Linköping University Post Print

In the eye of the beholder: bias and stochastic variation in cover estimates

Johan Bergstedt, Lars Westerberg and Per Milberg

N.B.: When citing this work, cite the original article.

The original publication is available at www.springerlink.com:

Johan Bergstedt, Lars Westerberg and Per Milberg, In the eye of the beholder: bias and stochastic variation in cover estimates, 2009, PLANT ECOLOGY, (204), 2, 271-283.

<http://dx.doi.org/10.1007/s11258-009-9590-7>

Copyright: Springer Science Business Media

<http://www.springerlink.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-20396>

In the eye of the beholder: bias and stochastic variation in cover estimates

Johan Bergstedt¹, Lars Westerberg¹ & Per Milberg^{1,2*}

¹*IFM Biology, Division of Ecology, Linköping University, SE-581 83 Linköping, Sweden*

²*Current address: Department of Crop Production Ecology, SLU, Box 7043, 750 07 Uppsala, Sweden*

**Author for correspondence: per.milberg@vpe.slu.se*

Abstract

Cover estimates by eye is a prevailing method to assess abundance. We examined cover estimates with regard to bias and random variation. Ten observers working with a national forest vegetation survey estimated sixteen 100 m²-plots, placed in two different vegetation types. These had similar species composition but were clearly distinguishable in the field. In species-wise analyses, observer bias varied greatly, with *Dicranum* spp., *Vaccinium vitis-idaea* and *V. myrtillus* having the largest bias. Experience had a surprisingly small impact on variation. Power analysis revealed only small differences between observers in the ability to distinguish the two vegetation types, and little value in averaging the assessments from 2, 3 or 4 observers. Cover estimates did better than presence/absence data in separating the two vegetation types and multivariate analyses were more powerful than univariate ones.

Keywords: forest vegetation, observer bias, statistical power, Sweden, visual assessment

Introduction

Changes in the abundance of plants are often in focus in vegetation ecology. There are a number of approaches when estimating abundance of plants: biomass harvesting, frequency, point-frequency, line-intercept, graphical and cover estimates by eye (Sarukhán and Harper 1973, Kent & Coker 1992, Mueller-Dombois & Ellenberg 2002). Since plants differ in size, the number of individuals is of little relevance in many respects. Biomass might be a more appropriate measure but when change over time in permanent plots is under study, a destructive assessment of biomass is inappropriate. Cover estimates by eye has been most widely used, since it is less time consuming, requires less equipment and is thus less expensive, than other methods. Jukola-Sulonen and Salemaa (1985) also concluded that, in spite of differences between methods, they all resulted in similar abundance relationships between plant species.

Obvious problems with visual cover estimates are that it is a subjective method and that “true” covers are usually unknown. For this reason, some prefer presence/absence recording under the assumption that it is a more objective method (e.g. Økland 1988, Kercher et al. 2003). The reliability of visual cover estimates has been the focus for a number of studies (Sykes et al. 1983, Jukola-Sulonen and Salemaa 1985, Floyd and Anderson 1987, Kennedy and Addison 1987, Mitchell et al. 1988, Dethier et al. 1993, Bråkenhielm & Qinghong 1995, Carlsson et al. 2005, Vittoz & Guisan 2007). Still, the contribution of bias due to individual observers versus other unexplained sources of error has rarely been reported (van Hees & Mead 2000, Milberg et al. 2008).

We wanted to evaluate (i) the relative contribution of random error and bias of the observer error to the variation of cover estimates and (ii) the statistical power to separate two vegetation types, using the observers' individual estimates (to estimate the possibility to improve the data by training, calibration and/or eliminating observers). We also assessed the variation explained (iii) by observer experience and (iv) by inter-observer differences of species estimates (i.e. estimate differences in observer bias between species). Further, we wanted to investigate (v) how the use of more than one observer affected the statistical power to separate two vegetation types as well as (vi) the difference in statistical power of cover estimate data and presence/absence ditto to distinguish the two vegetation types. The study context is monitoring of vegetation in large (100 m²) permanent plots currently conducted in Sweden at 10-year-intervals. The survey has been running since 1983 and covers all land area in Sweden, but a complete

vegetation inventory is only conducted in a limited number of land use classes.

Material and methods

The inventory

The National Survey of Forest Soils and Vegetation (NFSV, <http://www-markinventeringen.slu.se/>) is performed together with the National Forest Inventory (NFI) of Sweden. In addition to the cover of a list of species, the inventory also estimates the area without vascular plants (field layer missing, hereafter referred as FLM) and without soil cover of bryophytes and lichens (bottom layer missing, hereafter referred as BLM). In the NFSV, the number of 100 m² plots visited per working day is eight, and apart from the vegetation records, the field worker also takes soil and humus samples and, if time allows, participates in records of forestry-related variables (Anon. 2003). The personnel executing the inventory are foresters and biologists with a special training in the tasks of the survey. Every year there is a pre-season meeting with training and information about changes since the previous year and also a mid-season meeting for calibration and further training. We had the opportunity to use the observers for one day during their mid-season meeting (July 2005), when they had worked for 5 weeks in the NFSV in the current field season.

Study area and observers

The study area chosen was situated some 20 km east of Mora (61°01'N 14°32'O), central Sweden, in a coniferous forest. Vegetation types resistant to trampling were chosen, to minimize possible wearing of the plots (a feature that we were also able to evaluate).

Half of the area had a canopy dominated by spruce (*Picea abies* (L.) H. Karst.) and the other half by pine (*Pinus sylvestris* L.). The difference between the two adjacent vegetation types consisted more of a difference in abundance than a difference in species composition. The *Picea abies* area was dominated by *Vaccinium myrtillus* L. in the field layer and mosses in the bottom layer, and was slightly more mesic. The *Pinus sylvestris* area had more of *Vaccinium vitis-idaea* L., *Calluna vulgaris* L. and *Empetrum nigrum* (L.) in the field layer and more lichens in the bottom layer.

The ten observers differed in educational background and field experience, the latter ranging from observers doing their first season to one observer doing his 26th (Table 1).

Table 1. *Number of seasons in the NSFV per observer*

Observer	Number of seasons
1	26
2	11
3	9
4	7
5	6
6	5
7	4
8	2
9	1
10	1

Study design

The outline of the fieldwork in the current study was as similar as possible to the regular work in the NSFV. The sample plots were circular with a radius of 5.64 m (i.e. 100 m²). Cover was estimated as the vertical projection of each of the species and group of species. The cover is estimated as the cover of all above ground, living parts of the plants, and recorded in classes of 1 m², except for the lowest class which is between 0 and 0.1 m². The total sum of the cover estimates for a given plot usually exceeds 100 m² since the species are independently assessed and partly overlap in cover.

To avoid possible deterioration of data due to exhaustion, the study was performed during two consecutive half-days, that is, one afternoon followed by the next morning. In this respect, the study differed from the regular work in the NSFV where observers usually estimate a maximum of eight plots in the course of a full day's work. Other differences were the concentration on cover estimates while in regular work observers perform other tasks as well, the use of paper for recording instead of digital media and the density of sample plots thus, making the estimates of cover closer in time.

Eight plots were placed in each of the two vegetation types. The plot center in the 16 plots was marked with a pole with a 5.64 m string attached. At the center of the sample plots ten protocols were placed, numbered 1-10 giving the order in which the estimates of the plots were made, making it possible to evaluate any possible trampling effect. To evaluate the possible tiring of the observers, they numbered their estimates from 1-16 according to the order in which they estimated the plots. To avoid waiting periods, the observers were allowed to choose the order in which to do the plots, that is, when finished with one plot they looked for the next plot without observer.

The result was that the order for an observer in which the plots was visited was not independent nor was the visiting order for the observers of a particular plot.

Analyses

First, we merged records of cover class 0.1 m^2 ($n=412$) and 1 m^2 ($n=371$), in order to have all cover classes of equal size for subsequent analyses.

Visual analyses

In order to illuminate whether difficulties in estimating cover differs between groups of species and/or specific cover ranges, diagrams with coefficient of variation (standard deviation/mean) and mean cover were done. The mean estimates for each species per plot were grouped in four groups: (i) FLM and BLM missing, (ii) lichens, (iii) bryophytes and (iv) dwarf-shrubs. For each group, the mean cover per species and plot was plotted against coefficient of variation. Plots where one or more observers had missed the species were excluded, i.e. we focused on how observers differed in their actual estimates of cover, without confounding this with possible errors in actually finding the species.

Univariate analyses

For the evaluation of species-specific cover estimates, only plots where at least two observers had found the species were included. Mean cover and standard deviation per plot and species were calculated. To assess the variation dependent on observer, residuals from a one-way ANOVA, with plot identity as explanatory variable, were used. The residuals were subjected to species-wise variance component analyses with “observer” as a random variable.

Each observer’s estimate of species cover was compared to the mean of the other observers. The purpose was to show to what extent an observer was biased. For this, a linear regression was done for each species and observer where each plot ($N=16$) contributed two observations: the target observers estimate (dependent variable) and the average estimate of the other nine observer (independent variable). The slope indicated whether the observer generally used a larger ($r>1$) or smaller ($r<1$) span of cover records of a specific species, relative to the other observers. Species for which variation in cover estimates spanned $<10\%$ were not included in this analysis.

Multivariate analyses

Multivariate analyses were done with the CANOCO 4.5 software (ter Braak & Šmilauer 2002) using default options (i.e. no standardization and no transformation). Two methods were used: Principal Component Analysis (PCA) and Redundancy Analysis (RDA). These so called linear methods were chosen because of low beta-diversity in the data (Lepš & Šmilauer 2003).

To assess the magnitude of exhaustion of the observer and wearing of the plots, two partial redundancy analyses (pRDA) with alternately inventory order (i.e. in which order an observer visited the plots) and plot order (i.e. the order a particular plot was visited by an observer) as explanatory variables and all other variables (plot identity, observer identity and plot order or inventory order) as covariables.

To quantify the contribution of plot and observer in the explanation of species composition, three separate pRDAs were executed. The first, with plot as explanatory variable and observer as covariable, the second, with observer as explanatory and plot as covariable and the third, with both plot and observer as explanatory variables.

To evaluate the impact of experience, we ranked the observers from 1 to 10 based on the number of seasons of field work. Then, a pRDA with experience as explanatory variable and inventory order, plot order and plot identity as covariables was performed.

For the comparison of cover estimates by eye and presence/absence data we did two principal components analyses (PCA) with inventory order and plot order as covariables. Using data from all 16 plots and individuals observer, in total 160 samples, the first PCA was conducted using cover data, the other using the same data transformed to presence/absence. These PCAs evaluates the multivariate variation of individual sample plots as well as the potential to separate the two vegetation types sampled.

Power analyses

All power analyses were made in R (R Development Core Team 2006) and with the "vegan" package (Oksanen et al. 2006). We estimated the power to reject a null hypothesis, that there is no difference between our two vegetation types, for two general types of data sets: cover of individual species (univariate), and when using the complete species list (multivariate test). Power was assessed by resampling subsets of the data (bootstrapping). A random selection of plots ($n=3,4 \dots 8$ from each vegetation type) was drawn 100 times with replacement. First, power for each observer at sample size n was estimated for a multivariate RDA with vegetation type as a

categorical explanatory variable. Significance was assessed by choosing the targeted “critical” p-value to be 0.001 (see “anova.cca” in the “vegan” package), which in this context gives the same result as using 999 free permutations. Power was calculated for each n as the proportion of tested subsets with a p-value ≤ 0.005 . Secondly, power for mean cover in resample data for 1 to 4 randomly selected (with replacement) observers was assessed using the same RDA scheme. Here, mean cover estimates for the observers in the n selected plots was used. Thirdly, power for cover data and presence/absence data was estimated using similar RDA scheme as above. Subsets were created by randomly selecting an observer for each replicate and then analyzing the cover data and the presence/absence-transformed data. The difference between the data sets was illustrated by compiling p-values in a box-plot. Finally, power for using single species and a t-test was assessed for *Cladina* spp., *Pleurozium schreberi*, *Hylocomium splendens*, *Vaccinium vitis-idaea* and *Vaccinium myrtillus*. These species were chosen because they had the highest occurrence in the data. Again, observer was randomly chosen for each replicate and the p-value from a t-test for each of the five species was calculated.

Results

The order in which an observer visited the plots (order for observer) and the order in which a plot was visited by an observer (order on plot) were inherent features of the study design. They made a small, but significant, contribution to the variation in vegetation description by pRDA (Table 2).

The difference between the two vegetation types was more clearly illuminated in multivariate (Figure 1c) than in univariate analysis (Figure 1d). In a univariate analysis, the species with highest occurrence, *Vaccinium myrtillus*, contributed most to the power in separating the separating the two (Figure 1d).

Table 2. Results from two pRDAs using plot identity and order on plot or order for observer as covariables. The analyses were made on original cover estimates.

	Eigenvalue	P-value
Order on plot	0.009	0.0014
Order for observer	0.005	0.0242

Figure 1. Statistical power to detect the known difference between the vegetation types. Power at $\alpha=0.005$ was evaluated by resampling subsets of the data (bootstrapping) for different n , using four different types of analyses. 1a; power for each observer at sample size n was estimated for a multivariate analysis (RDA) with vegetation type as the only explanatory variable. 1b, power for mean cover in resample data for 1 to 4 randomly selected (with replacement) observers was assessed using the same RDA scheme but with the mean for cover estimates for the observers. 1c, power for cover data vs. presence/absence (broken line) data was estimated using similar RDA scheme as above. Subsets were created by randomly selecting an observer for each replicate and then analyzing the cover data and the presence/absence transformed data. Finally 1d, power for using single species and a two independent sample t -test was assessed for four species and one genus. Again, observer was randomly chosen for each replicate and the p -value from a t -test for each of the five species was calculated.

Random error and bias

In pRDA, plot explained 5 times more (53 %) of the variation in the data than observer (10.5 %). The joint effect of observer and sample plot explained little of the variation (0.3 %).

In univariate analyses, the bias connected to observer showed a relative contribution between 8.2 and 47.8 percent (of total variance in residuals) for species, between 1.2 (not significant) and 50.9 for groups of species, 55.2 percent for estimates of BLM and 67.2 percent for FLM (Table 3).

Inter-observer differences in statistical power

In RDA, the statistical power of individuals' estimates, to separate between the two vegetation types, exhibited small differences between observers, reaching an acceptable power for all observers at the same sample size (Figure 1a).

The effect of experience on variation in estimates

Although the experience of the observer had a significant effect in a pRDA ($p=0.0001$), the variation explained by experience was low (1.8 %). Increased experience of the observer was correlated with lower estimates of BLM and higher estimates of *Cladonia* spp. cover; for other species and groups of species experience had a low effect on the estimates of cover (Table 4).

Inter-observer differences in species records

Among the dwarf shrubs, the highest observer-dependency of the cover estimates was found for the most common species, *Vaccinium myrtillus* and *V. vitis-idaea*, in contrast with other dwarf-shrubs like *Calluna vulgaris* and *Empetrum nigrum* (Table 3). Among bryophytes, the genus *Dicranum* had the highest observer-dependent variance, while the genus *Sphagnum* had the lowest. The species *Hylocomium splendens* had a relatively high variance attributed to observer while *Ptilium crista-castrensis* and *Pleurozium schreberi* had a lower value (Table 3). The highest observer-dependent variance was exhibited by the missing layers: FLM and BLM (Table 3).

All groups (FLM and BLM, lichens, bryophytes and dwarf-shrubs) had higher coefficient of variation in estimates at low cover and lower variation at higher. At low cover, the variation in dwarf-shrubs estimates was lower than for the other three groups (Figure 2).

There were differences in the observers' use of the range of possible estimates. The mean of the slope of the linear regression of all species per observer with confidence interval included the mean of the other observers'

Table 3. Species noted, and their characteristics in the data set. Variance component analyses, evaluating the relative contribution of observer bias and random error per species, were conducted on residuals from an ANOVA using plot identity as explanatory variable. Single occurrences (recorded by a single observer) are likely erroneous records, that were not found by the authors when searched for. Missed occurrences refer to cases where at least two observers have noted the species. - = not calculated. FLM is field layer missing and BLM is bottom layer missing.

	Cover				Number of single occurrences	Missed occurrences (%)	Number of registrations (max 160)	Plot occurrences (max 16)	Observer bias (%)	P-value
	Average	SD	Min	Max						
FLM	46.2	15.03	7	72	0	0	160	16	55.2	0.000001
BLM	5.5	6.00	0	37	0	0	160	16	67.6	0.000001
<i>Cetraria</i> spp	2.5	2.16	0	11	0	8.9	82	9	20.9	0.001502
<i>Cladina</i> spp	6.3	6.39	0	26	0	4.4	153	16	13.3	0.000667
<i>Sphagnum</i> spp	12.6	22.79	0	87	0	17.0	84	11	1.2	0.342770
<i>Dicranum</i> spp	4.8	5.64	0	35	0	4.4	153	16	50.9	0.000001
<i>Ptilium crista-castrensis</i>	0.91	2.31	0	22	1	53.3	71	16	8.2	0.012937
<i>Pleurozium schreberi</i>	41.5	25.76	2	96	0	0	160	16	8.2	0.000034
<i>Hylocomium splendens</i>	30.3	25.33	0	89	0	4.4	153	16	20.7	0.000004
<i>Calluna vulgaris</i>	3.8	3.46	0	18	1	7.7	121	14	11.2	0.005330
<i>Vaccinium vitis-idaea</i>	19.4	9.45	3	49	0	0	160	16	47.8	0.000001
<i>Vaccinium myrtillus</i>	32.9	16.92	6	81	0	0	160	16	39.6	0.000001
<i>Empetrum nigrum</i>	2.4	3.13	0	16	2	16.7	77	11	17.6	0.001262
<i>Melampyrum pratense</i>	0.86	0.43	0	2	0	17.5	132	16		
<i>Deschampsia flexuosa</i>	0.90	0.44	0	2	0	15.0	119	14		
<i>Cladonia</i> spp	0.60	0.49	0	1	0	40.0	96	16		
<i>Cladina stellaris</i>	0.83	1.06	0	6	4	22.5	66	12		
Other lichens	0.53	0.50	0	1	0	47.0	80	15		
Other bryophytes	0.30	0.47	0	2	0	63.6	44	15		
<i>Linnaea borealis</i>	1.4	1.28	0	5	0	23.3	23	3		
<i>Peltigera aphthosa</i>			0	1	1	-	1	1		
<i>Stereocaulon</i> spp			0	5	1	-	1	1		

Table 4. Species scores and experience-explained variation from a pRDA with “experience” as explanatory variable and plot identity, order on plot and, order for observer as covariables. FLM is field layer missing and BLM is bottom layer missing.

	Species score	Experience-explained variation %
<i>Cladonia</i> spp	-0.344	11.8
<i>Melampyrum pratense</i>	-0.193	3.7
<i>Dicranum</i> spp	-0.157	2.5
FLM	-0.151	2.3
<i>Hylocomium splendens</i>	-0.146	2.1
<i>Deschampsia flexuosa</i>	-0.145	2.1
<i>Cladina</i> spp	-0.087	0.8
<i>Ptilium crista-castrensis</i>	-0.076	0.6
Other lichens	-0.052	0.3
<i>Vaccinium vitis-idaea</i>	-0.025	0.1
<i>Cetraria</i> spp	-0.019	0.0
<i>Sphagnum</i> spp	0.025	0.1
<i>Cladina stellaris</i>	0.026	0.1
<i>Linnea borealis</i>	0.027	0.1
<i>Calluna vulgaris</i>	0.075	0.6
<i>Empetrum nigrum</i>	0.114	1.3
<i>Pleurozium schreberi</i>	0.129	1.7
Other bryophytes	0.140	2.0
<i>Vaccinium myrtillus</i>	0.163	2.7
BLM	0.359	12.9

estimates for all observers save one (Table 5). Though, there were large differences for single species, with BLM (from -0.11 to 1.58) and *Ptilium crista-castrensis* (from 0.284 to 4.146) having the widest range. For BLM, the mean was very far away from 1 (Table 5).

Number of observers

The statistical power, to separate between the two vegetation types, of cover estimates by one randomly chosen observer differed little from the mean of 2-4 observers (Figure 1b).

Cover estimates versus presence/absence

The two vegetation types used in the present study were possible to separate along the first axis in PCA (Figure 3) when using the centroid values for each plot. The sample plots with bottom layer dominated by *Sphagnum* spp. were separated along the second axis, when using cover data (Figure 3a). Vegetation types were better separated in ordination space when using original cover data (Figure 3a), than when using data transformed to

Figure 2. The mean of ten observers' estimates of cover for missing layers, lichens, bryophytes and dwarf-shrubs in each of 16 plots plotted against coefficient of variation. Only plots where all observers had found the species are included. FLM is field layer missing and BLM is bottom layer missing.

presence/absence (Figure 3b). In contrast, the power of the RDA differed only marginally when comparing cover estimates by eye to presence/absence data (Figure 1c). However, even if the median probability differed little, the variation in p-values was higher for presence/absence than for cover data (Figure 4). That is, using presence/absence instead of cover estimates by eye increased the risk of doing a type-II error (accepting a false null hypothesis).

Table 5. Slopes from linear regressions of cover estimates involving individual observer estimates (dependent) vs. the average of other observers (independent). If no difference, then slope would be 1. Observers are ranked from most to least experienced (see Table 1) Species for which variation in cover estimates spanned <10% are not included. FLM is field layer missing and BLM is bottom layer missing.

Observer	1	2	3	4	5	6	7	8	9	10	Mean	SD
FLM	1.03	1.24	0.77	1.47	0.62	0.98	0.86	1.09	0.20	1.20	0.95	0.357
BLM	0.60	0.44	-0.11	0.24	0.39	0.34	-0.079	0.020	0.40	1.58	0.38	0.483
<i>Calluna vulgaris</i>	1.08	1.23	0.96	0.33	0.86	1.20	1.01	0.75	0.80	1.46	0.97	0.312
<i>Cetraria</i> spp	1.29	0.72	0.85	0.63	1.73	0.55	0.53	1.97	0.67	0.90	0.98	0.512
<i>Cladina</i> spp	1.02	1.48	0.96	0.68	0.97	0.56	0.56	0.86	0.42	1.28	0.88	0.336
<i>Dicranum</i> spp	0.13	0.93	0.20	0.18	1.21	0.47	0.22	4.42	0.96	0.19	0.89	1.301
<i>Empetrum nigrum</i>	0.67	1.52	0.29	0.55	1.25	0.36	1.31	1.05	0.94	2.05	1.00	0.555
<i>Hylocomium splendens</i>	1.34	1.16	0.92	1.44	0.50	0.67	0.87	0.76	0.80	1.47	0.99	0.340
<i>Pleurozium schreberi</i>	1.28	1.13	1.00	1.44	0.44	0.85	1.01	0.31	0.96	1.48	0.99	0.385
<i>Ptilium crista-castrensis</i>	0.39	1.16	4.15	0.55	0.69	0.28	0.72	0.36	2.39	0.28	1.10	1.245
<i>Sphagnum</i> spp	1.09	0.90	0.70	1.10	0.98	0.90	1.25	0.86	0.88	1.34	1.00	0.193
<i>Vaccinium myrtillus</i>	1.07	0.76	0.84	1.15	0.66	1.03	0.96	1.07	0.86	1.42	0.98	0.217
<i>Vaccinium vitis-idaea</i>	0.77	0.57	0.67	0.81	0.42	0.38	1.14	1.05	0.39	2.43	0.86	0.609
Mean	0.90	1.02	0.94	0.81	0.82	0.66	0.80	1.12	0.82	1.31		
SD	0.369	0.333	1.022	0.463	0.398	0.303	0.403	1.097	0.537	0.610		
Confidence interval (+-)	0.201	0.181	0.555	0.252	0.216	0.165	0.219	0.597	0.292	0.332

Figure 3. First two principal components from a PCA of vegetation data, illustrating the estimates of 10 observers in the same 16 plots in central Sweden. Symbols and error bars illustrate mean ordination score ($N=10$) and 95% confidence intervals. Black diamonds are the xeric vegetation type while open squares are the mesic vegetation type. Ground layer missing and bottom layer missing were excluded from analyses. Figure 3a is based on cover estimates and 3b on the same data transformed to presence/absence.

Figure 4. Comparison of cover estimates by eye and presence absence. Original cover data was transformed to presence/absence and then a resampling of subsets of the data (bootstrapping) from both cover estimates and presence/absence took place. A random selection of samples ($n=3, 4 \dots 8$ from each vegetation type) were drawn 100 times with replacement and p -values and their variation was calculated for each sample size. The number indicates the number of plots and the letter “c”, cover data while a number without letter indicate presence/absence data.

Discussion

Effects of the study design

Although there was a significant effect of wearing of the plots, the low eigenvalues indicated that it is possible to use more than one observer without seriously affecting the estimates in dwarf-shrub dominated areas. Trampling is the most probable cause of the wearing, and in the present vegetation it is possibly the estimates of lichens that are affected most.

The significant effect of inventory order for the observer could be due to tiring or an effect of doing many estimates in similar vegetation, i.e. a tendency by observers to standardize the estimates. If tiring does affect the estimates, it is essential to give enough rest between estimates. Of course in the original survey, observers are conducting other tasks between the cover estimates and that could be enough to counter the effect of tiring. The second explanation implies that cover estimates from sample plots in similar

vegetation and made with short time intervals tend to be more similar than otherwise would be expected. Kennedy & Addison (1987) found that under intensive sampling, similarity indices improved but the improvement was transient and after a 1-month break they returned to the initial level. This means that, if surveyors are working in similar vegetation for a long period of time, the estimates could be expected to become more similar. The NFSV sample plots are, for logistic reasons, in clusters corresponding to one day's work. The distances between sample plots, in all Sweden except the far south, are ≥ 400 m and all land use classes are involved. Hence, within the current monitoring system, there is large variation in vegetation sampled, which would suggest that there should be no temporal pattern in the quality of data collected.

Not surprisingly, multivariate analysis (Figure 1c) had higher power in separating the two vegetation types than did univariate analysis (Figure 1d). In the present vegetation types, the difference between the vegetation types was mainly manifested in differences in abundance of ground layer species. Since multivariate analysis compares the whole plant communities, the difference was more clearly illuminated than with a univariate, species-wise approach.

Random error and bias

Observer identity explained nearly 20% of the variance in the data. The high observer-dependency of the estimates indicates that bias is a substantial part of the variability of cover estimates. Another study in the same context, i.e. the NFSV, came to the conclusion that bias in cover estimates in most cases was $<10\%$ (Milberg et al. 2008). The difference between that and the present study is the methodology. Here, plots were subjectively located in two different vegetation types, the estimates were produced in rapid succession and the ten observers only performed the cover estimates, which were in one percent classes. In contrast, Milberg et al. (2008) used data from the actual survey, using only two observers with longer period of time between assessments, when the abundance was scored according to a non-linear 15-point scale. Furthermore, the observers conducted many other tasks and spent more total time in each plot. Hence, one hypothesis for the differences might be that observers get a more complete picture of the vegetation and may continuously evaluate and reevaluate the estimates made or to be made when given more time on each plot. Another reason might be due to the different scoring; non-linear 15-point scale vs. continuous, but transforming the latter data to a 15-point scale did not affect the large differences much (unpublished). A more likely reason might be that the error in data does not

follow a normal distribution. If so, estimates of SD will increase with the number of observers. Indeed, for many species, error distribution was skewed (of 13 abundant species, seven had a skewness >1). Although this might explain some of the discrepancies between the estimates of observer bias in Milberg et al. (2008) and the current study, several species in the latter exhibited low skewness despite large bias (e.g. *Vaccinium vitis-idaea* 0.41 and 48%, respectively; data not shown).

Throughout the data, it is clear that relative variation is higher when abundance is low, indicating that abundance is difficult to assess with uniform error over the whole range (Figure 2). The differences between groups of species and the missing layer group (FLM, BLM) are small, even if the variation in estimates for dwarf-shrubs seemed lower at low abundance than other groups of plants. Bråkenhielm & Qinghong (1995) found inter-personal error slightly greater for small and wide-spread plants, i.e. mosses and lichens, while Kercher et al. (2003) could not identify any taxon-related variability in a study of cover estimates between teams of two observers.

High observer-dependency in cover estimates has been found in other experiments (Sykes et al. 1983, Floyd & Anderson 1987, Kennedy & Addison 1987, Mitchell et al. 1988, van Hees & Mead 2000, Klimes 2003, Kercher et al. 2003, Vittoz & Guisan 2007), but few have separated random variation and bias. In a survey context, random variation in data could be countered by increasing sample size, while bias has to be taken into account when presenting results. Sykes et al. (1983) found bias consistent for an observer in relation to individual species, thus, making the use of a mean bias correction factor possible. This is probably not possible in a survey covering large areas with many different types of vegetation and, due to seasonal and weather variations, a variety of phenological stages.

Inter-observer differences in statistical power

To identify the existing difference between the vegetation types (effect size) using RDA, about the same sample size was required for all observers (Figure 1a). It is noteworthy, and reassuring, that the observer bias described above seems to have so little influence on multivariate power when distinguishing vegetation types.

The effect of experience on variation in estimates

The observers' experience explained a surprisingly low part of the multivariate variation in the data considering the wide range of experience among the observers. The estimates of BLM and *Cladonia* spp. could possibly be explained by the experienced observer's better identification of

the soil-covering parts of *Cladonia* spp. while the less experienced registered the same area as BLM.

When using cover estimates in plant ecology, many researchers emphasize training, calibration and experience to increase the reliability in data. Smith (1944) showed that calibration did diminish the range of eight observers' estimates of density in the same plots, but differences in the estimates still remained. In our study the effect of experience was limited. Why experience seems to make such a small difference is hard to explain. One explanation could be that once a mental image of the cover of a species has been reached it does not change with increasing experience. Since the true value of cover is usually unknown, no objective calibration can be done, and the incitements for change are very small. There were also different working models among the observers (personal observation); some started by estimating BLM and FLM and then estimated the species present; others started with the species and finished by estimating BLM and FLM. When all species and groups of species were summed for the plot, the observer adjusted some of the estimates in the end to get a more appropriate overall cover. In this, the choice of working model could be crucial: whether to use BLM and FLM as buffers or to adjust other species. Altogether, the results imply that training, calibration and experience are complex factors that need further studies.

Inter-observer differences in species records

The bias between species varied substantially. The high values were not always a result of the high variance expected in abundant species. Both BLM and FLM had a high degree of observer-dependent variance but BLM had low absolute values in cover and FLM high. The high bias for FLM and BLM implies that it is a difficult feature to estimate and very doubtful to use when monitoring vegetation. The high degree of observer-dependency for *Vaccinium myrtillus* and *V. vitis-idaea* gives some cause for concern since these species are very common in the boreal forests of Sweden and many times dominate the field layer and are thus of prime interest when trying to monitor changes. The high observer dependence for *Dicranum* spp. is obvious from the observers' means in the linear regression (Table 5) as well as the variance components analysis (Table 3). It seems that some observers had difficulties to detect the genus, since cover estimates of other common mosses as *Hylocomium splendens*, *Pleurozium schreberi* and *Ptilium crista-castrensis* had a substantially lower observer-dependent variance.

It was obvious that individual observers had problems with different species (Table 5). If this is a consistent feature for the individual observer, it

is essential to identify the troublesome species and address the problem with individual training.

The observer-dependent variation with regard to different species is difficult to explain. For example, there seems to be no clear connection within groups like mosses or dwarf-shrubs, nor between abundance and observer-dependent variation. This issue needs more research in order to establish the transferability of the results to other observers, plot sizes, locations and/or vegetation types.

Number of observers

In the present study, observers' estimates did not differ substantially in statistical power for detecting the known difference in vegetation type. Maybe more surprisingly, the idea to use more than one observer to get a lower variation in cover estimates (e.g. Nilsson 1992, Bergfur et al. 2004) was not supported by our results (cf. Vittoz & Guisan 2007). The extra cost of using two observers is hardly justified in light of the small, or negligible, gain in power.

Cover estimates versus presence/absence

The comparison between cover estimates by eye and presence/absence should be interpreted with care since the original data had been transformed to presence/absence. A field work protocol aiming at presence/absence records might lead to a different search method. For example, Vittoz & Guisan (2007) found that more species were found in subplots where both a species list and cover estimates were produced, than in plots where only a species list was required. This could possibly be explained by the discovering of supplementary species during the cover estimation process. Transformation of the cover estimates into a species list can therefore lead to smaller differences than expected if data had been collected using different methods.

It is clear that there was substantial variation between the observers in their ability to find and identify species present in the current study, thereby corroborating the findings of Vittoz & Guisan (2007) who found differences between experienced botanists in detection of species. The frequency of presence in many small plots as a mean to estimate abundance has been advocated as being less biased (Økland 1988). Even if presence/absence data has been shown to carry rather low observer-dependent variation (Ringvall et al. 2005), Milberg et al. (2008) found a frequency of 20 % missed occurrences in presence/absence sampling. It is possible that plot size contributes to the discrepancy; Ringvall et al. (2005) used small plot sizes

(0.01 and 0.33 m²) while Milberg et al. (2008) used 100 m²-plots. Still, with small plots, the sample size has to increase to achieve a representative species list (Jalonen et al. 1998).

In the present study, it is clear that, in spite of the high observer-dependent variation, cover data do better when distinguishing different vegetation types. Since the species number per plot is rather low, it is not surprising that the abundance information in cover data is of importance when distinguishing vegetation types.

Conclusions and advices

The “take home” messages from this study are both negative and positive. On the negative side, we recorded

- substantial observer bias that varied greatly among species.

On the positive side, we noted that ...

- the power to detect the two vegetation types in multivariate tests differ only little between observers.
- increased experience of the observer had little influence on cover estimates. I.e., less experienced observers do not generate poorer data than experienced, and a high staff turnover might not constitute a major threat to data quality.
- there is little point in averaging the scores of more than one observer, as this only marginally, if at all, affects statistical power when using multivariate analysis.
- cover estimates, despite their poor reliability, is preferable to presence/absence records. Both presence/absence and cover estimates showed a high variation but when distinguishing between vegetation types, the extra information in cover estimates proved superior.

Acknowledgement

Financial support was partly provided by the Swedish Environmental Protection Agency (PM, LW) and Stiftelsen Oscar & Lili Lamms Minne (JB).

References

Anonymous (2005) *Fältinstruktion 2005. Riksinventeringen av skog*. Institutionen för skoglig resurshushållning och geomatik and Institutionen för skoglig marklära, SLU, Umeå. (In Swedish).

- Bergfur J, Carlsson ALM & Milberg P (2004) Phenological changes within a growth season in two semi-natural pastures in southern Sweden. *Annales Botanici Fennici*, 41, 15-25.
- Bråkenhielm S & Qinhong L (1995) Comparison of field methods in vegetation monitoring. *Water, Air and Soil Pollution*, 79, 75-87.
- Carlsson ALM, Bergfur J & Milberg P (2005) Comparison of data from two vegetation monitoring methods in semi-natural grasslands. *Environmental Monitoring & Assessment*, 100, 235-248.
- Dethier M, Graham ES, Cohen S & Tear LM (1993) Visual versus random-point percent cover estimations: 'objective' is not always better. *Marine Ecology Progress Series*, 96, 93-100.
- Floyd DA. & Anderson JE (1987) A comparison of three different methods for estimating plant cover. *Journal of Ecology*, 75, 221-228.
- 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.
- Kennedy KA & Addison PA (1987) Some considerations for the use of visual estimates of plant cover in biomonitoring. *Journal of Ecology*, 75, 151-157.
- Kent M & Coker P (1992) *Vegetation description and analysis. A practical approach*. John Wiley and Sons, Chichester, UK.
- Kercher SM, Frieswyk CB & Zedler JB (2003) Effects of sampling teams and estimation methods on the assessment of plant cover. *Journal of Vegetation Science*, 14, 899-906.
- Klimes L (2003) Scale-dependent variation in visual estimates of grassland plant cover. *Journal of Vegetation Science*, 14, 815-821.
- Lepš J & Šmilauer P (2003) *Multivariate analysis of ecological data using CANOCO*. Cambridge University Press, Cambridge, UK.
- Milberg P, Bergstedt J, Fridman J, Odell G & Westerberg L (2008) Observer bias and random variation in vegetation monitoring data. *Journal of Vegetation Science*, 19, 633-644.
- Mitchell JE, Bartling PNS & O'Brien R (1988) Comparing cover-class macroplot data with direct estimates from small plots. *American Midland Naturalist*, 120, 70-78.
- Mueller-Dombois D & Ellenberg H (2002) *Aims and methods of vegetation ecology*. Blackburn Press, Caldwell, NJ, US.

- Nilsson C (1992) Increasing the reliability of vegetation analyses by using a team of two investigators. *Journal of Vegetation Science*, 3, 565.
- Økland T (1988) An ecological approach to the investigation of a beech forest in Vestfold, SE Norway. *Nordic Journal of Botany*, 8, 375-407.
- Oksanen J, Legendre P & O'Brien RB (2006) *vegan: Community Ecology Package version 1.8-3*. <http://cran.r-project.org/>
- R Development Core Team (2006) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Ringvall A, Petersson H, Ståhl G & Lämås T (2005) Surveyor consistency in presence/absence sampling for monitoring vegetation in a boreal forest. *Forest Ecology and Management*, 212, 109-117.
- Sarukhán J & Harper JL (1973) Studies on plant demography: *Ranunculus repens* L., *R. bulbosus* L., and *R. acris* L.: I. Population flux and survivorship. *Journal of Ecology*, 61, 675-716.
- Smith AD (1944) A study of the reliability of range vegetation estimates. *Ecology*, 25, 441-448.
- Sykes JM, Horrill AD & Mountford MD (1983) Use of visual cover assessments as quantitative estimators of some British woodland taxa. *Journal of Ecology*, 71, 437-450.
- Ter Braak CJF & Smilauer P (2002) *CANOCO Reference manual and CanoDraw for Windows User's guide: Software for Canonical Community Ordination (version 4.5)*. Microcomputer Power, Ithaca, New York, 500 pp.
- van Hees WWS & Mead BR (2000) Ocular estimates of understory vegetation structure in a closed *Picea galuca*/*Betula papyrifera* forest. *Journal of Vegetation Science*, 11, 195-200.
- Vittoz P & Guisan A (2007) How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science*, 18, 413-422.