Agreement or Chance: How Exact Are Tree Markings in Forest Management?

Carlos Pallarés Ramos

Workreport 455 2016
Master thesis 30hp A2E
MSc European Forestry

Supervisor:
Arne Pommerening

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
901 83 UMEÅ
www.slu.se/srh
Tfn: 090-786 81 00

ISSN 1401–1204
ISRN SLU–SRG–AR–455–SE

Agreement or Chance: How Exact Are Tree Markings in Forest Management?

Carlos Pallarés Ramos

.

Keywords: Human selection behaviour; Fleiss' kappa; tree marking; marteloscope, thinnings.

Master thesis in forest management at the Dept. of Forest Resource Management, 30 credits
EX0768, A2E
Supervisor: Arne Pommerening, SLU, dept. of Forest Resource Management,
Examiner: Göran Ståhl, SLU, dept. of Forest Resource Management

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
Utgivningsort: Umeå
Utgivningsår: 2016

ISSN 1401–1204
ISRN SLU–SRG–AR–455–SE

Abstract

Tree marking activities are usually assumed to follow textbooks and management plan guidelines. However, initial experiments starting in the 1990s have shown that there is much room for personal interpretation by field and operational staff. This study presents an analysis of tree selection variability among people selecting trees in *marteloscopes* as part of silvicultural training activities. The analysis has been done for two different thinning types, i.e. low and crown thinning, as well as for the selection of frame trees in twelve different forests in the United Kingdom.

Two different methods have been used for assessing the agreement among participants, i.e. Fleiss' kappa and a method based on the test statistic of the χ^2 goodness-of-fit test. The Pearson correlation coefficient was calculated to identify possible relationships between the aforementioned two agreement characteristics and the structural parameters of the forests. In general the agreement between test persons was low for all types of tree selection. Overall, the agreement was higher for low thinning exercises followed by the selection of frame trees and the marked frame-tree competitors in crown thinnings. Both agreement characteristics indicated that test persons tend to agree more when selecting trees for low thinnings and when selecting frame trees. There were no consistent patterns that suggest that the structure of the forest influences the level of agreement between test persons.

Key words: Human selection behaviour; Fleiss' kappa; tree marking; marteloscope, thinnings.

Contents

Introduction

Modern forestry is the result of two centuries of experimentation at practical and research level. Current forest management textbooks owe much to a historical evolution of forest practices, partly based on tradition, partly on research results (Agnoletti & Anderson 2000). The latter stem from long-term monitoring and experiments and respective conclusions have been translated into silvicultural prescriptions and management guidelines (Agnoletti 2006; Kirby & Watkins 2015).

These give approximate recommendations and there is a lot of room for machine operators and field staff to implement them in a way that personally seems most appropriate to them. This personal interpretation of research results and prescriptions gives rise to much variation. Thinnings are typical management operations where some trees are selected for removal to favour others that remain in the forest stand under consideration (Helms 1998). In this process, trees have to be marked for eventual removal and this marking is typically based on the aforementioned prescriptions and guidelines.

In previous decades, silviculturists have often assumed that this kind of tree selection is almost unanimous given the same education and thinning instructions. Research starting in the 1990s has cast some doubt on this assumption (Daume et al. 1997, Zucchini and Gadow 1995, Füldner et al. 1996). Apparently there is a lot of uncertainty in tree marking resulting in a considerable variation in the selection of trees.

When selecting trees, regardless of the management objective, a major decision is taken that will affect the dynamics of a stand for many years if not decades (Swift et al. 2013). This is where the importance of selecting trees lies, i.e. the process is directly linked to management objectives and to the practical application of knowledge. Junod (2011, pers. comm.) referred to tree marking as a key responsibility at the interface between global management planning and local implementation. In the past, field staff and machine operators performed this work according to oral or written instructions. However, hardly been any research has been carried out on how this task was actually performed (Gadow, 1996).

As pointed out, a single management operation can severely affect the dynamics of a forest stand. It is possible to predict the growth and dynamics of a stand rather accurately (Wyckoff & Clark 2005; Clark et al. 2007; Graham et al. 1999) and thus to simulate the consequences of alternative tree removals, for example, in terms of growth and yield (Murray & von Gadow 1991; Peltola et al. 2002). For this reason, stand development and tree selection for thinnings or for selective harvesting are closely related; research in tree selection agreement can effectively complement forest modelling by including person-specific tree selection probabilities.

Education and subsequent training can profile people's choice in terms of tree selection behaviour (Vítková et al. 2016); forestry staff commonly base their decisions on the silvicultural education and training they have received and on their own interpretation of the literature or of prescriptions. In forestry schools around the world, it is increasingly common to evaluate tree marking skills as part of courses in silviculture and forest management. Students are asked to mark trees according to some guidelines and their performance is assessed using a number of simple statistics.

Literature review

Studying agreement between individuals selecting trees has so far largely been neglected. Until now research in this field has focused on modelling and simulation of different thinning interventions and intensities (Mäkinen & Isomäki 2004a; Mäkinen & Isomäki 2004b; Nilsen & Strand 2008; Vanclay 1989; Kariuki 2008; Crecente-Campo et al. 2009; Zhang et al. 2014) following the rationale of forest growth and yield experiments.

The modelling work of these studies was mainly concerned with the objective of predicting forest dynamics including thinnings and harvesting. However, the modelling of these two processes always focussed on textbook or best-practice scenarios and hardly ever attempted to quantify the differences in forestry staff charged with the same task of marking trees. In this master thesis, I therefore studied and discussed the agreement among individuals marking trees for thinnings.

Recently Spinelli et al. (2016) studied the silvicultural results (in terms of basal area and trees per hectare) performed by test persons with different professional backgrounds in mixed continuous-cover-forestry woodlands in Northern Italy. They found no significant difference in the marking of test persons from different professional groups, however, also identified a substantial lack of agreement in terms of the selection of individual trees.

Human tree selection research implies the collection of data from specific sites where a comparison between individuals is possible. To facilitate this, the marteloscope experiment was developed for monitoring human decisions in tree selection. In practice, the marteloscope is similar to a standard forest research plot, where tree data are collected and recorded: stem diameter at breast height (1.3 meters above ground level), tree species, tree locations (Cartesian coordinates) and perhaps some additional qualitative features like log quality and habitat suitability. In addition to these measurements, trees are labelled with numbers, for two purposes: (1) For the identification of trees in the field experiment and (2) for linking individual trees with measurements. The marteloscope is often used for practical training in forestry, where trainees are required to mark trees according to some instructions and objectives (Poore 2011; Pommerening et al. 2015)

Marteloscope-based tree-marking training is used in a number of European countries and also overseas in the United States and Canada (see Table 1). It is possible that the technique is also used in some Asian and South American countries; however, it is difficult to identify relevant information from those regions.

Table 1. Projects involving the use of marteloscopes for in-situ forestry training.

Project / Organization	Location/Country	Objective	Website/URL
AFI (Association Futaie Irrégulière)	France	Monitor Continuous Cover Forestry (CCF)	http://prosilva.fr/html/index.html
CPFC (Centre de la Propietat Forestal de Catalunya)	Barcelona, Spain	Forestry training	http://cpf.gencat.cat/en/index.html
CRPF Auvergne (Regional Center of Private Forest Property of Auvergne)	France	Forestry training for CCF	http://www.crpfauvergne.fr (In French)
Forestry Commission UK	United Kingdom	Forestry training for CCF	http://www.forestry.gov.uk
Joseph W.Jones Research Center	Georgia, United States	Forestry training for multi-aged forest stands	http://www.jonesctr.org
SelecFor	Wales, United Kingdom	Forestry training for CCF	http://www.selectfor.com
University of Lleida	Lleida, Spain	Forestry training	http://www.forestal.udl.cat/es (In Spanish)
University of Moncton	Canada	Forestry training	https://www.umoncton.ca (In French)
University of Valladolid	Valladolid, Spain	Research and forestry training	http://www.uva.es/export/sites/uva/ (In Spanish)
University of Warsaw	Warszawa, Poland	Forestry training	http://en.uw.edu.pl
Hammer Project	France, Finland, Italy, Belgium and Spain.	Build a digital platform for thinning simulations	http://www.hammer-project.eu
Integrate+ (EFI, EFICENT & BMEL)	Freiburg, Germany	Create a European network of demonstration sites and specific software to be used in portable devices.	http://www.integrateplus.org

AFI (Association Futaie Irrégulière) together with the AgroParisTech-ENGREF at Nancy (France) were the first organizations that put forward the idea of marteloscopes as a tool for monitoring human tree selection behaviour (Bruciamacchie et al. 2005; Pommerening et al. 2015). The organisation now includes a network of 82 research plots across Europe where tree-marking exercises are regularly conducted. Human behaviour is assessed individually with a special emphasis on the consequences of the participants' individual selection for forest stand development.

CRPF, CPFC and all universities mentioned in Table 1 regularly organise marking exercises with students, forest stakeholders (private forest owners, forest managers, forest workers…) and with members of the general public including an evaluation and a discussion of individual tree selection performance.

The Forestry Commission in the United Kingdom have a forest training centre (Ae Training Centre, Scotland) that regularly offers silvicultural and other training courses. These include tree quality assessments for timber or recreational aspects, and also the marking of trees for multi-purpose forestry. These marking exercises include an individual evaluation of tree marking as well as comparisons between course participants (Haufe, 2015, pers. comm.). The British company SelectFor offers services related to the installation of marteloscopes and to forestry training in CCF based on marteloscope exercises.

The Joseph W. Jones Research Center in Georgia (US) conducts workshops on managing longleaf pine (*Pinus palustris* Mill.) including tree marking exercises for obtaining a irregular forest structure. The variability between individuals in the marking process is discussed during these exercises (Steven, 2016, pers. comm.).

The Hammer and Integral+ projects are similar to the projects mentioned above in the sense that the implementation of the marking exercises is similar. However, they have developed specific software that can be installed on portable devices such as tablets and allows electronic data entry and an immediate analysis. Thus, the trees are selected by ticking trees on a portable computer screen. Once the exercise is completed the participant can view an evaluation of his or her marking work.

The objective of this study is to understand how much agreement and chance is there between test persons marking trees and if the level of agreement is related to structural properties of the forests where the marking takes place. For this purpose we analyse data from twelve sites in Britain.

Materials and methods

Study sites

For this study, data from twelve marteloscope sites managed by the Forestry Commission (FC) of Scotland and England and by Natural Resources Wales were used. In the remainder of this document, all three organisations are collectively abbreviated as FC. All sites have previously been unthinned, i.e. on all sites there was a thinning urgency providing sufficient incentives for tree marking. The sites are widely distributed in Great Britain as shown in Fig. 1.

Fig. 1. Location of the study sites.

Most of the sites include plantations of Sitka spruce (*Picea sitchensis* (Bong.) Carr.), Hybrid larch (*Larix x marchlinsii*), Japanese larch (*Larix kaempferi* (Lamb.) Carr.) and Scots pine (*Pinus sylvestris* L.) In some of these stands, other species have later colonised the site, but the previously mentioned species represent the main species in terms of density.

In this thesis, when referring to forest sites, each forest name is composed of two parts: The first part includes the name of the forest and the second the year when the exercise was held. For some forests there have been several groups of test persons. Thus, the group number follows the year. For example, CannockChase2014-1 refers to the forest Cannock Chase, year 2014 and group 1.

Each marteloscope had a size of 0.1 hectares and for each one of them, I calculated basic summary characteristics and presented them in Table 2. Climate and soil characteristics were obtained for each site from the Ecological Site Classification database (Pyatt et al. 2001) and I also included them in Table 2.

Table 2. Description of the sites included in this research. Source: Ecological Site Classification (ESC) software (http://www.forestry.gov.uk/esc). *dbh* (diameter at breast height), *Hg* (top height of the stand calculated as the mean height of the 100 largest trees per hectare), *Accum. T* (accumulated temperature, i.e. the annual temperature sum of values above 5° Celsius), *MD* (moisture deficit calculated from estimated evaporation minus rainfall), *DAMS* (measure of mean average wind speed plus the likelihood of gales. Values under 12 represent sheltered conditions, values over 18 are highly exposed), *CT* (continentality) and *P* (precipitation). [a]The stand was re- measured after some windthrow and silvicultural interventions

Site	Species	Age	*N* (trees/ha)	*V* (m^3/ha)	*BA* (m^2/ha)	*dbh*	*Hg* (m)	*Accum. T*	*MD*	*DAMS*	*CT*	*P* (mm)
Ae	Sitka spruce	54	1336	389	42.9	20.1	21.1	1241.3	99.1	13.1	6.3	1517
Ardross	Hybrid larch	18	2180	200	32.3	13.7	13.4	1011	74.8	13.2	4.2	707.6
Bin	Sitka spruce	30	1540	616	59.3	22.1	22.1	1083.9	105.9	9.7	5.7	844.3
Black Isle	Scots Pine	23	2010	136	26	12.8	10.8	1082.3	89.4	14.5	4.2	783
Cannock Chase	Hybrid larch	21	2040	263	36.7	14.9	14.6	1506.1	144.1	10.8	10.2	731.7
Craigvinean	Sitka spruce	17	3000	348	56.7	15	14.8	973.7	61.8	11	5.8	1110.4
Crychan	Hybrid larch	23	1930	356	41.2	16.5	16.1	1315.3	89.1	13.9	8.2	1489.2
Crychan[a]	Hybrid larch	23	1610	390	41.5	18.1	17.7	1315.3	89.1	13.9	8.2	1489.2
Dalby	Japanese larch	28	1900	443	46.2	17.6	18.6	1287.2	140.9	13.9	8.2	751.7
Glentress	Sitka spruce	30	1760	599	58.1	20.5	23.2	1165.2	96.4	10.4	6.6	841.8
Haldon	Sitka spruce	24	1780	365	43.9	17.7	18.6	1655	128	15	8	1217.5
Loch Ard	Sitka spruce	23	2450	322	43.3	15	17.9	1322	113	9	5	2006
Peckett Stone	Beech	54	830	351	34.7	23.1	24.5	1527.9	129.9	15	8.1	981.6

In every marteloscope, for each tree the following variables were measured: diameter at breast height (*dbh*) (measured in centimetres at 1.3m height), total tree height (m), location (*x* and *y* coordinates in metres) and stem quality subjectively classified as

A: No defects,
B: Minor defects,
C: Severe defects in timber quality.

The assessment of timber quality was carried out by experienced FC staff taking characteristics such as stem straightness, branchiness, evidence of harvesting and extraction damage, bark stripping and diseases (e.g. butt rot) into account. The crucial point in this classification is how much the defect limits the use of the corresponding stem as sawn timber. In this context, the severity and location of the damage, as well as the number of defects per stem is important.

Test persons

The study included nineteen groups of tree markers. Each group was comprised of a different number of test persons varying from a minimum size of 9 to a maximum of 20 participants.

About 95% of the participants were FC staff in different capacities ranging from machine operators to work supervisors and also included woodland officers and forest managers. The remaining 5% of the participants mainly worked as forestry contractors.

Experiments and data structure

The training sessions conducted on each site included two different thinning exercises. The first exercise involved a low thinning, otherwise known as ‘thinning from below’, where trees are removed mainly from the lower canopy and from among the smaller diameter trees (Helms 1998). The main objective of this type of thinning is to promote the growth of larger trees by removing smaller ones.

The second exercise involved a crown thinning, also referred to as ‘thinning from above’ where trees are removed that are part of the main canopy in order to favour the best trees of the main canopy (Helms 1998). The rationale of this thinning is to promote the growth of selected trees by removing their direct competitors.

The thinning instructions are comparatively specific and differ from site to site depending on local conditions.

As an example, the thinning instructions of the two exercises in Ae forest are described. In Appendix 1 an example of a typical marking is included.

Exercise 1- Low thinning
• The management objective aims at maximising volume production. • The thinning should create a uniform stand structure and improve overall timber quality. • Mark trees for a moderate to heavy low thinning. The marking should correspond to a removal of between 50 and 80 $m^3 ha^{-1}$ of standing volume. • Mark trees by ticking off their numbers on your marking sheet.

Exercise 2- Crown thinning
• The management objective aims to establish 200 frame trees per hectare for green logs and to create a relatively diverse stand structure with options for CCF management. • Selection criteria for frame trees are mainly vigour, stability and timber quality; regular spacing is less important. • Select between 150 and 250 frame trees per hectare and apply a crown thinning to release these trees from competition. Remove between one and three competitors per frame tree. • Mark trees by ticking off the corresponding numbers on your marking sheet. If you want to suggest a tree to be a frame tree, tick the corresponding cell in the "Frame?" column. If you want to remove a tree, tick the corresponding cell in the "Remove?" column. Otherwise, leave blank.

In each exercise, the test persons fill in one marking sheet, see Appendix 1. Ticks are converted to the numeric value "1" and blanks to the numeric value zero "0". This results in a sequence of binary data for each trainee where a value of "1" corresponds to a tree being marked and "0" to a tree, which was not marked.

Statistical theory and models

One of the first steps in any analysis is to check if there is any agreement among the test persons. This was also the main focus of my thesis. In the event that there is chance agreement, it would make no sense to continue a study. A significant departure from chance agreement justifies a more detailed analysis. This implies that the marking of the test persons, statistically called raters, can be trusted (Fleiss et al. 2003, p. 598).
Therefore Fleiss and other authors created measures of agreement that quantify levels of agreement.

I considered a binary case of tree marking where a tree with mark "1" was selected by a test person. The selection behaviour of a test person j regarding tree i is denoted by the indicator function $\mathbf{1}_{ij}$:

$$\mathbf{1}_{ij} = \begin{cases} 1 \text{ if test person } j \text{ marks tree } i, \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

Further notation

m	Number of test persons (index j),
n	Number of trees (index i),
x_i	Number of marks "1" of the i^{th} tree.

It is clearly

$$x_i = \sum_{j=1}^{m} \mathbf{1}_{ij} \tag{2}$$

and the x_i values are between 0 and m.
The marking process can be modelled by the marking probabilities π_{ij} for the case that test person j marks tree i. It is

$$\pi_{ij} = P(\mathbf{1}_{ij} = 1). \tag{3}$$

Trees as objects of human selection also influence the decision process by their properties: Some trees are more attractive than others because of their stem form, size or because they are more accessible than others.

The following three statistical models can be considered.

$\pi_{ij} = \pi_{..} = \text{const.}$		$\pi_{..}$ model	Simplest model, null model	(4)
$\pi_{ij} = \pi_{.j}$	for all i	$\pi_{.j}$ model	Test persons differ in their active marking probability, which is the same for all trees.	(5)
$\pi_{ij} = \pi_{i.}$	for all j	$\pi_{i.}$ model	Trees differ in their passive marking probability, which is the same for all test persons.	(6)

In the $\pi_{..}$ model, the selection probability is constant, i.e. it is the same for all trees and test persons. This is a very simple and unrealistic model, serving as a kind of reference or null model.
The probability $\pi_{..}$ can be estimated by

$$\hat{\pi}_{..} = \sum_{i=1}^{n} x_i / mn\,, \tag{7}$$

i.e by the average proportion of marked trees.

In the $\pi_{.j}$ model, the test persons have different marking probabilities, which are constant for all trees. This assumes that the trees have no influence on the test persons' marking. This could describe the case of inactive test persons who do not physically visit the trees in question when marking them.
The probability $\pi_{.j}$ can be estimated by

$$\hat{\pi}_{.j} = \sum_{i=1}^{n} \mathbf{1}_{ij} / n \,, \tag{8}$$

i.e. by the proportion of trees marked by test person *j*.
In the $\pi_{i.}$ model, the trees have different marking probabilities, which are the same for all test persons. A group of lazy test persons, for example, walks through the forest and as a group they decide for every tree the marking probability and then finally select the trees using random numbers.
The probabilities $\pi_{i.}$ can be estimated by

$$\hat{\pi}_{i} = x_i / m \,, \tag{9}$$

i.e. by the proportion of marks for tree *i*.

Obviously the $\pi_{..}$ model is a special case both of the $\pi_{i.}$ and the $\pi_{.j}$ models. It can be expected that that in the case of the $\pi_{i.}$ model there is some agreement between the test persons, whilst in the case of the $\pi_{.j}$ model there is no agreement. An observed small degree of agreement obtained from data for the latter model would only be by chance.

Agreement measures

Fleiss' κ

Fleiss *et al.* (2003) describe various measures of agreement for the analysis of ratings on categorical scales. The book also includes a measure κ for multiple raters (Eq. 18.51, p. 614), which was introduced by Landis and Koch (1977). This characteristic is based on one-way analysis of variance.
The overall proportion of mark "1" is defined as

$$\overline{p} = \frac{\sum_{i=1}^{n} x_i}{nm} \,, \tag{10}$$

If the number of trees is large (e.g. $n \geq 20$), the mean square between trees (BMS) (in or case between trees) is approximately equal to

$$BMS = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - m\overline{p})^2}{m} \tag{11}$$

and the mean square within subjects (WMS) is equal to

$$WMS = \frac{1}{n(m-1)}\sum_{i=1}^{n}\frac{x_i(m-x_i)}{m} \tag{12}$$

κ corresponding to the so-called intraclass correlation coefficient is

$$\kappa = \frac{BMS - WMS}{BMS + (m-1)WMS} = 1 - \frac{\sum_{i=1}^{n}\frac{x_i(m-x_i)}{m}}{n(m-1)\overline{p}\overline{q}} \tag{13}$$

where $\bar{q} = 1 - \bar{p}$ (Fleiss et al. 2003, p. 610ff).

Kappa only uses summary information, i.e. only the x_i. The more detailed information which trees are marked by test person i and which are not, is not used in the calculation of κ in Eq. (13). As a consequence κ is not able to characterise all aspects of agreement between test persons. For example, κ correctly suggests “no agreement” for the $\pi_{.j}$ model, but it is also necessary to take the differences in the test persons’ marking activities into account. Also, kappa does not lead to unique values, i.e. it is possible to obtain the same values in quite different situations. Naturally it is always problematic to characterise a complicated multivariate distribution by one real number.

Landis and Koch (1977) standardised the labels attached to different scores of κ (see Table 3).

Table 3. Interpretation of κ values proposed by Landis and Koch (1977)

$\boldsymbol{\kappa}$	**Interpretation**
< 0	Poor agreement
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Apparently Table 3 is not universally accepted. Landis and Koch (1977) largely presented their personal opinion and did not provide evidence supporting these guidelines. Therefore Table 3 should be applied with great care.

Alternative measure of agreement χ

The differences in the test persons' marking behaviour can be characterised by the extent of deviation of the empirical distribution of the numbers n_j of trees marked by test person j ($n_j = \sum_{i=1}^{n} \mathbf{1}_{ij}$) from the distribution on the set $\{1,2,\dots,m\}$. I applied the test statistic of the χ^2 goodness-of-fit tests. This test is applied to the empirical distribution of the $\{n_j\}$ and tested against the uniform distribution. The test statistic is calculated according to Eq. (14).

$$\hat{\chi}^2 = \sum_{j=1}^{m} \frac{\left(n_j - \frac{\sum_{j=1}^{m} n_j}{m}\right)^2}{\frac{\sum_{j=1}^{m} n_j}{m}} \tag{14}$$

where

m	Number of test persons
n_j	Observed number of trees marked by test person j
n	Number of trees

I considered this characteristic because the uniform distribution of the numbers of marks is a necessary condition for the $\pi_{i.}$ model where the marking probabilities only depend on the trees and is the same for all test persons.

However, my aim is not to test this model but simply to characterise the extent of the deviation from the uniform distribution to identify significant differences in the making activity of the test persons.

The value of the $\hat{\chi}^2$ can be compared to the critical value $\chi_{m^2,0.005}$ (5% quantile of the $\hat{\chi}^2$ distribution with m^2 degrees of freedom) by calculating

$$\chi = \hat{\chi}^2 / \chi_{m-1,0.005}\,. \tag{15}$$

Small values of χ indicate small deviations from the uniform distribution, i.e. a high degree of agreement in terms of marking activity.

Site characteristics

Previous research has indicated that the level of agreement between different test persons selecting trees is comparatively low (Pommerening et al. 2015). I wondered whether the forests themselves might have had an influence on the different levels of agreement. In order to quantify investigate this possibility, correlations between the agreement measures described in the previous section and measurements of forest structure were studied.

Coefficient of variation (*cv*)

Using *dbh* measurements, the corresponding coefficient of variation, *cv*, is a measure of tree size diversity, i.e.

$$cv = \frac{\sqrt{\frac{\sum_{i=1}^{n}(dbh_i - \overline{dbh})^2}{n-1}}}{\overline{dbh}}, \tag{16}$$

where n is the number of trees in the marteloscope and $\overline{dbh}$ the arithmetic mean *dbh* (Porkress, 2004).

Skewness of the empirical diameter distribution (*sk*)

Sterba and Zingg (2006) proposed the skewness of the empirical diameter distribution as a measure of size diversity and this characteristic can be calculated as in Eq. (17).

$$sk = \frac{\sum_{i=1}^{n}(dbh_i - \overline{dbh})^3}{(n-1)\cdot sd^3}, \tag{17}$$

where n is the number of trees in the marteloscope, $\overline{dbh}$ the arithmetic mean *dbh* and *sd* is the standard deviation of *dbh* (Sterba and Zingg, 2006).

Herfindahl-Hirschman index

Another measure of tree size diversity used in this study was an adaptation of the Herfindahl index, also known as Herfindahl-Hirschman index. The Herfindahl index was originally developed as a measure of industry concentration in economics (Sun & Shao 2009). In this study, the index has been modified including basal area (*BA*), which is the cross-sectorial area of all stems of a given tree species or all stems in a stand measured at breast height and expressed per unit of land area (Helms 1998). I have defined this index as the sum of squares of the relative cross-sectorial areas of all trees in the marteloscope and computed it using Eq. (18).

$$H = \sum_{i=1}^{n}\left(\frac{BA_i}{BA}\right)^2 \tag{18}$$

n is again the number of trees in the marteloscope, BA_i is the basal area (m^2/ha) of tree i and *BA* is the sum of the basal areas of all trees per hectare (stand characteristic). This index has been found a useful measure of tree density because it incorporates tree size in the calculation (Peck et al. 2014).

A measure of tree density is the relative spacing index *RS* proposed by Hart (1928) and Becking (1953). It is defined as the ratio of the average growing space in a forest stand and stand top height. Relative spacing can be calculated according to Eq. (19).

$$RS = \frac{\sqrt{1000/N}}{H} \tag{19}$$

N is the number of trees per hectare and H is stand top height (m) which is the average height of the 100 trees per hectare of largest diameter (Helms 1998).
The stand density index (*SDI*) is a useful measure of the degree of site occupancy of a stand. It is based on the quadratic mean diameter and stand density (Reineke 1933, Eq. 20).

$$SDI = N\left(\frac{25}{d_q}\right)^{-1.605} \tag{20}$$

N is the number of trees per hectare and d_q the quadratic mean diameter (cm).
Pearson's correlation coefficient was used to check the correlations between κ and the previously described measures of size structure. All calculations were performed using R (R Core Team, 2014).

Results

Measure of agreement κ

The results using κ are shown in Fig. 2 and 3. Fig. 2 displays the κ values for both low and crown thinnings.

Fig. 2 indicates that all κ values relating to the low thinning exercises (blue dots) are larger than the κ values obtained from the crown thinning exercise. In eight out of fifteen sites, kappa related to low thinnings is larger than 0.4. The classification by Landis & Loch (1977) labels this value as "moderate agreement".

For the crown thinnings (red dots in Fig. 2), only the exercise at Loch Ard in 2015 shows a κ value larger than 0.4 For the rest of the sites, the agreement can be classified as slight or fair according to Landis & Loch (1977).

The results from the sites Loch Ard in 2015 and Cannock chase in 2013 and 2014 indicate comparatively small differences in the agreement between both thinning types. These differences are actually smaller than 0.1 but according to the Landis & Loch (1977) classification the overall agreement for these sites is still "moderate".

For the rest of the sites the differences between thinnings are greater (see Fig. 2) and this is due to low agreement values as a result of selecting trees for crown thinning and medium-sized values when trees for low thinning were selected.

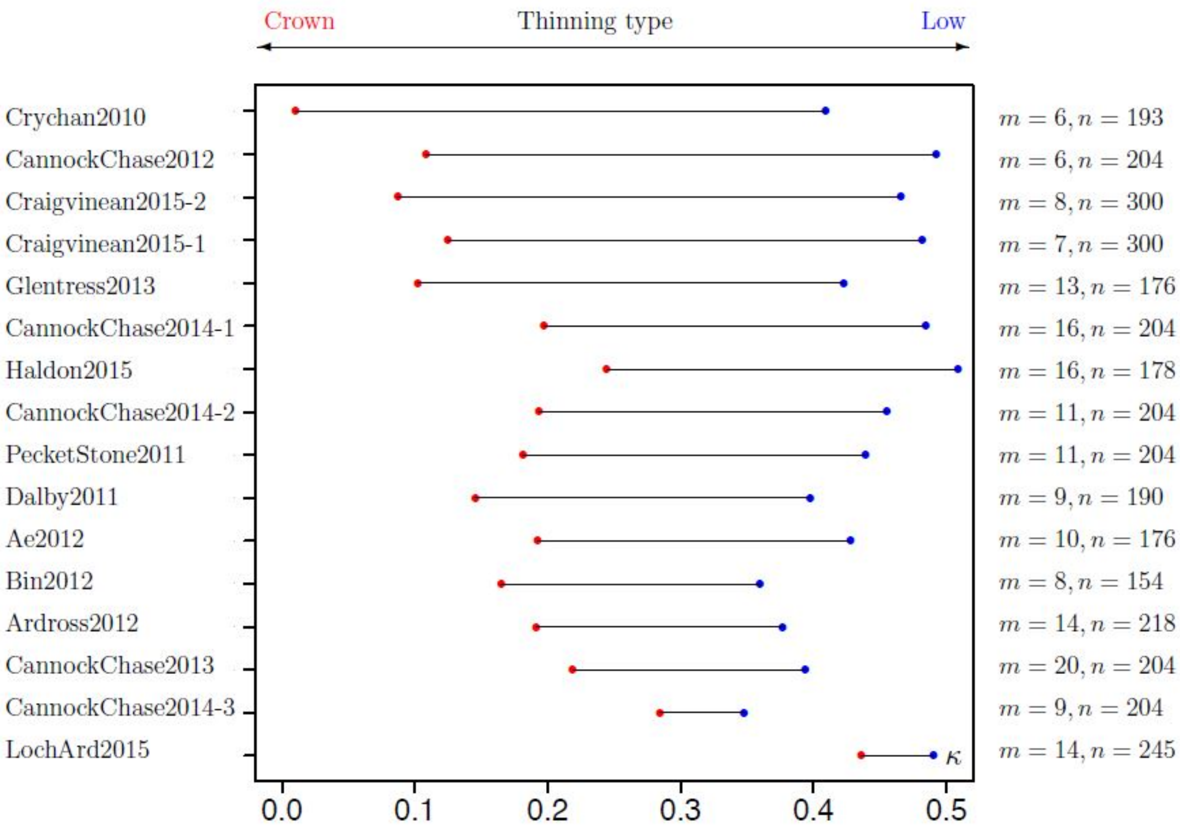

Fig. 2. κ values for the selection of trees in low and crown thinnings.

Fig. 3 shows the κ values for the exercises where trees were selected for removal as part of crown thinnings and for the selection of frame trees in the same exercises. From this figure it can be seen that the agreement in terms of selected frame trees is higher than the one for marking the competitors of these frame trees with the notable exception of Loch Ard in 2015. This is the only case in this study where participants agreed more when selecting competitor trees rather than frame trees.

Concerning the level of agreement when selecting frame trees, most κ values are between 0.2 and 0.4, corresponding to a "fair" agreement (Landis & Loch 1977). On the other hand, the corresponding agreement values relating to the selection of frame-tree competitors are in most cases markedly smaller. With the exception of site Loch Ard in 2015, at all sites κ values between 0.05 and 0.3 were achieved, corresponding to a "slight" agreement (Landis & Loch 1977).

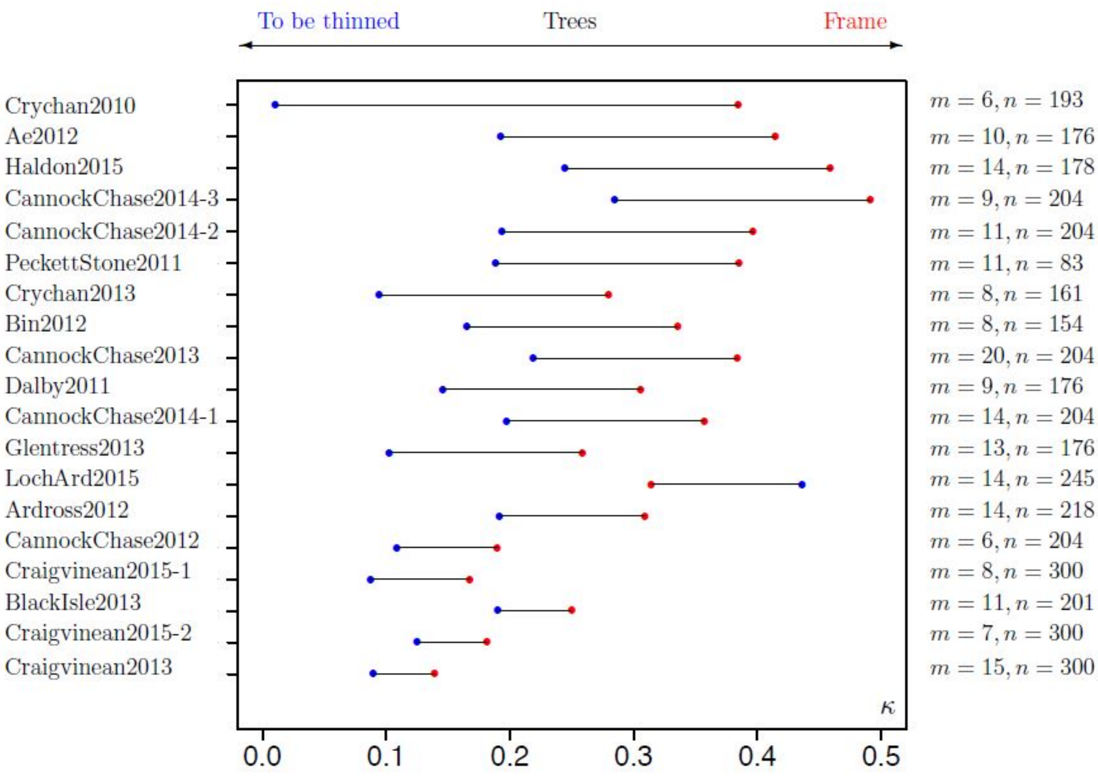

Fig. 3. κ values for the selection of frame trees and their competitors (crown thinning).

Alternative measure of agreement χ

Results using χ are shown in Fig. 4 and 5. Fig. 4 shows the results for both thinning types. The blue and red dots represent the χ values for low and crown thinning, respectively. In twelve of fifteen exercises the variability expressed by the χ values is larger for the crown thinning. Only for three exercises (Cannock chase in 2013 and 2014 [group 2] and Craigvinean in 2015 [group 2]) the variability of χ is larger for low thinnings.

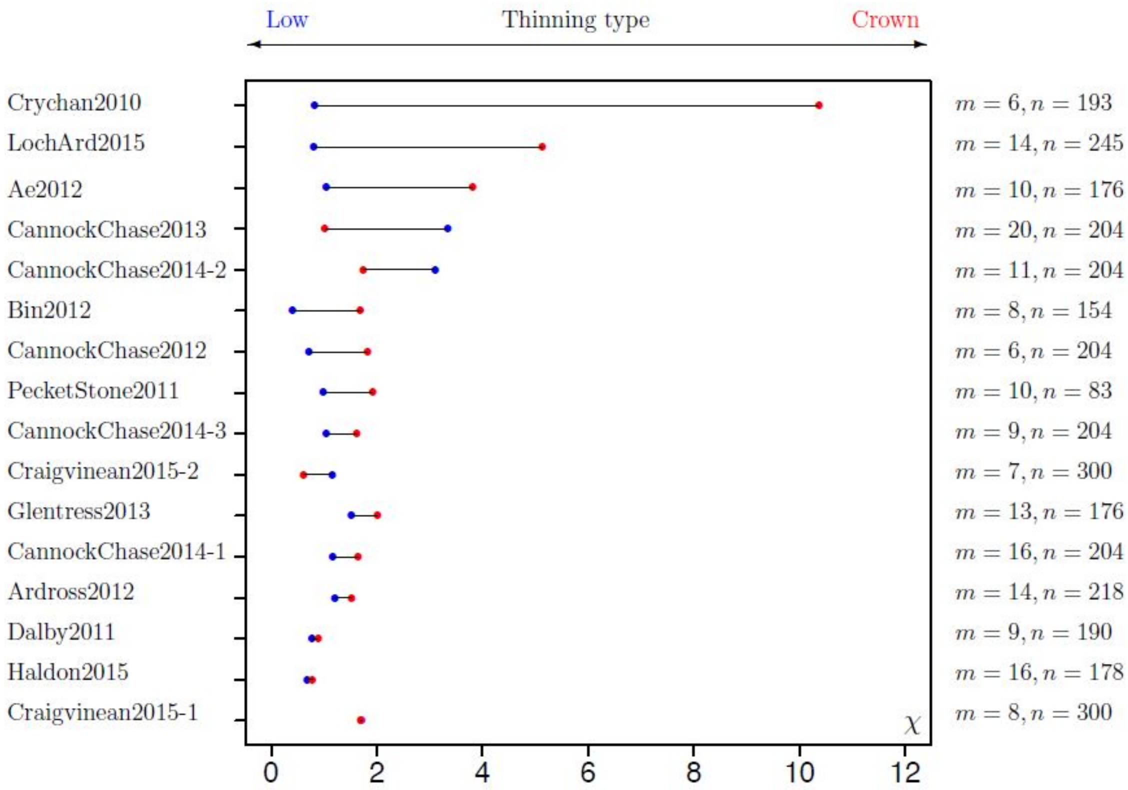

Fig. 4. χ statistic for the selection of trees in low and crown thinnings.

Fig. 5 shows the results of χ applied to the selection of frame trees and their competitors. The blue and red dots represent χ values for the marking of trees to be thinned and as frame trees, respectively. With the notable exception of the site Craigvinean in 2015 (group 2), the variability in χ is larger for crown thinnings than for the selection of frame trees.

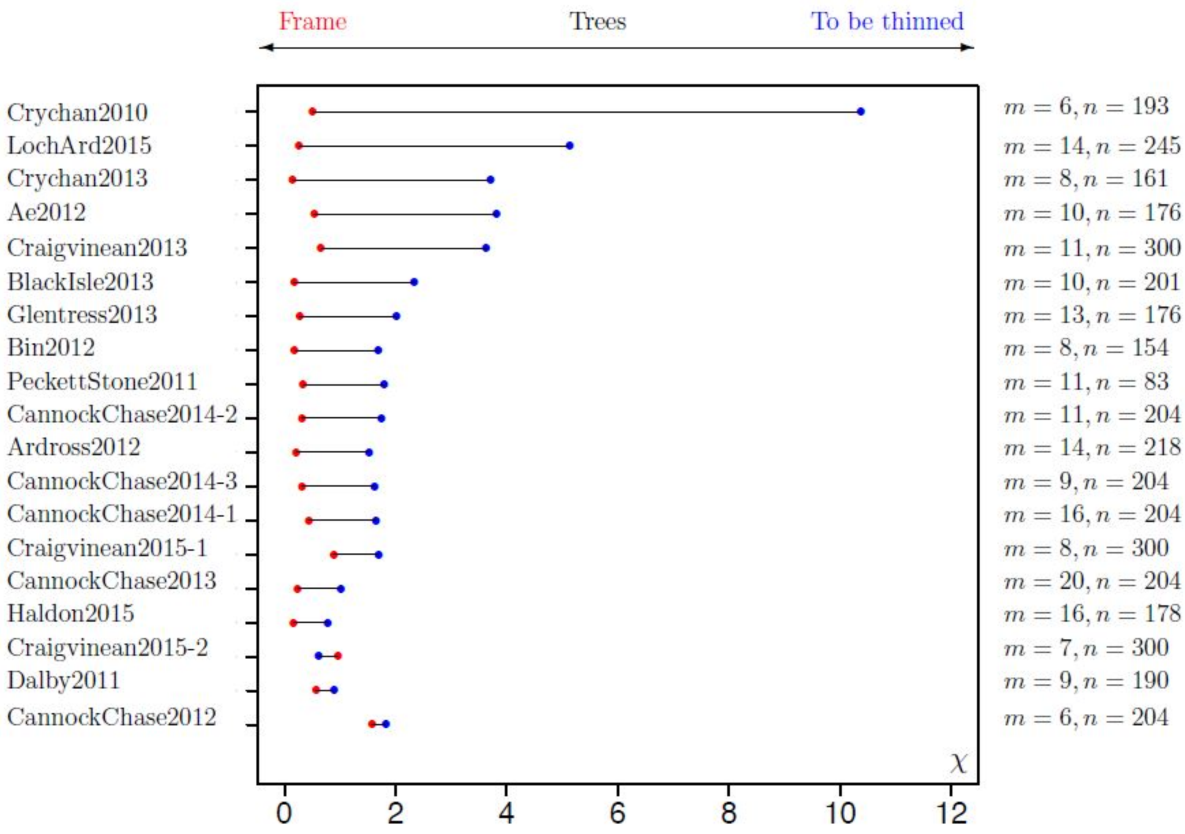

Fig. 5. χ statistic for the selection of frame trees and their competitors in crown thinnings.

Site characteristics

The structural parameters described in the Methods section were calculated and included in Table 4. All marteloscopes are located in plantations and this can be easily seen from the similarity of values of tree size diversity expressed with the coefficient of variation and the skewness of the diameter distribution. The skewness of the diameter distribution is positive for all the sites with the exception of the site Crychan, where the value is slightly negative. The values for the basal area (*BA*) vary between 32.2 m^2/ha (Ardross) and 58.1 m^2/ha (Glentress). The values of tree size density calculated with the Herfindahl-Hirschman index, are quite similar and rather low for all the exercises. The tree densities, measured with the relative spacing index (RS), vary from 0.11 (Glentress) and 0.18 (Ardross).

Table 4. Results of the structural measures: *cv* (coefficient of variation of diameter), *sk* (skewness of diameter), *SD* (standard deviation of diameter), *BA* (basal area expressed in m^2/ha), *H* (Herfindahl-Hirschman index), *RS* (relative spacing), *SDI* (stand density index)

Exercise	***cv***	***sk***	***SD***	***BA***	***H***	***RS***	***SDI***
Ae	0.35	0.17	6.68	42.9	0.01	0.12	1239.94
Ardross	0.37	0.49	4.78	32.3	0.01	0.18	834.18
Bin	0.30	0.12	6.33	59.3	0.01	0.13	1267.10
CannockChase	0.29	0.07	4.18	35.8	0.01	0.16	893.18
Crychan	0.28	-0.04	4.37	41.2	0.01	0.15	990.11
Dalby	0.28	0.31	4.81	46.2	0.01	0.13	1081.24
Glentress	0.29	0.06	5.87	58.1	0.01	0.11	1279.76
LochArd	0.35	0.36	4.95	43.3	0.01	0.13	1010.16
Haldon	0.35	0.39	5.85	43.9	0.01	0.14	1022.19
CannockChase	0.31	0.15	4.45	36.7	0.01	0.14	885.54
Craigvinean	0.24	0.07	3.68	56.7	0.00	0.12	1254.90
PeckettStone	0.29	0.33	6.48	34.7	0.02	0.00	730.01

Correlations between measures of agreement and structure

Pearson correlation coefficients were calculated using the measures of agreement and structure based on the tree data from the forest sites. The correlation values, as well as the level of significance, are shown in Table 5.

Table 5. Pearson correlations coefficients using κ and measures of forest structure. *cv* (coefficient of variation of diameter), *sk* (skewness of diameter), *SD* (standard deviation of diameter), *BA* (basal area expressed in m^2/ha), *H* (Herfindahl-Hirschman index), *RS* (relative spacing), *SDI* (stand density index), κ_{Low} (κ for trees selected to be removed in low thinnings), κ_{Crown} (κ for trees selected to be removed in crown thinnings), κ_{Frame} (κ for selecting frame trees in crown thinnings), χ_{Low} (χ value for low-thinning tree markings), χ_{Crown} (χ value for crown-thinning tree markings). Levels of significance of correlations given by *p*-value: ns if $p > 0.05$, * if $p \leq 0.05$, ** if $p \leq 0.01$, *** if $p \leq 0.001$ and 0 if $p = 0$.

	cv	***sk***	***SD***	***BA***	***H***	***RS***	***SDI***	κ_{Low}	κ_{Crown}	κ_{Frame}	χ_{Low}	χ_{Crown}	χ_{Frame}
cv	1.00	0.70	0.50	-0.31	0.22	0.24	-0.28	-0.05	0.61*	-0.01	0.00	0.29	0.13
sk	0.70	1.00	0.35	-0.34	0.36	-0.10	-0.37	0.04	0.57*	0.07	-0.14	-0.06	-0.04
SD	0.50	0.35	1.00	0.25	0.73	-0.49	0.14	-0.17	0.17	0.15	-0.25	0.10	-0.22
BA	-0.31	-0.34	0.25	1.00	-0.24	-0.14	0.98	0.02	-0.27	0.20	-0.30	-0.07	-0.40
HHI	0.22	0.36	0.73	-0.24	1.00	-0.77	-0.40	-0.14	0.06	0.10	-0.13	-0.02	-0.12
RS	0.24	-0.10	-0.49	-0.14	-0.77	1.00	0.04	-0.11	0.04	-0.21	0.22	0.05	0.21
SDI	-0.28	-0.37	0.14	0.98	-0.40	0.04	1.00	0.02	-0.28	0.17	-0.26	-0.05	-0.35
κ_{Low}	-0.05	0.04	-0.17	0.02	-0.14	-0.11	0.02	1.00	0.09	0.25	-0.10	0.06	0.20
κ_{Crown}	0.61*	0.57*	0.17	-0.27	0.06	0.04	-0.28	0.09	1.00	-0.03	0.15	0.20	0.26
κ_{Frame}	-0.01	0.07	0.15	0.20	0.10	-0.21	0.17	0.25	-0.03	1.00	0.01	0.25	-0.66
χ_{Low}	0.00	-0.14	-0.25	-0.30	-0.13	0.22	-0.26	-0.10	0.15	0.01	1.00	-0.10	0.01
χ_{Crown}	0.29	-0.06	0.10	-0.07	-0.02	0.05	-0.05	0.06	0.20	0.25	-0.10	1.00	0.04
χ_{Frame}	0.13	-0.04	-0.22	-0.40	-0.12	0.21	-0.35	0.20	0.26	-0.66	0.01	0.04	1.00

In general, the correlation coefficients are low. Just two significant correlations between the kappa measure for the competitor trees of exercise 2 (crown thinning) and the skewness and coefficient of variation of the diameter distribution could be identified. Among themselves, the measures of forest structure such as the coefficient of variation (cv), skewness (sk) and standard deviation (SD) of the diameter distribution show large values of correlation values (see Table 5).

Discussion

Based on previous work in this field, I started this research with the expectation that the agreement between forestry staff selecting trees would be moderate at best. However, in the past this finding has only come to light in very few, isolated studies. In this thesis, I had the opportunity to investigate a large data set covering a whole country and this is the first time that such a systematic analysis has ever been performed.

The results of this thesis have indeed confirmed the expected trend: Independent of the particular forest where the exercises were based and independent of the thinning type, the overall agreement was rather low. I can therefore conclude that independent of management type the variability of selected trees apparently is always considerable even among educated staff equipped with instructions.

I have quantified the agreement using two methods, i.e. Fleiss kappa and a measure based on the test characteristic of the χ^2 goodness-of-fit test. .

When answering the main question of this thesis, i.e. how much agreement and change is there between humans selecting trees, the significance of the kappa calculations has also been assessed. With the exception one site and exercise (site Crychan and year 2010, crown thinning), all κ values are significantly different from zero (see Table 6 and Hedderich and Sachs 2012, p. 671). Thus, it is possible to state that the weak agreement in tree selection on these sites is due to active human marking and not due to chance.

Table 6. Kappa results with the corresponding p-values, where κ_{Low} (κ for low-thinning tree selection), κ_{Crown} (κ for crown-thinning tree selection), κ_{Frame} (κ for frame tree selection), p-value $_{Crown}$ (p-value for κ of crown-thinning markings), p-value $_{Low}$ (p-value for κ of low-thinning markings) and p-value $_{Frame}$ (p-value for κ of frame-trees selection). Levels of significance of p-value: ns if $p > 0.05$, * if $p \leq 0.05$, ** if $p \leq 0.01$, *** if $p \leq 0.001$ and 0 if $p = 0$

Sites	κ_{Low}	κ_{Crown}	κ_{Frame}	p-value $_{Crown}$	p-value $_{Frame}$	p-value $_{Low}$
Ae2012	0.42	0.19	0.41	0	0	0
Ardross2012	0.37	0.19	0.30	0	0	0
Bin2012	0.35	0.16	0.33	0	0	0
CannockChase2012	0.49	0.10	0.18	**	0	0
CannockChase2013	0.39	0.21	0.38	0	0	0
Crychan2010	0.40	0.01	0.38	ns	0	0
Dalby2011	0.39	0.14	0.30	0	0	0
Glentress2013	0.42	0.10	0.25	0	0	0
BlackIsle2013	0.00	0.19	0.25	0	0	0
Craigvinean2013	0.00	0.08	0.13	0	0	0
CannockChase2014-1	0.48	0.19	0.35	0	0	0
CannockChase2014-2	0.45	0.19	0.39	0	0	0
CannockChase2014-3	0.34	0.28	0.49	0	0	0
Craigvinean2015-1	0.46	0.08	0.16	**	0	0
Craigvinean2015-2	0.48	0.12	0.18	0	0	0
Crychan2013	0.00	0.09	0.27	**	0	0
Haldon2015	0.50	0.24	0.45	0	0	0
LochArd2015	0.49	0.43	0.31	0	0	0
PeckettStone2011	0.43	0.18	0.38	0	0	0

Based on previous analyses, it was also expected that the agreement would be greater when selecting frame trees than when selecting trees to be removed from the stand. It seems plausible that test persons would agree more when selecting trees with certain, more or less well defined properties that remain in the stand until the end of the rotation. However, the results of this thesis made me re-consider this: I found consistently larger κ values for the low-thinning exercises followed by the selection of frame trees and the marking of frame-tree competitors in crown thinnings. This is an interesting and quite unexpected outcome of my study.

To interpret this result it is important to recall the British forestry context of this study. Thinnings from below have a very long tradition in British forestry, particularly in conifer plantations. Therefore it is very likely that British forestry staffs are most experienced in low thinnings and feel comfortable with this method. Abandoning this comfort zone when marking for crown thinnings apparently induces unease and confusion, particularly when doing this for the first time (as many test persons probably did).

As a consequence, the level of agreement is generally highest in low thinnings and lower in crown thinnings. However, my results also confirm that frame trees have an important didactic role to play, since the agreement in selecting frame trees is markedly higher than in selecting frame-tree competitors.

It may be of interest to mention the only case where the test persons agreed more when selecting frame-tree competitors as opposed to when marking frame trees. This is the case for Loch Ard in 2015. Dr. Jens Haufe, the person in charge of these marking exercises, informed me that this particular group mainly consisted of forest contractors and that at least 70% of them were machine operators. As part of their professional duties they are usually not in charge of selecting frame trees. Therefore this task must have been a total novelty for them and they may have focused their attention almost entirely on selecting larger trees to be removed (crown thinning) whilst not taking the frame tree selection seriously.

The second measure of agreement has put emphasis on assessing the variability in the number of trees selected in each exercise. The variability has been measured comparing the values of χ test. Smaller χ values indicate higher levels of agreement. In this sense the test χ has proven to provide results comparable to κ. Thus I have obtained the lowest values of χ for selecting frame trees and for marking crown thinnings. In addition, both cases represent the overall highest values of κ. Therefore both measures of agreement give consistent results. For instance, at Crychan in 2010 (Fig. 3 and 5) the κ value calculated from the crown thinning selection is the lowest of all exercises whereas the χ value is the highest.

With the exception of the study published by Pommerening et al. (2015), this thesis is the first attempt of quantifying variability in human tree selection by using Fleiss' kappa. Whilst working on this thesis I came across papers that critically discuss Fleiss' kappa (see for example Viera & Garrett 2005). Though Fleiss' kappa is a very popular measure of agreement, particularly in medicine, there are clear limitations. It seems that there is a need to carry out more basic research for identifying measures that are better suited for assessing the level of agreement.

The Pearson correlation coefficients have not revealed any clear relationships between the agreement measures and the structure of the forest. As one might expect, some of the measures of forest structure show large correlation values. This is due to the fact that the diameter measurements are included in all of their calculations.
There were only two weak correlations between κ related to the crown thinnings and the skewness and coefficient of variation of the diameter distribution. With the exception of one exercise, the skewness is always positively correlated with κ or almost zero. Although the correlation values are rather low, these findings offer the possibility of continuing research in terms of relating agreement to forest structure. Further research should also include sites with more diverse forest structures.
One limitation of this thesis has been the lack of information regarding the test persons. Given legal requirements in the UK, data regarding gender, age, profession and residence among other factors were not in the public domain. Future research should try to include this information in order to look for possible influences on the agreement results.

Acknowledgements

I am indebted to the supervisor of my project Prof. Dr. Arne Pommerening for his advice, constructive suggestions and repeated manuscript revisions which have contributed to this dissertation. As part of this project he also helped me to secure a short internship at Córdoba University (Spain) which allowed me to carry out my very own marteloscope and thus to make valuable practical experience in this field. Prof. Pommerening also involved me in the "Forest Facts" research information entitled "Towards understanding human tree selection behaviour". My gratitude also goes to Prof. Dr. Dietrich Stoyan (TU Bergakademie Freiburg, Germany) for his input regarding the Fleiss' kappa characteristic and the theoretical marking models. Dr. Jens Haufe (Ae Forestry Commission Training Centre, UK) organised the marking exercises in the UK, collected the data and kindly made them available to this study. He also provided additional information on the study sites and the test persons. I am also grateful to Dr. Jaime Uría-Díez and Xin Zhao (both Swedish University of Agricultural Sciences) for their help with the statistical calculations.

References

Agnoletti, M. and Anderson, S., 2000. Forest History: International Studies on Socio-economic and Forest Ecosystem Change. CAB International, Wallingford, UK.

Agnoletti, M., 2006. Man, forestry, and forest landscapes. Trends and perspectives in the evolution of forestry and woodland history research. *Schweizerische Zeitschrift für Forstwesen* 157, 384–392.

Becking J.H., 1953. Einige Gesichespunkte für die Durchführung von Durchforstungsversuchen in gleichaltrigen Beständen. *IUFRO-Congress,* Roma, 580-582.

Bruciamacchie, M., Pierrat, J.-C. and Tomasini, J., 2005. Modèles explicatif et marginal de la stratégie de martelage d'une parcelle irrégulière. [Explicative and marginal models for a marking strategy of an unevenaged stand.] *Annals of Forest Science* 62, 727-736.

Clark, J.S., Wolosin, M., Dietze, M., Ibáñez I., Shannon, L., Welsh, M. and Kloeppel B., 2007. Tree growth inference and prediction from diameter censuses and ring widths. *Ecological Applications* 177, 1942–1953.

Crecente-Campo, F., Pommerening, A. and Rodríguez-Soalleiro, R., 2009. Impacts of thinning on structure, growth and risk of crown fire in a Pinus sylvestris L. plantation in northern Spain. *Forest Ecology and Management* 257, 1945–1954.

Daume, S., Füldner and Gadow, K. v., 1997. Zur Modellierung personenspezifischer Durchforstungen in ungleichaltigen Mischbeständen. [Modelling person-specific thinnings in uneven-aged mixed stands.] *Allgemeine Forst- und Jagdzeitung* 169, 21-26.

Fleiss, J., Levin, B. & Cho Paik, M., 2003. *Statistical Methods for Rates and Proportions*, John Wiley & Sons. New York.

Füldner, K., Sattler, S., Zucchini, W. and Gadow, K. v., 1996. Modellierung personenabhängiger Auswahlwahrscheinlichkeiten bei der Durchforstung. [Modelling person-specific tree selection probabilities in a thinning.] *Allgemeine Forst- und Jagdzeitung* 167, 159-162.

Gadow, K. v., 1996. Modelling growth in managed forests — realism and limits of lumping. The Science of the Total Environment 183 167–177.

Graham, R.T. et al., 1999. The effects of thinning and similar stand treatments on fire behavior in western forests. *General Technical Reports of the US Department of Agriculture, Forest Service*, (September), 1–25.

Hart, H.M., 1928: *Stamtal en Dunning.* Thesis. The Netherlands.

Hedderich, J. and Sachs, L., 2012. Angewandte Statistik: Methodensammlung mit R. Heidelberg, Springer.

Helms, J.A., 1998. The dictionary of forestry. Society of American Foresters, Bethesda, MD.

Kariuki, M., 2008. Modelling the impacts of various thinning intensities on tree growth and survival in a mixed species eucalypt forest in central Gippsland, Victoria, Australia. *Forest Ecology and Management* 256, 2007–2017.

Kirby, K.J. and Watkins, C., 2015. *Europe's changing woods and forests: From*

wildwood to managed landscapes. CABI, Wallingford, UK.
Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
Mäkinen, H. and Isomäki, A., 2004a. Thinning intensity and growth of Norway spruce stands in Finland. *Forestry* 77, 349–364.
Mäkinen, H. and Isomäki, A., 2004b. Thinning intensity and long-term changes in increment and stem form of Scots pine trees. *Forest Ecology and Management* 203, 21–34.
Murray, D.M. and von Gadow, K., 1991. Relationships between the diameter distributions before and after thinning. *Forest Science* 372, 552–559.
Nilsen, P. and Strand, L.T., 2008. Thinning intensity effects on carbon and nitrogen stores and fluxes in a Norway spruce (Picea abies (L.) Karst.) stand after 33 years. *Forest Ecology and Management* 256, 201–208.
Peck, J.E., Zenner, E.K., Brang, P. and Zingg, A., 2014. Tree size distribution and abundance explain structural complexity differentially within stands of even-aged and uneven-aged structure types. *European Journal of Forest Research* 133, 335–346.
Peltola, H., Miina, J., Rouvinen, I. and Kellomäki, S., 2002. Effect of early thinning on the diameter growth distribution along the stem of Scots pine. *Silva Fennica* 364, 813–825.
Pommerening, A., Vítková, L., Zhao, X. and Pallarés, C., 2015. Towards understanding human tree selection behaviour. *Forest Facts* 9, SLU.
Poore, A., 2011. The marteloscope. *Woodland Heritage*, 28–29.
Porkess, R., 2004. Collins dictionary statistics. Collins dictionary series, second ed. Harper Collins.
Pyatt, G., Ray, D. and Fletcher, J., 2001. An ecological site classification for forestry in Great Britain. Forestry Commission Bulletin 124.
R Core Team, 2014. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL http://www.R-project.org/.
Reineke, L.H., 1933. Perfecting a stand-density index for even-aged forests. *J. Agric. Res.* 46, 627–638.
Spinelli, R, Magagnotti, N., Pari, L. and Soucy, M., 2016. Comparing tree selection as performed by different professional figures. *Forest Science* 62, 213-219.
Sterba, H. and Zingg, A., 2006. Abstandsabhängige und abstandsunabhängige Bestandesstrukturbeschreibung. [Distance-dependent and distance independent description of stand structure.]. Allg. Forst-u. J.-Ztg. [German Journal of Forest Research] 177, 169–176.
Sun, H. and Shao, R., 2009. Herfindahl index applied to Fourier analysis. *Digital Signal Processing: A Review Journal* 194, 726–730.
Swift, E., Duchesne, I., Ung, C.H., Wang, X. and Gagné, R., 2013. Impacts of partial harvesting on stand dynamics and tree grades for Northern hardwoods of the Acadian forest region, NRCan, Canadian Forest Service - Canadian Wood Fibre Centre.

Vanclay, J.K., 1989. Modelling Selection Harvesting in Tropical Rain Forests. *Journal of Tropical Forest Science* 1, 280–294.

Viera, A.J. & Garrett, J.M., 2005. Understanding interobserver agreement: The kappa statistic. *Family Medicine* 375, 360–363.

Vítková, L., Ní Dhubháin, A. and Pommerening, A., 2016. Agreement in tree marking: What is the uncertainty of human tree selection in selective forest management? Forest Science. In press.

Wyckoff, P.H. and Clark, J.S., 2005. Tree growth prediction using size and exposed crown area. *Canadian Journal of Forest Research* 351, 13–20.

Zhang, C., Zhao, X. and Gadow, K. v., 2014. Analyzing selective harvest events in three large forest observational studies in North Eastern China. *Forest Ecology and Management* 316, 100–109.

Zucchini, W. and Gadow, K. v. 1995. Two indices of agreement among foresters selecting trees for thinning. *Forest and Landscape Research* 1, 199-206.

Appendix 1

Marking sheet for crown thinnings								
No.	Frame?	Remove?	No.	Frame?	Remove?	No.	Frame?	Remove?
1			**46**			**91**		
2			**47**			**92**		
3			**48**			**93**		
4			**49**			**94**		
5			**50**			**95**		
6			**51**			**96**		
7			**52**			**97**		
8			**53**			**98**		
9			**54**			**99**		
10			**55**			**100**		
11			**56**			**101**		
12			**57**			**102**		
13			**58**			**103**		
14			**59**			**104**		
15			**60**			**105**		
16			**61**			**106**		
17			**62**			**107**		
18			**63**			**108**		
19			**64**			**109**		
20			**65**			**110**		
21			**66**			**111**		
22			**67**			**112**		
23			**68**			**113**		
24			**69**			**114**		
25			**70**			**115**		
26			**71**			**116**		
27			**72**			**117**		
28			**73**			**118**		
29			**74**			**119**		
30			**75**			**120**		
31			**76**			**121**		
32			**77**			**122**		
33			**78**			**123**		
34			**79**			**124**		
35			**80**			**125**		
36			**81**			**126**		
37			**82**			**127**		
38			**83**			**128**		
39			**84**			**129**		
40			**85**			**130**		
41			**86**			**131**		
42			**87**			**132**		
43			**88**			**133**		
44			**89**			**134**		
45			**90**			**135**		

Appendix 2

This appendix includes a collection of pictures taken during my stay in Spain as part of a short mission of the COST Action FP1206 entitled "EuMIXFOR" (European Network of Mixed Forests) group. My research proposal for installing and implementing a marteloscope plot at the Natural Park of Sierras de Cazorla, Segura y las Villas (Jaén, Spain) was approved and I succesfully conducted this mission from 4th until the 14th of March 2016.

Picture 1. Landscape view from the marteloscope location. Photographer: Carlos Pallarés Ramos.

Picture 2. Details of species composition and structure inside the martelocope. Photographer: Carlos Pallarés Ramos.

Picture 3. Setting the limits of the marteloscope using yellow spray. Photographer: Carlos Pallarés Ramos.

Picture 4. Temporary tag used for numbering a trees. Photographer: Carlos Pallarés Ramos.

Picture 5. Temporary number tag attached to an evergreen oak (Quercus ilex subsp. *ilex) stem. Photographer: Carlos Pallarés Ramos.*

Picture 6. The author measuring a tree location with the field map device. Photographer: Pedro José Pérez Moreno.

Picture 7. The author recording a tree location using the scope of the field map device. Photographer: Pedro José Pérez Moreno.

Picture 8. The author holding pole and reflector next to a tree for recording its location. In the foreground: The scope and compass of the field map device. Photographer: Pedro José Pérez Moreno.

Picture 9. Map showing the recorded tree locations and their numbers on the field map screen. Photographer: Carlos Pallarés Ramos.

Picture 10. The author assessing the stem quality of a Corsican pine (Pinus nigra subsp. *Salmannii). Photographer: Pedro José Pérez Moreno.*

Results from the Swedish University of Agricultural Sciences Nr 9 2015

SLU

FOREST FACTS

Students of Göttingen University selecting trees in a marteloscope near Reinhausen. Photo: Arne Pommerening

Towards Understanding Human Tree Selection Behaviour

Arne Pommerening, Lucie Vítková, Xin Zhao and Carlos Pallarés Ramos

Understanding how humans select trees for professional and non-professional purposes is crucial to sustainable forest management.

Since the mid 1990s data collection methods and experiments have been designed to measure human tree selection behaviour.

This research fills the gap between natural sciences and psychology; it also constitutes citizen science at the same time.

First results indicate that agreement between different test persons is comparatively low.

Humans tend to behave conservatively when asked to select trees differently and disagreement increases in such situations.

Our research has brought to light that many people have wrong perceptions about their tree selection styles and that their marking behaviour can differ markedly when selecting the same trees repeatedly.

Research into human tree selection behaviour has great potential for novel, interdisciplinary studies.

In the preface of his textbook "Planning in a Forest Enterprise" from 1972 Gerhard Speidel wrote that the environment we currently live in owes much to the decisions of human beings. He concluded that "decision making is therefore among the most fascinating and most responsible activities in this world". Along the same lines Gadow (1996) asserted that "the modifications of forest structure caused by management often have a far greater effect on forest development than natural growth".

In the past, it was often assumed that humans marking trees for thinnings or as habitat trees, do this more or less precisely according to textbook opinions, forest plans or other instructions. However, the selection of supposedly "desirable", "undesirable" or indifferent trees for a given purpose, may that be timber production or conservation, is a challenging task for any human being and a serious limitation on the potential for Continuous Cover Forestry in some countries. Naturally, no matter how detailed instructions are, there is likely to be some variation between the decision making of different individuals. In addition to this variability between individuals it has also been noted that one and the same experienced person can mark trees in the same forest differently on separate occasions. As any human behaviour, also tree marking behaviour depends on weather, general mood and starting point in the forest to name but a few factors. This second variability component can be referred to as variability within individuals.

Idealised textbook forest management is for example often assumed in tree growth models that allow the forest manager to project the current forest state into the future. Such models are an essential tool for identifying the best course of action in long-term sustainable forest management. However, how many of these forecasts are really helpful if they do not account for the uncertainty introduced by human decision making?

There are a number of fundamental research questions associated with human tree selection behaviour. First, it is useful to establish how much agreement there is actually among test persons. After quantifying the general agreement, clusters of similar behaviour and outliers must be identified, in order to uncover those individuals that made the largest contribution to the lack of agreement. Covariates such as tree size, timber quality, habitat value but also personal background information can then help to explain individual behaviour. Finally, the question must be addressed if the lack of agreement matters or whether there is a sufficiently common pattern in the selection behaviour that is compatible with the corresponding forest management objectives.

Specific research questions include

* Can different tree selection behaviour be attributed to, for example, gender, age or occupation?
* Do forestry staff educated in different institutions select trees differently?
* Is there any geographic relationship?
* How do individuals respond to training and are willing to change their behaviour?
* What effect do tree species composition and structure have on tree selection?

Research in human tree selection behaviour has a large potential as it bridges natural sciences and psychology. In an ideal way it combines basic research with science that is highly relevant to the forestry sector at the same time.

Example results

There are many options for analysing data of human tree selection behaviour. In this section we briefly illustrate but a few possibilities. Marteloscope-based research involving human choices coded as binary data is new to forest science. Such data, however, are not uncommon in social choice theory, particularly in one of its applications, approval voting. In approval voting, voters approve of a certain number of candidates. In human tree selection research, the test persons are "voters" and the trees are "candidates". A tree selected may mean "approved". In contrast to elections, the number of voters is clearly smaller than the number of candidates (Stoyan and Pommerening 2015).

Origins

Research in forest growth and yield focused on ideal textbook treatments and delivered the first fundamental results in the 1950s. In the mid 1990s, Prof. Klaus von Gadow initiated a research group at Göttingen University in Germany investigating the tree selection behaviour of forest managers and machine operators in different forest ecosystems. The group designed a special survey method for data gathering which went by the names of *"thinning-event inventory"* and *"harvesting-event inventory"*. The key idea of this survey method was to schedule the data gathering at the time of tree marking prior to the actual tree removal. In contrast to traditional forest inventory methods, the thinning-event inventory captured both the initial forest stand conditions and the residual stand. The changes in forest structure could then be analysed and decisions could be revised if necessary (see Figure 1).

Towards the end of the 1990s, a group at AgroParisTech-ENGREF at Nancy (France) around Max Bruciamacchie realised the potential of this research idea for training and education in silviculture. The group at Nancy offered field-based training courses to forestry professionals and students. In these courses, the participants were then asked to mark trees for thinnings on a sheet of paper similar to a questionnaire. Their choices were analysed using specialised software or MS Excel and personalised result and feedback sheets were handed out to every participant at the end of the training session. In order to attach a new brand to this type of research plot the group at Nancy coined the name *marteloscope* (from French *martelage* – marking) for these training sites.

In the first decade of this century, the first author and his Tyfiant Coed project team at Bangor University used the marteloscope idea in a training project in Wales (UK).

The use of marteloscopes in training and education has since then spread from France to Switzerland, Britain and Ireland. Poore (2011) described the application of marteloscopes for education and training. The popularity of marteloscopes is on the increase, however, its original research purpose has only been pursued at few institutions including the Swedish University of Agricultural Sciences.

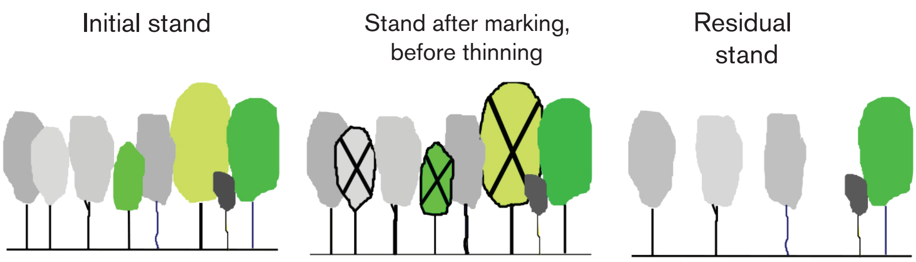

Figure 1. The principle of the thinning-event analysis: The initial stand, the stand after marking but before thinning and the residual stand are sampled and analysed at the same time (original drawing by Klaus von Gadow).

Simple statistics include bar charts showing the ordered proportions of trees marked by the test persons and the marking frequency of the trees (see Figure 3 for the Welsh marteloscope at Coed y Brenin in a mixed broadleaved-conifer woodland). Despite instructions the marking proportions range from 10 % to 48 % and test person #1 can be considered as an outlier, i.e. s/he shows a somehow extreme behaviour of selecting many trees. In class 0 (Figure 3, right), 30 trees have never been marked by any test person. Although this is seemingly a complete agreement, it is kind of pseudo or passive, since the agreement is in discarding these trees. Classes 14 and 15 are empty, i.e. the maximum score a tree has received

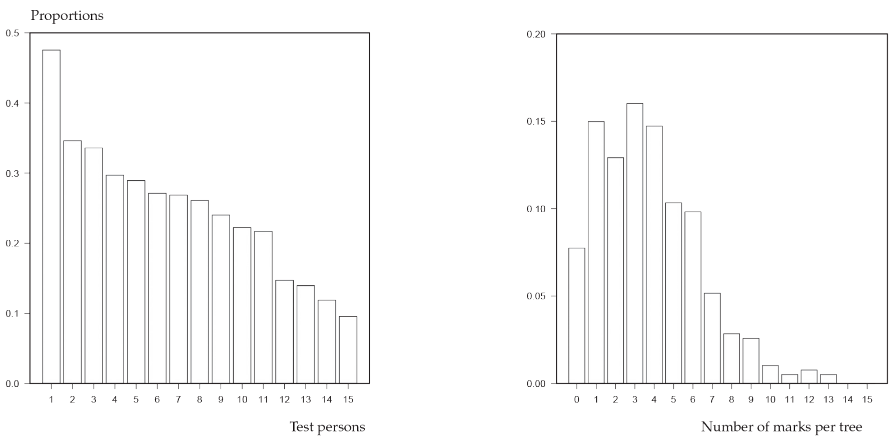

Figure 3. Bar charts of the marking activities of fifteen test persons in the Coed y Brenin marteloscope (left). The marking frequencies of the corresponding trees (right).

Marteloscope experiments

A marteloscope is typically set up as a research plot with rectangular boundaries. 100 × 100 m is often a suitable size, where the number of trees to be included should ideally be between 150 and 500 trees. A marteloscope of this size takes a test person approximately three hours to complete. The area should be sufficiently large so that the test persons do not influence each other's decision making. It can also be recommended to select a site where thinnings have not occurred during the last ten years, so that the thinning urgency is high. Every tree has a unique number, which is painted on the stem surface with waterproof paint as clearly as possible and should be visible from afar. If feasible, one could consider painting tree numbers twice on each tree from opposite directions for better identification. Minimum tree measurements should include stem diameters (min 5–7 cm). Ideally the surveying should also include tree locations, total heights (at least on a sample basis), volume/biomass, habitat value and timber quality. Marteloscopes are not different from research plots commonly used in silviculture and growth & yield science. In fact, plots from the latter two subject areas can often be re-used as marteloscopes. What is different are the research objectives and the analysis.

Figure 2 (left) shows a map of the Clocaenog marteloscope in North Wales (UK) including the marks of one test person from 2006: An extraction rack runs through the centre of this Sitka-spruce marteloscope. Black colour highlights frame trees[1], grey the trees to be thinned and the unmarked trees are white. The corresponding stacked empirical diameter distribution is also shown (Figure 2, right). Obviously both frame trees and trees to be thinned were mostly selected from larger diameter classes resulting in a crown thinning.

Table 1 is an example of a typical marking sheet. The design of the marking sheet may vary depending on the purpose of the training session or experiment. For example, it is possible to leave the tree-number column blank and to ask the test persons just to note those trees that they intend to select. Species and stem-diameter information can also be omitted if this serves the experimental purpose.

The design of the marking sheet surely has an influence on the marking behaviour. It is therefore advisable to consider the design carefully and to be creative in order to improve the outcome of the experiment.

Usually basic instructions are given to all 15–30 test persons along with a brief qualitative and quantitative description of the forest stand. The details of these instructions depend on the purpose of the experiment and can even be completely omitted, when testing the test person's intuitive management skills.

It is also good practice to record the test person's names, gender, work affiliation, professional and geographic backgrounds. Any additional information can potentially turn out as useful covariates or for post-stratification and aid the interpretation of the results.

In the data processing, the marking sheets are digitalised. A cross or tick indicating tree selection is converted to a '1'. No selection results in '0', so that a typical test person data column consists of a continuous sequence of 0's and 1's.

1) Frame trees (also referred to as future crop trees) are trees that are selected for their specific characteristics (e.g. economic or conservation value) and retained for a long time until final harvesting (or even beyond). At the time of final harvesting only the frame trees are left in the forest.

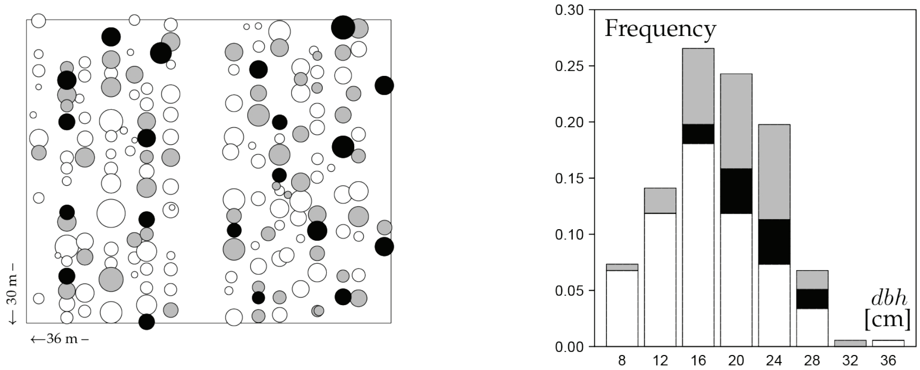

Figure 2. Map and proportions of frame and competitor trees marked by one test person in the Clocaenog marteloscope experiment in North Wales in 2006. Details are provided in the text.

Table 1. Design of a typical, basic tree marking sheet for use in marteloscope research. *DBH* is diameter at breast height and measured in centimetres. "Frame" is short for frame tree.

Tree#	Species	*DBH*	Frame	Thin
1	Birch	55.4		X
2	Pine	60.6	X	
3	Pine	61.5		X
4	Birch	33.5		
5	Birch	42.1		
6	Pine	52.3		
7	Birch	15.6		
8	Birch	57.2		
9	Pine	64.3		X
10	Birch	24.2	X	

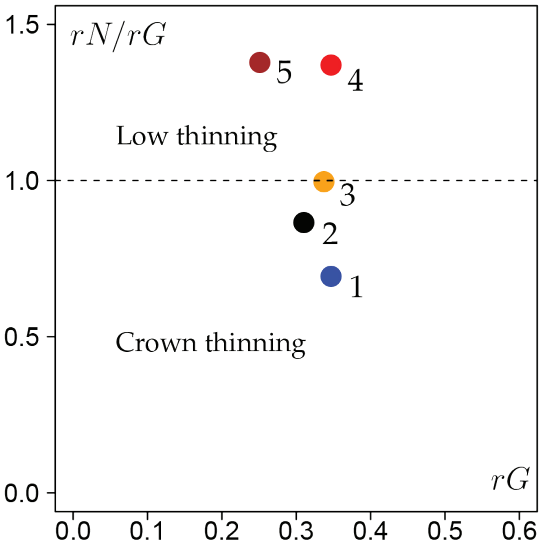

Figure 4. The *NG* ratio of five test persons in the Coed y Brenin marteloscope.

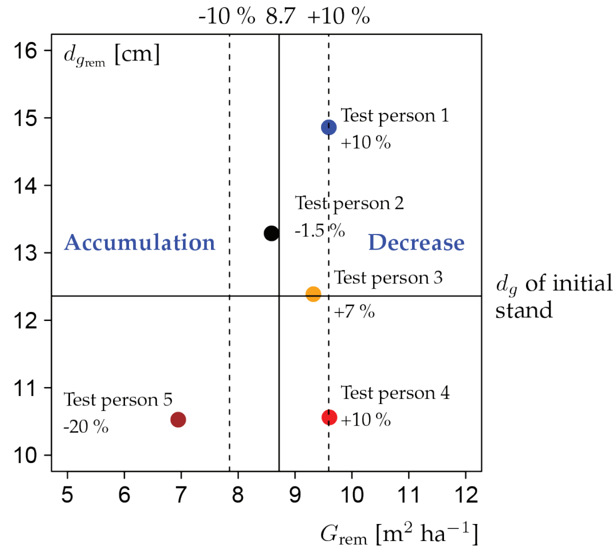

Figure 5. Quadratic mean diameter over basal area of trees marked for thinning by five test persons in the marteloscope experiment at Coed y Brenin. The stand growth rate of 8.7 m^2 ha^{-1} was recorded for the period 2006–2011.

was 13 and this implies that two trees were marked by 13 test persons out of 15. The distribution suggested by the bar chart is reminiscent of a binomial distribution and the shape indicates little agreement.

Fleiss et al. (2003) described measures of agreement used for ratings on categorical scales in psychological and medical research. Among them is a measure κ (kappa) for multiple test persons.

$\kappa = 1$ indicates complete agreement, values close to zero, on the other hand, are almost chance agreements. In contrast to many applications in medicine and psychology, kappa applied to marteloscope experiments usually scores comparatively low indicating poor agreement. For example, in the mixed broadleaved-conifer plantation at Coed y Brenin in Wales, planted in 1985, including 387 trees and 15 test persons, κ was 0.102. Slightly better results were found in an Irish marteloscope including 131 trees and 24 test persons in a pure Sitka spruce plantation with $\kappa = 0.194$. When applied to the selection of frame trees in the same marteloscope κ was 0.310. Apparently it is easier to agree on final crop trees that remain in the forest until final harvesting than on the competitors of these frame trees. The concept of identifying potentially harmful trees for eviction seems to be a highly abstract and difficult one for the human mind. This may explain why the overall results are poor. In recent research, we also found that κ was on average around 0.40 for thinnings from below at nine different marteloscopes across Great Britain and around 0.15 for crown thinnings on the same sites using the same test persons. Thinnings from below have so far been usual practice in Great Britain and the recent introduction of crown thinnings as part of Continuous Cover Forestry possibly inspired confusion in tree selection behaviour.

When moving on to analyse the behaviour of individual test persons, we can for example quantify the thinning type the different test persons have applied. This is very useful, since thinning types, e.g. crown thinning and thinning from below[2], in silviculture have a strong effect on forest development and are often handled as qualitative concepts. In many marteloscope training sessions it even turned out that some test persons were convinced of marking for a crown thinning, however, the statistics, much to their surprise, indicated a thinning from below: The perception of what humans sometimes think they do can be different from what they actually do. A suitable indicator of thinning type is the *NG* ratio (Kassier 1993), defined as the relative number of trees removed divided by the relative basal area removed.

If $NG = 1$, the mean size of the trees marked is near the mean basal-area tree, d_g. This can often be the result of a natural disturbance such as windthrow or snow damage. If $NG < 1$, proportionally less trees were removed than basal area. This typically indicates a crown thinning. Finally, proportionally more trees were removed than basal area if $NG > 1$ indicating a thinning from below. The more an observed *NG* ratio differs from 1, the clearer the thinning-type trend. Figure 4 shows the distribution of the *NG* ratio

"Research in human tree selection behaviour is an inspiring, novel research direction."

[2]) A *crown thinning*, also referred to as *high thinning* or *thinning from above* involves the removal of big, dominant trees to favour trees of roughly the same size. A *thinning from below* or *low thinning* implies the removal of smaller and less dominant trees than those that are favoured and usually dominate the main canopy (after Helms 1998).

of five test persons in the Coed y Brenin marteloscope. We can clearly see that the thinning intensity (measured on the abscissa in relative basal area, *rG*) is almost similar (with the notable exception of test person 5), however, two test persons have marked for a thinning from below, one person for an indifferent and two persons for a crown thinning, although all test persons were briefed to aim at a crown thinning. Using the *NG* ratio as part of a marteloscope experiment in Tikincor Forest (County Tipperary, Ireland) indicated that inexperienced test persons were able to learn and implement new forest management techniques more readily than experts. Interestingly, in the same study, the agreement among experts was larger before rather than after training (Vítková et al. 2015). This was not a totally unexpected outcome, but the finding is crucial for the future design of education and training in forestry.

If growth rates can be calculated from past survey records, it can also be useful to check how much the tree volume or basal area suggested by the test persons in their markings coincides with the increment of the last five or ten years. This provides crucial information about the sustainability of a suggested intervention. Figure 5 depicts the same five test persons as in Figure 4. All test persons above the horizontal line representing the initial quadratic mean diameter performed a tree marking corresponding to a crown thinning. The two participants 4 and 5 below this line made decisions leading to a thinning from below.

The vertical solid line marks the stand increment of the last five years and is accompanied by two dashed vertical lines that give a region of allowance of ±10 %. Assuming that no additional thinning is carried out within the next five years other than the one defined by the test persons, basal-area values smaller than 8.7 m^2 lead to an accumulation of basal area. Values larger than 8.7 m^2 result in a decrease of stand basal area. The marking of test person 2, a university student without forestry education, is closest to the observed stand increment whilst test person 5, a professional forester, has marked trees so that basal area will further accumulate. All others carried out a marking that decreases stand basal area, which makes sense in this very dense forest stand. Interestingly, test persons 1 and 4 both mark trees to amount to 10 % more than the basal area increment of the last five years, however, one of them carries out a crown thinning and the other a thinning from below.

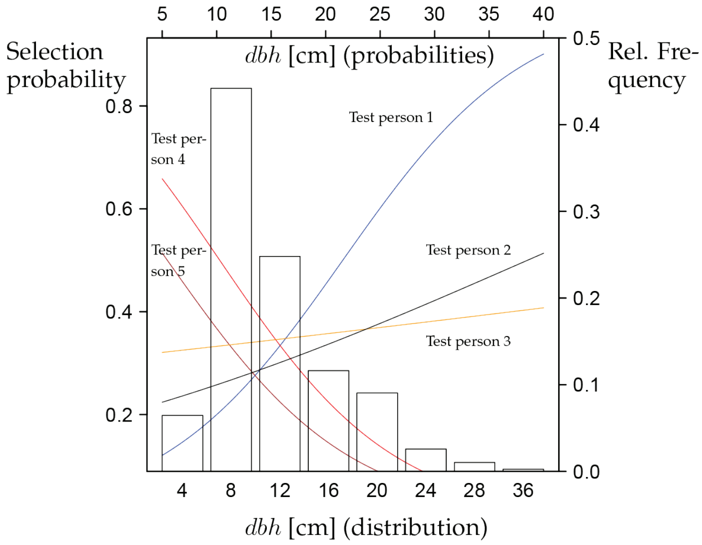

Figure 6. Tree selection probabilities of five test persons participating in a marteloscope experiment at Coed y Brenin and their dependence on *dbh*.

Finally, we can use covariates to learn about factors that may have influenced human decision making. In analysing them we can draw on experience and knowledge from mortality and survival analysis. The technique used to analyse tree selection probabilities is that of logistic regression with binary response. The probability of selection success can now be related to a set of linear predictors. In the simplest case we can start with stem diameter and later involve other covariates to see which of them had the greatest influence on the marking decisions of a particular person.

Figure 6 featuring again the same five test persons as in the previous figures clearly shows that test persons 1, 2, 4 and 5 have been influenced by stem diameter whilst test person 3 has not responded much to *dbh*. For this test person stem diameter has not been an important criterion. With test persons 1 and 2 the selection probability increases with increasing stem diameter. This suggests a tendency towards a crown thinning. Test persons 4 and 5 show a tendency reflecting thinning from below, as the tree selection probability increases with decreasing stem diameters. Interestingly, we also found that the person-specific parameters of the logistic regression model were strongly correlated with the respective *NG* ratios.

Implications and future directions

Research in human tree selection behaviour is an inspiring, novel research direction at the interface between natural sciences, social sciences and ecology. It uses a strictly interdisciplinary approach by combining statistical methods from outside forest science with forestry knowledge. At the same time, the research described in this fact sheet can also be considered as impact analysis contributing to transparency of decision making in forestry and thus to professional credibility. Research in human tree selection behaviour also provides useful information for modelling person-specific thinning and harvesting strategies as well as on their variability to be incorporated in tree and forest simulators. The analysis of human choices in forests helps to predict the consequences of interventions and keeps data up to date. Marteloscope-based research is also citizen science, where frequently non-scientists collect data and make them available to professional researchers for detailed analysis.

Marking exercises and training sessions promote forestry education, life-long learning and continuing professional development. Thus forestry staff can re-confirm skills and knowledge in regular sessions similar to hunters and stalkers who have to train the command of their weapons on a regular basis and acquire new skills. In this sense, marteloscopes are also tools

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

for – as Klaus von Gadow put it – preventive sustainability control and adaptive management. Finally, as part of open-day events at universities and colleges, marking exercises and experiments can be organised to inspire prospective students. In the same way other groups of people outside forestry with a very different professional background can carry out tree markings as team-building events. However, the topic of human tree selection is much bigger and includes the selection of habitat trees in nature conservation, the marking of trees by harvester drivers in forest operations, the virtual removal of trees in simulator software, the selection of trees by indigenous people as opposed to professionals as well as the selection of trees for burial in forest cemeteries or the selection of Christmas trees in a Christmas tree plantation. In all these activities, trees are selected for various different purposes and the selection process is affected by a wide range of different factors, which partly relate to the trees and the forest ecosystem which they are part of and partly to implicitly human factors. Anybody interested in this research or in the practical application of marteloscopes is welcome to contact the authors for advice.

A student from the Estonian University of Life Sciences selecting trees in a marteloscope at Järvselja Forest (Estonia). Photo: Arne Pommerening

Acknowledgements
This research note is dedicated to Klaus von Gadow, Professor in Forest Planning and Growth & Yield Science at Göttingen University (Germany), who initiated this research direction, on the occasion of his upcoming 75th birthday in 2016. This publication also supports the work of the Cost Action FP1206 "EuMixFor".

Keywords
Marteloscope, marking, forest management, agreement, logistic regression.

Read more:

- **Fleiss, J. L, Levin, B. & Paik, M. C. 2003.** Statistical methods for rates and proportions. 3rd edition. J. Wiley & Sons, Chichester.
- **Gadow, K. v. 1996.** Modelling growth in managed forests – realism and limits of lumping. The Science of the Total Environment 183, 167–177.
- **Helms, J. A. (ed.) 1998.** The dictionary of forestry. Society of American Foresters. CABI Publishing, Oxon.
- **Kassier, H. W. 1993.** Dynamics of diameter and height distributions in commercial timber plantations. PhD thesis University of Stellenbosch. Stellenbosch.
- **Poore, A. 2011.** The marteloscope – a training aid for continuous cover forest management. Woodland Heritage 2011, 28–29.
- **Stoyan, D. & Pommerening, A. 2015.** Set-theoretic analysis of voter behaviour in approval voting applied to forest management. In preparation.
- **Vítková, L., Ní Dhubháin, A. & Pommerening, A. 2015.** Competence in tree marking: How do field staff select trees for continuous cover forest management? Forest Science. Submitted.

Authors:

Arne Pommerening
Professor in Mathematical Statistics Applied to Forest Sciences, Department of Forest Resource Management, SLU, SE-901 83 Umeå, Sweden.
+46-90-786 82 47
arne.pommerening@slu.se

Lucie Vítková
Dr, European Forest Institute, Central European Regional Office (EFICENT), Wonnhaldestraße 4, D-79100 Freiburg, Germany.
+49-761-4018 142
lucie.vitkova@efi.int

Xin Zhao
PhD student, Department of Forest Resource Management, SLU, SE-901 83 Umeå, Sweden.
+46-90-786 86 16
xin.zhao@slu.se

Carlos Pallarés Ramos
MSc student, Department of Forest Resource Management, SLU, SE-901 83 Umeå, Sweden.
carlospa@student.uef.fi

FOREST FACTS – Results from the Swedish University of Agricultural Sciences
ISSN: 1400-7789. **Production:** SLU, Faculty of Forest Sciences 2015.
Responsible editor: Karin.Ljung@slu.se.
Editor: Goran.Sjoberg@slu.se.
Illustrator: Fredrik Saarkoppel, Kobolt Media AB.
Subscription: www.slu.se/faktaskog.