

Observer error in vegetation surveys: a review

Lloyd W. Morrison^{1,2,*}

¹ Department of Biology, Missouri State University, 901 S. National Avenue, Springfield, MO 65897, USA

² National Park Service, Heartland Inventory and Monitoring Program, 6424 W. Farm Road 182, Republic, MO 65738, USA

*Correspondence address. Department of Biology, Missouri State University, 901 S. National Avenue, Springfield, MO 65897. Tel: +1-417-836-3119; Fax: +1-417-836-8886; E-mail: lloydwmorrison@missouristate.edu

Abstract

Aims

Vegetation sampling employing observers is prone to both inter-observer and intra-observer error. Three types of errors are common: (i) overlooking error (i.e. not observing species actually present), (ii) misidentification error (i.e. not correctly identifying species) and (iii) estimation error (i.e. not accurately estimating abundance). I conducted a literature review of 59 articles that provided quantitative estimates or statistical inferences regarding observer error in vegetation studies.

Important Findings

Almost all studies (92%) that tested for a statistically significant effect of observer error found at least one significant comparison. In surveys of species composition, mean pseudoturnover (the percentage of species overlooked by one observer but not another) was 10–30%. Species misidentification rates were on the order of 5–10%. The mean coefficient of variation (CV) among observers in surveys of vegetation cover was often several hundred % for species

with low cover, although CVs of 25–50% were more representative of species with mean covers of >50%. A variety of metrics and indices (including commonly used diversity indices) and multivariate data analysis techniques (including ordinations and classifications) were found to be sensitive to observer error. Sources of error commonly include both characteristics of the vegetation (e.g. small size of populations, rarity, morphology, phenology) and attributes of the observers (e.g. mental fatigue, personal biases, differences in experience, physical stress). The use of multiple observers, additional training including active feedback approaches, and continual evaluation and calibration among observers are recommended as strategies to reduce observer error in vegetation surveys.

Keywords: interobserver error, intraobserver error, misidentification, pseudoturnover, vegetation sampling

Received: 4 June 2015, Revised: 3 November 2015, Accepted: 21 November 2015

INTRODUCTION

Vegetation sampling involving observers is characterized by an inherent degree of error. Some of the earliest studies of observer error in vegetation sampling reported errors of ‘startling magnitude’ (Hope-Simpson 1940) that were ‘seemingly insurmountable’ (Smith 1944). Yet many vegetation studies involving observers, and particularly involving subjective estimates of observers, have been conducted and continue to be done. Although digital imagery may represent an alternative in some situations, the use of visual estimation in vegetation sampling is widespread because it is usually less time-consuming and expensive.

One way of quantifying error in vegetation studies is by comparing observer estimates to true values. This has rarely been done because true values are almost never known. Comparisons have also been made among different methods,

assuming one method is more precise than another (Everson and Clarke 1987; Smith 1944; Sykes *et al.* 1983). Yet it is well known that different methods are to some degree sampling different variables and will produce different results (Bråkenhielm and Liu 1995; Everson and Clarke 1987; Floyd and Anderson 1987). Thus, observer error has usually been quantified by making comparisons among observers or among different observations made by the same observer over time. These comparisons reflect the amount of precision among the observers’ estimates, rather than their accuracy. As Gotfryd and Hansell (1985) state: ‘Although a researcher may justifiably ignore accuracy in many situations, it is indefensible to ignore lack of precision.’

It is critical to have some estimate of the amount of uncertainty associated with the sampling process, to be able to make the appropriate inferences. The greater the uncertainty, or error, involved in sampling, the greater the possibility that

the resulting inferences will be erroneous. Unfortunately, many studies do not include estimates of observer error, and thus implicitly assume such error is either nonexistent or trivial in relation to study goals. Such assumptions may often be unsubstantiated.

In this review, I present a broad overview of the studies that have quantified observer error in vegetation sampling. Although the focus is on observer error per se, I include some comparisons among methods, when they were incorporated into the studies of observer error. The following main themes are pursued: (i) In general, what magnitude of observer error is associated with vegetation sampling? (ii) What are the sources of error? (iii) How does observer error vary with the basic elements of sampling design? (iv) How may this error be minimized in future studies, or what sorts of caveats should be attached to any resulting inferences?

MATERIALS AND METHODS

I attempted to include all studies on observer error in sampling plant populations or communities that have been published in the scientific literature. I did not include reports published in the gray literature, articles in journals that were not in English, or theses or dissertations. I excluded articles in which mere anecdotal accounts were given, including only those with quantitative estimates of the magnitude of, or statistical inferences regarding, observer error. I did not include articles that focused on large areas (i.e. many hectares) for the purposes of constructing plant atlases (Rich and Woodruff 1992), as many of the differences among observers in this case are simply due to observers visiting different subsets of the overall area of interest. A total of 59 studies met these criteria (supplementary Appendix 1).

Most community-level studies included entire vegetation communities, although some focused on subsets of communities, or in a few cases a select group of species. Some studies only sought to produce lists of species present, whereas others also attempted to estimate the abundance of each species. In general, three primary types of 'mistakes' could be made that resulted in erroneous characterization of vegetation: (i) not observing species (or individuals) actually present (i.e. overlooking error), (ii) not correctly identifying species present (i.e. misidentification error) and (iii) not accurately estimating abundance (i.e. estimation error).

Two types of observer error were considered: inter-observer and intra-observer. Inter-observer error results from two or more observers (or teams of observers) obtaining different results. Intra-observer error occurs when the same observer (or team of observers) obtains different results at different times (assuming the variable of interest has not changed).

Many different statistics and summary measures have been used to evaluate observer error, making comparisons among studies problematic. Many studies of species composition, however, have reported differences within or among observers as rates of pseudoturnover, and many studies of

vegetation cover have reported coefficients of variation. Thus, in this review, I focus on these two metrics, which do allow for direct comparisons to be made among many studies.

Pseudoturnover, following the work of Nilsson and Nilsson (1985) and based on classical island biogeographic theory (McArthur and Wilson 1967), is calculated as:

$$(A + B/S_A + S_B) * 100 \quad (1)$$

where A and B are the number of species recorded exclusively by each of two observers, and S_A and S_B are the total number of species recorded by each observer. Pseudoturnover refers to the percentage of species missed either by an observer or team in one of two sampling periods (intra-observer) or missed by one observer or team but not another observer or team (inter-observer). Pseudoturnover values may range between 0 (if all species are seen by both observers or in both periods) and 100 (if an entirely different subset of species is recorded by each observer or in each time period). Pseudoturnover is a measure that encompasses both overlooking and misidentification errors.

Some studies reported the degree of similarity among observers or time periods calculated by the Sørensen Index. Pseudoturnover is the inverse of similarity, and pseudoturnover can easily be determined as: 1 - the Sørensen Index. To facilitate comparisons, the Sørensen Index was converted to pseudoturnover in comparing errors associated with species lists.

Failure to correctly estimate abundance (e.g. percent cover) has often been quantified by the coefficient of variation (CV) to document variability among observers, or among different intervals for the same observer. The CV is simply the standard deviation expressed as a percentage of the mean ([standard deviation/mean] × 100).

RESULTS AND DISCUSSION

Temporal and spatial overview

The earliest studies on the subject date to the late 1930s–early 1940s. Hope-Simpson's (1940) study of intra-observer error in the UK and Smith's (1944) study of both intra- and inter-observer error in the western U.S. are classics that have been frequently cited. These studies did not generate many similar efforts, however, and only a handful of studies were conducted prior to 1980 (Figure 1). Much greater interest in this subject is evident in this century, based on the relatively large number of published studies since 2000.

Geographically, almost half of the studies in this review were conducted in Europe (47%). A little over a quarter (27%) were done in North America, and 17% were conducted in Australia. Africa and Asia are both greatly under-represented, with only four studies and one study conducted on each continent, respectively. I could find no such studies from South America. Thus, a strong geographical bias exists as the majority of studies are from the Northern Hemisphere (75% from Europe and North America). Only two studies

were conducted within the tropics (Hall and Okali 1978; Walker 1970).

As a function of habitat types, over a third of the studies (36%) were done in habitats described as woodland or forest. Only a single study was done in a tropical forest (Hall and Okali 1978). Pasture or grassland (grazed or with a grazing history) represented 21% of the surveyed literature, while rangeland or sagebrush represented 10%. Meadows (predominantly ungrazed) or clearcuts accounted for 12%. Other habitats represented were (number of studies in parenthesis): shrubland (3), islands (3), bogs (2), high-alpine summits (1), limestone glade (1), lichen community (1) and wasteland (1).

The above proportions may be representative of the total number of vegetation studies actually conducted in these geographic regions or habitat types. The inferences derived from such studies of observer error, however, may not apply to understudied regions or habitats. For example, observer error is likely to be greater in tropical regions due to the higher species diversity and greater structural complexity of many habitats.

Basic components of design

The number of observers evaluated (or groups of observers working together, in some cases) varied widely among studies (Figure 2), ranging from a single observer (intra-observer comparisons) to as many as 41 individual observers (in Ringvall *et al.* 2005). Of the 59 studies reviewed, a third (33%) evaluated >6 observers. In theory, assuming the existence of inherent variability in the overall population of observers, studies employing larger numbers of observers would be more likely to find evidence of variability among observers, yet they would also be more likely to be representative of this overall population.

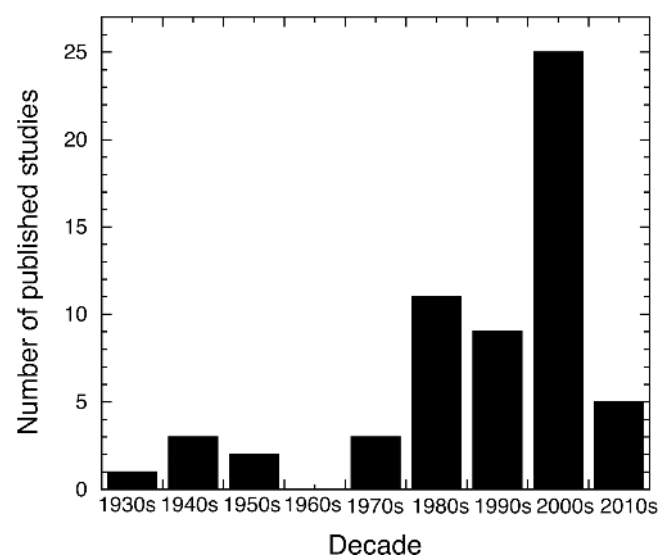


Figure 1: number of published articles on observer error in vegetation sampling by decade.

Most studies (78%) evaluated inter-observer error. Only 12% focused on intra-observer error, and 10% addressed both. A number of different variables were evaluated. Percent cover was the most frequently included variable (44 studies), followed by species composition (32 studies). Other variables used included frequency, richness, detectability, biomass, yield and a variety of different indices or metrics. Ten studies evaluated multiple variables.

Statistical analyses

In the studies of inter-observer error, 34 included tests of statistical significance. Of these, 32 indicated at least one statistically significant difference to exist between or among observers. The two studies in which no statistically significant differences were found were detectability studies (see below). Only three of the studies of intra-observer error tested for statistical significance, and only one failed to find any significant differences. This study was by Oredsson (2000), who conducted only a single significance test of intra-observer error.

Although statistical significance has become the golden standard by which to judge ecological hypotheses, biological importance is of greater practical utility (McBride *et al.* 1993; Yoccoz 1991). In some studies that reported statistically significant results, the authors also stated that the magnitude of the difference was relatively small (Carlsson *et al.* 2005; Symstad *et al.* 2008).

Pseudoturnover in species composition studies

Four studies evaluated intra-observer error in estimates of species composition. Pseudoturnover values in these studies were <15% (Table 1). An intra-observer pseudoturnover rate of 15% would mean that 15% of the species were not observed in both survey intervals. Such error may be expected

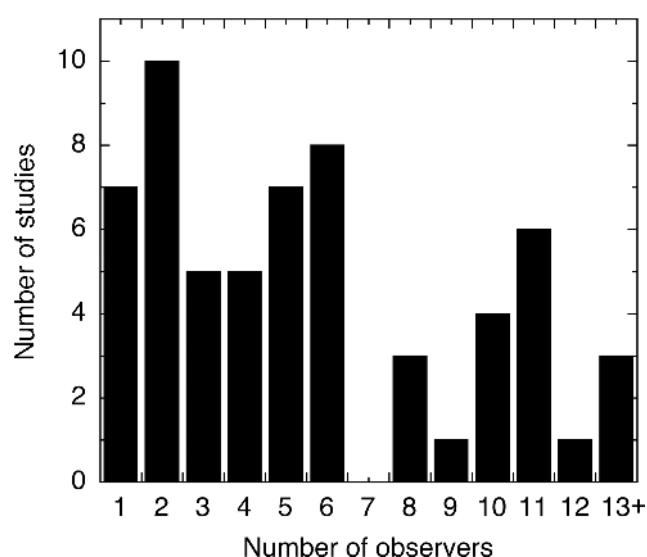


Figure 2: histogram of the number of observers (or teams of observers) evaluated in studies of observer error in vegetation sampling. The '13+' category includes $n = 16, 17$ and 41 observers.

Table 1: values of pseudoturnover reported from published studies

Study	Pseudoturnover (%) mean \pm SD (range)	Sample area	No. of observers or teams
Hope-Simpson (1940)	8.8	0.25 ha	Intra-observer
Nilsson and Nilsson (1982)	13.6 ^a	0.03–2.19 ha	Intra-observer
Nilsson and Nilsson (1983)	7.9 \pm 2.1 (6.3–11.6)	0.03–1.04 ha	Intra-observer
Nilsson and Nilsson (1985)	11.4 (4.2–19.4)	0.03–2.19 ha	2
Kirby <i>et al.</i> (1986)	(23–36) ^b	200 m ²	6
Kennedy and Addison (1987)	(4–12) ^c	1 m ²	Intra-observer
Lepš and Hadincová (1992)	13	25 m ²	2
Scott and Hallam (2002)	24 (0–69) ^d	0.16 m ²	2
Kercher <i>et al.</i> (2003)	19.1 \pm 1.2 (9.3–30.5)	1 m ²	2
Gray and Azuma (2005)	29.4 ^c	170 m ²	2
Gray and Azuma (2005)	33.4 ^c	1 m ²	2
Vittoz and Guisan (2007)	15.7 \pm 2.8 ^c	0.4 m ²	4–7
Vittoz and Guisan (2007)	10.4 \pm 5.8 ^c	4 m ²	4–7
Vittoz and Guisan (2007)	11.6 \pm 2.1 ^c	40 m ²	4–6
Symstad <i>et al.</i> (2008), visual estimate	20.1 \pm 2.1 (6–57)	5 m ²	5
Symstad <i>et al.</i> (2008), point frequency	27.2 \pm 3.5	100 points	5
Archaux <i>et al.</i> (2009)	10.9 (2.2–30) ^c	100 m ²	11
Vittoz <i>et al.</i> (2010), Scotland	28.0 \pm 2.9 (25–32) (all species)	100 m ²	8
Vittoz <i>et al.</i> (2010), Switzerland	16.0 \pm 3.2 (12–18) (vascular species)	100 m ²	9
Vittoz <i>et al.</i> (2010), Scotland	12.5 \pm 3.1 (8–15) (all species)	1 m ²	8
Vittoz <i>et al.</i> (2010), Switzerland	9.0 \pm 2.7 (5–11) (vascular species)	1 m ²	9
Burg <i>et al.</i> (2015)	13.5 \pm 1.14 ^f (0–33.3)	637–16 720 m ²	2

Units of replication are sample plots unless indicated otherwise.

^aDetermined by subtracting ‘certain + probable’ from ‘crude’ turnover.

^bFrom Scott and Hallam (2002).

^cConverted from the Sørensen Index.

^dRange refers to pseudoturnover across species types rather than across sites.

^eStandard deviations are from Vittoz *et al.* (2010)

^fUncertainty represents 1 SE rather than 1 SD.

to be relatively low, assuming the observers retained at least a partial memory of which species were present or where they were located.

Pseudoturnover in studies of inter-observer error was generally higher, with mean values ranging from ~10% to 30% (Table 1). An inter-observer pseudoturnover rate of 30%, e.g. would mean that 30% of the species recorded were not observed by both observers (or teams). Most calculations of pseudoturnover were based on visual searching and identification. Symstad *et al.* (2008) compared visual searching to a point frequency method (100 points), and reported that mean turnover was higher by the point method (27% compared to 20%), as fewer species were sampled.

The ubiquity of pseudoturnover indicates that most species lists are probably incomplete, by anywhere from 10% to 30%. This shortcoming does not appear to be generally acknowledged in most studies of species composition. Only a few studies of observer error explicitly came to this conclusion. For example, Ringvall *et al.* (2005) reported a ‘tendency toward general underestimation’ of species composition. Archaux *et al.* (2009) suggested the existence of a ‘systematically biased,

underestimate of true species richness’ in vegetation studies because some species are ‘unavoidably missed’ during surveys.

Detectability

Many recent studies focused primarily at the population level have addressed the issue of detectability. Detectability is often <100%, and in this review I have included studies in which differences in detectability among observers have been documented, rather than simply imperfect detectability per se. Kéry and Gregg (2003) found no differences between two observers in detectability of an orchid. Likewise, Chen *et al.* (2009) reported no significant differences between two observers in detection probabilities for six plant species. All species were fairly conspicuous, however (>1 m high with relatively large leaves).

In contrast, Moore *et al.* (2011), employing an experimental approach, found detection probabilities did vary significantly among observers searching for transplanted plants, ranging from 9% to 100% of the plants actually encountered. Similarly, Clarke *et al.* (2012) examined detectability for seven of the ‘most visible, persistent, and easily identifiable

Table 2: coefficients of variation (CV) from cover estimates reported from published studies

Study	Method (sample area)	Coefficient of variation (%) mean \pm SD (range)	Number of observers or teams
West (1938)	VE (1 m ²)	11.6 ^a	3
Friedel and Shaw (1987b)	VE (2.6–9 ha)	15.8 \pm 10.8 (5.4–28.0) ^{b,c}	6
Friedel and Shaw (1987b)	Wheel point (2000 points)	10.7 \pm 5.7 (4.2–15.1) ^{b,c}	6
Everson and Clarke (1987)	Step point	31.66 \pm 11.31 (17.6–47.3)	6
Tonteri (1990)	VE (2 m ²)	25.8 \pm 7.6 (15–37) ^d	11
Stampfli (1991)	Fixed point (176 points)	28.8 \pm 16.3 (6.4–316.2)	10
Van Hees and Mead (2000)	VE (100 m ²)	85.1 \pm 26.7 (50–124)	6
Klimeš (2003)	VE (9.8–4 m ²)	(0–225)	5
Helm and Mead (2004)	VE (1 m ²)	172.4 \pm 121.6 (61–533)	6
Vittoz and Guisan (2007)	VE (4 m ²)	80.8 \pm 40.0	4–7
Vittoz and Guisan (2007)	Point method (100 points)	38.9 \pm 16.0	4–6
Gorrod and Keith (2009)	VE (100 m ²)	(40–300 for means \leq 20; <75 for means >20)	10
Bergstedt <i>et al.</i> (2009)	VE (100 m ²)	99.98 \pm 52.6	10
Vittoz <i>et al.</i> (2010)	VE (100 m ²)	108.2 \pm 50.7	8
Vittoz <i>et al.</i> (2010)	VE (1 m ²)	66.2 \pm 6.1	8
Vittoz <i>et al.</i> (2010)	Frequency count (100 cells)	19.4 \pm 7.1	8
Vittoz <i>et al.</i> (2010)	Point method (200 points)	20.4 \pm 3.3	8

Units of replication for calculation of means represent individual species, unless otherwise noted. VE, visual estimation.

^aBased on estimation of all vegetation rather than individual species.

^bBased on cumulative aerial cover of shrubs and trees rather than individual species.

^cReplicates are different sites.

^dBased on six target species.

perennial vegetation species,’ and found that non-detection errors varied among six different surveyors.

Misidentification rates

Only four studies attempted to quantify rates of misidentification. Klimeš *et al.* (2001), sampling grasslands, reported a 10–20% discrepancy among observers, (33% in the smallest plots), which they attributed primarily to misidentification. Working in a variety of habitat types, Scott and Hallam (2002) reported 5.9% of records were misidentified at the species level and 1.9% of records were misidentified at the genus level.

In a study of French lowland forests, Archaux *et al.* (2006) found misidentification errors ranging from 5.6% to 10.5%, depending upon the observer. Misidentification rates averaged 2.3% at the species level and 0.9% at the genus level for the tree layer, with corresponding mean rates of 5.3% and 1.3%, respectively, for the ground vegetation layer (Archaux *et al.* 2009). The rate of overlooking was found to be 6.7 and 3.6 times that of the misidentification rates for the tree level and ground vegetation level, respectively.

Thus, based on a very small number of studies, misidentification rates appear to be higher for ground vegetation than tree species, and higher for grasslands than forests. Misidentification error also generally appears to be of a smaller magnitude than overlooking error, based on comparison of these misidentification rates to the pseudoturnover rates discussed above.

Coefficients of variation in percent cover estimates

The coefficients of variation (CV) for estimates of percent cover in published studies are presented in Table 2. A wide range of CVs is evident. In general, CVs from studies in which a more objective point method was used were relatively low (< 40%). CVs from studies involving visual estimation revealed greater variability and mean values approached or exceeded 100% in some studies (Bergstedt *et al.* 2009; Helm and Mead 2004; Vittoz *et al.* 2010). The CV often varied greatly among species within individual studies, as evidenced by the large ranges in Table 2 (Gorrod and Keith 2009; Helm and Mead 2004; Klimeš 2003).

Typically, species with the lowest cover values (<~10%) had the highest CVs. Plotting the relationship between CV and mean percent cover often results in a curvilinear relationship, with CV values rapidly declining as cover increases for small values of cover, and then more gradually declining or flattening out for larger cover values (Gorrod and Keith 2009, Figure 1; Gorrod *et al.* 2013, Figure 1). CVs of several hundred percent for species with low cover were not uncommon, whereas CVs of 25–50% were more representative of species with mean covers of >50%.

This relationship does not necessarily mean that greater variability exists among observer estimates for less abundant species, however. In fact, greater variation among observer estimates has been associated with more moderately abundant

species (~40–60% cover) (Sykes *et al.* 1983; Vittoz *et al.* 2010, their Figure 2; Gorrod *et al.* 2013, see supplementary Appendix 1). Hahn and Scheuring (2003), using computer simulations, reported that subjects made the greatest error when cover was estimated to be between 25% and 75%.

The resolution to this apparent dilemma is attributable to a mathematical artifact in the calculation of the CV. For a given standard deviation (i.e. amount of variability among observers), the CV will vary curvilinearly with the mean, yielding a curve that decreases in slope less rapidly as the mean increases (supplementary Figure S1). Thus, caution must be taken in interpreting such high CV values for species with low cover.

Cover was estimated to the nearest 5% or even 1% in many studies, while in others cover categories were employed. The use of categories for cover estimation is common due to the inability to visually estimate cover precisely, and accurate assessment simply requires estimation within a few predetermined categories. A number of studies evaluated differences in category estimation among observers. Such evaluations provided remarkably similar results, with a relatively large number of species (or plots, depending on the design) differing by one category, and a much smaller number differing by more than one category. The exact percentages were: 39.5% and 3% (Lepš and Hadincová 1992), 46% and 4% (Klimeš 2003), 41% and 6% (Gray and Azuma 2005), 33% and 'rare' (MacDonald 2010) and 47.5% and 11.5% (Archaux *et al.* 2007), respectively. Cheal (2008) found variation that spread over three categories. Similarly, in a study of intra-observer error utilizing subjective frequency as a measure of abundance, Hope-Simpson (1940) reported that 36% of species differed by one category and 12% differed by two categories. Thus, the available evidence suggests that, in studies of vegetation abundance utilizing category estimation, between one-third and one-half of all estimates are erroneous, but most of these are probably off by only one category.

Other types of effect sizes

Although almost all studies reported some statistically significant findings of observer error, the size of the effect varied. This 'effect size' was determined in various ways by different authors, and were not always directly comparable even when similar variables were studied. The effect sizes reported, however, do provide insight into the relative amounts of overall variability that have characterized the respective studies. Here, I summarize some of the reported effect sizes not included within rates of pseudoturnover, misidentification, or coefficients of variation discussed above. The results of these studies, along with documented pseudoturnover (Table 1) and CVs of cover estimates (Table 2), indicate most studies that included entire plant communities found observer error in the 10–30% range.

Sykes *et al.* (1983), employing confidence intervals, concluded that variation in cover estimates were in the range of 10–20% for inter-observer error, and 5–15% for intra-observer error. Using a multivariate approach, Gotfryd and

Hansell (1985) found evidence for an overall observer effect in the 20–30% range. Kennedy and Addison (1987), in a study of intra-observer error, concluded that changes in vegetation cover should be '> 20% before they can be attributed to factors other than annual fluctuation and measurement error'. Klimeš *et al.* (2001) estimated that discrepancies in species number estimates between individual observers ranged from 10 to 20% in large plots and up to 33% in smaller plots.

Working in lowland forests, Archaux *et al.* (2006) found that single observers overlooked 20–30% of species actually present. Archaux *et al.* (2009) reported that the mean overlooking rate was 19.2% for the ground vegetation layer and 15.5% for the tree layer. Bergstedt *et al.* (2009), working in coniferous forests, concluded that observer identity explained nearly 20% of the variance in their data set. Vittoz *et al.* (2010) concluded that unless large numbers of plots were used, changes in cover or frequency in alpine vegetation were only likely to be detected for abundant species (>10% cover), or if changes were large (50% or more). Gorrod and Keith (2009) found an average coefficient of variation of 15–18% in estimated values of biological indices for grassy woodlands, and suggested this was likely an underestimate of real world variation.

Few studies reported small amounts of observer error. For example, Carlsson *et al.* (2005) concluded that variation attributable to inter-observer error was <3%. Only two observers were tested, however, whom the authors acknowledged were 'unrepresentative by being very similar in experience and perception'. Ringvall *et al.* (2005) reported 'good consistency between surveyors and high accuracy according to a reference survey'. This study had the largest number of individual observers (41!), although surveyors were only asked to record the presence or absence of six different species or species groups. The only two studies of inter-observer error that failed to find any significant differences were both detectability studies of similarly limited scope (see above).

Effect of plot size

Plot size varied greatly among studies, ranging from 0.001 m² to several ha, although more studies were conducted on relatively smaller plots (Figure 3). Attempting to evaluate the effect of plot size by comparing among studies is fraught with peril, due to the great diversity of study techniques, habitat types and geographical locations. A few studies did include multiple plot sizes, however, and direct comparisons may be made from these. Sykes *et al.* (1983), working with plot sizes of 4, 50 and 200 m², reported greater variation in cover estimates among observers for larger plot sizes. In contrast, Klimeš (2003), using plots ranging from 0.001 to 4 m², found that the CV in cover estimates among observers was higher for smaller plots.

A resolution for this discrepancy is provided by Vittoz and Guisan (2007), who suggest that with plots larger than 4 m² it is difficult to obtain a 'global view' of the entire plot, which makes estimation of cover more difficult. Furthermore,

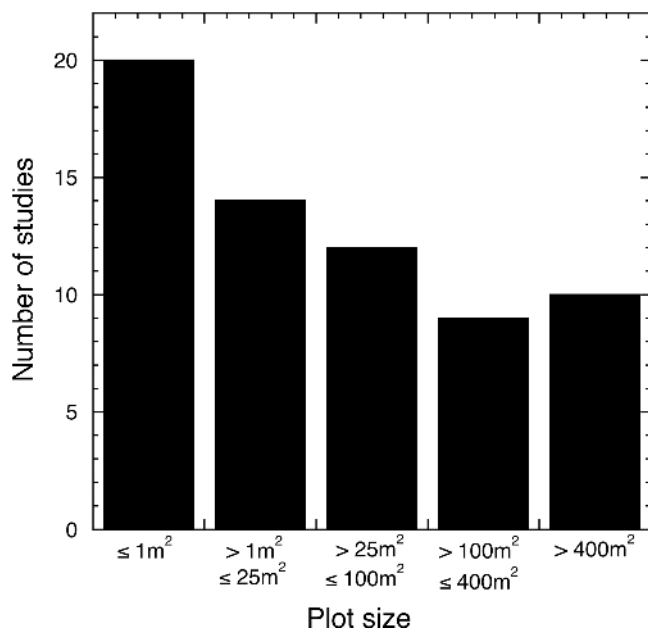


Figure 3: histogram of plot sizes evaluated in studies of observer error in vegetation sampling. Some studies evaluated multiple plot sizes. Nilsson and Nilsson's (1982, 1983, 1985) studies of islands in lakes covered a wide range of island areas and are not included in this graph.

Klimeš' (2003) smallest plots (0.001 m^2) are too small to be useful in most ecological studies, and Klimeš noted that the vegetation was actually higher than the plot diameter, so that 'a small change from the exactly vertical projection to an inclined one may markedly change the perception of plant cover'. Thus high variability in cover estimates for such small plots is likely an artifact of unrealistically small plot size in relation to plant size.

Concerning species composition, Ringvall *et al.* (2005) reported no effect of two different sized plots, although both were relatively small (0.01 and 0.33 m^2). Likewise, Vittoz and Guisan (2007) concluded that the area of plots (0.4 , 4 and 40 m^2) had an 'unclear' influence on their species lists. Based on the use of five different variables (including species composition and species cover), Archaux *et al.* (2007) concluded that plant censuses carried out on small quadrats (2 and 4 m^2) were not more reliable than larger ones (400 m^2), and advocated using larger quadrats primarily because they contain more species. The study with the broadest range of plot sizes was that of Nilsson and Nilsson (1985) on islands (0.03 – 2.19 ha), who reported—somewhat surprisingly—a negative correlation of pseudoturnover with island area (i.e. greater similarity in species lists between teams of observers for larger islands). This could be an artifact of working on islands: If the species pool that is able to colonize and survive on islands is limited (Morrison 2013), increases in island size may result in greater abundances of populations rather than additional new species (i.e. there may be fewer rare species on larger islands).

Thus, the limited evidence available suggests error associated with cover estimates may be minimized by the use of plots ranging from ~ 1 to 4 m^2 in size. There is no clear consensus, however, on the relation of plot size and error associated with species lists.

Effect of time spent searching

As time spent searching for species increases, the discovery rate is usually higher near the beginning of the search period, resulting in a curvilinear species accumulation curve that flattens out with additional searching (Archaux *et al.* 2006; Gray and Azuma 2005; Kirby *et al.* 1986; Klimeš *et al.* 2001). The rate of species discovery may vary among observers, however, producing curves that are higher for some observers than others (Archaux *et al.* 2006; Kirby *et al.* 1986). This reflects inherent differences among observers in their sampling efficiencies.

In some cases, depending upon the design of the study, different observers may spend different amounts of time surveying a plot, which may lead to inter-observer variability. Burg *et al.* (2015), surveying the flora of high-alpine summits, concluded that the 'major cause of a high pseudoturnover was a large difference in botanizing time between observers'. Archaux *et al.* (2006), working in forests, concluded that 'sampling time is certainly an underestimated biasing factor in vegetation studies' and recommended it would often be better to fix the sampling time rather than using minimum or maximum time limits. Klimeš *et al.* (2001), however, working in grasslands, found that time limitation was not the main factor causing incompleteness of species lists, as correcting for differences in sampling time did not explain the discrepancies among observers in their study.

Some time limits employed in vegetation studies may simply be too short. For example, Hope-Simpson (1940) reported that 47% more species were found with a 2 to 3-fold increase in sampling time, suggesting their normal protocol yielded only about two-thirds of the species actually present. For large plots, it will usually be impractical to attempt to discover all species. If large plots are to be employed, examination of species accumulation curves as a function of time spent sampling could help determine the optimum amount of time to spend per plot in balancing the need for completeness of species lists versus controlling costs, as well as identify differences in species discovery rates among observers.

Effect of experience or training

The effect of prior experience or training was evaluated in a number of studies. Twelve studies specifically evaluated whether variation among observers could be attributed to differences in experience. The results were mixed, as seven found some effect of experience (Bergstedt *et al.* 2009; Hall and Okali 1978; McCune *et al.* 1997; Oredsson 2000; Ringvall *et al.* 2005; Scott and Hallam 2002; Vittoz and Guisan 2007) whereas six did not (Burg *et al.* 2015; Cheal 2008; Chen *et al.* 2009; Kéry and Gregg 2003; Moore *et al.* 2011; Sykes *et al.* 1983). In the studies by Kéry and Gregg (2003) and Chen

et al. (2009), some training was also given, which could have obfuscated the effects of prior experience. Seven studies evaluated whether training had an effect. Only one study—*Archaux et al.* (2009)—failed to find any evidence for a training effect. All others (*Campbell and Arnold 1973; Kennedy and Addison 1987; Murphy and Lodge 2002; Smith 1944; Stapanian et al. 1997; Symstad et al. 2008*) found at least some evidence that training increased the precision or accuracy of estimates.

Whether an effect of experience can be detected will likely depend upon the range of variation in experience among observers, as well as the relevance of the experience, which differed widely among studies. Training is likely to have a greater potential to increase precision among less experienced observers. For example, *Symstad et al.* (2008) reported that in the early stages of training, the first visual estimates of individual grass cover (for the same plot) ranged from 2% to 30% across observers. Thus while the overall effect of the experience of observers was equivocal and varied by study, there was strong evidence that training did increase precision.

Random variation or systematic bias?

A number of studies addressed the question of whether observed error represented random variation or a systematic bias. Some found no evidence of a systematic bias (*Klimeš 2003; Lepš and Hadincová 1992; Smith 1944*). *Gorrod and Keith (2009)*, in a study with 10 observers, reported that each observer made the most extreme estimate on at least one site, in each of two types of vegetation assessment. *Archaux et al.* (2007), in a study of five different variables, estimated that the magnitude of random variation was twice as high as that of a systematic bias.

In contrast, other studies have reported that some observers or teams consistently tended to record over- or underestimates relative to other observers or teams (*Bråkenhielm and Liu 1995; Carlsson et al. 2005; Kercher et al. 2003; Sykes et al. 1983; Tonteri 1990*). Some studies indicated that a single observer (or team) was relatively far from the group mean. For example, *Goodall (1952)* reported that two observers agreed consistently, while a third's estimates frequently either exceeded or fell below those of the other two. *McCune et al.* (1997) found that a single observer found very few species and inflated the between-crew variance. *Oredsson (2000)* reported that one observer recorded significantly fewer species than five others. *Gorrod and Keith (2009)* found that one observer's total site scores were consistently different from the group mean. Finally, *Tonteri (1990)* reported that one observer tended to consistently overestimate cover.

Overall, whether error is random or systematic will likely depend upon the design of the study, and the backgrounds and biases of the observers. Evaluation of a potential systematic bias is useful, because if one is found it may be possible to provide a correction factor. *Sykes et al. (1983)* suggested that in cases in which individual observers either consistently underestimate or overestimate cover, a correction factor

could be used to reduce inter-observer bias. A separate correction factor would be necessary for each species, however, and I could find almost no evidence that later studies have attempted this approach. A study in this survey in which a correction factor was successfully employed was that by *Young et al. (2008)*, who estimated abundance of a single species—a cryptic winter annual (*Physaria filiformis* [Rollins] O'Kane and Al-Shehbaz). Observers were found to consistently underestimate the abundance of this species, so that it was possible to apply correction factors to increase overall population estimates *post hoc*.

Effects on multivariate analyses

An important question is whether observer bias in estimating variables in the field is substantial enough to result in misleading inferences from multivariate data analyses. Most studies that addressed this aspect have indicated that such error can substantially affect the interpretation of multivariate analyses. For example, *Hall and Okali (1978)* found that, using PCA ordination, observer bias obscured a seasonal gradient in species richness. *Gotfryd and Hansell (1985)* reported that a number of multivariate techniques were found to be sensitive to observer effects, and that 'transformation of variables or use of non-parametric techniques did not mitigate observer effects whatsoever'. *Block et al. (1987)* concluded that interpretations of multivariate axes can vary 'depending on the person who collects the data'. *Tonteri (1990)* found that multivariate methods could effectively reduce noise in the data, but an observer effect could still be recognized.

In contrast, *Lepš and Hadincová (1992)* found that ordination and classification were generally insensitive to sampling errors, but depended on the transformations used. *Bergstedt et al. (2009)* reported a significant effect of measurement bias on multivariate analyses, but of a relatively small magnitude. *McCune et al. (1997)* found gradient scores obtained from nonmetric multidimensional scaling to be more consistent among observers than species lists. They attributed this to a 'redundancy of information provided by different species'. This study was somewhat unique in the survey, as it focused on lichen communities, and was conducted with relatively large plots (0.378 ha).

Thus the majority of studies indicate that multivariate analyses are likely to be affected by observer error in collecting field data, although in some cases the effects may be relatively small. Such errors may also bias the calculation of ecological indices. For example, *Archaux (2009)* reported that species richness estimators did not correctly account for differing degrees of completeness of species lists among teams, and were highly sensitive to misidentifications. *Gorrod et al. (2013)* concluded that measurement error can result in systematic underestimation of ecological indices.

Suitability for long-term monitoring

If repeated vegetation surveys are to be useful for long-term monitoring, the amount of error associated with the surveys

must be of a small magnitude relative to the degree of change expected or desired to detect. A number of studies in this review have evaluated observer error in relation to long-term monitoring, and their conclusions vary. For example, [Stapanian *et al.* \(1997\)](#), evaluating inter-observer error in field methods for the Forest Health Monitoring Program, found that the proportion of variation due to measurement error and temporal variability was generally <13%, for both species richness and cover. They concluded the methodology was 'suitable and practical for a large-scale ecological monitoring program'. [Murphy and Lodge \(2002\)](#), investigating inter-observer error in cover estimates of grasslands, concluded that inexperienced observers 'could be trained quickly and easily to estimate cover'. Their standards for the precision of results were relatively low, however, as their goal was to obtain 'sufficient accuracy to identify broad categories of low (<40%), medium (40–70%) and high (>70%)' cover.

[Gray and Azuma \(2005\)](#) quantified inter-observer error in species composition and percent cover in evaluating a vegetation indicator for the national Forest Inventory and Analysis Program. The observed values of pseudoturnover were on the high-end of the range observed for all studies in this review, and error rates for categories of cover estimation were similar to those of other studies (see Coefficients of variation in percent cover estimates). Nevertheless, they concluded that the vegetation indicator 'provides a robust and valuable tool for assessing forest health'. [Symstad *et al.* \(2008\)](#) evaluated inter-observer error for eight different variables in a long-term grasslands monitoring program. Although they reported 'substantial' pseudoturnover (6–57%), they concluded that the inter-observer error documented was 'within precision levels for many variables'.

[Plattner *et al.* \(2004\)](#) compared differences in mean species richness between botanists who double sampled plots as part of the Biodiversity Monitoring in Switzerland Programme. Differences in mean species richness were small—0.1 species at the local (10 m²) and 5.0 species at the landscape (12 500 m²) scales—and they concluded that 'systematic methodological errors were negligible' and random errors were also small. They did not take into account species identities, however, ignoring the potential for pseudoturnover in the species lists.

In contrast, [van Hees and Mead \(2000\)](#), evaluating inter-observer error in a horizontal/vertical profiling method of forest monitoring, concluded that observers 'did not consistently and repeatedly estimate vegetation cover'. Estimates of observers were not consistent relative to each other from one plot to the next or from one measurement period to the next. They concluded that the measurement error component was 'large enough to question the validity of cover estimates for use as a classification tool in addition to its utility for change detection'. Furthermore, they called for 'a re-evaluation of the goals and a search for more appropriate sampling methods'. Finally, in a study evaluating inter-observer error in species richness estimators derived from surveys in French lowland forests, [Archaux \(2009\)](#) concluded that estimators obtained

from different sampling teams, even if misidentifications are removed, 'may not be sufficiently reliable to confidently conduct analyses of spatial or temporal changes in plant richness'.

Sources of error

A number of potential sources of error in vegetation surveys have been identified in the literature. These can be grouped into three main categories: (i) characteristics of the vegetation, (ii) the environment associated with the sampling and (iii) attributes associated with the observers. In the first category, errors may result from characteristics of the plants that make them difficult to notice, such as small size (of individuals or populations), geographical rarity, morphological traits such as narrow leaves or a 'winding' growth form, similarities with other species, immature individuals, or a change in appearance due to herbivory ([Carlsson *et al.* 2005](#); [Hope-Simpson 1940](#); [Kennedy and Addison 1987](#); [Kéry and Gregg 2003](#); [Kercher *et al.* 2003](#)). The second category includes environmental conditions that may make sampling more difficult and are largely outside the influence of observers. Examples are bad weather, steep topography, varying light levels over the course of a day, and a changing overall vegetation matrix ([Klimeš *et al.* 2001](#); [Kéry and Gregg 2003](#); [Moore *et al.* 2011](#)). Even the colors of the vegetation relative to that of the soil may have an effect ([Hahn and Scheuring 2003](#)).

The third category includes attributes of observers that are largely under control of the observers or their supervisors, such as mental or physical fatigue, collaborators who may either disturb or stimulate the observers, personal biases, differences in experience and expectations among observers, spatial scale of observation and assessment, physical stress, and lack of enthusiasm ([Bråkenhielm and Liu 1995](#); [Burg *et al.* 2015](#); [Klimeš *et al.* 2001](#); McDonald 2010; [Moore *et al.* 2011](#); [Walker 1970](#)). Differences among observers in the amount of time spent sampling may also result in discrepancies ([Archaux *et al.* 2006](#); [Burg *et al.* 2015](#)). [Lepš and Hadincová \(1992\)](#) note, with refreshing honesty, that because of disturbance to concentration of the observer, 'in some cases species were observed in the field but not entered on the list'. The importance of the mental state of the observer should not be underestimated in vegetation surveys. As [Walker \(1970\)](#) stated: 'Every method is entirely dependent on the integrity and attitude of the operator.' Most authors have assumed that observers are proficient in the taxonomy of the flora; if this is not the case, this represents another obvious source of error. Physical fitness of the observer may be an overlooked, yet important attribute in arduous environments; [Burg *et al.* \(2015\)](#) found that the number of species missed in surveys of high-alpine summits increased with the length of ascent.

It should be kept in mind that the three types of error defined above (i.e. overlooking, misidentification, and estimation) result from these sources. Any given source may primarily result in one type of error (e.g. the similarity of one species with other species may result in misidentification errors in a species inventory), although many of the sources

can potentially result in any of the three error types (e.g. mental fatigue could result in plants being overlooked, misidentified, or their abundance incorrectly estimated).

Solutions

The authors of many studies in this review offered a variety of suggestions for the reduction of observer error. The optimal (and obvious) solution would be to use the same observer or team of observers for all data collection. This would eliminate inter-observer error, but not intra-observer error, although the former was usually found to be of greater magnitude than the latter. Unfortunately, this is impractical for many studies involving large spatial areas or a large number of plots, and for studies spanning multiple years.

Given that multiple observers will be necessary, many authors stressed the importance of training prior to collection of field data (Archaux *et al.* 2009; Bråkenhielm and Liu 1995; Burg *et al.* 2015; Chen *et al.* 2009; Friedel and Shaw 1987a, b; Gorrod and Keith 2009; Kercher *et al.* 2003; Stapanian *et al.* 1997; Sykes *et al.* 1983). In addition to training in the field, Bråkenhielm and Liu (1995) suggested the use of test figures with exactly known cover to evaluate the accuracy of cover estimates, since true cover in the field is often unknown. Stapanian *et al.* (1997) recommended rigorous certification tests and field audits. Kercher *et al.* (2003) concluded that all vegetation surveys should include a quality control component. They suggested double sampling of plots to calculate pseudoturnover and including quality control results in any publication.

Sykes *et al.* (1983) suggested 'screening' observers, and rejecting individuals who cannot produce 'acceptably consistent results' after training. Given that a number of studies found evidence that estimates of single observers or teams represented outliers relative to the group mean (Gorrod and Keith 2009; Goodall 1952; McCune *et al.* 1997; Oredsson 2000; Tonteri 1990), rejecting such individuals could dramatically increase overall precision. Sykes *et al.* (1983) also suggested that 'calibrations' could be used to correct for biases after the survey. This could be done either in an attempt to correct for biases among observers (more common) or for a generalized bias of over- or underestimation (Ringvall *et al.* 2005; Young *et al.* 2008). It would first need to be demonstrated that errors are systematic, rather than random, however. Furthermore, such a correction factor may need to be employed on a per species basis.

Some authors suggested ways of reducing the potential for making errors. For example, Kercher *et al.* (2003) recommended conducting a reconnaissance of sites to generate an initial species list, and reaching a consensus among observers on the level of identification required for difficult taxa. They also suggested consulting with professional taxonomists to reduce the possibility of misidentifications. Archaux *et al.* (2009) suggested employing only botanists familiar with the local flora and noting any doubts of field identifications. Gorrod and Keith (2009) suggested that observers should

specify 'plausible bounds', or a range of values around point estimates, rather than a single value, to convey the appropriate associated uncertainty. Several authors recommended using smaller sampling units or longer recording times (Archaux *et al.* 2009; Bråkenhielm and Liu 1995; Gorrod and Keith 2009; Kercher *et al.* 2003). Others suggested making only a few surveys per day, to reduce observer fatigue (Archaux *et al.* 2009; Walker 1970). Burg *et al.* (2015) recommended frequent changes in observer team composition to help equalize observer skills.

A common recommendation was the use of multiple observers (Archaux 2009; Archaux *et al.* 2009; Gorrod and Keith 2009; Klimeš *et al.* 2001; Klimeš 2003; Symstad *et al.* 2008; Vittoz *et al.* 2010). The advantage of multiple observers in studies of species composition is that combining lists of individual observers usually results in an increased number of species found (Klimeš *et al.* 2001). The advantage of multiple observers in studies of percent cover is that extreme estimates may be adjusted (Klimeš 2003), and average estimates may be closer to true values because errors associated with individual estimates are canceled out (Wintle *et al.* 2013).

The use of at least three observers has been recommended (Archaux 2009; Klimeš *et al.* 2001; Klimeš 2003), although little information exists on the marginal decrease in error due to additional observers. Sykes *et al.* (1983) determined that the confidence interval could be reduced by half using groups of four observers. Vittoz and Guisan (2007), however, found no significant improvement with pairs compared to single observers. The primary disadvantage of adding observers would be additional cost.

Wintle *et al.* (2013) suggested that feedback methods should be introduced into the training of observers in vegetation studies. In experiments, knowledge of other observer's estimates and of group averages improved the performance of observers. They provided intriguing evidence that incorporating active feedback techniques in training allowed observers to more appropriately reflect uncertainty in their estimates.

Of the three types of 'mistakes' listed above, misidentification errors could be, in theory, almost completely eliminated by enhanced taxonomic training and consultation with professional taxonomists. Overlooking errors could be minimized by a number of the suggestions above, particularly the use of multiple observers, smaller plot size, and efforts to reduce observer fatigue. Characteristics of the vegetation such as small stature, or cryptic growth form or phenological state, will always make some species prone to being overlooked.

Estimation error could also be minimized with the suggested solutions above, although it appears to be inevitable to some degree, reflecting a basic limitation of our ability to accurately quantify objects subjectively. For example, using computer simulations, Hahn and Scheuring (2003) found that estimation error was minimal when the range of cover was divided into 10 equal categories. Most subjects divided the cover level into from 10 to at most 20 intervals in their

minds, even if they had the opportunity to make more precise estimates.

The use of digital imagery analysis is becoming more widespread in vegetation sampling, and may provide more precise estimates of abundance than can human observers. A variety of methods have been used, including nadir photography from cameras suspended above the vegetation (Macfarlane and Ogden 2012) as well as ground-based photography (Jorgensen *et al.* 2013). Different digital imagery vegetation analysis techniques do yield different results, however (Jorgensen *et al.* 2013). Additionally, photographic techniques may produce underestimates when layers of vegetation are present, and may be susceptible to problems of shading.

Moreover, a digital imagery method that works well at one site may not produce accurate results at another. For example, Limb *et al.* (2007) described a digital imagery method that resulted in greater precision in measuring biomass compared to human observers. Leis and Morrison (2011), however, found the method did not yield accurate estimates in denser grasslands. Thus, while digital imagery may avoid the subjectivity associated with observer estimates, its accuracy should still be evaluated for each application.

Although digital imagery may not be suitable for many field applications, it may be useful in training observers. For example, Gallegos (2005) describes a computer aided calibration program that allows the comparison of observer estimates to true cover. In experimental trials ‘even a short time of calibration greatly improved the estimations’. Such feedback on the accuracy of observer estimates (or when true cover values are not available, on group averages as discussed above) represent relatively inexpensive and efficient methods of training and re-evaluating observers over the course of the field season.

CONCLUSIONS

As Kéry and Greg (2003) put it: ‘Although plants stand still and wait to be counted, they sometimes hide.’ Almost all studies of vegetation sampling considered found evidence of observer error. In almost all cases, (92% of studies) statistical tests revealed significant observer effects. The magnitude of the effect varied greatly among studies, however. In many studies, the authors considered the size of the observer effect to be serious enough to cast doubt on any resulting inferences.

Observer error or bias is not a unique feature of vegetation sampling. It is encountered in many other fields of biological study (Elphick 2008), and often not adequately evaluated (Burghardt *et al.* 2012). I am unaware, however, of any other comprehensive reviews such as this for any other biological discipline.

Given the large number of studies published on vegetation sampling, it is likely that some papers containing data on observer error have been unintentionally left out of this review. In many papers, documentation of such error was a minor part of the study, and it is difficult to identify these papers in

computer searches of index terms. The results obtained from the studies that are included, however, appear to be robust, and are unlikely to be biased by any overlooked papers.

From a retrospective point of view we may ask: were the observer errors described as being of ‘startling magnitude’ by Hope-Simpson (1940) and ‘seemingly insurmountable’ by Smith (1944) unusual relative to later studies? From Hope-Simpson’s (1940) study of intra-observer error, I calculated pseudoturnover to be 8.8%, which is on par with later studies of intra-observer error, and toward the low end of the range of studies of inter-observer error (Table 1). Hope-Simpson (1940) estimated relative abundance in categories and reported that 36% of species differed by one category, and 12% differed by two categories, which (as discussed above) is similar to succeeding studies using categories for abundance estimation. Smith (1944) reported that density estimates of different observers ranged from 71.2 to 139.8% of the group average. This is no worse than the variability reported in many of the studies estimating percent cover (Table 2). Thus, there is little evidence that the magnitude of observer error has declined in vegetation sampling since these classic studies, although later studies have rarely described the error in such dramatic tones. This leaves us with two principal conclusions: Over the intervening decades there has either been growing acceptance of the relatively large degrees of observer error in vegetation studies, or growing disregard.

Based on this literature survey, the following generalized conclusions may be drawn:

- 1) Observer error is a pervasive component of all vegetation sampling involving human observers.
- 2) A portion of this error is due to characteristics of the vegetation and the environment associated with the sampling, and is unavoidable. The component of error due to attributes associated with observers, however, may be minimized.
- 3) There are many potential sources of observer error, as well as many potential methods of reducing the error. Practitioners should carefully consider the options available to obtain more precise results.
- 4) The degree of observer error associated with vegetation sampling can be, and should be, quantified and reported along with the results in every study. Although it may not be possible to determine the accuracy of observer estimates, precision can and should be documented. Double sampling 10% of plots may represent a reasonable trade-off between limited resources and the need for error reporting in many studies.
- 5) In studies evaluating change in vegetation populations or communities (in the absence of error validation), any differences smaller than ~25% should be viewed with skepticism, as all of the documented change could simply be due to observer error.

SUPPLEMENTARY DATA

Supplementary material is available at *Journal of Plant Ecology* online.

ACKNOWLEDGEMENTS

M. DeBacker and C. Young provided helpful comments on the manuscript. Views, statements, findings, conclusions, recommendations and data in this report are those of the author(s) and do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

Conflict of interest statement. None declared.

REFERENCES

- Archaux F (2009) Could we obtain better estimates of plot species richness from multiple-observer plant censuses? *J Veg Sci* **20**:603–11.
- Archaux F, Berges L, Chevalier R (2007) Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecol* **188**:179–90.
- Archaux F, Camaret S, Dupouey J-L, *et al.* (2009) Can we reliably estimate species richness with large plots? An assessment through calibration training. *Plant Ecol* **203**:303–15.
- Archaux F, Gosselin F, Berges L, *et al.* (2006) Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *J Veg Sci* **17**:299–306.
- Bergstedt J, Westerberg L, Milberg P (2009) In the eye of the beholder: bias and stochastic variation in cover estimates. *Plant Ecol* **204**:271–83.
- Block WM, With KA, Morrison ML (1987) On measuring bird habitat: influence of observer variability and sample size. *Condor* **89**:241–51.
- Bråkenhielm S, Liu Q (1995) Comparison of field methods in vegetation monitoring. *Water Air Soil Pollut* **79**:75–87.
- Burg S, Rixen C, Stöckli V, *et al.* (2015) Observation bias and its causes in botanical surveys on high-alpine summits. *J Veg Sci* **26**:191–200.
- Burghardt GM, Bartmess-LeVasseur JN, Browning SA, *et al.* (2012) Perspectives – minimizing observer bias in behavioral studies: a review and recommendations. *Ethology* **118**:511–7.
- Campbell NA, Arnold GW (1973) The visual assessment of pasture yield. *Australian J Exp Agric Anim Husb* **13**:263–7.
- Carlsson ALM, Bergfur J, Milberg P (2005) Comparison of data from two vegetation monitoring methods in semi-natural grasslands. *Environ Monit Assess* **100**:235–48.
- Cheal D (2008) Repeatability of cover estimates? *Ecol Manag Restor* **9**:67–8.
- Chen G, Kery M, Zhang J, *et al.* (2009) Factors affecting detection probability in plant distribution studies. *J Ecol* **97**:1383–9.
- Clarke KD, Lewis M, Brandle R, *et al.* (2012) Non-detection errors in a survey of persistent, highly-detectable vegetation species. *Environ Monit Assess* **184**:625–35.
- Elphick CS (2008) How you count counts: the importance of methods research in applied ecology. *J Appl Ecol* **45**:1313–20.
- Everson CS, Clarke GPY (1987) A comparison of six methods of botanical analysis in the montane grasslands of natal south africa. *Vegetatio* **73**:47–52.
- Floyd DA, Anderson JE (1987) A comparison of three methods for estimating plant cover. *Journal of Ecology* **75**:221–8.
- Friedel MH, Shaw K (1987a) Evaluation of methods for monitoring sparse patterned vegetation in arid rangelands. I. Herbage. *J Environ Manag* **25**:297–308.
- Friedel MH, Shaw K (1987b) Evaluation of methods for monitoring sparse patterned vegetation in arid rangelands. II. Trees and shrubs. *J Environ Manag* **25**:309–18.
- Gallegos A (2005) Design and evaluation of a computer aided calibration program for visual estimation of vegetation cover. *Arbetsrapport*, Vol. 142. SLU, Institutionen för skoglig resurshushallning och geomatik.
- Goodall DW (1952) Some considerations in the use of point quadrats for the analysis of vegetation. *Austr J Biol Sci* **5**:1–41.
- Gorrod EJ, Bedward M, Keith DA, *et al.* (2013) Systematic underestimation resulting from measurement error in score-based ecological indices. *Biol Conserv* **157**:266–76.
- Gorrod EJ, Keith DA (2009) Observer variation in field assessments of vegetation condition: implications for biodiversity conservation. *Ecol Manag Restor* **10**:31–40.
- Gotfryd A, Hansell RIC (1985) The impact of observer bias on multivariate analyses of vegetation structure. *Oikos* **45**:223–34.
- Gray AN, Azuma DL (2005) Repeatability and implementation of a forest vegetation indicator. *Ecol Indic* **5**:57–71.
- Hahn I, Scheuring I (2003) The effect of measurement scales on estimating vegetation cover: a computer-assisted experiment. *Commun Ecol* **4**:29–33.
- Hall JB, Okali DUU (1978) Observer bias in a floristic survey of complex tropical vegetation. *J Ecol* **66**:241–50.
- Helm DJ, Mead BR (2004) Reproducibility of vegetation cover estimates in south-central alaska forests. *J Veg Sci* **15**:33–40.
- Hope-Simpson JF (1940) On the errors in the ordinary use of subjective frequency estimations in grassland. *J Ecol* **28**:193–209.
- Jorgensen CF, Stutzman RJ, Anderson LC, *et al.* (2013) Choosing a diva: a comparison of emerging digital imagery vegetation analysis techniques. *Appl Veg Sci* **16**:552–60.
- Kennedy KA, Addison PA (1987) Some considerations for the use of visual estimates of plant cover in biomonitoring. *J Ecol* **75**:151–8.
- Kercher SM, Frieswyk CB, Zedler JB (2003) Effects of sampling teams and estimation methods on the assessment of plant cover. *J Veg Sci* **14**:899–906.
- Kéry M, Gregg KB (2003) Effects of life-state on detectability in a demographic study of the terrestrial orchid *Cleistes bifaria*. *J Ecol* **91**:265–73.
- Kirby KJ, Bines T, Burn A, *et al.* (1986) Seasonal and observer differences in vascular plant records from British woodlands. *J Ecol* **74**:123–32.
- Klimeš L (2003) Scale-dependent variation in visual estimates of grassland plant cover. *J Veg Sci* **14**:815–21.
- Klimeš L, Dancak M, Hajek M, *et al.* (2001) Scale-dependent biases in species counts in a grassland. *J Veg Sci* **12**:699–704.
- Leis SA, Morrison LW (2011) Field test of digital photography biomass estimation technique in tallgrass prairie. *Rangeland Ecol Manag* **64**:99–103.
- Lepš J, Hadincová V (1992) How reliable are our vegetation analyses? *J Veg Sci* **3**:119–24.

- Limb RF, Hickman KR, Engle DM, *et al.* (2007) Digital photography: Reduced investigator variation in visual obstruction measurements for southern tallgrass prairie. *Rangeland Ecol Manag* **60**:548–52.
- MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography*. Princeton: Princeton University Press.
- MacDonald AJ (2010) Testing the reliability of assessment of land management impacts on scottish upland vegetation. *Plant Ecol Divers* **3**:301–12.
- Macfarlane C, Ogden GN (2012) Automated estimation of foliage cover in forest understorey from digital nadir images. *Methods Ecol Evol* **3**:405–15.
- McBride GB, Loftis JC, Adkins NC (1993) What do significance tests really tell us about the environment? *Environ Manag* **17**:423–32.
- McCune B, Dey JP, Peck JE, *et al.* (1997) Repeatability of community data: species richness versus gradient scores in large-scale lichen studies. *The Bryologist* **100**:40–6.
- Moore JL, Hauser CE, Bear JL, *et al.* (2011) Estimating detection-effort curves for plants using search experiments. *Ecol Appl* **21**:601–7.
- Morrison LW (2013) Nestedness in insular floras: Spatiotemporal variation and underlying mechanisms. *J Plant Ecol* **6**:480–8.
- Murphy SR, Lodge GM (2002) Ground cover in temperate native perennial grass pastures. I. A comparison of four estimation methods. *Rangeland J* **24**:288–300.
- Nilsson IN, Nilsson SG (1982) Turnover of vascular plant species on small islands in lake mockeln, south sweden 1976–1980. *Oecologia* **53**:128–33.
- Nilsson IN, Nilsson SG (1985) Experimental estimates of census efficiency and pseudoturnover on islands: error trend and between-observer variation when recording vascular plants. *J Ecol* **73**:65–70.
- Nilsson SG, Nilsson IN (1983) Are estimated species turnover rates on islands largely sampling errors? *Am Nat* **121**:595–7.
- Oredsson A (2000) Choice of surveyor is vital to the reliability of floristic change studies. *Watsonia* **23**:287–91.
- Plattner M, Birrer S, Weber D (2004) Data quality in monitoring plant species richness in switzerland. *Commun Ecol* **5**:135–43.
- Rich TCG, Woodruff ER (1992) Recording bias in botanical surveys. *Watsonia* **19**:73–95.
- Ringvall A, Petersson H, Stahl G, *et al.* (2005) Surveyor consistency in presence/absence sampling for monitoring vegetation in a boreal forest. *For Ecol Manag* **212**:109–17.
- Scott WA, Hallam CJ (2002) Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecol* **165**:101–15.
- Smith AD (1944) A study of reliability of range vegetation estimates. *Ecology* **25**:441–8.
- Stampfli A (1991) Accurate determination of vegetational change in meadows by successive point quadrat analysis. *Vegetatio* **96**:185–94.
- Stapanian MA, Cline SP, Cassell DL (1997) Evaluation of a measurement method for forest vegetation in a large-scale ecological survey. *Environ Monit Assess* **45**:237–57.
- Sykes JM, Horrill AD, Mountford MD (1983) Use of visual cover estimates as quantitative estimators of some british woodland taxa. *J Ecol* **71**:437–50.
- Symstad AJ, Wienk CL, Thorstenson AD (2008) Precision, repeatability, and efficiency of two canopy-cover estimate methods in northern great plains vegetation. *Rangeland Ecol Manag* **61**:419–29.
- Tonteri T (1990) Inter-observer variation in forest vegetation cover assessments. *Silva Fennica* **24**:189–96.
- van Hees WWS, Mead BR (2000) Ocular estimates of understory vegetation structure in a closed picea glauca/betula papyrifera forest. *J Veg Sci* **11**:195–200.
- Vittoz P, Bayfield N, Brooker R, *et al.* (2010) Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *J Veg Sci* **21**:1035–47.
- Vittoz P, Guisan A (2007) How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *J Veg Sci* **18**:413–22.
- Walker BH (1970) An evaluation of 8 methods of botanical analysis on grasslands in Rhodesia. *J Appl Ecol* **7**:403–16.
- West O (1938) The significance of percentage area determinations yielded by the percentage area or density list method of pasture analysis. *J Ecol* **26**:210–7.
- Wintle BC, Fidler F, Vesk PA, *et al.* (2013) Improving visual estimation through active feedback. *Methods Ecol Evol* **4**:53–62.
- Yoccoz NG (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bull Ecol Soc Am* **72**:106–11.
- Young CC, Morrison LW, Kelrick MI, *et al.* (2008) Monitoring *Lesquerella filiformis* Rollins (Missouri bladderpod): application and evaluation of a grid-based survey approach. *Nat Areas J* **28**:370–8.