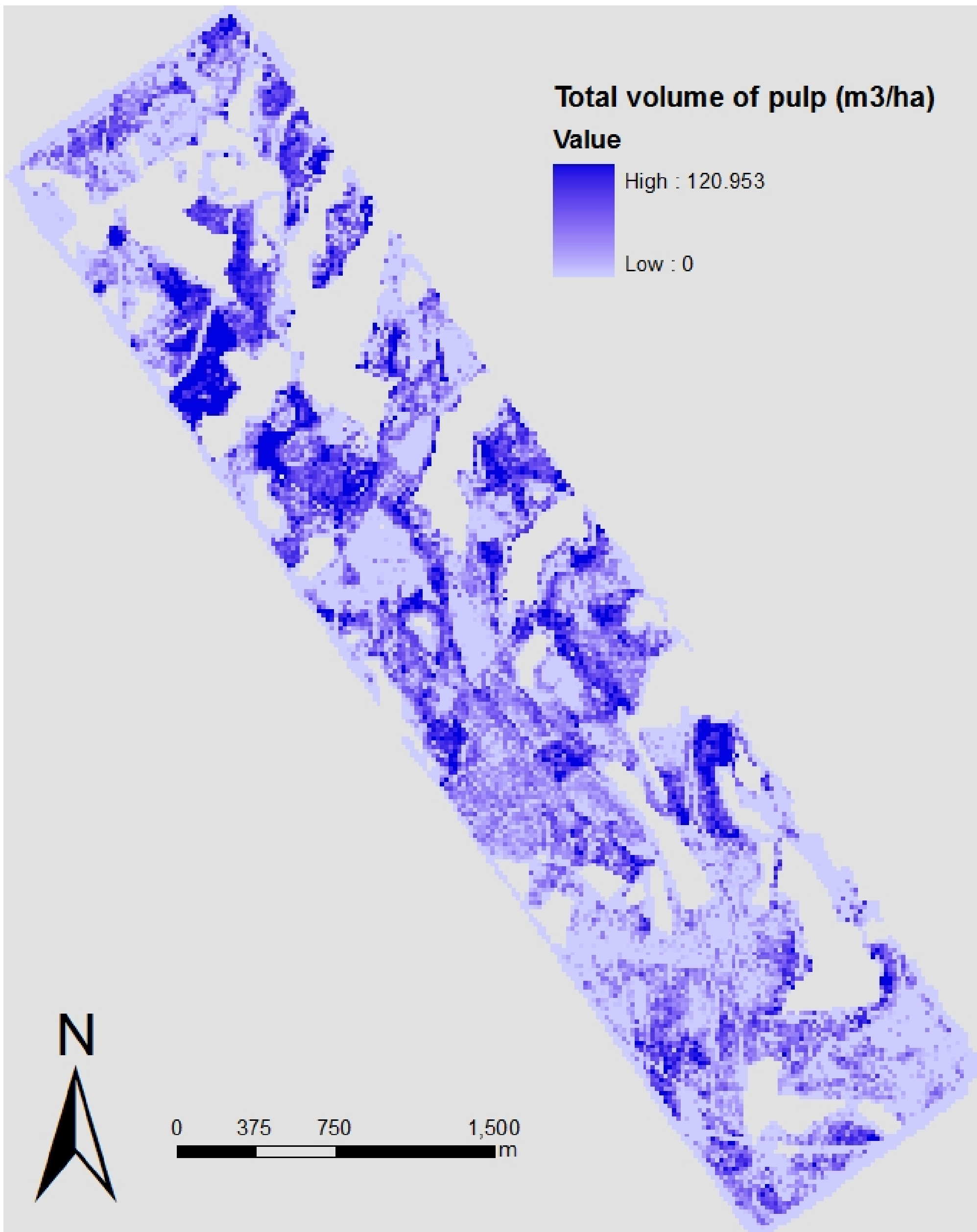
Appendix II


LIDAR & ToSIA
Predicted total volume
Total volume (m3/ha)
Value
High : 771.174
Low : 0
N
0
375
750
1,500
m

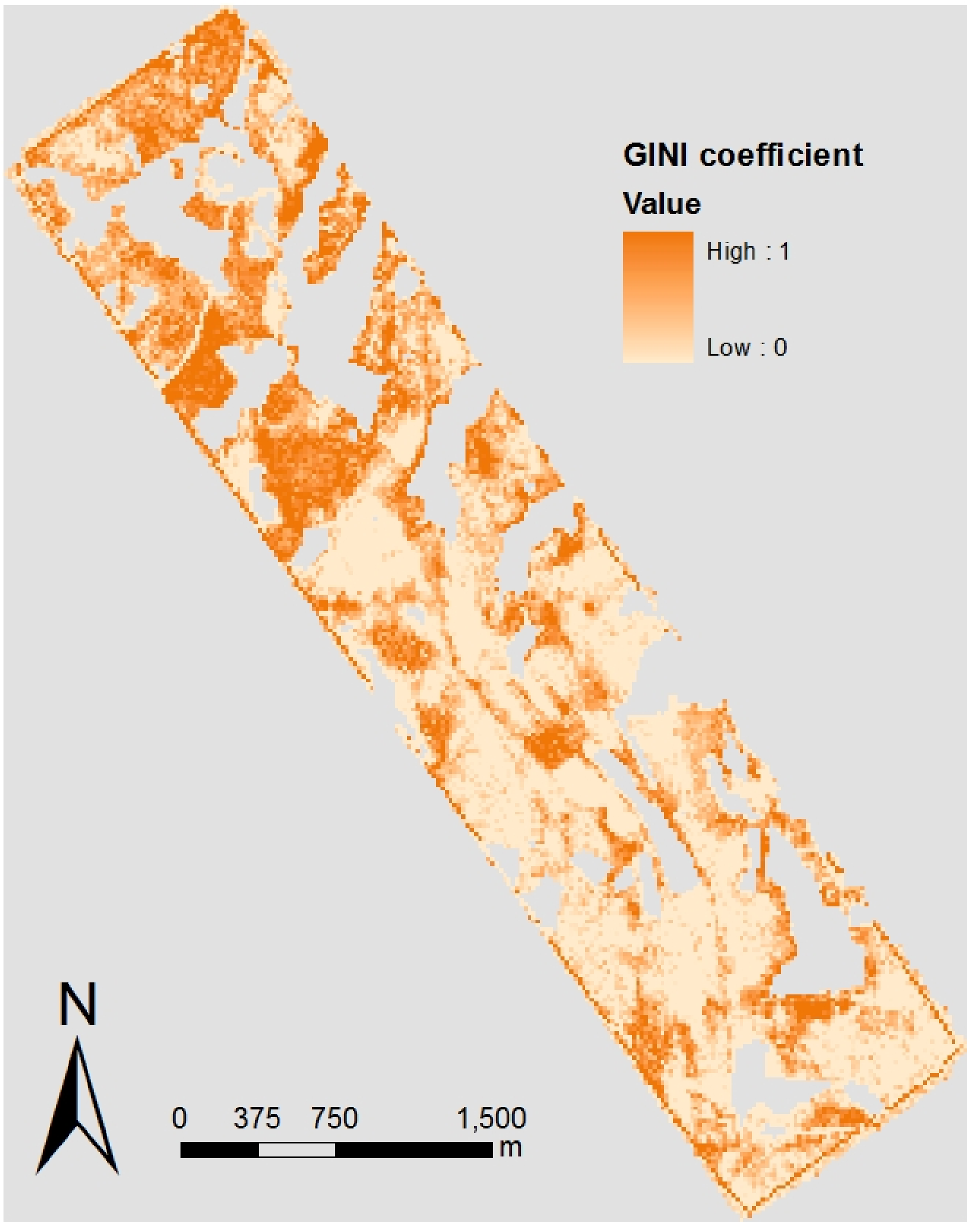

LIDAR & ToSIA
Predicted total volume of logs
Total volume of logs (m3/ha)
Value
High : 396.118
Low : 0
N
0
375
750
1,500
m

LIDAR & ToSIA

Predicted total volume of pulp

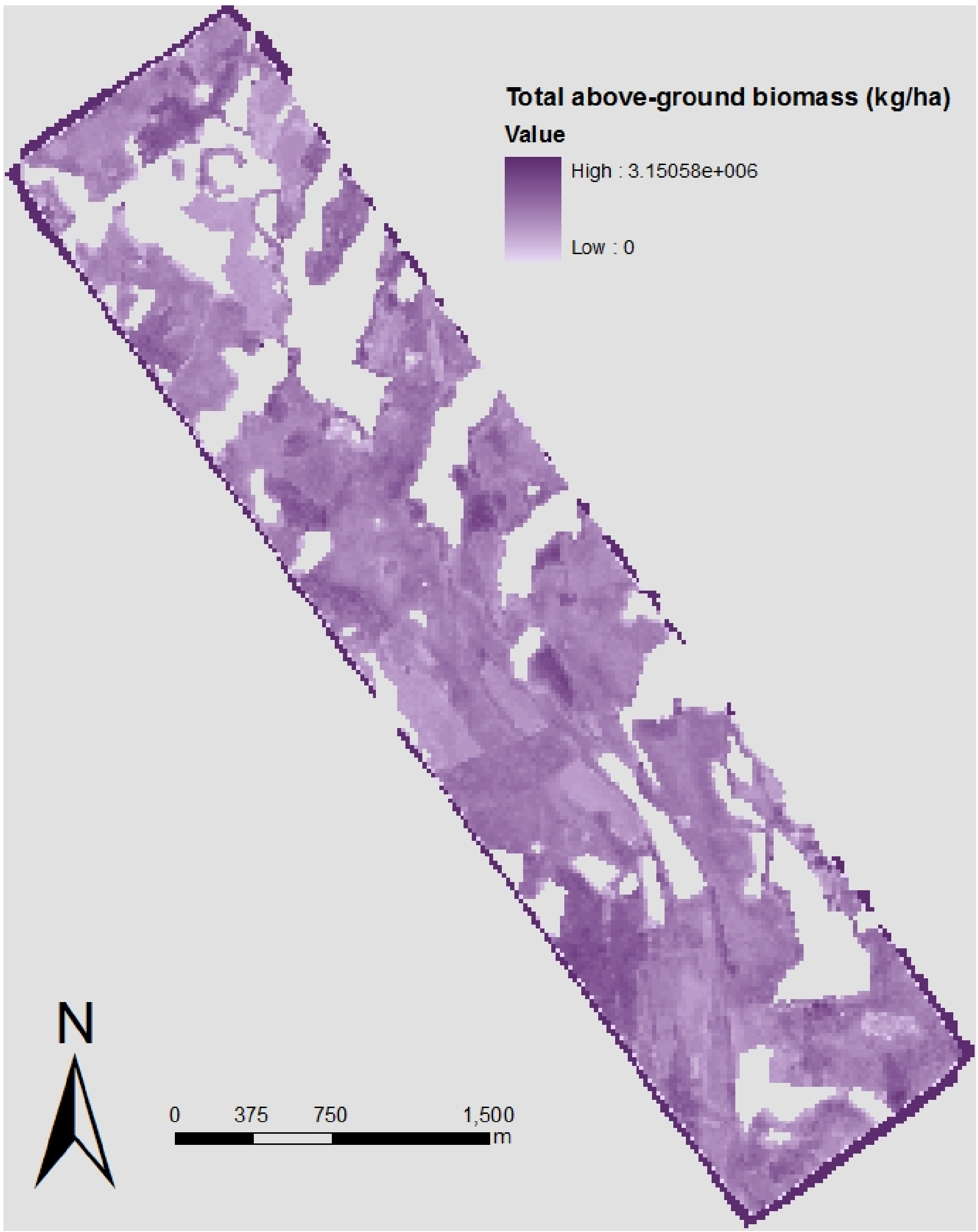


LIDAR & ToSIA
Predicted Gini coefficient



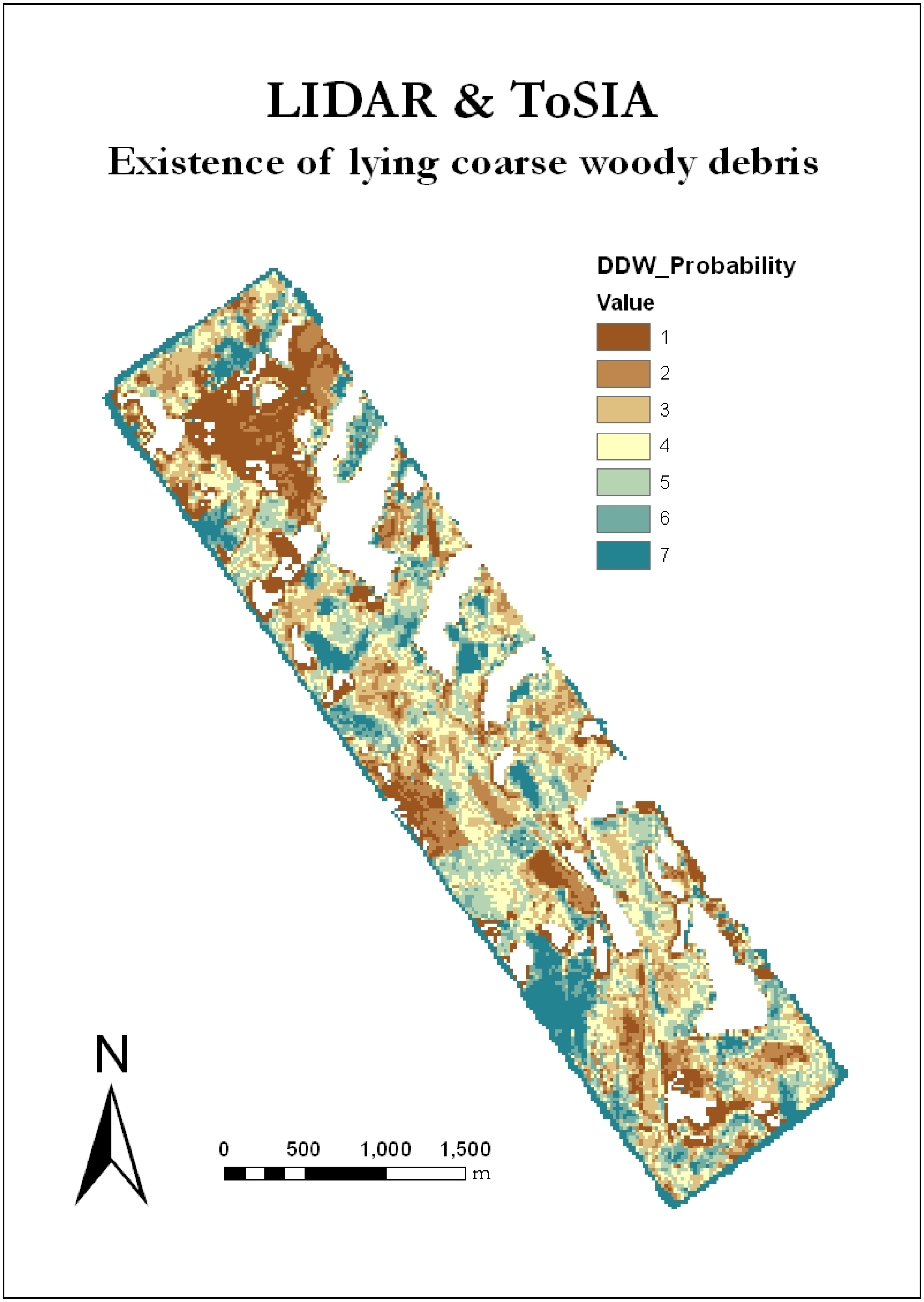
LIDAR & ToSIA

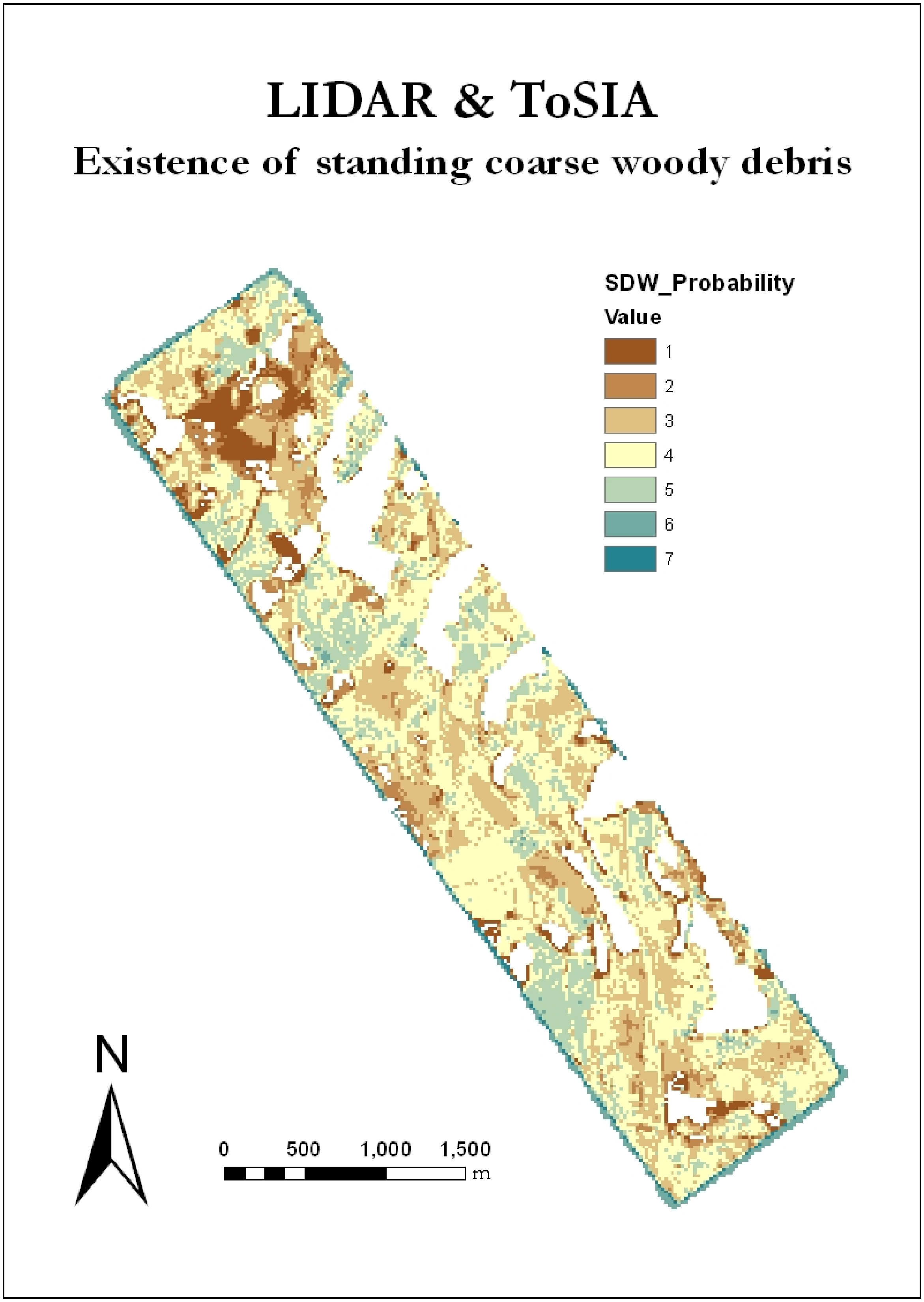
Predicted total above-ground biomass



LIDAR & ToSIA

Predicted total below-ground biomass


LIDAR & ToSIA
Existence of lying coarse woody debris
DDW_Probability
Value
1
2
3
4
5
6
7
N
0
500
1,000
1,500
m


LIDAR & ToSIA
Existence of standing coarse woody debris
SDW_Probability
Value
1
2
3
4
5
6
7
N
0
500
1,000
1,500
m

Appendix III: 1-Total volume of logs; 2-Total volume of pulp; 3-Gini coefficient; 4-Probability of existence of Coarse Woody Debris

Class ID 1 Number of Cells 183

Means 3.24237 3.96189 1.77585 5.77596

Covariance

1	22.13520	-2.87031	1.12173	5.92717
2	-2.87031	30.22076	-0.19400	2.13067
3	1.12173	-0.19400	27.45294	-0.96285
4	5.92717	2.13067	-0.96285	4.57041

Class ID 2 Number of Cells 70

Means 28.77259 3.15701 7.40896 9.40000

Covariance

1	78.04348	-3.23147	-22.19686	1.23292
2	-3.23147	35.22125	17.31625	4.35641
3	-22.19686	17.31625	80.92136	-4.10853
4	1.23292	4.35641	-4.10853	9.02609

Class ID 3 Number of Cells 92

Means 5.14743 25.23847 14.73106 7.25000

Covariance

1	36.90389	-13.53885	12.97806	8.30465
2	-13.53885	84.60917	-11.78061	-1.36407
3	12.97806	-11.78061	104.60118	0.43472
4	8.30465	-1.36407	0.43472	3.06868

Class ID 4 Number of Cells 58

Means 8.49524 42.72372 51.50512 8.36207

Covariance

1	66.38082	-17.93472	-2.72208	12.97703
2	-17.93472	118.55005	39.04666	2.36690
3	-2.72208	39.04666	108.00369	-4.06796

4	12.97703	2.36690	-4.06796	4.02450

Class ID 5 Number of Cells 44

Means 35.34444 3.07939 42.07282 8.06818

Covariance

1	112.29281	-12.34531	20.55422	11.79787
2	-12.34531	29.11814	9.00531	5.66754
3	20.55422	9.00531	103.55170	-0.77943
4	11.79787	5.66754	-0.77943	8.62315

Class ID 6 Number of Cells 52

Means 15.32100 13.76955 66.74054 6.53846

Covariance

1	116.15757	7.94546	11.65752	19.38899
2	7.94546	128.31227	-11.74174	16.78034
3	11.65752	-11.74174	151.01891	-5.02393
4	19.38899	16.78034	-5.02393	6.76320

Class ID 7 Number of Cells 57

Means 14.06011 48.30201 91.26760 8.43860

Covariance

1	124.44171	-2.89183	-0.35539	20.50673
2	-2.89183	153.35814	17.98589	12.66592
3	-0.35539	17.98589	98.52361	-0.43447
4	20.50673	12.66592	-0.43447	5.50063

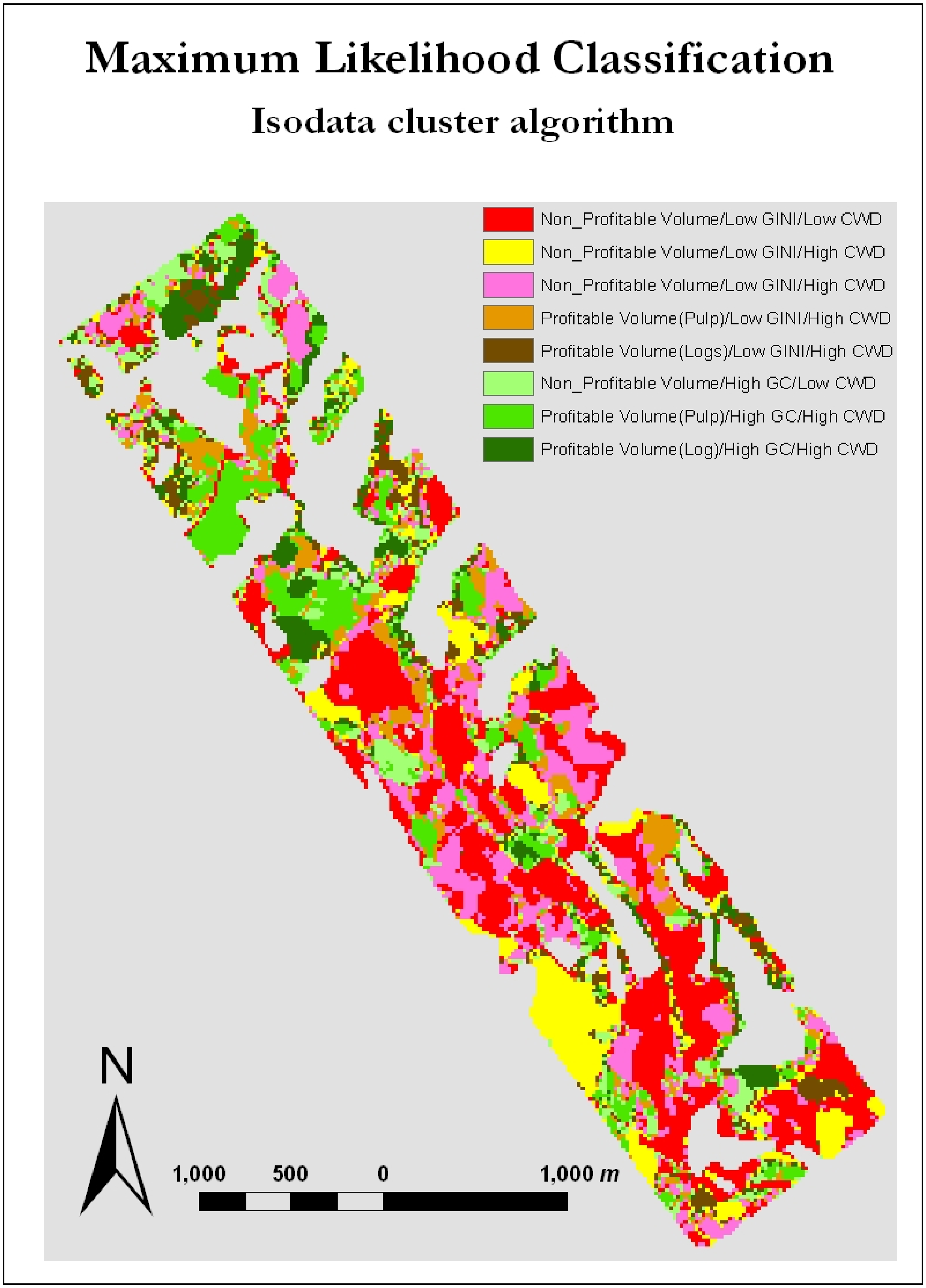
Class ID 8 Number of Cells 54

Means 49.11256 11.56710 91.64449 9.40741

Covariance

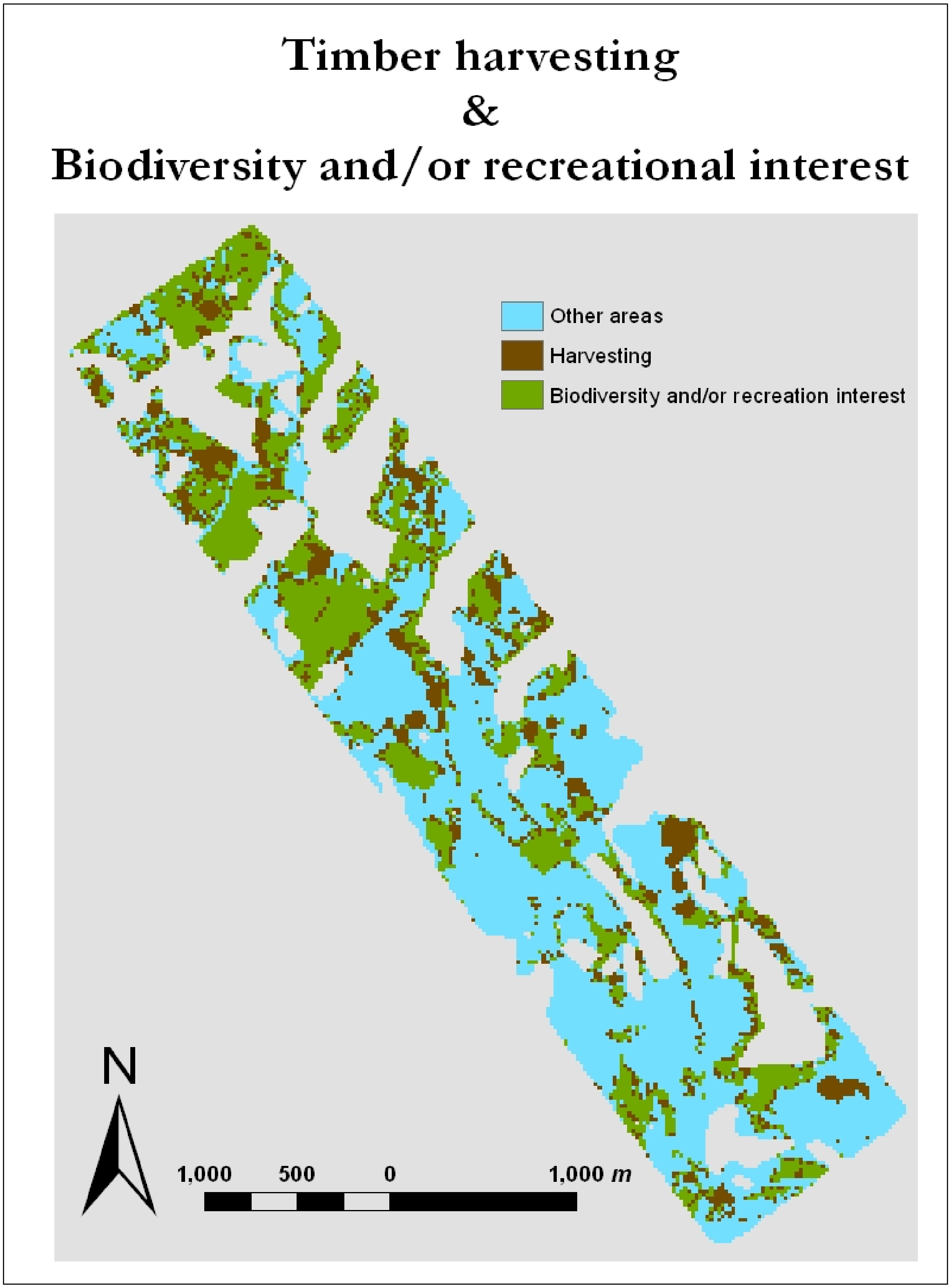
1	139.06345	-48.09203	-17.44329	6.70367
2	-48.09203	143.92722	30.13732	15.43061
3	-17.44329	30.13732	94.34750	-0.99782
4	6.70367	15.43061	-0.99782	6.92523

Appendix IV

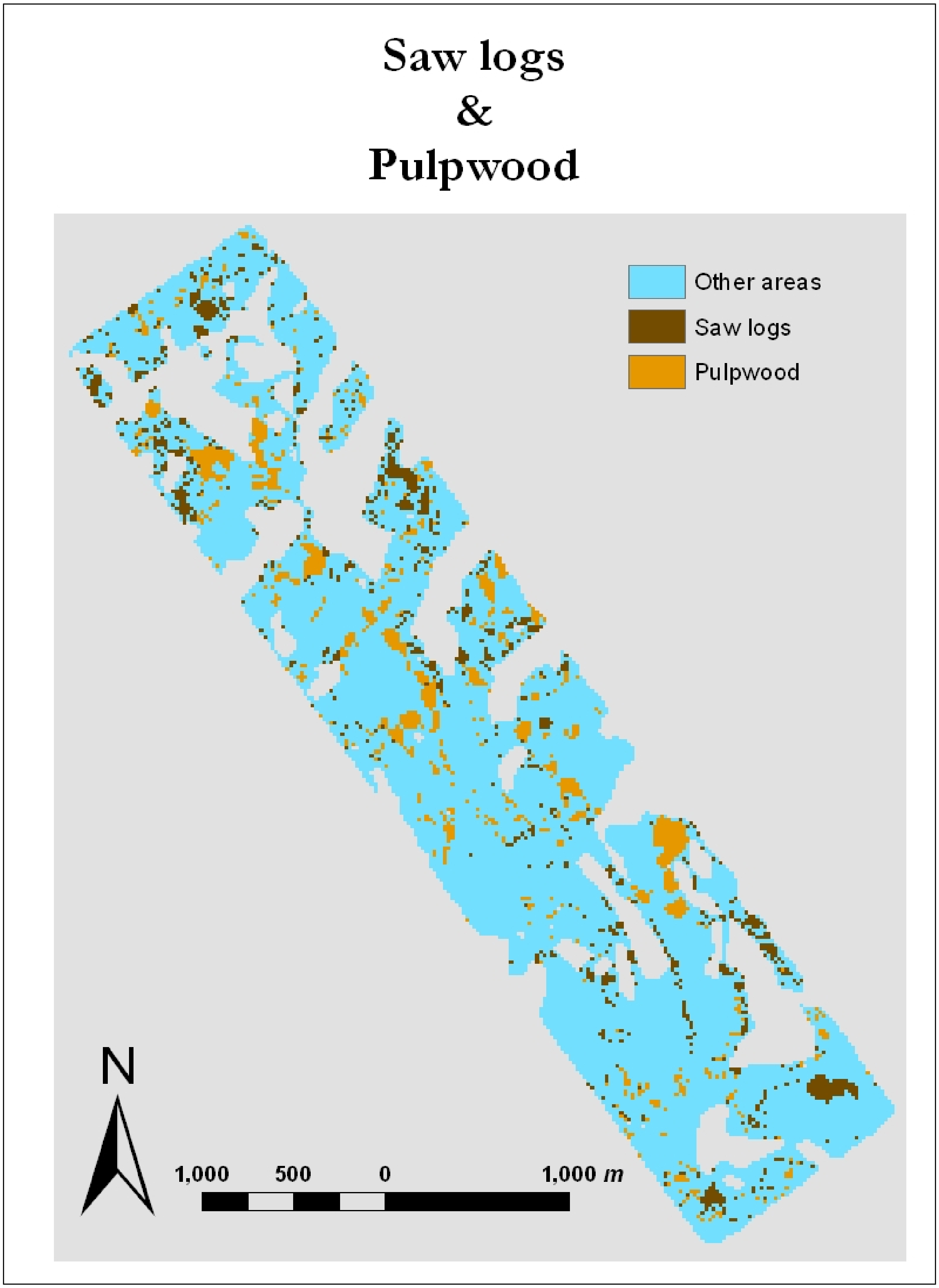


Maximum Likelihood Classification
Isodata cluster algorithm
Non_Profitable Volume/Low GINI/Low CWD
Non_Profitable Volume/Low GINI/High CWD
Non_Profitable Volume/Low GINI/High CWD
Profitable Volume(Pulp)/Low GINI/High CWD
Profitable Volume(Logs)/Low GINI/High CWD
Non_Profitable Volume/High GC/Low CWD
Profitable Volume(Pulp)/High GC/High CWD
Profitable Volume(Log)/High GC/High CWD
N
1,000
500
0
1,000 m

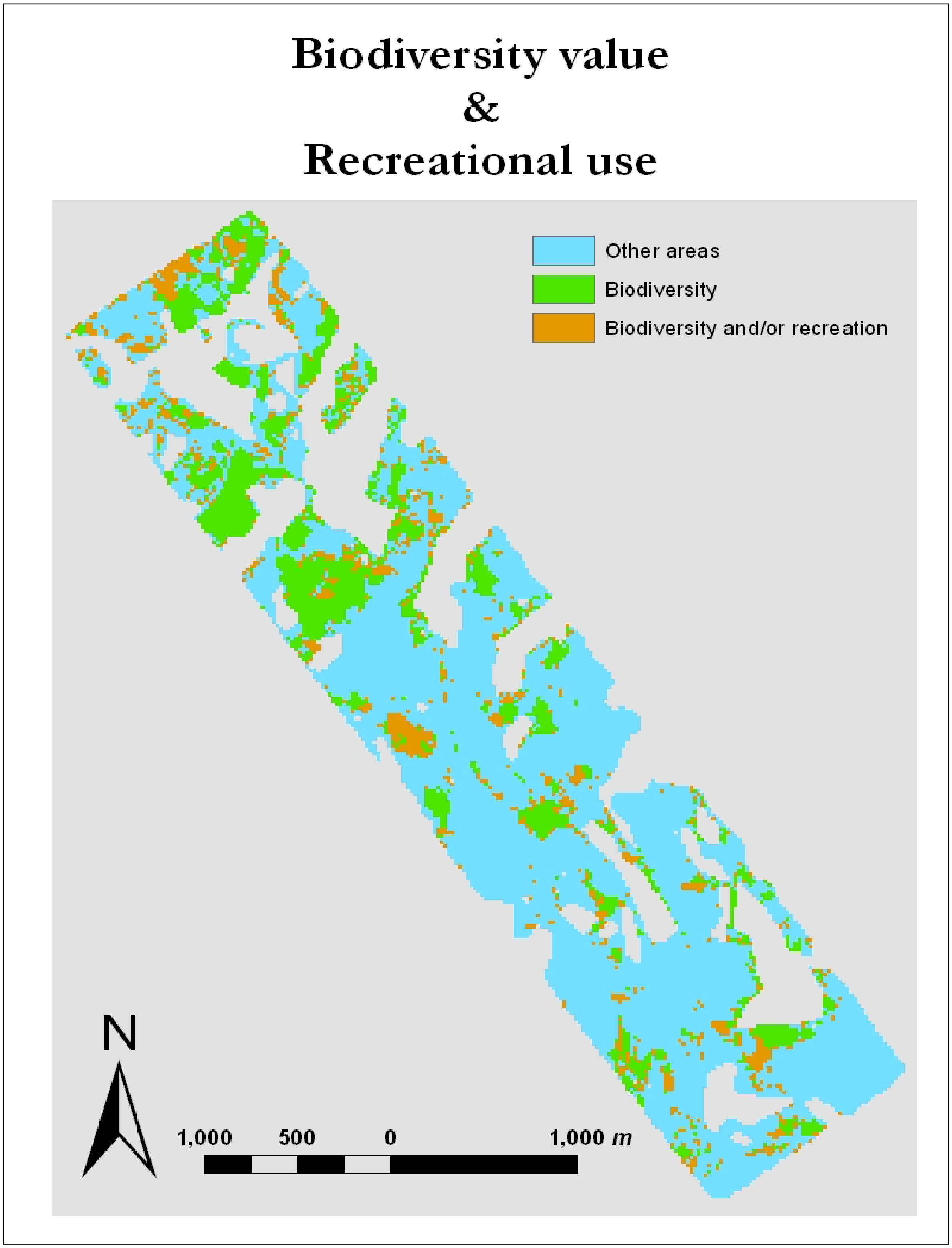
Appendix V


Timber harvesting
&
Biodiversity and/or recreational interest
Other areas
Harvesting
Biodiversity and/or recreation interest
N
1,000
500
0
1,000 m

Appendix VI


Saw logs
&
Pulpwood
Other areas
Saw logs
Pulpwood
N
1,000
500
0
1,000 m

Appendix VII



Appendix VIII

Proportions of species mixture