Adaptive Cluster Sampling in Inventorying Forest Damage by the Common Pine Sawfly (*Diprion pini*)

Mervi Talvitie*, Tuula Kantola*, Markus Holopainen* and Päivi Lyytikäinen-Saarenmaa*

ABSTRACT

Climate change and biological invasions are threats to healthy forest environments throughout the world. Some species that have previously caused only small-scale damage have now become serious pests, causing massive outbreaks and yield losses including in Scandinavia. The spatial scale of outbreaks and intensity of defoliation caused by the common pine sawfly (*Diprion pini* L.) can vary between years, due to fluctuation in population dynamics. The study area is situated in Ilomantsi, eastern Finland, where *D. pini* has caused vast needle losses in managed Scots pine stands. We aimed at developing an accurate and cost-efficient inventory method for insect damage, in which we compared stratified adaptive cluster sampling, random adaptive cluster sampling and simple random sampling. Stratified adaptive cluster sampling proved to be the most accurate method and was a promising candidate for inventorying and monitoring pest insect damage in the study. Adaptive cluster sampling is a promising method for inventorying and monitoring such phenomena when area does not remain constant all the time.

Keywords: adaptive cluster sampling, insect outbreak, defoliation, *Diprion pini*

INTRODUCTION

Climate change has caused rising temperatures throughout the world. Ecological balance has been interrupted, causing among other things pest damage in managed forests (Dale et al., 2001; Evans et al., 2002). The speed of change is higher at higher latitudes, as in Scandinavia. Forest insects, formerly regarded as harmless organisms during recent decades, are now causing serious damage in Finland (Lyytikäinen-Saarenmaa and Tomppo, 2002; De Somville et al., 2007). Economic losses can be considerable, appr. 300 - 1000 eur ha⁻¹, depending on intensity of needle loss. It can require over a decade for a tree to recover fully (Lyytikäinen-Saarenmaa et al., 2005).

The common pine sawfly (*Diprion pini* L.) (Hymenoptera, Diprionidae) is a univoltine species in Scandinavia. The gregarious larvae actively consume all the needle age-classes of Scots pine (*Pinus sylvestris* L.) during August and September, which could lead to total defoliation of host trees in the climax phase of population dynamics (Virtasaari and Varama, 1987; Geri, 1988). Mature and maturing trees have the highest defoliation risk, but at peak densities seedlings can also suffer from defoliation (Geri, 1988; De Somville et al., 2004). *Diprion pini* hibernates in the cocoon stage and adult insects emerge early in the season. Diapause is typical for this species, lasting up to several years and prolonging outbreaks. Outbreaks have usually covered only some hundreds or thousands of hectares during the last 150 years, according to published records (e.g. Kangas, 1963, De Somville et al., 2007). A massive outbreak by *D. pini* occurred on dry and dryish pine forests in central Finland between 1997 and 2001, covering 300,000 ha (Lyytikäinen-Saarenmaa and Tomppo, 2002). The outbreak reached thus Ilomantsi district (eastern Finland) in 1999, where sawfly densities have fluctuated since then, showing a chronic nature.

Monitoring of sawfly outbreaks and defoliation is typically based on field sampling, i.e. collecting different life stages and performing analyses of this material in a laboratory (e.g. Juttimen and Varama, 1986). This kind of monitoring consumes extensive amounts of resources, and results may be biased. The method is still in use at the Finnish Forest Research Institute (Metla). The latest innovation in monitoring is based on semiochemicals, i.e. sex pheromones of sawflies, but

Corresponding author: Mervi Talvitie
E-mail: mervi.talvitie@helsinki.fi

*Department of Forest Science, University of Helsinki, P.O. Box 27, 00014 University of Helsinki, Finland

*For. Plann. 16: 141-147 (2011)
the method results in population and risk estimates only at the stand level (Lyytikäinen-Saarensmaa et al., 1999, 2006). Furthermore, estimates for forthcoming defoliation and yield losses may be indicative. Metla carries out the National Forest Inventory (NFI) every 10 years, when information on forest health as a side product of permanent and temporary field clusters is also collected at the coarse level (Tomppo et al., 2006). The NFI applies satellite data, e.g. Landsat Thematic Mapper (TM) images, but not for monitoring of forest disturbances. Long inventory periods and sparse layout of clusters result in problems for annual needs of precise information on forest disturbances and risk estimates for the coming years.

Deciding on an optimal inventory method is often the crucial phase for a certain research problem. The natural distribution of ecosystems is fragmented, mainly due to human intervention. Ecological distributions are very rarely randomly distributed (Levin, 1992). With conventional forest inventory methods, large areas must be covered to achieve acceptable precision of rare objects (Green, 1993, Green and Young, 1993). Estimates need to be reliable; imprecise estimates may skew the real situation in the population and result in bad forest management decisions. Thus, prior information on the population of interest is crucial to decision-making (Kangas and Maltamo, 2006).

Adaptive cluster sampling (ACS) was suggested as a method for estimating rare and clustered populations that are difficult to estimate with high accuracy, using conventional sampling methods (Thompson, 1990, 1991; Roelsch, 1993; Magnussen et al., 2005). Many studies concerning ACS have been undertaken with simulated data (e.g. Roelsch, 1993; Brown and Manly, 1998; Brown, 2003; Su and Quinn, 2003; Magnussen et al., 2005). Few ACS studies, to our knowledge, have been done with actual data. Acharya et al. (2000) examined the density of three different rare tree species in Nepal with ACS, and Talvitie et al. (2006) used ACS in inventorying coarse woody debris in the city forests of Helsinki.

The aim of the present study was to develop an accurate and cost-efficient inventory method for defoliating forest insects. ACS was chosen as the field inventory method. The defoliation estimate and variance of stratified ACS (ACSs) were compared with those of random ACS (ACSRs) and simple random sampling (SRS) inventories. ACSs, ACSrs, and SRS samples were all based on the same field inventory and sample plots.

**SAMPLING THEORY AND DESIGN**

In sampling rare clustered phenomena, ACS has the advantage over other, more conventional sampling methods (e.g. Thompson, 1990; Roelsch, 1993). The initial set of units is selected, using a probability sampling procedure, and additional units are added to the sample from the neighbourhood of unit $i$ in case the variable of interest $y_i$ satisfies a given criterion, e.g. if $y_i \in C, C = \{x|x \geq c\}$, prior to sampling. The procedure is repeated until no new additional plots can be found whose variable exceeds the given value. Thus, sampling effort is focused on limited areas with a high number of interesting variables.

The population consists of $N$ units $(1, 2, \ldots, N)$ and has variables of interest $y_i = (y_{i1}, y_{i2}, \ldots, y_{ik})$. In forest applications, a tree or sample plot is normally taken as a basic unit. For every unit $i$ in the population, a physical neighbourhood $A_i$ will be defined that consists of a collection of units that includes unit $i$. If unit $j$ is in the neighbourhood of unit $i$, then unit $i$ is also in the neighbourhood of unit $j$ (Thompson, 1990).

A network is defined as a group of units whose values are all at least as large as the critical value $C$. A unit that does not satisfy this condition, but is in the neighbourhood of the unit that does, is referred to as an edge unit (Thompson, 1990). A network within its edge units is called a cluster. The difference from conventional sampling designs is that the procedure for selecting an adaptive sample is dependent on the population values observed in the field. The efficiency of the ACS method is dependent on the density and clustering degree, neither of which is known before the initial sampling. All the above-mentioned factors increase the uncertainty in choosing which study method to use in an investigation.

Thompson (1990) presented a modified Horwitz-Thompson

---

**Table 1. Simple random sampling and adaptive cluster sampling estimators and variances applied in the study**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple random sampling (SRS)</td>
<td>$\bar{y}<em>{SRS} = \frac{1}{n} \sum</em>{i=1}^{n} y_i$</td>
<td>$\text{var}(\bar{y}<em>{SRS}) = \frac{\sum</em>{i=1}^{n} (y_i - \bar{y}_{SRS})^2}{n-1}$</td>
</tr>
<tr>
<td>Adaptive cluster sampling (ACS)</td>
<td>$\bar{y}<em>{ACS} = \frac{1}{N} \sum</em>{k=1}^{K} y_j^k / \alpha_j$</td>
<td>$\text{var}(\bar{y}<em>{ACS}) = \frac{1}{N^2} \left[ \sum</em>{k=1}^{K} \frac{y_j^k / \alpha_j}{\alpha_j} \left( \frac{1}{\alpha_j} - 1 \right) \right]$</td>
</tr>
</tbody>
</table>

$N$ = number of units in the population, $n$ = number of units in the initial sample, $K$ = number of distinct networks represented in the initial sample, $y_j^k$ = observation in unit $i$, $\alpha_j$ = sum of observations in network $k$, $\alpha_j$ = the probability that the initial sample intersects network $A_i$, $\alpha_j$ = the probability that the initial sample includes at least one unit in each of networks $j$ and $k$. 

---

NII-Electronic Library Service
Adaptive Cluster Sampling in Inventoring Forest Damage by the Common Pine Sawfly (Diprion pini)

Table 2 Stand characteristics of the Palokangas study area

<table>
<thead>
<tr>
<th>Variable</th>
<th>Average</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D (cm)</td>
<td>14.7</td>
<td>11.5</td>
</tr>
<tr>
<td>H (m)</td>
<td>11.6</td>
<td>8.7</td>
</tr>
<tr>
<td>V (m² ha⁻¹)</td>
<td>72.8</td>
<td>71.1</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>53</td>
<td>37</td>
</tr>
<tr>
<td>Basal area (m² ha⁻¹)</td>
<td>13.5</td>
<td>7.7</td>
</tr>
<tr>
<td>Trees ha⁻¹</td>
<td>1,394</td>
<td>1,336</td>
</tr>
</tbody>
</table>

Fig. 1 Location of Palokangas study area. Areas infected by D. pini are seen as grey

contrast to the results from previous studies in other countries, elevated defoliation has also been found in younger stands in the eastern parts of Finland, without clear evidence for a preference for only well-drained mineral soils (de Somviele et al., 2004).

The sample plots were situated 50 m from each other in the north-south and east-west directions. The position of each plot was obtained with a Trimble GPS Pathfinder Pro XH (Trimble Navigation Ltd., Sunnyvale, CA, USA), which can reach up to 30-cm accuracy. The locations were postprocessed with local base station data, resulting in an average error of app. 1-1.5 m. If a sample plot fell on the border of two stands, it was transferred towards a stand it more belonged to, thus avoi-
ding edge effects. All trees, having a diameter over 5 cm at breast height (1.3 m) were measured, using relascope sample plots (factor \(q = 1\)). For each tree in a sample plot, the species, diameter at breast height, distance and degree from the plot centre were measured. For pines, the defoliation percentage and canopy strata were determined. Every seventh tree of each species was taken as a sample tree. The height of the median trees for each sample plot and the sample trees were determined. Moreover, the age of the stand in a sample plot was determined from the median tree.

The inventory methods used were SRS, ACSnr and ACSas. All three methods were based on the initial or total sample plots: SRS only initial sample plots, ACSas total sample size without stratification and ACSnr total sample size with stratification. Stratification was performed according to site class and soil type, using the information on compartmentwise inventory, consisting of three strata: 1) high-productivity (HP) class, 2) low-productivity (LP) class, and 3) peatland (Table 3). The HP class contained dryish mineral soil site and sites more fertile than it, and the LP class contained sites poorer than the dryish mineral soil site. The third stratum included peatland plots. Within each stratum, sample plots were chosen, using SRS. There were a total of 55 initial sample plots taken in the sample, representing 0.5% of the total population.

The variable defoliation intensity of a sample plot was taken as the critical variable. Defoliation intensity was an indicator of the former D. pini outbreak intensity. The intensity of defoliation of a tree was visually estimated from different directions according to EICHHORN (1996), comparing the tree under investigation to a reference tree with full healthy foliage in the same site type; an accuracy of 10% was used. The defoliation intensity of a sample plot was calculated as the mean value of the defoliation intensity of each tree in the dominant and intermediate crown layers. The underlying trees were omitted from calculation of the plot mean due their limited ability to compete, which could skew the total plot defoliation. The critical value \(C\) was 20% in the present study; if the mean defoliation intensity in a sample plot was 20% or more, four additional sample plots were measured north, south, west and east of the unit. The defoliation intensity of 20% can be seen as a threshold value. Intensities less than 20% cause no serious harm to tree increment or mortality; however, a 20% intensity may decrease radial growth by 40-50% after needle consumption by D. pini (LYYTIKÄINEN-SAARENMAA et al., 2006).

RESULTS

The initial sample consisted of 55 and the total sample of 180 plots. Four additional units were on a road, and two sample plots were impossible to measure due to such narrow stands. There were three clusters in the HP stratum and three clusters in the LP stratum with at least one sample plot having a defoliation intensity over 20% (Table 4). No defoliation rate over 20% was found in the peatland class. The largest cluster, a total of 43 units, was found in HP (Fig. 2). The ACS meth-

Table 4 Clusters in the final sample, number of plots and their defoliation intensity in comparison to the critical value \(C = 20\%\)

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Cluster</th>
<th>Plots</th>
<th>(C &gt; 20%)</th>
<th>(C &lt; 20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>HP</td>
<td>2</td>
<td>43</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>HP</td>
<td>3</td>
<td>21</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>LP</td>
<td>4</td>
<td>21</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>LP</td>
<td>5</td>
<td>37</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>LP</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

![Fig 2](image_url) Arrangement of a cluster. Initial sample plots (dark grey diamond), additional neighbouring plots satisfying the condition \(C \geq 20\%\) (black circle) and additional plots, called edge units, with less than 20% defoliation (light grey circle)
Table 5: Number of units in the initial sample (m) and final sample (m'), estimates (μ), variances (σ²) and coefficients of variation (CV) in each stratum and in the total population

<table>
<thead>
<tr>
<th>Method</th>
<th>Stratum</th>
<th>n</th>
<th>m</th>
<th>μ</th>
<th>σ²</th>
<th>CV%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACSn</td>
<td>total</td>
<td>55</td>
<td>180</td>
<td>8.3</td>
<td>21.4</td>
<td>55.7</td>
</tr>
<tr>
<td></td>
<td>HP</td>
<td>29</td>
<td>95</td>
<td>6.4</td>
<td>20.1</td>
<td>68.6</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>20</td>
<td>79</td>
<td>9.2</td>
<td>26.4</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>Peatland</td>
<td>6</td>
<td>6</td>
<td>11.4</td>
<td>15.6</td>
<td>34.6</td>
</tr>
<tr>
<td>ACSn</td>
<td>total</td>
<td>55</td>
<td>180</td>
<td>8.3</td>
<td>30.1</td>
<td>66.2</td>
</tr>
<tr>
<td>SRS</td>
<td>total</td>
<td>55</td>
<td>55</td>
<td>8.5</td>
<td>109.3</td>
<td>123.2</td>
</tr>
</tbody>
</table>

ods were calculated, using the total sample plots (180 plots), while the SRS calculations were based on the initial sample plots.

The SRS, ACSn and ACSn estimators for the defoliation mean and variance were calculated (Table 5). The mean values for all the estimators were very similar: 8.3% using ACSn and ACSn to 8.3% using SRS. The mean value represents the mean defoliation index of the total study area. Within ACSn strata, the mean values varied between 6.4% in the richest HP soil class and 11.4% in peatland.

The coefficient of variation (CV) describes how much the variation is in relation to the estimate. It is a normalized measure of dispersion of a probability distribution and, thus, offers a way to compare the means and variances of different scales. The CV values are represented as percentages (Table 5). The smallest percentage for CV was found in ACSn (55.7%). Again, that of SRS (123.2%) was significantly larger.

**DISCUSSION**

In inventorying rare and clustered forest phenomena, ACS has been considered as an effective design for estimating density and total amount of the variable of interest (Thompson, 1990; 1991). Using ACS, one can focus measurements on areas with high damage intensity. If population groups are larger than sample plots, it is more efficient to measure and estimate the variables of interest using the ACS method (Acharya et al., 2000). In many situations, the costs of a unit that do not satisfy the critical value C are lower than those of a plot that do. Moving from one sample plot to another takes time; if the plots are situated near each other the time spent on walking is decreased rapidly. However, one reason why ACS is seldom used in practice is that the sampling size is not known in advance. The risk that the sample size would become too large and thus, time-consuming, can be solved with restricted ACS (Brown and Manly, 1998). The restricted design allows the final sample size to be approximately determined before field sampling (Brown and Manly, 1998).

Here, the defoliation intensity estimates were quite similar in all sampling methods, which increases the need for careful evaluation of the effectiveness of the methods. The total population variance of ACSn was smallest because the sampling method takes into account prior information from the stratiﬁcation. Both the ACSn and ACSn methods allow measurement of additional units; in ACSn, the outcome was calculated for the entire sample and the result showed a larger variance. Thus, the influence of prior information is signiﬁcant. When each stratum is calculated separately, the results achieved are more precise. The SRS variance was nearly twice as large as that of ACSn and was not capable of competing with ACS methods for the effectiveness of inventorying insect-caused defoliation. The SRS method suffered from small sample sizes and the loss of taking adaptive measuring units. On the other hand, even if the SRS took the same number of sample plots into calculation, it would suffer from its ineffectiveness.

The CV describes the dispersion of the variable so that the variable's measurement unit is no object. Higher CV implies that there is greater dispersion in the variable. Here, CV also conﬁrmed the superiority of ACSn in inventorying D. pini-caused damage in comparison to ACSn and SRS. The CV% of ACSn was larger to some extent; again, that of SRS was signiﬁcantly larger.

In ACSn, comparison between strata shows that peatland had the most homogenous sample plots. There were no seedling plots in the stratum, in which no or only slight defoliation intensity was found. If prior information is available, it is worth considering how to use it best in selecting the initial sample. Our results agree with those of de Somviele et al. (2004) in Palokangas; defoliation is lower in stands situated in more fertile forest site types. The largest cluster was found in the HP stratum, mainly situated in the seedling tree position, where there are appr. 20 trees per hectare. Such low densities can also constitute a risk of defoliation in more fertile productivity stands.

During late summer 2008, D. pini defoliation was relatively minor. There were only two plots in the final sample with defoliation intensities over 40%, and most of the initial sample plots had intensities under 20%. Diprion pini-caused defoliation has occurred in the Palokangas area for over a decade, fluctuating since that time. If the ACS method is used to inventorying the phenomenon, more sample plots would be measured in years when the defoliation is more intense.

Some of the sources of error in the fieldwork may have included measurement and positioning errors in locating and measuring the relascope sample plot. Yet, the effects of these errors could not be analytically observed in the results. Positioning errors can result in misidentified strata in plots when ACSn is used. However, due to the accurate positioning achieved with the global positioning system (GPS) device, there were no such problems in our field measurements.

The above-mentioned source of error can lead to the risk of misidentifying the variable of interest. In SRS, misidentifications may cancel each other, but in adaptive sampling they

**J. For. Plann. 16: 141-147 (2011)**
accumulate. Thus, an adaptive sampling method should be employed for variables that can be identified with full confidence. Here, the critical variable was estimated visually, which could easily cause deviation in results if the surveyor were not a professional. Naked-eye calibration is essential when two or more researchers are estimating the critical variable.

Further studies will be focused on acquiring additional information by aerial photographs and laser-scanning data, thus bringing more information to the initial sample selection (LYTTYräinen-Saarenmaa et al., 2006). Remote-sensing materials are valuable means for detecting and mapping insect damage in rare and clustered phenomena. In this study, stratification was performed according to the compartment-wise information from the study area. With remote-sensing data, the stratification could be carried out by focusing on more plots in areas where pest damage could already be detected from the material.

LITERATURE CITED


Adaptive Cluster Sampling in Inventorying Forest Damage by the Common Pine Sawfly (Diprion pini)


(Received 23 September 2009)
(Accepted 24 March 2010)

J. For. Plann. 16: 141-147 (2011)