

Computer-aided calibration for visual estimation of vegetation cover

Gallegos Torell, Åsa¹ & Glimskär, Anders^{2*}

¹*Department of Forest Resource Management, Swedish University of Agricultural Sciences (SLU), SE-901 83 Umeå, Sweden; E-mail asa.gallegos.torell@srh.slu.se;* ²*Department of Ecology, SLU, P.O. Box 7044, SE-75007 Uppsala, Sweden;*
**Corresponding author; E-mail anders.glimskar@ekol.slu.se*

Abstract

Question: What precision and accuracy of visual cover estimations can be achieved after repeated calibration with images of vegetation in which the true cover is known, and what factors influence the results?

Methods: Digital images were created, in which the true cover of vegetation was digitally calculated. Fifteen observers made repeated estimates with immediate feedback on the true cover. The effects on precision and accuracy through time were evaluated with repeated proficiency tests. In a field trial, cover estimates, before and after calibration, were compared with point frequency data.

Results: Even a short time of calibration greatly improves precision and accuracy of the estimates, and can also reduce the influence of different backgrounds, aggregation patterns and experience. Experienced observers had a stronger tendency to underestimate the cover of narrow-leaved grasses before calibration. The field trial showed positive effects of computer-based calibration on precision, in that it led to considerably less between-observer variation for one of the two species groups.

Conclusions: Computer-aided calibration of vegetation cover estimation is simple, self-explanatory and time-efficient, and might possibly reduce biases and drifts in estimate levels over time. Such calibration can also reduce between-observer variation in field estimates, at least for some species. However, the effects of calibration on estimations in the field must be further evaluated, especially for multilayered vegetation.

Keywords: Environmental monitoring; observer error; plot inventory; vegetation cover.

Nomenclature: Karlsson (1998).

Introduction

When working in the field with plant inventories, observers will often be required to describe vegetation cover. The most common meth-

od is visual estimation, where an observer assesses the cover percentage of (one or a group of) plant species in a plot of a predetermined size (Økland 1990). This is a method that can be executed efficiently in the field, requires little equipment and seems to be quite easy to learn as well as to teach (Klimeš 2003).

Many empirical studies report quite large variations in estimated cover between measurements, and therefore low precision of such estimates. Kennedy & Addison (1987) determined that the “measurement error” (the arithmetic mean of absolute difference in cover between two successive cover estimates) was around 10%, while Tonteri (1990) found a “measurement error” of 15–40%. van Hees & Mead (2000) found no increase in precision after three separate measurements, even though the observers discussed the visual assessment after each day’s work. However, in these studies, there was no feedback for the observers, and they had no way of knowing who had the most accurate result. A further disadvantage if there is no way of knowing the “right” answer is the risk of erroneous drifts over time. Thus, the most common criticism of this method lies in it being subjective (Jukola-Sulonen & Salemaa 1985; Tonteri 1990, cf. Dethier et al. 1993). Carlsson et al. (2005) argue that the visual estimation method is often so much quicker than other methods that, for monitoring purposes, one can afford a higher number of plots, thereby increasing the statistical power and compensating for possible lower precision.

Normally, monitoring schemes that require results of the above type provide training for employees in estimating cover, usually including calibration, in order to reduce variation between observers (Nilsson 1992; Jalonen et al. 1998). The most common calibration methods are either to compare estimates among the group and nominate the group mean value as the “correct” result (Floyd & Anderson 1987; van Hees & Mead 2000), or to compare results with a reference value estimated

by an experienced observer. Such calibrations are useful, but provide no method to determine an accurate answer. Another method is to calculate “true” values based on photographs (Bråkenhielm & Liu 1995; Vanha-Majamaa et al. 2000).

The work described here is a part of the environmental monitoring programme, *National Inventory of Landscapes in Sweden* (NILS), initiated by the Swedish Environmental Protection Agency (Inghe 2001; Esseen et al. 2007). The data collection includes both colour-infrared aerial photo interpretation (Allard et al. 2005) and extensive fieldwork in a grid of permanent plots (Esseen et al. 2007). Cover is estimated for groups of species or life forms, which reduces potential problems with species (mis-)identification.

The main aim of this study is to test a computer-aided calibration programme with true cover calculated from digital images, which can be used before as well as during the field season to increase data quality. We investigate if observers can improve their precision and accuracy when given a strictly organised calibration scheme and rapid feedback on correct results. According to Dethier et al. (1993), the most important sources of error in estimating cover are leaf morphology, colour/contrast, aggregation and species identification. By constructing digital images, it is possible to vary such factors independently, in a statistically sound design.

Methods

Many variables may influence precision and accuracy in cover estimation. In this test, we focussed on learning and structural aspects, excluding problems with species identification. Accuracy was evaluated by comparing visual estimation to (i) cover calculated by image analysis in computer-generated images and (ii) point frequency estimates of cover in field plots. Precision was evaluated from among-observer variation for each image or field plot. Six treatment factors were included in the analysis of computer-generated images: experience, calibration time (successive proficiency tests), cover, leaf form, leaf aggregation and background.

Construction of images

The images were designed to simulate the circular sample plots of 0.25 m² used in NILS field inventory (Esseen et al. 2007). This means that the size of the simulated leaves was similar to that of true leaves of comparable species in relation to

the total area of the plot. In both cases, the area is sufficiently small that it is easy for the observer to obtain an overview from one position, without moving around. However, for multilayered vegetation in the field, the eye-view must be moved to also include leaves that are concealed by leaves of other species. No attempt has been made in this study to evaluate the effects of multilayered vegetation on precision and accuracy, and multilayered vegetation was not included in the images.

The effect of differences in leaf form was tested by using images with broad leaves (shoots of *Vaccinium vitis-idaea*) and narrow leaves (leaf blades of grass). *Vaccinium vitis-idaea* shoots were collected and fastened semi-upright in bouquets of various sizes on white paper using Blu-tack. Digital photos were taken of the shoots with a Nikon Coolpix 4500 camera with flash and macro lens from a height of 40 cm. The background was digitally removed from the photographs using Adobe Photoshop (Ver. 7.0). New images (1477×1477 pixels) were constructed using copy-and-paste techniques in Adobe Photoshop. A black circular frame covered each image and defined the estimation area as a circle with a diameter of approximately 1475 pixels. Attempts were made to use photos of different species of grass, but this presented too great a problem for digital cutting; grass images were instead produced digitally with the *dune grass brush* (400–425 pixels) in Adobe Photoshop.

In order to establish the correct cover percentage of the constructed images, an unsupervised classification was performed with ERDAS Imagine (Ver. 8.7) on single-layer tiff images with a white background constructed in Adobe Photoshop. The cover of plants in the image was calculated by dividing the amount of green pixels by the amount of green + white pixels.

Three different backgrounds were used in this trial, one completely white (referred to as *B-white*), one photograph with scattered wooden twigs and a few green shoots of *Anemone nemorosa* on a dark background (*B-dark*), and one photograph with scattered lichens (*Cladonia* sp.) and a tussock of moss (*Polytrichum* sp.) on a grey background of rock (*B-light*).

Design of the computer trial

The trial encompassed four tests, where each observer involved in a test estimated the cover of several images without being given the correct result (hereafter called proficiency test), alternated with three calibration sessions, where observers

would immediately be given the correct result after estimating the cover. The observers were asked not to round up their estimates but to assess the cover as closely as possible to the nearest percentage point. The maximum time for each calibration session was 15 min to ensure a level of uniformity in the trial.

All four proficiency tests contained 24 images with a white background (*B-white*), while the first and last proficiency tests (1 and 4) also contained 24 images with *B-dark* and some *B-light* images. In total, proficiency tests 1 and 4 contained 48 images, whereas proficiency tests 2 and 3 contained 24 images. Each of the three calibration sessions (which were alternated with the proficiency tests, see above) had an equal amount of *B-white*, *B-dark* and *B-light* images; in essence, the calibration sessions consisted of one of each type of picture combining species, background, aggregation and quantity class (true cover 0-32%, 33-65% and 66-100%, respectively), with a total of 36 images in each session.

The test observers consisted of 15 individuals selected on the basis of their previous experience in estimating visual cover. They were evenly split into three categories; *Novices (Group N)*, with no previous experience; *Semi-experienced (Group S)*, with a minimum of 1 week and maximum two seasons of fieldwork; *Experts (Group E)*, with a minimum of three seasons of fieldwork or similar. The observers were of mixed ages, genders and occupations. Some of the novices had no background in biology or natural sciences.

Design of the field trial

To test the validity of the calibration method for cover estimations in the field, nine observers performed a trial with cover estimation in 0.25-m² circular plots, before and after calibration. The observers performing the field trial were all experienced field personnel who had used the calibration programme as a part of their training. Before the field trial, they were instructed not to use the calibration programme for a month, so as to make a comparison of before- and after-effects possible. That the observers had used the calibration routines before means that some of the differences that may otherwise have been caused by differences in training or tradition were probably reduced. Therefore, the results of the trial are, to some extent, conservative, in that it would be expected to show less effect of calibration than if the observers had never used the calibration routines before.

Two species groups were chosen: (i) Broad-leaved: dwarf shrubs of the genus *Vaccinium*

(*V. myrtillus*, *V. vitis-idaea* and *V. uliginosum*) and (ii) Narrow-leaved: graminoids with flat leaves (including *Carex* spp., *Luzula* spp., *Calamagrostis* spp., *Deschampsia cespitosa*, *Molinia coerulea*, *Milium effusum*, but excluding species with thread-like leaves such as *Festuca ovina*, *F. rubra*, *Nardus stricta* and *Deschampsia flexuosa*). The field trial was performed in subalpine heaths and grasslands in the Hemavan area, in northwest Sweden, in August 2007.

For each species group, nine plots were inventoried. To avoid introducing any errors caused by the observers remembering the estimated values from the first round, two sets of plots were laid out within the same area. The plots were selected to contain as similar vegetation and range of cover values as possible, between 5% and 67% cover. The plots were randomly assigned to the observers, and the observers visited the other plots on a second occasion. Between the first and the second round of field estimations, the observers used the same calibration images and routines (excluding the proficiency tests) as in the computer trial. All three of the calibration sessions were performed with 36 images in each session.

In the field plots, the “true” value of cover for each species group was estimated with the help of a point frequency frame with 100 points. To avoid bias and to minimize disturbance of the vegetation, a device sending out a narrow, vertical laser beam was placed in the frame at each point. The laser beam had a diameter of ca. 3 mm. If the centre of the beam (as narrowly defined as possible) hit any part of the shoot of a plant in the species group, it was counted as a hit. The plant part of the first hit was then carefully bent aside, so that the beam had the possibility to hit any plant shoot of the other species group included in the test. Each of the two species groups was counted no more than once for each of the 100 points in the plot. This procedure was used to make the estimates comparable with those of the images, in which only one species was included in each image. The vegetation in the subalpine heath where the trial was performed is low and not very layered.

Data analysis

Data analysis was based on difference between “true” cover and visually estimated cover. Two separate Type III GLM-ANOVAs were calculated for subsets of the total dataset to achieve a balanced design for the analyses. *Leaf form*, *Aggregation*, *Experience*, *Background* and *Test* were treated as fixed factors, and *Observer* (nested within *Experience*) as a random factor. The effects and interactions of

Table 1. Type III GLM-ANOVA with the difference between estimated and “true” cover as the dependent variable, tested for all main factors (except *Background*) and all two-way interactions that include *Experience* and *Test*, for all four proficiency tests. Only images with a white background are included. *Observer* (nested within *Experience*) is treated as a random factor, and all other factors as fixed.

Source	<i>F</i>	<i>P</i>
Intercept	2.911 _{1,12}	0.114
Leaf form	49.806 _{1,1404}	<0.001
Aggregation	1.010 _{1,1404}	0.315
Experience	2.517 _{2,12}	0.122
Test	72.189 _{3,1404}	<0.001
Observer (Experience)	11.025 _{12,1404}	<0.001
Leaf form×Experience	4.929 _{2,1404}	0.007
Leaf form×Test	4.029 _{2,1404}	0.007
Aggregation×Experience	8.245 _{3,1404}	<0.001
Aggregation×Test	7.739 _{3,1404}	<0.001
Experience×Test	5.092 _{6,1404}	<0.001

Table 2. Repeated contrasts for proficiency tests within Type III GLM-ANOVA (cf. Table 1), with the difference between estimated and “true” cover as the dependent variable. Hypothesized value of contrast = 0.

	Contrast estimate	<i>P</i>
Test 1 versus Test 2	-5.056	< 0.001
Test 2 versus Test 3	-1.622	0.003
Test 3 versus Test 4	-0.669	0.226

these factors on difference were tested for all images with white background for all four proficiency tests (24 images for each; Table 1). This analysis included *repeated contrasts*, in which the changes in difference between successive proficiency tests (coded as a factor with four levels) were tested (Table 2). A separate GLM-ANOVA was calculated for images of all three backgrounds. As *B-light* images were created only for the quantity class with true-cover interval 33-65% in proficiency tests 1 and 4, only images of the same tests and quantity class were selected for all three backgrounds. This yielded the same number of images (16) in each class (cf. Fig. 4). In both analyses, two-way interactions with all combinations including *Experience* and *Test* were included, for all independent variables except *Observer* (cf. Table 1).

GLM-ANOVAs were also performed with the difference calculated from arcsine (square root) and logit transformed cover values, but neither of these transformations changed the results more than marginally, and none of the significance levels differed from the analysis based on non-transformed values.

For each observer in the field study, the estimated cover was compared with the point frequency

cover for each plot using paired *t*-tests. This comparison was made separately for estimates before and after calibration. To test the relationship between precision and “true” cover (cf. Figs 2 and 5), GLM analyses of covariance were performed with *Coefficient of variation* among observers per plot as a dependent variable, *Test* (before and after calibration) as a fixed factor and *True cover* as a covariate. To investigate any dependence of using the same images (but in different order) on proficiency tests 1 and 4, analyses of difference between the two tests were compared with and without repeated measures (*Linear Mixed Model*, SPSS 14.0).

The dependent variable fitted well to a normal distribution. Levene’s test for homogeneous variances showed a significant deviation for the main effect+two-way interactions model ($F_{239,1920} = 2.560$; $P = <0.001$). A significant deviation also occurred for arcsine and logit transformed data. However, there was no visible relation between the error variance and the mean, neither for transformed nor for untransformed data. The analyses were therefore considered robust to the assumption of homogeneous variances, since sample sizes are equal between groups in both analyses. However, caution should be used in accepting *P*-values close to 0.05.

Results

Computer trial

All test groups underestimated the cover of both narrow and broad leaves during the first proficiency

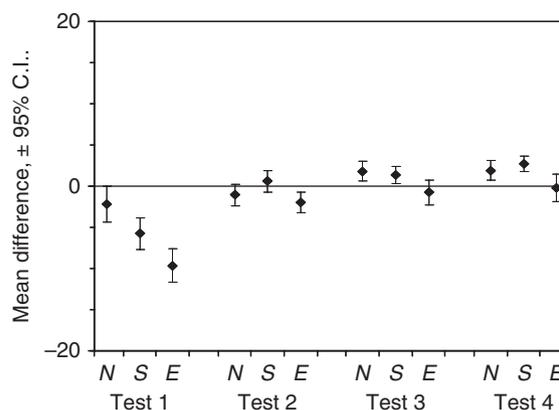


Fig. 1. Mean difference between true and estimated vegetation cover ($\pm 95\%$ confidence interval) in relation to observer experience. Images with a white background are included, for the whole range of “true” cover, for all four proficiency tests. *N*, novice; *S*, semi-experienced; *E*, experienced. Tests 1-4 are the four successive proficiency tests, which were alternated with calibration sessions.

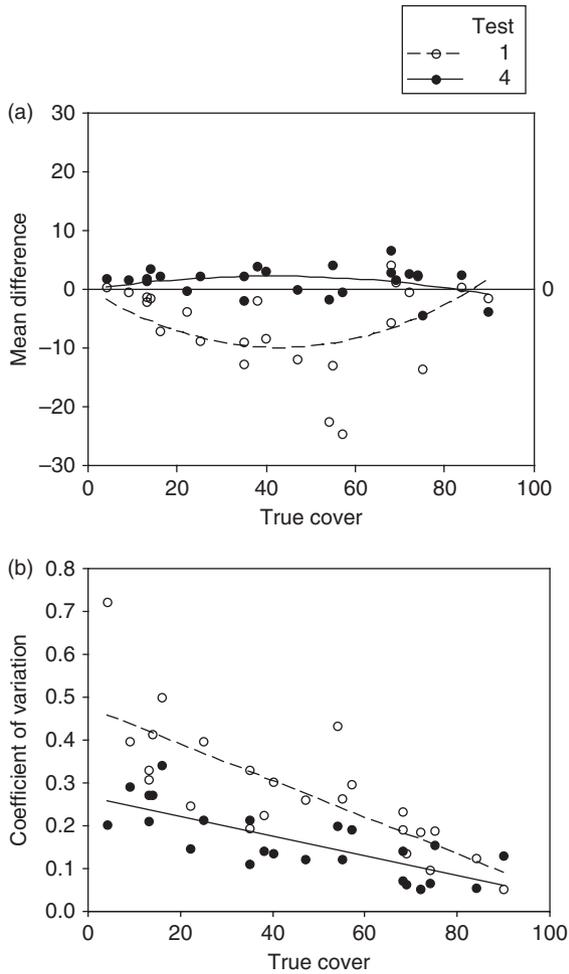


Fig. 2. Mean difference between true and estimated vegetation cover and coefficient of variation (*CV*) of the difference for each image (average among 15 observers) in relation to “true” cover, for proficiency tests 1 and 4. Images with a white background are included. Fitted trend lines of (a) are quadratic with intercept = 0. (a) Mean difference; (b) coefficient of variation.

test (Fig. 1). *Group N* had the least amount of underestimation, while *Group E* had the highest underestimation. Already after the first calibration session, all groups showed a substantial improvement in accuracy, i.e. a decrease in underestimation (Fig. 1; Table 2). From test 1 to test 2, there was a marked change in accuracy (as indicated by the difference between “true” cover and visually estimated cover), whereas the change between test 3 and test 4 was small and non-significant (Fig. 1; Table 2). The difference between groups was largest at intermediate cover of 40-60% (Fig. 2a).

The effect of *Test* (before and after calibration) on *CV* was highly significant ($F_{1,44} = 22.7$; $P < 0.001$) in an analysis of covariance in which *True cover* as a covariate was also significant

($F_{1,44} = 59.2$; $P < 0.001$). *CV* was almost halved after calibration (Fig. 2b). However, there was a significant interaction between *Test* and *True cover* ($F_{1,44} = 5.43$; $P = 0.024$) in the model, which indicates that the slopes of the curves differ. The fitted trend line for *Mean difference* versus *True cover* is presented as quadratic fit with intercept = 0 (Fig. 2a). This is because r^2 was considerably higher for quadratic fit than for linear fit (*Test 1*: 0.562 versus 0.311; *Test 4*: 0.293 versus 0.133). However, for *CV*, the r^2 was very similar for quadratic and linear fits (*Test 1*: 0.603 versus 0.599; *Test 4*: 0.603 versus 0.580), so in that case a linear fit was used (Fig. 2b). The analysis does not show if the fit for *True cover* of the quadratic model is significantly better than the linear model, or if it is just an effect of adding another parameter. However, it is obvious that this pattern is much stronger for *True cover* than for *CV* (where the additional parameter adds little to r^2), and this conclusion of a unimodal relationship for *True cover* and a linear relation for *CV* is also supported by a visual inspection of Fig. 2a and b.

The results indicate that narrow leaves are more difficult to estimate than broad leaves, and that this is true for both images with scattered and images with clustered leaves. Experienced observers underestimated cover most strongly (Fig. 3a). Cover in scattered-leaf images was more difficult to estimate both for narrow and broad leaves. However, after calibration, every test group strongly decreased their estimation error (as indicated by mean difference) in all categories of *Leaf form* and *Aggregation* (Fig. 3b). Initially, the underestimation was larger for images with a white background (*B-white*) and lower for the darker, more heterogeneous images (*B-dark*) (Fig. 4a). However, after calibration there seemed to be no substantial difference between the three backgrounds of the images. Consistent with other factors in the study, all groups decreased their estimation error after calibration (Fig. 4b).

In the test of main effects of independent variables, all factors except *Aggregation* and *Experience* were strongly significant (Table 1). Finally, all five two-way interactions including *Test* and *Experience* were significant (Table 1), which indicates that time for calibration and observer experience influence the effects of other factors. In the second GLM-ANOVA including *Background*, similar results were obtained for most factors. For *Background*, the main effect was significant ($F = 22.254_{2,686}$; $P < 0.001$), as well as the interactions with *Experience* ($F = 2.398_{4,686}$; $P = 0.049$) and *Test* ($F = 3.067_{2,686}$; $P = 0.047$). Other results that differed compared to the analysis without the

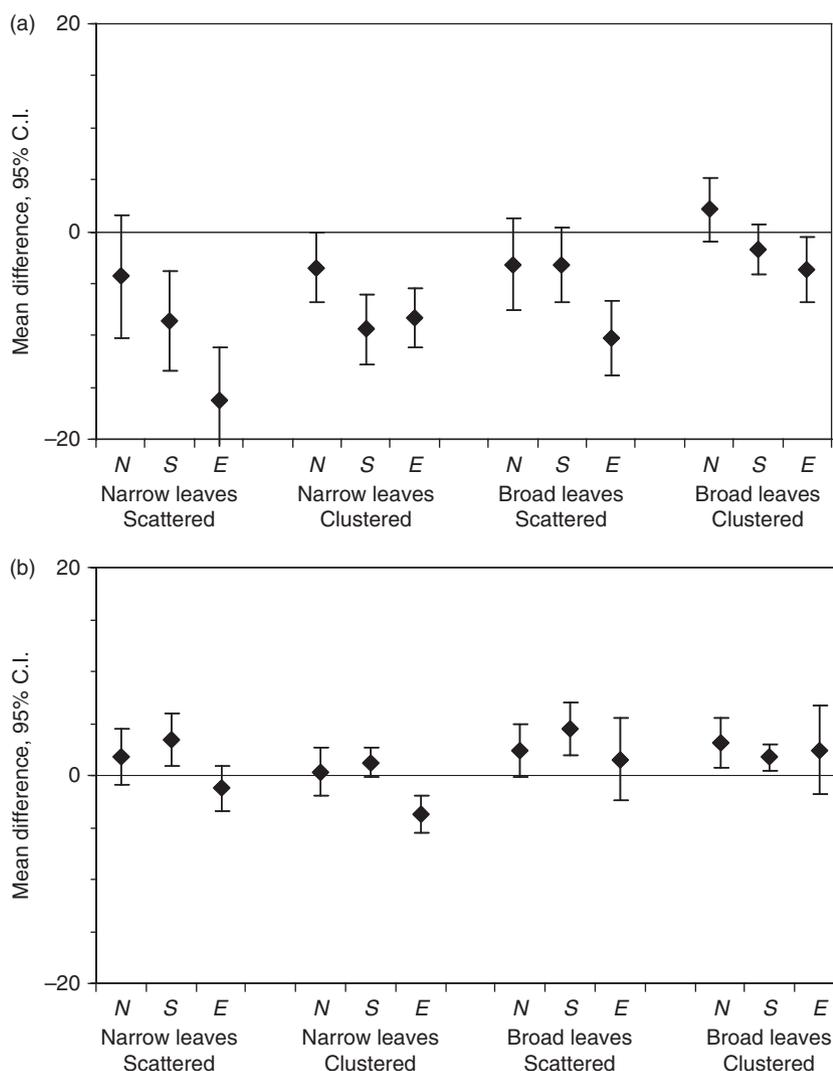


Fig. 3. Mean difference between true and estimated cover, depending on aggregation level and leaf form, for proficiency tests 1 and 4. *N*, novice; *S*, semi-experienced; *E*, experienced. Images with a white background are included for the whole range of “true” cover. (a) Proficiency test 1; (b) proficiency test 4.

Background factor (Table 1) demonstrated that the main effect of *Aggregation* was significant ($F = 5.876_{1,686}$; $P = 0.016$), whereas the interaction *Leaf form* × *Experience* was not significant.

For proficiency tests 1 and 4 (but not for other tests or calibration sessions), the same images were used, but in a different order. However, the improvement on the information criteria and the standard error of including repeated measures was small (2% lower for Akaike’s information criterion), which means that there was no strong effect of dependence between the two tests because of this. The effect of *Test* was strongly significant in both cases. Furthermore, the covariance between tests was considerably less than the covariance within any of the tests. Therefore, we considered the improve-

ments when using a more complex model including repeated effects to be negligible.

Field trial

Visually, the calibration seemed to change the general level of estimated cover in the field, if compared with point-frequency estimates for the same plots (Fig. 5a and b). An average underestimation of -3.2% before calibration changed to an overestimation of $+5.0\%$ after calibration. This can be compared to the results of the computer trial, where the average underestimation of -7.1% changed to an overestimation of only $+1.6\%$ after calibration. However, neither *Test* nor the linear or quadratic

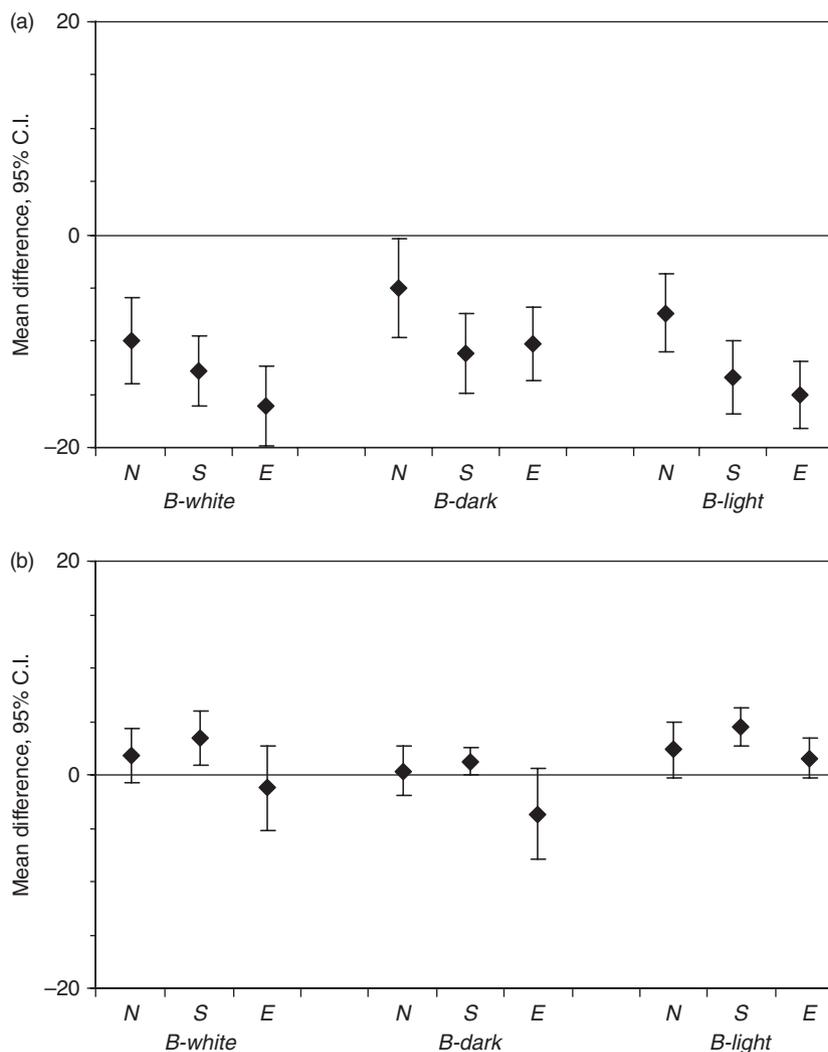


Fig. 4. Mean difference between “true” and estimated cover, depending on the photo background: white (*B-white*), dark background with branches (*B-dark*) and light image with lichens (*B-light*) for proficiency tests 1 and 4. *N*, novice; *S*, semi-experienced; *E*, experienced. Only images within the “true” cover interval 33-65% are included. (a) Proficiency test 1; (b) proficiency test 4.

curve fits had any statistically significant effect (Fig. 5c and d).

For individual observers in the field trial before calibration, two observers significantly underestimated cover compared with the point frequency estimates for narrow-leaved plants (not shown), whereas for broad-leaved plants, four observers significantly underestimated and one overestimated cover. After calibration, all significant differences for individual observers were overestimations in relation to the point frequency estimates: three for narrow-leaved plants and one for broad-leaved plants.

Similar to estimations based on computer images, the *Coefficient of variation (CV)* for the difference between visually estimated and “true”

cover (estimated with point frequency) between observers and within species groups decreased almost linearly with plant cover (Fig. 5a and b). The effect of *Test* on *CV*, with point-frequency cover estimate (*True cover*) as a covariate, was strongly significant for broad-leaved species ($F_{1,15} = 18.3$; $P < 0.001$; Fig. 5b), but not for narrow-leaved species ($F_{1,15} = 0.192$; $P = 0.668$; Fig. 5a). The interaction between *Test* and *True cover* was not significant, and therefore was removed from the model. The average *CV* for narrow-leaved plants was 0.270 and 0.259 before and after calibration, respectively (Fig. 5a). However, the decrease of *CV* after calibration for broad-leaved plants was stronger, with a decrease in average *CV* from 0.296 to 0.203, i.e. a decrease in *CV* of more than 30% (Fig. 5b). The relative change

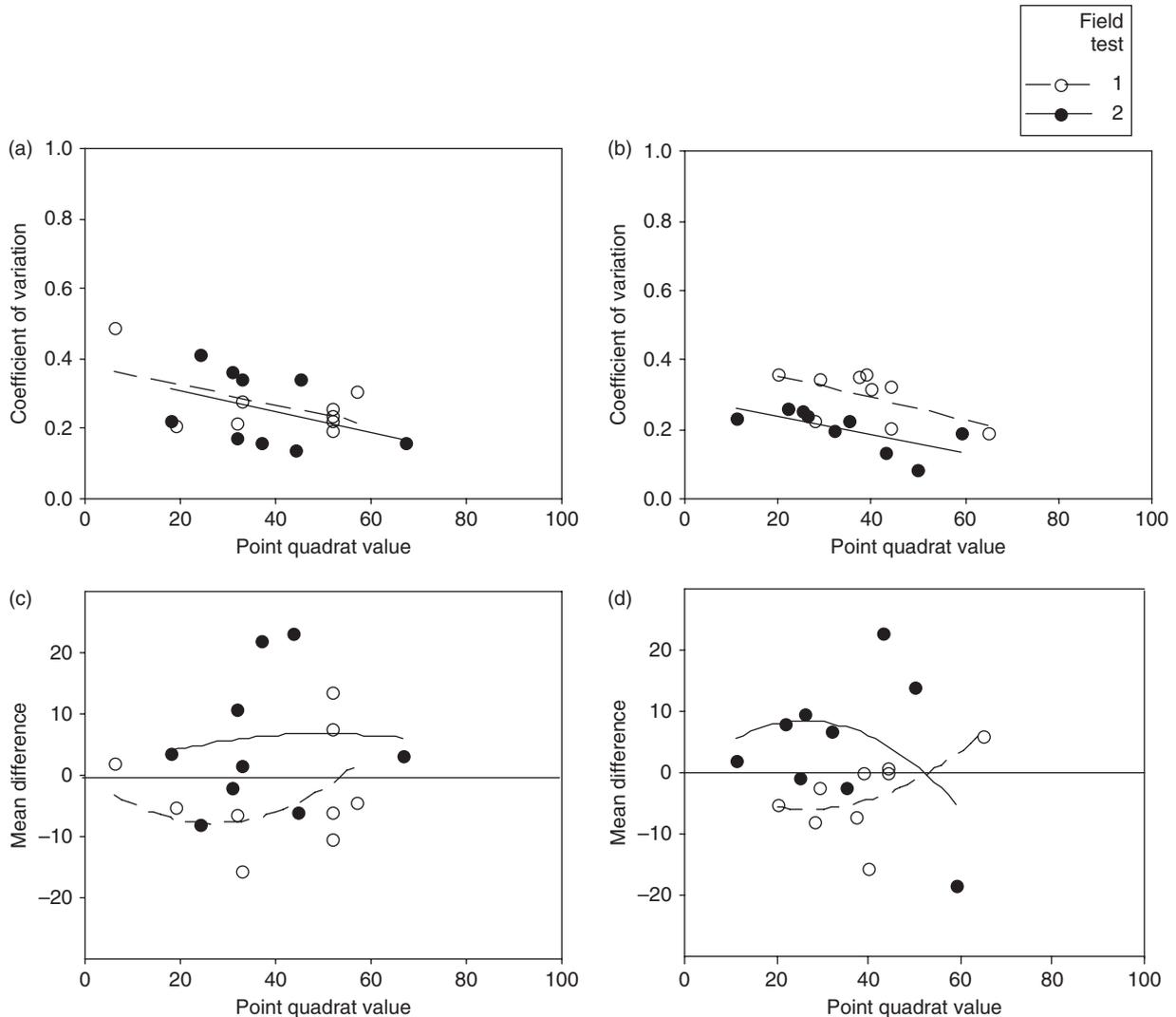


Fig. 5. Mean difference between visually estimated cover in the field and “true” cover estimated from point frequency sampling in the same field plots, and coefficient of variation (CV) among observers before (Field test 1) and after (Field test 2) computer calibration. (a) CV for narrow-leaved species; (b) CV for broad-leaved species; (c) Mean difference for narrow-leaved species (graminoids); (d) mean difference for broad-leaved species (*Vaccinium* dwarf shrubs).

in CV at 50% cover (calculated from the regression equation) was even more marked: from 0.258 to 0.158, which is a decrease of almost 40%.

Discussion

The results show that even a relatively short time spent on calibration can considerably reduce both systematic and random errors in cover estimates. Initially, the difference in estimates between “true” cover and the group mean was between -10% and -3% , and after calibration this changed to between 0% and $+3\%$. Inter-observer

variation is important in all method evaluations, and is a good indicator of precision. In the study of Sykes et al. (1983), differences between observers were always significant, and Dethier et al. (1993) showed that between-observer variation was greater than within-observer variation. In the present study, we found that the between-observer variation (estimated as coefficient of variation) decreased substantially after calibration, both in the computer trial and in the field trial, and was in almost all cases much lower than the mean CV of 50% reported by Tonteri (1990) and Klimeš (2003). These cover estimates were made for species groups or individual dominant species in species-poor vegetation,

whereas for sparse species in species-rich vegetation, the CV is expected to be higher. However, Klimeš (2003) reported that the CV is about 50% even for very abundant species, which would correspond more to the situation in this study. In fact, the expected higher reliability of the visual cover estimations for species groups and life forms of field layer plants, instead of for individual species, is one reason why the life-form approach is used for estimating cover in the field and bottom layer in the NILS monitoring programme (Esseen et al. 2007). With efficient tools, improved performance in estimating cover can obviously be acquired in a very short time. Furthermore, the current computer programme was self-explanatory and easy to use, so that even inexperienced observers could start to use it efficiently with the help of only very short written instructions.

Observer experience

The results clearly demonstrate that experienced observers without calibration tended to have the largest estimation errors (proficiency test 1). One possible explanation for this could be that unsubstantiated advice on visual estimation might be imprinted in experienced field observers. In field training in Sweden, observers are frequently reminded that some grass species consist of very narrow leaves with very low total cover. Consequently, experienced observers might provide lower estimates of narrow leaves to be on the safe side. Inexperienced observers have no prejudices as to how different species types should be treated. Also, according to Dethier et al. (1993), inexperienced observers do not produce results of significantly lower quality than experienced observers.

True cover

The highest discrepancy in estimation (absolute error) occurs where there is intermediate cover, especially between 40% and 60% true cover. This is consistent with the results of Sykes et al. (1983), who estimated that the most extensive discrepancies would occur in the 50% cover region and be less at the two extremes. Consequently, the coefficient of variation (relative error) decreases with increasing cover. This means that it is of little use to provide a statement on expected coefficient of variation unless the range of cover values is defined. The results of this study are not consistent with the results of Jukola-Sulonen & Salemaa (1985), who found that observers tend to overestimate low cover and underestimate high cover.

In the current study, cover was estimated to the nearest percentage point. This method is also recommended by Tonteri (1990) and Carlsson et al. (2005), see also Bråkenhielm & Liu (1995). With an expected standard deviation after calibration of only 6.2%, there is no reason whatsoever to introduce new sources of error by using crude class estimates. Our experience, from several years of field surveys, is that observers feel comfortable with these estimation routines after a very short time of practise.

Leaf form, aggregation and background

As for the impact of pattern on estimation accuracy, all interactions with *Test* were significant, which indicates that the accuracy (amount of under- or overestimation) changes after calibration. Narrow-leaf images resulted in a high degree of underestimation, which may be attributed to the reasons discussed above. It is reasonable to assume that with even narrower, thread-like leaves, estimation errors would have been even greater. According to Kennedy & Addison (1987), species that are easily seen are easiest to estimate. The broad-leaved *Vaccinium* plants used in this study clearly fall into this category. Kennedy & Addison (1987) found that when observers increased their familiarity with the vegetation, precision of sampling increased. In their study, species identification was an important factor, which adds another source of uncertainty. It may be important to evaluate these two aspects separately.

Aggregation is known to be highly important in field estimation (e.g. Dethier et al. 1993). In this study, there is a clear tendency that underestimations are less likely for images with clustered leaves than for those with scattered leaves. After calibration, the difference between aggregation types decreased substantially.

It could be hypothesised that a more heterogeneous background would increase estimation error in terms of both accuracy and precision, or that more contrast would lead to overestimation of cover. None of these hypotheses are supported by the results, since underestimation before calibration is strongest for images with a white background. Also for *Background*, the results indicate that computer calibration can be used to improve accuracy and precision of cover estimations.

Comparing computer and field estimations

The deviating effect of computer calibration on precision of field estimations, as estimated by the

coefficient of variation, for graminoids (narrow leaves) and *Vaccinium* dwarf shrubs (broad leaves) is not so easy to explain. One possible reason is that the group of the three *Vaccinium* species in the broad-leaf class is more homogeneous, with similar leaf form and size. The graminoid group in the field plots, on the other hand, consists of several species with varying leaf size and form. For future studies, more effort should be spent to increase the realism of the digital images.

The effects of computer calibration on the accuracy of field cover estimates are somewhat ambiguous. In the field trial, the underestimation before calibration changed into a slightly larger overestimation after calibration, in comparison to the “true” values of the point frequency measurements. The reasons for this are not known. However, the precision (as estimated by the coefficient of variation) for broad-leaved species was very similar in the field and the computer trial, both before and after calibration. This indicates that for easily seen species and life forms, there may be no fundamental differences between estimating cover on a computer screen and in the field. The strong effect of decreasing between-observer variation for broad-leaved species is encouraging, and should inspire further studies on how to increase the efficiency of calibration routines. Since there are no other known field methods that can compete with the time-efficiency and flexibility of visual cover estimations for large-scale vegetation monitoring, this is indeed an urgent task.

Obviously, it is much more difficult to compare the effects on accuracy than on precision in the field, because it is more difficult to obtain a true value from field plots, and also the point frequency data are not real “true” cover, but estimates with their own built-in errors (e.g. Goodall 1952; Vanha-Majamaa et al. 2000). One way to avoid this could be to calculate cover from photographs taken in the field, but this is realistic only for shoots and leaves that are either sufficiently broad and clear-cut (manual delineation) or sufficiently different in colour from the background (automatic methods) so that reliable “true” cover can be calculated (Vanha-Majamaa et al. 2000; cf. Bråkenhielm & Liu 1995). In multi-layered vegetation, all such methods are unreliable, but even a simple calibration procedure will make the results more reliable than when there is no calibration at all (cf. Vanha-Majamaa et al. 2000). The sparse and haphazard use of calibration in many field studies may be an important cause for the unjustified poor reputation of visual cover estimation in relation to other methods.

Conclusions

The results from this study show that calibration is important and that it can significantly decrease the error of visual cover estimation. It also seems to work in a short period of time – all test observers had a significant decrease in estimation error even after the first calibration session. This means that this type of calibration can be used frequently during a field season without taking valuable time away from the inventories. Such calibration may also generate increased confidence in the field observers. If the observers are not confident that the results are sufficiently reliable, this can lead to ‘it-doesn’t-matter-what-I-write-nobody-knows-if-it’s-right-anyway’ attitude. The type of calibration suggested here gives very real feedback, which the majority of the observers indicated as a good incentive to try to improve. This boosts confidence as well as increasing the trustworthiness of the results.

Acknowledgements. Many thanks to Göran Ståhl, Per Milberg, Leoš Klimeš, Onno van Tongeren and Zuzana Münzbergová for constructive criticism of earlier versions of the manuscript. Thanks to Liselott Marklund, Åsa Hagner and Sture Sundquist for planning and leading the fieldwork. Above all, thanks to the persons who volunteered for the first calibration trial and to personnel performing the field trial during the Hemavan field excursion.

References

- Allard, A., Nilsson, B., Pramborg, K., Ståhl, G. & Sundquist, S. 2005. *Manual for aerial photo interpretation in the National inventory of landscapes in Sweden (NILS)*. Swedish University of Agricultural Sciences, Department of Forest Resource Management, Umeå, SE.
- Bråkenhielm, S. & Liu, Q. 1995. Comparison of field methods in vegetation monitoring. *Water, Air, and Soil Pollution* 79: 75–87.
- Carlsson, A.L.M., Bergfur, J. & Milberg, P. 2005. Comparison of data from two vegetation monitoring methods in semi-natural grasslands. *Environmental Monitoring and Assessment* 100: 235–248.
- Dethier, M., Graham, E.S., Cohen, S. & Tear, L.M. 1993. Visual versus random-point percent cover estimations: ‘objective’ is not always better. *Marine Ecology Progress Series* 96: 93–100.
- Esseen, P.-A., Glimskär, A., Ståhl, G. & Sundquist, S. 2007. *Field instruction for National Inventory of the Landscape in Sweden, NILS, year 2007*. SLU, Department of Forest Resource Management, Umeå, SE.
- Floyd, D.A. & Anderson, J.E. 1987. A comparison of three methods for estimating plant cover. *Journal of Ecology* 75: 221–228.

- Goodall, D.W. 1952. Some considerations in the use of point quadrats for the analysis of vegetation. *Australian Journal of Scientific Research, Series B* 5: 1–41.
- Inghe, O. 2001. The Swedish landscape monitoring programme – current status and prospects for the near future. In: Groom, G. & Reed, T. (eds.) *Strategic landscape monitoring for the Nordic countries. TemaNord 2001*. 523, pp. 61–67. Nordic Council of Ministers, Copenhagen, DK.
- Jalonen, J., Vanha-Majamaa, I. & Tonteri, T. 1998. Optimal sample and plot size for inventory of field and ground layer vegetation in a mature *Myrtillus*-type boreal spruce forest. *Annales Botanici Fennici* 35: 191–196.
- Jukola-Sulonen, E-L. & Salemaa, M. 1985. A comparison of different sampling methods of quantitative vegetation analysis. *Silva Fennica* 19: 325–337.
- Karlsson, T. 1998. Förteckning över Svenska Kärlväxter. *Svensk Botanisk Tidskrift* 91: 241–560 (in Swedish with English summary).
- Kennedy, K.A. & Addison, P.A. 1987. Some considerations for the use of visual estimates of plant cover in biomonitoring. *Journal of Ecology* 75: 151–157.
- Klimeš, L. 2003. Scale-dependent variation in visual estimates of grassland plant cover. *Journal of Vegetation Science* 14: 815–821.
- Nilsson, C. 1992. Increasing the reliability of vegetation analyses by using a team of two investigators. *Journal of Vegetation Science* 3: 565.
- Økland, T. 1990. Vegetational and ecological monitoring of boreal forests in Norway, SE Norway. *Sommerfeltia* 10: 1–52.
- Sykes, J.M., Horrill, A.D. & Mountford, M D. 1983. Use of visual cover assessments as quantitative estimators of some British woodland taxa. *Journal of Ecology* 71: 437–450.
- Tonteri, T. 1990. Inter-observer variation in forest vegetation cover assessments. *Silva Fennica* 24: 189–196.
- Vanha-Majamaa, I., Salemaa, M., Tuominen, S. & Mikkola, K. 2000. Digitized photographs in vegetation analysis – a comparison of cover estimates. *Applied Vegetation Science* 3: 89–94.
- van Hees, W.W.S. & Mead, B.R. 2000. Ocular estimates of understory vegetation structure in a closed *Picea glauca/Betula papyrifera* forest. *Journal of Vegetation Science* 11: 195–200.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Image for calibration of cover estimation, with broad-leaved plants (shoots of *Vaccinium vitis-idaea*), scattered leaves and a white background.

Figure S2. Image for calibration of cover estimation, with narrow-leaved plants (created by Dune grass brush in Adobe Photoshop), scattered leaves and a light background.

Figure S3. Image for calibration of cover estimation, with broad-leaved plants (shoots of *Vaccinium vitis-idaea*), aggregated leaves and a dark background.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 24 August 2008;

Accepted 9 July 2009.

Co-ordinating Editor: A. Chiarucci.