Working Papers of the Finnish Forest Research Institute 307

http://www.metla.fi/julkaisut/workingpapers/2014/mwp307.htm
ISBN 978-951-40-2495-5 (PDF)
ISSN 1795-150X

Accuracy of visual tree defoliation assessment: a case study in Finland

Petteri Muukkonen, Martti Lindgren and Seppo Nevalainen

METLA www.metla.fi

Working Papers of the Finnish Forest Research Institute publishes preliminary research results and conference proceedings.

The papers published in the series are not peer-reviewed.

The papers are published in pdf format on the Internet.

http://www.metla.fi/julkaisut/workingpapers/
ISSN 1795-150X

Office
Post Box 18
FI-01301 Vantaa, Finland
tel. +358 29 532 2111
e-mail julkaisutoimitus@metla.fi

Publisher
Finnish Forest Research Institute
Post Box 18
FI-01301 Vantaa, Finland
tel. +358 29 532 2111
e-mail info@metla.fi
http://www.metla.fi/

Authors
Petteri Muukkonen, Martti Lindgren & Seppo Nevalainen

Title
Accuracy of visual tree defoliation assessment: a case study in Finland

Year	Pages	ISBN	ISSN
2014	22	ISBN 978-951-40-2495-5 (PDF)	1795-150X

Regional Unit / Research programme / Projects
Vantaa Research Unit

Accepted by
Raisa Mäkipää, Senior researcher, 13th October 2014

Abstract
Defoliation (crown thinning) is widely used as a rapid method of tree condition assessment. As a method that is based on subjective visual observation it might be influenced by statistically significant observer bias. Significant observer bias has been discovere in some countries. We analyzed the significance of observer bias occurring in the Finland's forest condition monitoring system. We analysed the data of three training courses held to the field personnel (2006, 2007, 2008). Our results indicate that some inconsistencies occur between observers, but these are still not systematic in nature. In conclusion, the detected observer biases are independent incidents, caused mainly by the observer perception during the single events. Therefore there is no need to make any systematic corrections for Finnish national visual tree defoliation assessments. We suggest that the best way to improve field assessments is the proper education and guidance of field personnel.

Keywords
Crown condition, Forest condition, Forest health, Forest monitoring

Available at
http://www.metla.fi/julkaisut/workingpapers/2014/mwp307.htm

Replaces

Is replaced by

Contact information
Petteri Muukkonen, Finnish Forest Research Institute, P.O. Box 18, FI-01301 Vantaa, Finland. E-mail petteri.muukkonen@metla.fi

Other information

Contents

1. Introduction

Pan-European forest condition monitoring became an interesting and topical issue few decades ago when large scale decline in forest vitality, connected with the effects of air pollution, occurred in Europe (e.g. Salemaa et al., 1991; Redfern and Boswell, 2004). For this reason monitoring of forest condition in relation to effects of anthropogenic pollution has been performed in Europe since the mid-1980s. Since then, public interest in the subject has waned and it no longer draws the same level of political interest (Innes, 1993). Yet, nowadays the need to observe changes in forest biodiversity and carbon stock has expanded the thinking of forest condition monitoring (Moffat et al., 2008). One example of international harmonisation of forest condition monitoring is the European forest monitoring programme ICP Forests (the International Co-operative Programme on the Assessment and Monitoring of Air Pollution Effects on Forests) which was initiated in 1985 and which was established under the UN/ECE Convention on Long-Range Transboundary Air Pollution (CLRTAP) (Innes, 1993; Derome et al., 2007). Finland has participated in this programme from the launch.

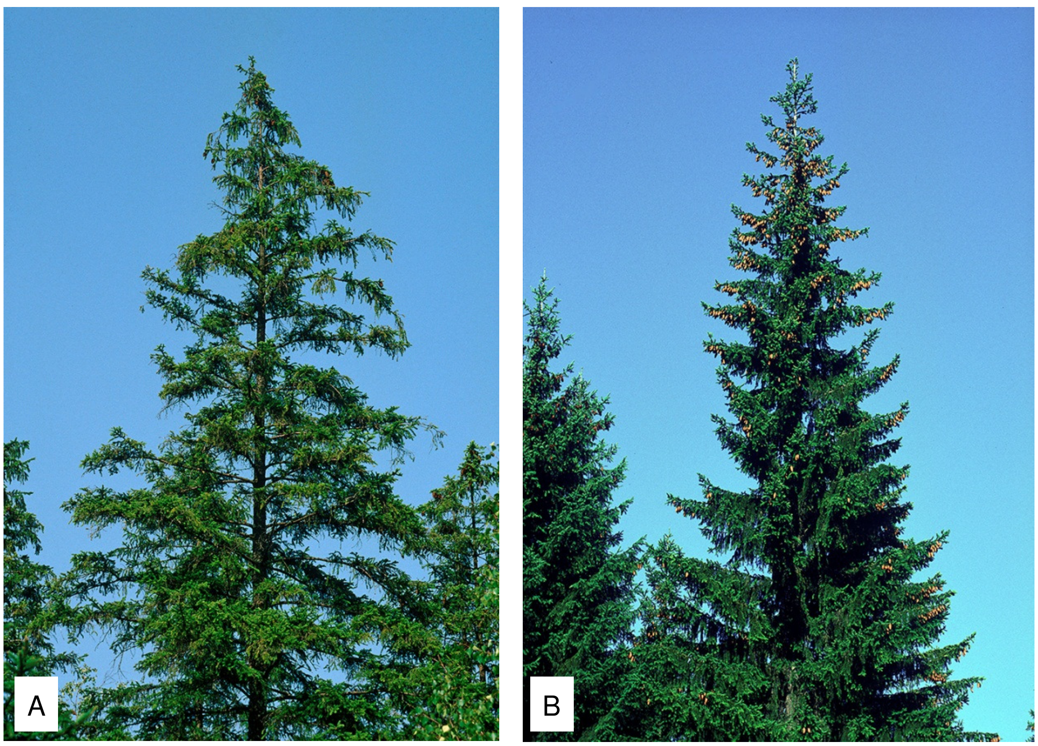

Figure 1. The defoliation is estimated as percent share of defoliated leaf- or needle-loss according to real or fictional non-defoliated reference tree at same age. The defoliated tree A has a defoliation degree of 61–70%, while the health tree B has a defoliation of 0–10%. (Photo. Erkki Oksanen, Finnish Forest Research Institute)

Figure 2. The defoliation of tree crown is typically surveyed by binoculars from the living crown (Photo. Antti Pouttu, Finnish Forest Research Institute)

To perform this monitoring the Finnish Forest Research Institute carries out the annual tree crown condition surveys on a national grid (Derome et al., 2007). This consists of internationally standardised methods, which is characterised by large number of trees assessed for a few parameters (Ghosh and Innes, 1995). The most important variable is leaf- and needle-loss (i.e. defoliation[1]) (Ghosh and Innes, 1995; Salemaa and Lindgren, 2000) because tree crown observations are typically the first signs indicating the natural or anthropogenic stresses (Zarnoch et al., 2004). Many site and damage factors reduce needle age, which can be seen as premature needle shedding and crown defoliation (Salemaa and Lindgren, 2000). The defoliation may be caused by aging of trees, properties of habitat site, climate and weather, outbreaks of pests or diseases or anthropogenic influence (Westman and Lesinski 1986; Salemaa et al., 1993; Metzger and Oren, 2001).

The value of defoliation is measured as percent share of defoliated leaf- or needle-loss according to real or fictional non-defoliated reference tree at same age (see Figure 1) (Salemaa et al., 1993). The defoliation of tree crown is typically surveyed by binoculars from the living crown (Figure 2). This method is widely used and internationally approved standard to survey forest condition and it is especially practical indicator for large scale monitoring (Strand, 1996).

[1] Strictly speaking, the term 'defoliation' is misleading because it is not meaning the actual loss of foliage, but rather the transparency of a tree in comparison to a fully foliated tree of the same species, the branching type and the same age growing under similar site conditions (Dobbertin et al., 2005).

Figure 3. Field personnel of the Finnish Forest Research Institute are training their skills to observe tree defoliation according to international guidelines. (Photo. Seppo Nevalainen, Finnish Forest Research Institute)

Though there are very strict international guidelines and intensive training programs for the assessment of tree defoliation (see Figure 3), a subjective component still exists (Gertner and Köhl, 1995; Salemaa and Lindgren, 2000). Visual assessment of crown condition and defoliation is susceptible to a number of different sources of error (Salemaa and Lindgren, 2000). Therefore, defoliation assessment is widely criticized (e.g. Innes, 1988a; Innes, 1988b; Innes et al., 1993; Salemaa et al., 1993; Gertner and Köhl, 1995; Ghosh et al., 1995; Metzger and Oren, 2001). It is also criticised due the fundamental weakness of being a nonspecific symptom of the change in tree condition (Innes, 1988a; Rehfuess, 1989; Innes, 1992).

Subjectivity in defoliation assessment has many causes. Sometimes it is difficult to separate the effect of phenotype, tree age and growing conditions from the effects of e.g. air pollutants (Salemaa et al., 1993). Again, sometimes the perception of observers (Innes, 1988b; Salemaa and Lindgren, 2000) and sometimes even weather and lighting conditions (Salemaa and Lindgren, 2000; Metzger and Oren, 2001) might cause observation bias. If such a bias occurs, it is often increased by weather condition, the visibility of the crown, tree species, tree age and social position (Wulff, 2002). It is also evident that crown dimensions affect to crown condition assessments; eg. crowns of trees with path-lengths more than 10 m are always likely to be rated less than 30% defoliated and thus considered healthy, although their crowns may be as unhealthy as those of tree with path-lengths less than 4 m and rated more than 80% defoliated (Metzger and Oren, 2001). These might be caused by the set-up of an imaginary reference tree (Solberg and Strand, 1999). The observers or observer teams might also have an individual style of assessments, which can be turn up e.g. as a reluctance of using the lowermost or uppermost parts of the scale or as a preference of using rounded scores (Solberg and Strand, 1999).

An observer bias is defined as the difference between true and observed tree defoliation (Gertner and Köhl, 1995; Wulff, 2002). Sampling error is another source of variation, but in our opinion, the observer error has a special importance in three reasons. Firstly, Finland's forest condition monitoring is currently a part of National Forest Inventory and its national systematic grid, which means that the sampling error is already well studied and quantified (see Tomppo et al., 2001). Secondly, observer bias can be larger than the components due the sampling error (Gertner and Köhl, 1995). Thirdly, observer bias can result an artificial impression of geographical patterns if observers are operating regionally (Gertner and Köhl, 1995; Strand, 1996). Herewith, it may result in inconsistent or even false reports about forest condition. Due the primary objective of tree crown condition monitoring is to provide information about changes in crown condition, it is extremely important to keep the assessment level of the individual observer constant (Salemaa and Lindgren, 2000).

Several previous national case studies have evaluated the observer bias and they have two opposite conclusions. Statistically significant differences between observers and observer teams have been found in most of studies (Innes, 1988b; Strand, 1996; Metzger and Oren, 2001; Wulff, 2002) while some of the reporters have not found any significant observer bias that in their national monitoring system (Salemaa et al., 1993; Redfern and Boswell, 2004). Yet, Strand (1996) concluded that his data did not provide conclusive evidence of the observer bias because the trees might have been assessed at different phenological state.

Because it is evident that the occurrence of significant observer bias is a case specific issue, it is essential to examine that in detail also in Finland, which is not done before extensively. In Finland, previously only Salemaa and Lindgren (2000) have studied the assessment error, but they compared only the defoliation assessment made by two expert observers and the supervising survey team. Therefore, the aim of this article was to study the difference between several individual expert observers in the defoliation assessment of the Finland's forest condition monitoring system. It is important to recognize and analyse the observer bias to achieve adequate reliability of forest condition monitoring. This is essential because it is not easy to determine confidence limits for defoliation assessments and therefore it is difficult to assess small changes in forest condition inventories (Innes, 1988b).

2. Material and Methods

2.1 Data

Our study is based on the data of annual training course held to field personnel of Finland's forest condition monitoring in 2006, 2007 and 2008. During the course the reliability of the defoliation levels between different observers was studied on the basis of visual assessment of the same trees independently (Lindgren, 2002). The test material of the course consists of Norway spruce (Picea abies), Scots pine (Pinus sylvestris) and birch (Betula pendula and B. pubescens) individuals that are not growing on actual sample plots of national survey grid (Lindgren, 2001). The data consists of 20 individual trees in each of eleven study site (for study sites see Table 1).

The defoliation of tree crown is typically surveyed by binoculars (see Figure 2) from the upper half of living crown (Norway spruce) or from the upper 2/3 part (Scots pine and birches) (Salemaa et al., 1993; Salemaa and Lindgren, 2000). This means that in first phase observers determined the lower limit of living crown base.

Defoliation is assessed according to normal foliage cover of a tree (Hanisch and Kilz, 1990). The reference normal tree can be either i) a real, non-defoliated tree of the same age, same type of crown and growing under similar conditions in the vicinity of the sample tree, or ii) an imaginary tree with a degree of defoliation of 0% (Salemaa and Lindgren, 2000). Observations of foliage should never be confined to individual needles and leaves or branches. Hanisch and Kilz (1990) have stated that, instead, they should cover stand as a whole, taking in the tree in its enirety, the sun and shade crowns and going right through to individual boughs and branches from different parts of crown. Defoliation is assessed in 10%-classes (Jukola-Sulonen et al., 1990; Salemaa et al., 1991).

Table 1. General description of the study sites.

	Tree species	ID	No. of observers			No. of trees		
			2008	2007	2006	2008	2007	2006
Monoculture forest	Betula pendula	BetPe	11	10	12	20	20	20
	Betula pubescens	BetPu	–	10	12	–	20	20
	Picea abies	PicAb 1	11	10	12	20	20	20
	Picea abies	PicAb 2	11	10	12	20	20	19
	Picea abies	PicAb 3	–	–	12	–	–	20
	Pinus sylvestris	PinSy 1	11	10	12	20	20	20
	Pinus sylvestris	PinSy 2	11	10	–	20	20	–
	Pinus sylvestris	PinSy 3	–	–	11	–	–	20
Mixed forest[2]	Betula spp.	BetMix	11	10	12	20	20	20
	Picea abies	PicMix	11	10	12	20	20	20
	Pinus sylvestris	PinMix	11	10	12	20	20	20

[2] Tree species in question growing in the mixed forest.

2.2 Statistical analyses

First we detected the differences in tree-wise defoliation scores across multiple observer attempts. We used the Friedman's test, which is a nonparametric statistical test for k-related samples developed by Friedman (1937; 1939; 1940). The computational formula for the Friedman test is

$$\chi_r^2 = \frac{12}{Nk(k+1)}\sum_{j=1}^{k} R_j^2 - 3N(k+1) \qquad \text{(Equation 1)}$$

where k is the number of ranked observers (columns), N is the number of trees (rows), and Rj is the sum of the ranked scores in each column. Under the null hypotheses (H0:) the independent variable, the individual observer, is assumed to have no effect on the dependent variable, scores of defoliation, the scores from different observers come from the same population (H0: Rj = R'j = R''j ...). Thus, the alternate hypothesis (H1:) is that at least one set (observer) of scores is not from the same population. The Friedman's test was performed with the SPSS 16.0 package.

Secondly, when the null hypothesis (H0:) was rejected, we tested bivariate nonparametric post hoc analysis with individual observers as testing units given by

$$\left|R_j - R'_j\right| \geq z\sqrt{\frac{Nk(k+1)}{6}} \qquad \text{(Equation 2)}$$

where Rj and R'j are the sums of rank sums being compared, and z is the z score from the standard normal curve corresponding to p/[k(k–1)] (Sheldon et al., 1996).

3. Results

In the year 2006 training course of field personnel of visual tree defoliation assessment, there occur statistically significant ($p < 0.05$) observer differences in all study sites and in all tree species based on the Friedman's test statistics (Table 2). Yet, when comparing results of bivariate post hoc analysis, there does not exists such a pair-wise difference in the case of one pure Scots pine stand (PinSy 1) and one mixed growing Scots pine stand (PinMix) (Table 3). On the whole, in the all other study sites, the visual defoliation observations between different observers are rather congruent. However, there still exist some dissenting judgments between few observers. In the Norway spruce study site PicAb 1, observer coded as letter B is divergent according almost all other observers. Also observer J has quite often different opinion in that same study site PicAb 1. In addition, observer J has several occasional disagreements with other observers, but those do not compose any clear pattern. In the Scots pine study PinSy 3 and in the mixed growing Norway spruce stand PicMix, the observer D has repeatedly different judgements than others.

Table 2. The Friedman's test statistics of dissimilarity of observers' decisions (H_0: $R_j = R'_j = R''_j$ ...). Significant *p*-values (< 0.05) are set in boldface.

			2008		2007		2006	
	Tree species	ID	χ_r^2	*p*-value	χ_r^2	*p*-value	χ_r^2	*p*-value
Monoculture forest	Betula pendula	BetPe	83.18	**<0.001**	–	**0.001**	39.64	**<0.001**
	Betula pubescens	BetPu	–	–	48.436	**<0.001**	73.20	**<0.001**
	Picea abies	PicAb 1	66.58	**<0.001**	28.72	**0.001**	91.21	**<0.001**
	Picea abies	PicAb 2	66.96	**<0.001**	42.11	**<0.001**	31.19	**<0.001**
	Picea abies	PicAb 3	–	–	–	–	38.60	**<0.001**
	Pinus sylvestris	PinSy 1	30.52	**0.001**	14.99	0.091	19.47	**<0.001**
	Pinus sylvestris	PinSy 2	45.03	**<0.001**	17.73	**0.038**	–	–
	Pinus sylvestris	PinSy 3	–	–	–	–	63.49	**<0.001**
Mixed forest[3]	Betula spp.	BetMix	47.01	**<0.001**	14.48	0.106	57.95	**<0.001**
	Picea abies	PicMix	45.07	**<0.001**	43.97	**<0.001**	68.77	**<0.001**
	Pinus sylvestris	PinMix	41.79	**<0.001**	31.38	**<0.001**	36.37	**<0.001**

[3] Tree species in question growing in the mixed forest.

Table 3. Matrix of nonparametric *post hoc* analysis for the data of year 2006. Coefficients of difference $\left|R_j - R'_j\right|$ and significances (*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, – = not significant) were derived from bivariate procedure (see Equation 2) between observers (A–L). Those study sites without any statistically significant differences are not shown.

		A	B	C	D	E	F	G	H	I	J	K	L
BetPe	A		–	–	*	–	–	–	–	–	–	–	–
	B	9.0		–	*	–	–	–	–	–	–	–	–
	C	16.5	7.5		**	–	–	–	–	–	–	–	–
	D	**77.0**	**86.0**	**93.5**		–	–	*	–	–	**	–	**
	E	2.5	11.5	19.0	74.5		–	–	–	–	–	–	–
	F	25.5	34.5	42.0	51.5	23.0		–	–	–	–	–	–
	G	0.5	8.5	16.0	**77.5**	3.0	26.0		–	–	–	–	–
	H	27.0	36.0	43.5	50.0	24.5	1.5	27.5		–	–	–	–
	I	26.5	35.5	43.0	50.5	24.0	1.0	27.0	0.5		–	–	–
	J	18.5	9.5	2.0	**95.5**	21.0	44.0	18.0	45.5	45.0		–	–
	K	6.0	15.0	22.5	71.0	3.5	19.5	6.5	21.0	20.5	24.5		–
	L	18.0	9.0	1.5	**95.0**	20.5	43.5	17.5	45.0	44.5	0.5	24.0	
BetPu	A		–	–	–	–	–	–	–	–	–	–	–
	B	74.0		–	***	–	–	–	–	*	–	–	–
	C	60.5	13.5		***	–	–	–	–	–	–	–	–
	D	40.0	**114.0**	**100.5**		***	–	–	–	–	***	***	**
	E	60.0	14.0	0.5	**100.0**		–	–	–	–	–	–	–
	F	8.0	66.0	52.5	48.0	52.0		–	–	–	–	–	–
	G	18.5	55.5	42.0	58.5	41.5	10.5		–	–	–	–	–
	H	8.0	66.0	52.5	48.0	52.0	0.0	10.5		–	–	–	–
	I	4.5	**78.5**	65.0	35.5	64.5	12.5	23.0	12.5		–	–	–
	J	63.0	11.0	2.5	**103.0**	3.0	55.0	44.5	55.0	67.5		–	–
	K	70.5	3.5	10.0	**110.5**	10.5	62.5	52.0	62.5	75.0	7.5		–
	L	48.0	26.0	12.5	**88.0**	12.0	40.0	29.5	40.0	52.5	15.0	22.5	
PicAb 1	A		***	–	–	–	–	–	–	–	*	–	–
	B	**116.5**		–	**	***	***	***	***	***	–	***	*
	C	69.5	47.0		–	–	**	–	–	–	–	–	–
	D	28.0	**88.5**	41.5		–	–	–	–	–	–	–	–
	E	0.5	**117.0**	70.0	28.5		–	–	–	–	*	–	–
	F	25.0	**141.5**	94.5	53.0	24.5		–	–	–	***	–	–
	G	0.5	**117.0**	70.0	28.5	0.0	24.5		–	–	*	–	–
	H	6.5	**123.0**	76.0	34.5	6.0	18.5	6.0		–	*	–	–
	I	11.5	**105.0**	58.0	16.5	12.0	36.5	12.0	18.0		–	–	–
	J	**77.0**	39.5	7.5	49.0	**77.5**	**102.0**	**77.5**	**83.5**	65.5		*	–
	K	2.0	**118.5**	71.5	30.0	1.5	23.0	1.5	4.5	13.5	**79.0**		–
	L	32.0	**84.5**	37.5	4.0	32.5	57.0	32.5	38.5	20.5	45.0	34.0	
PicAb 2	A		–	–	–	–	–	–	–	–	–	–	–
	B	56.5		–	–	–	–	**	–	–	–	–	–
	C	10.0	46.5		–	–	–	–	–	–	–	–	–
	D	8.0	48.5	2.0		–	–	–	–	–	–	–	–
	E	15.5	41.0	5.5	7.5		–	–	–	–	–	–	–
	F	17.5	39.0	7.5	9.5	2.0		–	–	–	–	–	–
	G	28.5	**85.0**	38.5	36.5	44.0	46.0		–	–	*	–	–
	H	14.0	42.5	4.0	6.0	1.5	3.5	42.5		–	–	–	–
	I	24.5	32.0	14.5	16.5	9.0	7.0	53.0	10.5		–	–	–
	J	53.0	3.5	43.0	45.0	37.5	35.5	**81.5**	39.0	28.5		–	–
	K	34.0	22.5	24.0	26.0	18.5	16.5	62.5	20.0	9.5	19.0		–
	L	35.5	21.0	25.5	27.5	20.0	18.0	64.0	21.5	11.0	17.5	1.5	

Table 3. Continued.

		A	B	C	D	E	F	G	H	I	J	K	L
PicAb 3	A		–	–	–	–	–	–	–	–	*	–	–
	B	54.5		–	–	–	–	–	–	–	–	–	–
	C	29.5	25.0		–	–	–	–	–	–	–	–	–
	D	16.0	38.5	13.5		–	–	–	–	–	–	–	–
	E	31.5	23.0	2.0	15.5		–	–	–	–	–	–	–
	F	3.5	58.0	33.0	19.5	35.0		–	–	–	*	–	–
	G	12.5	42.0	17.0	3.5	19.0	16.0		–	–	–	–	–
	H	43.0	11.5	13.5	27.0	11.5	46.5	30.5		–	–	–	–
	I	7.5	62.0	37.0	23.5	39.0	4.0	20.0	50.5		**	–	–
	J	**79.5**	25.0	50.0	63.5	48.0	**83.0**	67.0	36.5	**87.0**		–	–
	K	43.0	11.5	13.5	27.0	11.5	46.5	30.5	0.0	50.5	36.5		–
	L	49.5	5.0	20.0	33.5	18.0	53.0	37.0	6.5	57.0	30.0	6.5	
PinSy 3	A		–	–	–	–	–	–	–	–	–	–	–
	B	30.5		–	–	–	–	–	–	–	–	–	–
	C	11.0	19.5		–	–	–	–	–	–	–	–	–
	D	–	–	–		–	–	–	–	–	–	–	–
	E	24.0	6.5	13.0	–		–	–	–	–	–	–	–
	F	24.5	6.0	13.5	–	0.5		–	–	–	–	–	–
	G	34.5	65.0	45.5	–	58.5	59.0		–	***	–	–	–
	H	37.0	6.5	26.0	–	13.0	12.5	71.5		–	–	–	–
	I	73.5	43.0	62.5	–	49.5	49.0	**108.0**	36.5		***	–	–
	J	37.0	67.5	48.0	–	61.0	61.5	2.5	74.0	**110.5**		–	–
	K	23.5	7.0	12.5	–	0.5	1.0	58.0	13.5	50.0	60.5		–
	L	23.5	7.0	12.5	–	0.5	1.0	58.0	13.5	50.0	60.5	0.0	
BetMix	A		–	*	–	–	–	–	*	–	–	–	–
	B	22.5		–	–	–	–	–	–	–	–	–	–
	C	**79.0**	56.5		–	–	–	–	–	***	–	–	–
	D	6.0	16.5	73.0		–	–	–	*	–	–	–	–
	E	7.0	15.5	72.0	1.0		–	–	–	–	–	–	–
	F	36.0	13.5	43.0	30.0	29.0		–	–	–	–	–	–
	G	53.5	31.0	25.5	47.5	46.5	17.5		–	*	–	–	–
	H	**83.0**	60.5	4.0	**77.0**	76.0	47.0	29.5		***	–	–	–
	I	23.5	46.0	**102.5**	29.5	30.5	59.5	**77.0**	**106.5**		*	–	–
	J	56.0	33.5	23.0	50.0	49.0	20.0	2.5	27.0	**79.5**		–	–
	K	31.5	9.0	47.5	25.5	24.5	4.5	22.0	51.5	55.0	24.5		–
	L	15.0	7.5	64.0	9.0	8.0	21.0	38.5	68.0	38.5	41.0	16.5	
PicMix	A		–	–	**	–	–	–	–	–	–	–	–
	B	44.5		–	–	–	–	*	–	*	–	–	–
	C	37.0	7.5		–	–	–	–	–	–	–	–	–
	D	**93.5**	49.0	56.5		**	***	***	**	***	*	***	–
	E	4.0	48.5	41.0	**97.5**		–	–	–	–	–	–	–
	F	6.5	51.0	43.5	**100.0**	2.5		–	–	–	–	–	–
	G	34.5	**79.0**	71.5	**128.0**	30.5	28.0		–	–	–	–	–
	H	7.0	37.5	30.0	**86.5**	11.0	13.5	41.5		–	–	–	–
	I	35.0	**79.5**	72.0	**128.5**	31.0	28.5	0.5	42.0		–	–	–
	J	8.5	36.0	28.5	**85.0**	12.5	15.0	43.0	1.5	43.5		–	–
	K	19.5	64.0	56.5	**113.0**	15.5	13.0	15.0	26.5	15.5	28.0		–
	L	29.0	15.5	8.0	64.5	33.0	35.5	63.5	22.0	64.0	20.5	48.5	

In the year 2007 training course, there do not occur so often statistically significant ($p < 0.05$) observer differences (Table 2). Only in the 7 study sites from totally 9 have significant differing visual observations, while in the year 2006 all 10 study sites contain some inconsistencies. The observers are delightfully compatible in the study sites BetMix, which is birches growing on mixed forest, and pure Scots pine stand PinSy 1. The bivariate post hoc analysis detects generally some little occasional pair-wise dissimilarity (Table 4). Only in the case of pure Norway spruce stand PicAb 2, the observer B has several times different opinion than almost all other observers. It must keep in mind that even single dissimilarity between two observers detected in the bivariate post hoc analysis is enough to signal statistically significant over all disagreements in the Friedman's test of dissimilarity.

Table 4. Matrix of nonparametric *post hoc* analysis for the data of year 2007. Coefficients of difference $\left|R_j - R_j'\right|$ and significances (*** = p<0.001, ** = p<0.01, * = p<0.05, – = not significant) were derived from bivariate procedure (see Equation 2) between observers (A–L). Those study sites without any statistically significant differences are not shown.

		A	B	C	D	E	F	G	H	K	L
BetPu	A		–	–	–	–	–	*	**	–	–
	B	42.5		–	–	–	–	–	–	–	–
	C	10.0	32.5		–	–	–	–	*	–	–
	D	51.0	8.5	41.0		–	–	–	–	–	–
	E	55.5	13.0	45.5	4.5		–	–	–	–	–
	F	12.5	30.0	2.5	38.5	43.0		–	*	–	–
	G	**69.5**	27.0	59.5	18.5	14.0	57.0		–	–	–
	H	**77.0**	34.5	**67.0**	26.0	21.5	**64.5**	7.5		–	–
	K	38.0	4.5	28.0	13.0	17.5	25.5	31.5	39.0		–
	L	34.0	8.5	24.0	17.0	21.5	21.5	35.5	43.0	4.0	
PicAb 2	A		*	–	–	–	–	–	–	–	–
	B	65.0		***	*	*	*	–	**	–	*
	C	19.0	84.0		–	–	–	–	–	–	–
	D	5.0	70.0	14.0		–	–	–	–	–	–
	E	1.5	66.5	17.5	3.5		–	–	–	–	–
	F	2.5	67.5	16.5	2.5	1.0		–	–	–	–
	G	42.5	22.5	61.5	47.5	44.0	45.0		–	–	–
	H	15.0	80.0	4.0	10.0	13.5	12.5	57.5		–	–
	K	13.5	51.5	32.5	18.5	15.0	16.0	29.0	28.5		–
	L	2.0	63.0	21.0	7.0	3.5	4.5	40.5	17.0	11.5	
PicMix	A		–	–	**	–	–	–	–	–	–
	B	15.5		–	–	–	–	–	–	–	–
	C	29.0	13.5		–	–	–	–	–	–	–
	D	**71.5**	56.0	42.5		–	–	***	*	–	–
	E	40.5	25.0	11.5	31.0		–	*	–	–	–
	F	18.5	3.0	10.5	53.0	22.0		–	–	–	–
	G	28.0	43.5	57.0	**99.5**	**68.5**	46.5		–	–	–
	H	6.5	9.0	22.5	**65.0**	34.0	12.0	34.5		–	–
	K	29.5	14.0	0.5	42.0	11.0	11.0	57.5	23.0		–
	L	12.0	3.5	17.0	59.5	28.5	6.5	40.0	5.5	17.5	
PinMix	A		–	–	–	–	–	–	–	–	–
	B	29.0		***	–	–	–	–	–	–	–
	C	54.5	**83.5**		–	–	–	–	–	–	–
	D	0.5	28.5	55.0		–	–	–	–	–	–
	E	13.0	42.0	41.5	13.5		–	–	–	–	–
	F	0.5	29.5	54.0	1.0	12.5		–	–	–	–
	G	1.5	27.5	56.0	1.0	14.5	2.0		–	–	–
	H	4.5	33.5	50.0	5.0	8.5	4.0	6.0		–	–
	K	9.5	38.5	45.0	10.0	3.5	9.0	11.0	5.0		–
	L	1.0	28.0	55.5	0.5	14.0	1.5	0.5	5.5	10.5	

In the year 2008 training course, just like at 2006, there occur statistically significant ($p < 0.05$) observer differences in all study sites and in all tree species (Table 2). Highest inconsistencies between observers occur for birch and Norway spruce. The bivariate post hoc analysis discovers pair-wise dissimilarities between single observers in all cases expect one – pure Scots pine stand PinSy 1 (Table 5). That stand (PinSy 1) had the lowest significance in the Friedman's test of dissimilarity (see Table 2). All other stands have some occasional pair-wise dissimilarity, which can be deployed to single observer just like in the case of Norway spruce stand and observer coded as J. That anonymous observer did not participated to the field test at 2007, but results of the 2006 indicate that observer J has some inconsistent observations especially in the case of Norway spruce.

Table 5. Matrix of nonparametric *post hoc* analysis for the data of year 2008. Coefficients of difference $\left|R_j - R'_j\right|$ and significances (*** = p<0.001, ** = p<0.01, * = p<0.05, – = not significant) were derived from bivariate procedure (see Equation 2) between observers (A–M). Those study sites without any statistically significant differences are not shown.

		A	C	D	E	F	G	H	J	K	L	M
BetPu	A		–	**	–	–	–	–	–	–	–	–
	C	22.5		–	–	–	–	–	*	–	–	–
	D	**85.5**	63.0		–	–	**	–	***	*	***	–
	E	29.0	6.5	56.5		–	–	–	**	–	–	–
	F	38.5	16.0	47.0	9.5		–	–	***	–	–	–
	G	1.0	21.5	**84.5**	28.0	37.5		–	–	–	–	–
	H	65.5	43.0	20.0	36.5	27.0	64.5		***	–	**	–
	J	54.5	**77.0**	**140.0**	**83.5**	**93.0**	55.5	**120.0**		–	–	**
	K	13.5	9.0	**72.0**	15.5	25.0	12.5	52.0	68.0		–	–
	L	19.0	41.5	**104.5**	48.0	57.5	20.0	84.5	35.5	32.5		–
	M	32.5	10.0	**53.0**	3.5	6.0	31.5	33.0	**87.0**	19.0	51.5	
PicAb 1	A		–	–	–	–	–	–	***	–	–	–
	C	51.0		–	–	–	–	–	–	–	–	–
	D	62.0	11.0		–	–	–	–	–	–	–	–
	E	10.0	41.0	52.0		–	–	–	***	–	–	–
	F	12.5	38.5	49.5	2.5		–	–	***	–	–	–
	G	32.5	18.5	29.5	22.5	20.0		–	**	–	–	–
	H	23.5	27.5	38.5	13.5	11.0	9.0		***	–	–	–
	J	**119.0**	68.0	57.0	**109.0**	**106.5**	**86.5**	**95.5**		***	***	**
	K	24.5	26.5	37.5	14.5	12.0	8.0	1.0	**94.5**		–	–
	L	10.0	41.0	52.0	0.0	2.5	22.5	13.5	**109.0**	14.5		–
	M	34.5	16.5	27.5	24.5	22.0	2.0	11.0	**84.5**	10.0	24.5	
PicAb 2	A		–	–	–	–	–	–	**	–	–	–
	C	37.0		–	–	–	–	–	***	–	–	–
	D	9.0	28.0		–	–	–	–	***	–	–	–
	E	19.5	17.5	10.5		–	–	–	***	–	–	–
	F	14.5	51.5	23.5	34.0		–	–	*	–	–	–
	G	52.5	15.5	43.5	33.0	67.0		–	***	–	–	–
	H	0.5	37.5	9.5	20.0	14.0	53.0		**	–	–	–
	J	**85.5**	**122.5**	**94.5**	**105.0**	**71.0**	**138.0**	**85.0**		*	*	*
	K	14.5	51.5	23.5	34.0	0.0	67.0	14.0	**71.0**		–	–
	L	11.5	48.5	20.5	31.0	3.0	64.0	11.0	**74.0**	3.0		–
	M	13.5	50.5	22.5	33.0	1.0	66.0	13.0	**72.0**	1.0	2.0	

Table 5. Continued.

PinSy 2	A		–	–	–	–	–	–	–	–	–	–
	C	22.5		–	–	–	–	–	–	–	–	–
	D	61.5	39.0		–	–	**	–	***	–	–	–
	E	32.0	9.5	29.5		–	–	–	*	–	–	–
	F	20.0	2.5	41.5	12.0		–	–	–	–	–	–
	G	23.5	46.0	**85.0**	55.5	43.5		–	–	–	–	–
	H	8.5	14.0	53.0	23.5	11.5	32.0		–	–	–	–
	J	40.5	63.0	**102.0**	**72.5**	60.5	17.0	49.0		–	–	–
	K	17.5	5.0	44.0	14.5	2.5	41.0	9.0	58.0		–	–
	L	20.0	2.5	41.5	12.0	0.0	43.5	11.5	60.5	2.5		
	M	25.0	2.5	36.5	7.0	5.0	48.5	16.5	65.5	7.5	5.0	
BetMix	A		–	–	–	–	–	–	–	–	–	–
	C	11.5		–	–	–	–	–	–	–	–	–
	D	9.5	35.5		–	–	–	–	–	–	–	–
	E	0.0	26.0	9.5		–	–	–	–	–	–	–
	F	2.5	23.5	12.0	29.0		–	–	–	–	–	–
	G	6.0	20.0	15.5	25.5	17.5		–	–	–	–	–
	H	14.0	12.0	23.5	17.5	25.5	0.0		–	–	–	–
	J	14.0	12.0	23.5	17.5	25.5	0.0	8.0		*	**	***
	K	16.5	9.5	26.0	15.0	28.0	2.5	10.5	**78.5**		–	–
	L	26.0	0.0	35.5	5.5	37.5	12.0	20.0	**88.0**	12.0		–
	M	31.5	5.5	41.0	0.0	43.0	17.5	25.5	**93.5**	17.5	15.0	
PicMix	A		–	–	–	–	–	–	–	–	–	–
	C	30.0		–	–	–	–	–	*	–	–	–
	D	12.5	42.5		–	–	–	–	–	–	–	–
	E	8.0	22.0	20.5		–	–	–	–	–	–	–
	F	20.5	50.5	8.0	28.5		–	–	–	–	–	*
	G	40.0	10.0	52.5	32.0	60.5		–	**	–	–	–
	H	24.0	6.0	36.5	16.0	44.5	16.0		*	–	–	–
	J	46.5	**76.5**	34.0	54.5	26.0	**86.5**	**70.5**		–	–	***
	K	18.5	11.5	31.0	10.5	39.0	21.5	5.5	65.0		–	–
	L	14.0	16.0	26.5	6.0	34.5	26.0	10.0	60.5	4.5		–
	M	49.5	19.5	62.0	41.5	**70.0**	9.5	25.5	**96.0**	31.0	35.5	
PinMix	A		–	–	–	–	–	–	–	–	–	–
	C	26.5		–	–	–	–	–	–	–	–	–
	D	40.0	13.5		–	–	–	*	–	–	–	–
	E	43.5	17.0	3.5		–	–	**	–	–	–	–
	F	15.5	11.0	24.5	28.0		–	–	–	–	–	–
	G	32.0	5.5	8.0	11.5	16.5		*	–	–	–	–
	H	38.0	64.5	**78.0**	**81.5**	53.5	**70.0**		*	**	–	–
	J	39.0	12.5	1.0	4.5	23.5	7.0	**77.0**		–	–	–
	K	44.0	17.5	4.0	0.5	28.5	12.0	**82.0**	5.0		–	–
	L	7.5	19.0	32.5	36.0	8.0	24.5	45.5	31.5	36.5		–
	M	15.5	11.0	24.5	28.0	0.0	16.5	53.5	23.5	28.5	8.0	

4. Discussion

Although we detected some inconsistencies between observers, those are still occasional in nature because those inconsistencies are detected by pair-wise comparison between two observers. Only in some cases, we detected that single observer has disagreements with several other observers simultaneously. Yet, our opinion is that these are independent incidents and are caused mainly by the observer perception during that specific event. Only longer follow-up study can ascertain if some observer constantly have statistically significantly different observations. Our results don't give any affirmation and empower to produce correction tools to future. If some consistent dissimilarities exist and can be detected, those should be handled by proper advisement and training because all discernible inconsistencies are non-systematic and unpredictable.

Herewith, these minor differences discovered in the present study do not affect to accuracy of national level defoliation surveys of Finland's forest condition monitoring. Under and over estimates might compensate each other in large scale inventories (e.g. national level). Solberg and Strand (1999) have also concluded that even if tree defoliation assessments contain some bias, they believe those have the ability to provide crude, but reliable estimates of spatial and temporal trends, when these trends are not too weak. They highlighted that trends and changes should be clearly higher than the rate of bias. On the basis of this study, we can approve the conclusion of Strand (1996) that the tree defoliation assessment is especially practical method for large scale monitoring irrespective of its weaknesses and multifarious sources of bias. However, visual observations and surveys always include some sources of uncertainty which must keep in mind when analyzing national defoliation data (Salemaa et al., 1993).

Our results indicate that Scots pine trees are more consistently observed than Norway spruce trees and birches. This might be related to fact that the crown of Scots pine consists of low number of living needle cohorts (Muukkonen, 2004), which make field monitoring easier. In addition, Scots pine does not have so evident branching type strains in different ecotones in Finland (Salemaa et al., 1993), which affects to working experience of observers living in different locations. On the contrary, Norway spruce is famous about the fact that it has large number of climatic and site strains with different branching shapes (Hanisch and Kilz, 1990; Innes, 1993). In addition, Norway spruce is quite difficult to observe because it has large amount of substitute shoots produced by dormant bud, which have not produced shoots in the first year after their development (Hanisch and Kilz, 1990). These buds can produce a large amount of the branching. Basically we can say that crown architecture varies markedly between species (Innes, 1993), which effects to easiness of visual tree defoliation assessment.

If some inconsistencies between observers exist this might have several reasons. Firstly, our conclusion is that the visual assessment of tree defoliation itself is more accurate part of whole assessment chain than the visual estimation of the lower limit of living crown, which effects directly to overall defoliation estimate of single tree. In the other words, there might occur more inconsistencies in crown determination than in defoliation assessment. Yet, we have not tested this hypothesis in the current study. Secondly, field observers are working and living in different parts of Finland and it is evident that boreal tree species have extraordinary large number of climatic and site strains with different branching shapes; especially Norway spruce. Our data is collected from single location, which might consist of trees with unfamiliar branching types for some observers living and working in other ecotones.

We concluded on same way than Solberg and Strand (1999) that the statistically significant differences between observers can be attenuated by two ways. Firstly, the same observers should survey the same study sites at every round of survey, so that temporal trends are more reliable. Secondly, large scale averages should, if possible, be based on several observers or observer teams. Increasing the number of observers or observer teams per spatial unit can reduce the effect of observer bias on the large scale average (Gertner and Köhl, 1995; Solberg and Strand, 1999). Yet, forest condition monitoring is an annual long term assessment and the field crew (observers) may vary during it and consequently there might occur inconsistent practices throw the long time series (Gertner and Köhl, 1995).

Although the visual method to survey tree defoliation has been abundantly criticized for a decades, it has been and it still is a widely used method. Innes (1988b) has already suggested few decades ago that, in the future, remote sensing techniques may provide useful tools to replace observer depended visual tree defoliation assessments or, as Stone et al. (2003) discussed, to cover extensive forest areas. Subsequently, Dobbertin et al. (2004) have discussed that some new procedures have been proposed and are currently being evaluated. Results are promising (Mizoue and Dobbertin, 2003, 2004), but semi-automatic image analysis are still affected by some operator error (Mizoue et al., 2004) and airborne or satellite remote sensing methods are too coarse to detect particular targets (Coops et al., 2004). Nonetheless, before these methods became common, the suggestion of Innes (1988b) and Salemaa et al. (1993) is still topical; the education and calibration courses of individual observers is essential in observer-based defoliation survey.

Based on our results, we don't see any reason for systematic correction of observations of individual observers in the Finnish forest condition monitoring, because individual inconsistencies are adventitious and non-systematic. Our conclusion is that the proper education and guidance of field personnel is essential for providing as reliable tree defoliation assessments as possible. Although, every summer before the start of the field work period, the observers undergo a one-week training course where they receive practical training in the assessment procedures and the assessment level is calibrated (Salemaa and Lindgren, 2000), this can be still improved.

Acknowledgements

The Finland's forest condition monitoring is carried out by the Finnish Forest Research Institute. The authors express their thanks particularly to the field personnel of the Finnish forest condition monitoring for enabling the data used in this study. This study and the forest condition monitoring programme is being carried out as a part of the FutMon project, co-financed by the Life+ programme of the European Union. This study was partially financed also by EU Life+ -project Climate change induced drought effects on forest growth and vulnerability – Climforisk (LIFE09 ENV/FI/000571).

References

Coops, N.C., Stone, C., Culvenor, D.S.,Chisholm, L., 2004. Assessment of crown condition in Eycalypt vegetation by remotely sensed optical indices. Journal of Environmental Quality 33, 956–964.

Derome, J., Lindgren, M., Merilä, P., Beuker, E.,Nöjd, P., 2007. Forest condition monitoring under the UN/ECE and EU programmes in Finland. In: Merilä, P. et al. (Eds), Forest condition monitoring in Finland: national report 2002–2005. Working Papers of the Finnish Forest Research Institute. p. 11–20.

Dobbertin, M., Hug, C.,Mizoue, N., 2004. Using slides to test for changes in crown defoliation assessment methods. Part I: visual assessment of slides. Environmental Monitoring and Assessment 98, 295–306.

Dobbertin, M., Hug, C.,Mizoue, N., 2005. Using slides to test for changes in crown defoliation assessment methods part II: application of the image analysis system CROCO. Environmental Monitoring and Assessment 102, 167–178.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association. 32, 675–701.

Friedman, M., 1939. A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 34, 109.

Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics 11, 86–92.

Gertner, G.,Köhl, M., 1995. Correlated observer errors and their effects on survey estimates of needle-leaf loss. Forest Science 41, 758–776.

Ghosh, S.,Innes, J.L., 1995. Combining field and control team assessments to obtain error estimates for surveys of crown condition. Scandinavian Journal of Forest Research 10, 264–270.

Ghosh, S., Innes, J.L.,Hoffmann, C., 1995. Observer variation as a source of error in assessments of crown condition through time. Forest Science. 41, 235–254.

Hanisch, B.,Kilz, E., 1990. Monitoring of forest damage. Eugen Ulmer & Co, Stuttgart, 334 p.

Innes, J.L., 1988a. Forest health surveys – a critique. Environmental Pollution 54, 1–15.

Innes, J.L., 1988b. Forest health surveys: problems in assessing observer objectivity. Canadian Journal of Forest Research. 18, 560–565.

Innes, J.L., 1992. Forest decline. Progress in Physical Geography 16, 1–64.

Innes, J.L., 1993. Forest health: its assessment and status. Centre for Agriculture and Bioscience International, Oxon, 677 p.

Innes, J.L., Landmann, G.,Mettendorf, B., 1993. Consistency of observations of forest tree defoliation in three European countries. Environmental Monitoring and Assessment 25, 29–40.

Jukola-Sulonen, E.-L., Mikkola, K.,Salemaa, M., 1990. The vitality of conifers in Finland, 1986–88. In: Kauppi, P. et al. (Eds), Acidification in Finland. p. 523–561.

Lindgren, M., 2001. Results of the 2000 national crown condition survey (ICP Forests/Level I). In: Ukonmaanaho, L.,Raitio, H. (Eds), Forest condition monitoring in Finland: national report 2000. The Finnish Forest Research Institute, Research Papers. p. 57–65.

Lindgren, M., 2002. Results of the 2001 national crown condition survey (ICP Forests/Level I). In: Rautjärvi, H. et al. (Eds), Forest condition monitoring in Finland: national report 2001. The Finnish Forest Research Institute, Research Papers. p. 42–50.

Metzger, J.M.,Oren, R., 2001. The effect of crown dimensions on transparency and the assessment of tree health. Ecological Applications 11, 1634–1640.

Mizoue, N.,Dobbertin, M., 2003. Detecting differences in crwon transparency assessements between countries using the image analysis system CROCO. Environmental Monitoring and Assessment 89, 179–195.

Mizoue, N.,Dobbertin, M., 2004. Within country accuracy of tree crown transparency estimates using the image analysis system CROCO: a case study from Switzerland. Environmental Modelling and Software 19, 1089–1095.

Mizoue, N., Dobbertin, M.,Sugawara, D., 2004. Operator errors in tree crown transparency estimates using the image analysis system CROCO. Computers and Electronics in Agriculture 44, 247–254.

Moffat, A.J., Davies, S.,Finér, L., 2008. Reporting the results of forest monitoring – an evaluation of the European forest monitoring programme. Forestry 81, 75–90.

Muukkonen, P., 2004. Needle biomass turnover rates of Scots Pine (Pinus sylvestris L.) derived from the needle-shed dynamics. Trees 19, 273–279.

Redfern, D.B.,Boswell, R.C., 2004. Assessment of crown condition in forest trees: comparison of methods, sources of variation and observer bias. Forest Ecology and Management 188, 149–160.

Rehfuess, K.E., 1989. Acidic deposition – extent and impact on forest soils, nutrition, growth and disease phenomena in Central Europe: a review. Warer, Air, and Soil Pollution 48, 1–20.

Salemaa, M., Jukola-Sulonen, E.-L.,Lindgren, M., 1991. Forest condition in Finland, 1986–1990. Silva Fennica 25, 147–175.

Salemaa, M., Jukola-Sulonen, E.-L., Nieminen, T.,Nöjd, P., 1993. Latvatunnukset ja puun kasvu elinvoimaisuuden ilmentäjinä. In: Hyvärinen, A. et al. (Eds), Metsäluonto ja ilmansaasteet. p. 77–92.

Salemaa, M.,Lindgren, M., 2000. Crown condition. In: Mälkönen, E. (Ed), Forest condition in a changing environment – the Finnish case. p. 121–132.

Sheldon, M.R., Fillyaw, M.J.,Thompson, W.D., 1996. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. Physiotherapy Research International 1, 221–228.

Solberg, S.,Strand, L., 1999. Crown density assessments, control surveys and reproducibility. Environmental Monitoring and Assessment 56, 75–86.

Stone, C., Wardlaw, T., Floyd, R., Carnegie, A., Wylie, R.,de Little, D., 2003. Harmonisation of methods for the assessment and reporting of forest health in Australia – a starting point. Australian Forestry 66, 233–246.

Strand, G.-H., 1996. Detection of observer bias in ongoing forest health monitoring programmes. Canadian Journal of Forest Research 26, 1692–1696.

Tomppo, E., Henttonen, H.,Tuomainen, T., 2001. Valtakunnan metsien 8. inventoinnin menetelmä ja tulokset metsäkeskuksittain Pohjois-Suomessa 1992-94 sekä tulokset Etelä-Suomessa 1986-92 ja koko maassa 1986-94. Metsätieteen aikakauskirja 1B/2001, 99–248.

Westman L., Lesinski, J. 1986. Kronutglesning och andra förändringar I grankronan. Morfologisk beskrivning. Naturvårdsverket, Rapport 3262. 96 p.

Wulff, S., 2002. The accuracy of forest damage assessments – experiences from Sweden. Environmental Monitoring and Assessment 74, 295–309.

Zarnoch, S.J., Bechtold, W.A.,Stolte, K.W., 2004. Using crown condition variables as indicators of forest health. Canadian Journal of Forest Research 34, 1057–1070.