



Missouri State[™]
U N I V E R S I T Y

BearWorks

College of Natural and Applied Sciences

5-3-2016

Observer error in sampling a rare plant population

Lloyd W. Morrison
Missouri State University

Craig C. Young

Follow this and additional works at: <https://bearworks.missouristate.edu/articles-cnas>

Recommended Citation

Morrison, Lloyd W., and Craig C. Young. "Observer error in sampling a rare plant population." *Plant Ecology & Diversity* 9, no. 3 (2016): 289-297.

This article or document was made available through BearWorks, the institutional repository of Missouri State University. The work contained in it may be protected by copyright and require permission of the copyright holder for reuse or redistribution.

For more information, please contact [BearWorks@library.missouristate.edu](mailto: BearWorks@library.missouristate.edu).

Observer error in sampling a rare plant population

Lloyd W. Morrison^{a,b,*} and Craig C. Young^a

^aNational Park Service, Heartland Inventory and Monitoring Program, Republic, MO, USA; ^bDepartment of Biology, Missouri State University, Springfield, MO, USA

(Received 24 March 2016; accepted 2 August 2016)

Background: Estimation of abundance in vegetation sampling involving observers is almost always characterised by observer error, although such error is rarely reported.

Aims: To quantify observer error in population estimation of the rare plant species *Physaria filiformis* in Missouri, USA.

Methods: The abundance of *P. filiformis* was estimated within 25-m² plots by six trained observers with varying experience levels over 10 years. Observers assigned plots to six predefined density classes. A total of 477 plots were estimated annually, and actual counts were conducted on ca. 10% of the plots to assess per cent agreement of estimates with counts.

Results: Over a third of the estimates of plant abundance evaluated for accuracy (36.4%) deviated from exhaustive counts. The majority of the misestimates were underestimates by one density class (29.4%). The number and type of misestimates varied systematically with density class.

Conclusions: Observer error could be explained to some degree by variation in population density, but not by experience. It appears that inherent differences exist among observers that cannot be entirely compensated for by experience or training. Observer error in this system represents a systematic bias, and can be compensated for by use of correction factors, which would ideally be both density class-dependent and observer-specific.

Keywords: density class; monitoring; observer error; *Physaria filiformis*; rare plant sampling

Introduction

Vegetation sampling is almost always, if not always, characterised by some degree of error. A recent review of vegetation studies that employed observers found that 92% of the 59 total studies that tested for a statistical effect of observer error found at least one significant comparison (Morrison 2016). The magnitude of error often accounted for as much as 20–30% of the value of the measured variable. Even in smaller scale studies when observer error may be reduced by the use of digital imagery, other sources of error may arise (e.g. underestimates of abundance may result when layers of vegetation overlap, and shading may introduce problems in analyses) (Morrison 2016). Thus, all vegetation ecologists should realise the potential for error in field studies, and attempt to quantify it and report it in any publications (Kercher et al. 2003). Data quality is a critical consideration in any monitoring programme (Elzinga et al. 2001; Legg and Nagy 2006).

Quantifying abundance in plant populations or communities is usually accomplished by estimating cover-abundance. For most species, per cent cover is a more appropriate measure than number of individuals, since the size of individuals may vary greatly with age and other factors. Most published estimates of observer error associated with quantifying plant abundances are related to per cent cover (Morrison 2016). Observer error, however, may invalidate inferences because the magnitude of observer error can exceed statistically observable minimum

differences. Moreover, the vast majority of studies published in the literature do not quantify observer error.

Here, we evaluate observer error in the estimation of abundance within six density categories of a rare species of flowering plant in the mustard family (Brassicaceae) – *Physaria filiformis* (Rollins) O’Kane and Al-Shehbaz, the Missouri bladderpod. *P. filiformis* is a small (10–20 cm tall) winter annual that is listed as threatened and monitored annually as part of the Natural Resource Inventory and Monitoring Program of the US National Park Service. To obtain a demographically robust measurement, sampling has focused on estimating the number of individuals, rather than per cent cover, as a measure of abundance (Young et al. 2008a). Because populations may reach tens of thousands of individuals, numbers of individuals within plots have been estimated within density classes rather than counted. Assigning the wrong abundance category to a particular plot represents an obvious potential source of observer error.

The rationale for, and evaluation of, the sampling protocol for this particular rare plant has been described elsewhere (Young et al. 2008b). Here, we focus on the observer error associated with the estimation of abundance. The following questions were addressed: (1) What is the magnitude of observer error associated with the estimation of abundance? (2) Is there an effect of the experience level of the observer? (3) Are errors random or systematically biased? and (4) What, if anything, can be

*Corresponding author. Email: LloydMorrison@MissouriState.edu

done to increase the precision of estimates once errors have been documented?

Materials and methods

Focal species

P. filiformis inhabits limestone, dolomite, and shale glades in Alabama, Arkansas and Missouri (Al-Shehbaz 2010). A winter annual, the plant germinates in late summer to autumn, overwinters as a rosette, produces yellow flowers (5–9 mm petals) in April–May, sets seed and senesces by late spring. The rosette leaves are 1–2.4 cm in length. Each rosette may support multiple erect to decumbent flowering stems with heights up to 25 cm. The plant produces globose fruits with a 3–4 mm diameter. Fruits dehisce upon drying. A substrate endemic, the plant is rare and listed by the US Fish and Wildlife Service as a threatened species. Long-term studies in Missouri have shown that population size fluctuates widely among years, suggesting an overriding influence of density-independent factors (Young et al. 2008b). This species prefers open microhabitats and is threatened by woody plant encroachment (Young et al. 2009).

Study site and sampling design

This study was conducted on a large limestone glade located in Wilson's Creek National Battlefield, Republic, Missouri. Various sampling designs have been evaluated for this *P. filiformis* population (Morrison et al. 2008); see Young et al. (2008b) for a rationale for the design employed, which utilises density classes. Each year from 2006 to 2015, we assessed the abundance of *P. filiformis* in a contiguous grid of 477 permanently marked 25-m² plots. The plots included the glade area with the highest concentration of *P. filiformis*.

We conducted surveys in April during peak flowering. Plots were marked with flags at each corner. In each plot, we estimated the density of plants using a density-class scale: 0 = no plants; 1 = 1–9 plants; 2 = 10–49 plants; 3 = 50–99 plants; 4 = 100–499 plants; 5 = 500–999 plants; and 6 = 1000–4999 plants. To calculate a point estimate of the population size for each year, we summed the mid-points of the density classes attributed to each plot (Young et al. 2008b).

The low and high ends of the estimated density classes assigned to each cell can be summed, respectively, to calculate the end points of a population size interval, which (assuming correct density class assignment) contains the actual population size (i.e. is similar to a 100% confidence interval). These population size intervals were found to be very similar in width and location to 95% confidence intervals determined from a sampling approach (Young et al. 2008b). The midpoints of population size intervals, however, may not be good approximations of sample means, and counts may provide more accurate point estimates (Young et al. 2008b). The primary

advantage of estimating all plots rather than counting a sample of plots is that the former method allows for mapping relative densities across the glade.

Training of observers – harmonising abundance estimates

We practised a series of informal quality control measures during the study. Observers frequently discussed how to systematically observe a plot and how *P. filiformis* growth forms differed among microhabitats. New observers were always trained, which involved plant identification in its various growth forms in the field. Each trainee then estimated density in conversation with an experienced observer. Finally, each trainee estimated density while being supervised by an experienced observer. Once counts between the observer and trainee matched, the trainee was allowed to make independent observations of abundance. Initial training typically lasted 30–45 min. It should be noted that as the study progressed, multi-year observers became aware of the tendency to underestimate density and may have tried to hedge against this tendency, especially in cases where density was marginal between two categories.

Accuracy assessments

Accuracy assessments were conducted by dividing plots into five 1-m wide lanes and counting all plants at ground level within these lanes. In each year, a random sample of 60 plots was selected for accuracy assessments. In some years, additional plots were added haphazardly to increase the sample size, from density classes that were under-represented in the random sample. At least 10 plots for each density class were included when available. Although some degree of error may characterise the count data, it is of a relatively small magnitude and because of the relatively wide density classes used, would affect a very small proportion of the accuracy assessments.

In 2006 and 2010, plots to be used for accuracy assessments were selected by using slightly different procedures. In 2006, the 477 plots, representing the core area, were first grouped into fifty-three 15-m × 15-m strata. A single 5-m × 5-m plot was then randomly selected from each stratum. Nine additional plots were haphazardly selected to increase the total number of plots assessed for each density class to at least 10. The selection process employed in 2006 was designed to increase the spatial balance of the plots selected; this was later determined to be unnecessary (Morrison et al. 2008). In 2010 (a low-abundance year), 10 plots to be used for accuracy assessments were selected randomly from each density class after estimates were made. Only three density classes were represented (0, 1 and 2) and only three plots were estimated as class 2. The percentage of the 477 plots evaluated for accuracy ranged from 13% to 15% in all years except 2010, in which it was 5% (Table 1).

Overall, 609 accuracy assessments were conducted over the 10 years. In 500 of the assessments, an

Table 1. Population estimates of *Physaria filiformis* inhabiting a glade in south-western Missouri, along with numbers and percentages of plots assessed for accuracy.

Year	Population point estimate	Percentage of plots occupied	Number of plots counted	Percentage of plots counted
2006	29,269	70.9	61	12.8
2007	16,822	72.1	73	15.3
2008	3918	54.1	60	12.6
2009	1937	44.2	60	12.6
2010	794	30.2	23	4.8
2011	49,043	85.7	68	14.3
2012	36,212	79.2	63	13.2
2013	48,040	77.6	69	14.5
2014	465	18.4	60	12.6
2015	19,014	67.7	72	15.1

Percentage of plots counted is based on 477 plots for all years (2006–2015).

independent observer made the estimates, and in 109 cases the estimates were made by a team of two observers. Because it was not known with certainty which observer of each team made the estimates, or if there were any conferrals, the team estimates were included in the overall accuracy assessment, but excluded from analyses of observer error. Accuracy was determined including all assessments ($n = 609$) and including only assessments with *P. filiformis* present ($n = 437$).

In the analyses of observer error, all cases in which actual counts were zero were excluded, as these were rarely misclassified, to better focus on observer ability to estimate non-zero abundance. The final sample size for assessing observer error was 365 estimates, made by 7 observers. One observer only had five total estimates evaluated, and was not included in the comparisons among observers. In comparisons of observers among years, cases in which an observer had <6 estimates evaluated were not included; this resulted in an overall sample size of 346. Overall, years of experience ranged from 0 to 10, and two observers made estimates in five or more years.

Data analysis

Per cent agreement of estimates with counts was calculated as the percentage of abundance estimates that agreed with the exhaustive counts. Per cent agreement of estimates with counts was summarised by observer as a function of population size, years of experience and density class. Statistical associations were evaluated by the use of Pearson correlation coefficients. SPSS Version 20.0 (IBM Corp 2011) was used for all analyses.

Results

Accuracy assessments

Overall (for all observers in all years), in almost two-thirds (63.6%) of all accuracy assessments, the estimated

Table 2. Types and frequencies of errors for each actual density class, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015).

Actual	Density class		Error	Per cent within each actual class
	Estimate	Frequency		
0	0	158	None	91.9
0	1	13	Overestimate by 1 class	7.6
0	2	1	Overestimate by 2 classes	0.6
1	0	37	Underestimate by 1 class	27.2
1	1	96	None	70.6
1	2	3	Overestimate by 1 class	2.2
2	0	5	Underestimate by 2 classes	4.6
2	1	40	Underestimate by 1 class	36.7
2	2	63	None	57.8
2	3	1	Overestimate by 1 class	0.9
3	1	1	Underestimate by 2 classes	2.2
3	2	29	Underestimate by 1 class	63.0
3	3	16	None	34.8
4	2	14	Underestimate by 2 classes	13.2
4	3	46	Underestimate by 1 class	43.4
4	4	46	None	43.4
5	3	3	Underestimate by 2 classes	11.5
5	4	15	Underestimate by 1 class	57.7
5	5	8	None	30.8
6	4	2	Underestimate by 2 classes	14.3
6	5	12	Underestimate by 1 class	85.7
6	6	0	None	0

category was in agreement with the exhaustive counts (Table 2). Almost all estimates that were not in agreement with counts were underestimates: 29.4% were underestimated by one density class, and 4.1% were underestimated by two density classes. Only 2.8% were overestimated by one density class, and a single assessment (0.2%) was overestimated by two density classes. No estimates were off by three or more density classes.

Errors in estimation were not consistent across density classes, however. Agreement of estimates with counts generally decreased with the higher density classes (Table 2). Overestimates were restricted to density classes 0–2. Underestimates became more prevalent as density class increased (Figure 1). For density class 3 and higher, the number of underestimates exceeded cases of agreement with counts. For density class 6, all estimates were underestimates.

Plots without *P. filiformis* were rarely misclassified (91.9% agreement with counts). If plots without

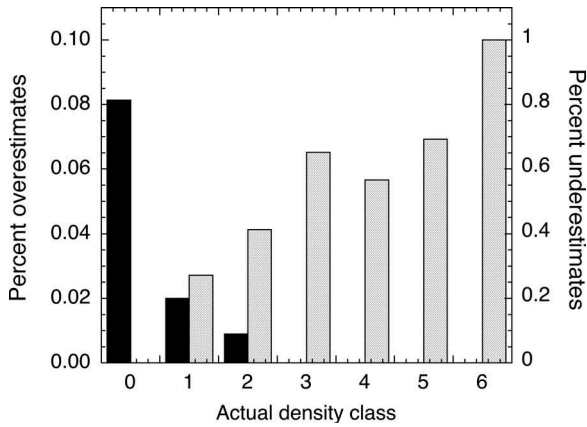


Figure 1. The percentage of overestimates (solid bars) and underestimates (stippled bars) for each actual density class, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015). Note the differences of scale on the y-axes; the scale for underestimates is an order of magnitude higher than the scale for overestimates.

P. filiformis were excluded, only 52.4% of accuracy estimates agreed with counts. Almost all misestimates were underestimates: 41.0% were underestimated by one density class and 5.7% were underestimated by two density classes. Only four plots were overestimated, all by one density class.

Patterns in observer error

The per cent agreement of estimates with counts summarised over all years varied widely among observers, ranging from 40% to 68% (Figure 2). The number of estimates made also ranged widely among observers, and the relationship between per cent agreement of estimates with counts and number of estimates was marginally significant and negative ($r = -0.79, P = 0.06, n = 6$).

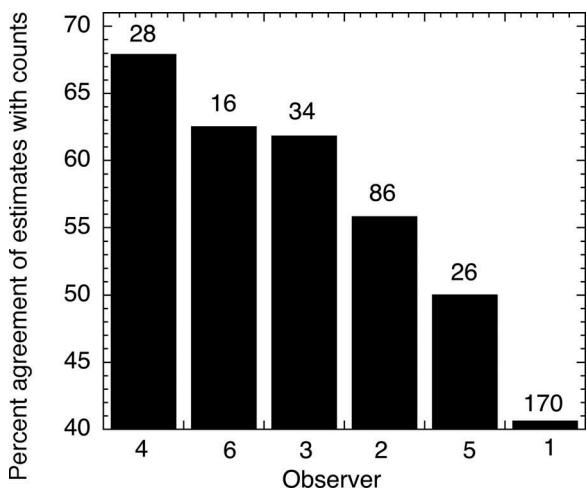


Figure 2. Overall per cent agreement of estimates with counts by observer, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015). Numbers above bars represent number of plots for which accuracy was evaluated for each observer.

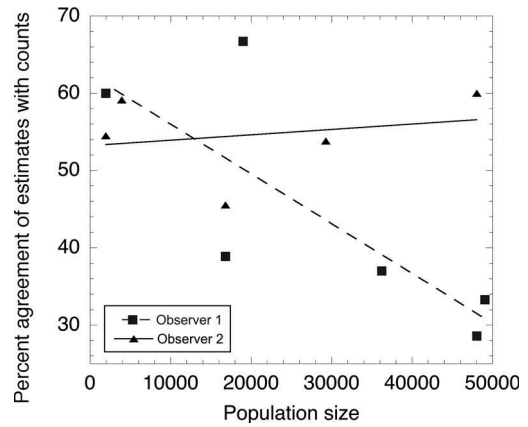


Figure 3. Per cent agreement of estimates with counts for two observers in different years, plotted as a function of total population size, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015) (observer 1: $y = 62.43 - 0.0006x, R^2 = 0.62$; observer 2: $y = 53.19 + 0.0001x, R^2 = 0.05$).

Two observers estimated abundances over multiple years (observer 1: 6 years; observer 2: 5 years). The per cent agreement of estimates with counts for observer 1 varied widely over time, ranging from 28.6% to 66.7% (range: 38.1) (Figure 3). In contrast, the per cent agreement of estimates with counts for observer 2 varied much less, from 45.5% to 60% (range: 14.5). Population abundance varied among years, however, and this could have affected agreement of estimates with counts. The two observers made estimates in eight total years, both making estimates in only three of the same years, although both made estimates over a similar range of population sizes. For observer 1, the relationship between per cent agreement of estimates with counts and population size was marginally significant and negative ($r = -0.79, P = 0.06, n = 6$). For observer 2, there was no association ($r = 0.23, P = 0.71, n = 5$).

Effects of experience and plant density

Per cent agreement of estimates with counts is plotted by years of experience of the observer in Figure 4. Much variability is obvious in the data. For the two observers with the most data points (observers 1 and 2), associations were tested but neither was significant (observer 1: $r = 0.18, P = 0.75, n = 6$; observer 2: $r = 0.59, P = 0.30, n = 5$). Much more variability in this relationship characterised observer 1 compared to observer 2.

The overall per cent agreement of estimates with counts for each observer (calculated over all years) was plotted as a function of the actual density in the plot (Figure 5). Per cent agreement of estimates with counts ranged from 0% to 100%, although it tended to be higher for the lower density classes. For the two observers with the most observations (observers 1 and 2), associations were tested and both were significant and negative (observer 1: $r = -0.83, P = 0.04, n = 6$; observer 2: $r = -0.95, P = 0.048, n = 4$). Observer 2 always had higher

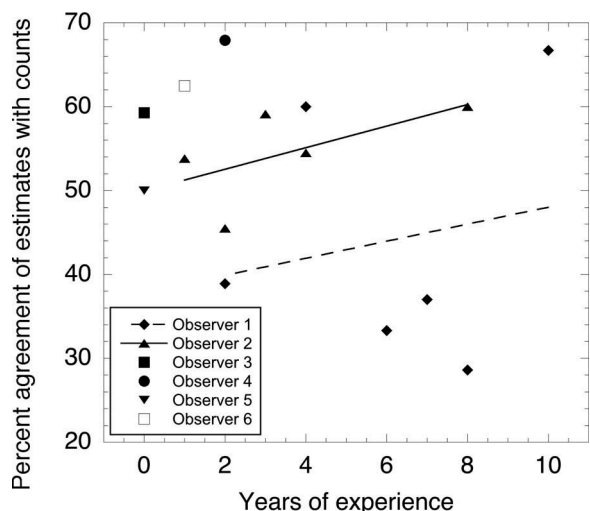


Figure 4. Per cent agreement of estimates with counts as a function of years of experience, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015). Six different observers are represented. Lines of best fit are shown for the two observers with the most years of data (observer 1: $y = 37.84 + 1.01x$, $R^2 = 0.03$; observer 2: $y = 49.94 + 1.29x$, $R^2 = 0.35$).

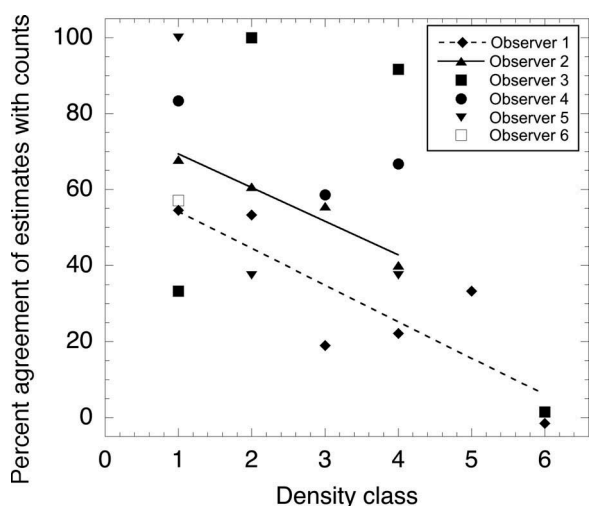


Figure 5. Per cent agreement of estimates with counts as a function of density class, from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015). Six different observers are represented. Lines of best fit are shown for the two observers with the most years of data (observer 1: $y = 63.31 - 9.41x$, $R^2 = 0.69$; observer 2: $y = 78.25 - 8.73x$, $R^2 = 0.90$).

agreement of estimates with counts than observer 1, even though observer 1 had more experience.

Because observer 1 estimated a larger number of plots than any of the other observers, and in more years, it is possible that the lower overall agreement of estimates with counts for observer 1 was due to estimating more of the high-density plots (density class 4 or higher), which were more difficult to classify accurately. Observer 1 did estimate a slightly larger proportion of high-density plots than

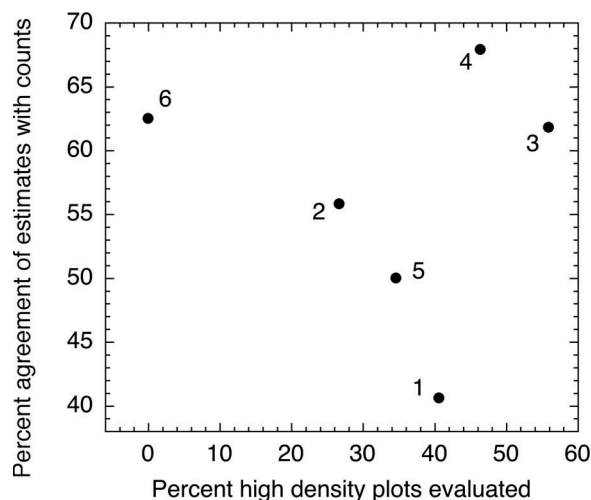


Figure 6. Per cent agreement of estimates with counts versus the percentage of estimated high-density plots (density class 4 or higher), from estimates of abundance of *Physaria filiformis* inhabiting a glade in south-western Missouri over 10 years (2006–2015). Each point represents a different observer. Numbers refer to the different observers.

existed in the data set overall (40.0% vs. 33.3%). There was, however, no relationship between agreement of estimates with counts and the percentage of high-density plots estimated among observers ($r = -0.05$, $P = 0.92$, $n = 6$) (Figure 6).

Discussion

Comparisons with other studies

In general, when evaluating observer error in vegetation sampling, one may attempt to measure accuracy (i.e. how close estimates are to true values) or precision (i.e. how close different observer’s estimates are to each other). In most studies of observer error, accuracy is unknown, as there exists no practical method of obtaining an unbiased value of the parameter in question. Thus, most studies of observer error have evaluated precision rather than accuracy (Morrison 2016). By carefully counting the numbers of individuals in a subset of plots, we were able to evaluate the accuracy of observer estimates.

Although the overall accuracy of observer estimates may seem low (only 63.6% of all plots counted agreed with estimates), it is very similar to that documented in studies of observer error using per cent cover as an indicator of abundance and evaluating precision. Lepš and Hadincová (1992), sampling relevés in meadows, clear cuts and peat bogs in the Czech Republic and using the Braun-Blanquet scale for cover, reported that between two observers 39.5% of estimates were different by one category, and 3% were different by more than one category. Klimeš (2003), working in 4-m² plots in grasslands also in the Czech Republic found that, after transformations of per cent cover data to the Braun-Blanquet scale, 46% of estimates between two observers were different by one

category, and 4% were different by two categories. Klimeš (2003) employed five total observers and calculated discrepancies based on all possible pairs of observers.

Archaux et al. (2007), working in 400 m² quadrats in French lowland forests, reported that estimates of cover-abundance using the Braun-Blanquet scale between two observers were different by one category for 47.5% of cases, and different by two categories for 11.5% of cases. Archaux et al. (2007) also used the Braun-Blanquet scale and, like Klimeš (2003), calculated discrepancies based on all possible pairs of observers (four total). Gray and Azuma (2005), sampling forests in Oregon, reported that for cover estimates, 41% of quadrats differed by one category and 6% differed by two categories. Cheal (2008), in an evaluation of the ability of 16 experienced observers to evaluate cover of a single grass species in open shrubland in Victoria, Australia, found variation that spread over three Braun-Blanquet categories.

Similarly, in a study of intra-observer error conducted in grasslands in the UK with subjective frequency as a measure of abundance, Hope-Simpson (1940) found that 36% of species differed by one category and 12% differed by two categories. Thus, for vegetation studies using category estimation of abundance in which precision has been evaluated, between one-third and one-half of all estimates are erroneous, but most are off by only one category. These are directly comparable to our estimates of accuracy – 32.2% of all plots were misestimated by one density class, and 4.3% were misestimated by two density classes – and suggest the magnitude of such error rates are ubiquitous in vegetation studies given the normal number and range of categories chosen to estimate abundance.

Systematic bias vs. random variation

Accuracy assessments revealed a strong bias in estimation of *P. filiformis* abundance – 33.5% of all estimates were underestimates, whereas only 3% were overestimates. Moreover, all overestimates occurred in the lower density classes (0–2), and the majority (78% of all overestimates) occurred when actual densities were zero. In terms of the effect on estimation of total population size, overestimates are trivial. For example, if the abundance of a plot was actually zero, but it was overestimated to be density class one (1–9 individuals), the point estimate of the total population would be erroneously overestimated by only 5. In contrast, underestimates were more prevalent for the higher density classes. Given that higher density classes were wider, underestimation in the higher density classes results in much greater error in the total population estimate. For example, if a plot was actually in density class 6 (1000–4999 individuals), but it was underestimated as density class 5 (500–999 individuals), the point estimate of the total population would be erroneously underestimated by 2250.

The mechanisms underlying the observed bias appear primarily to be associated with the growth form

of the plant and the habitat matrix. *P. filiformis* consists of multiple stems arising from a basal rosette, 1–4 cm in diameter (Young et al. 2008a). On average, an individual has two or three flowering stems (Thomas and Willson 1992), but as many as thirty flowering stems have been documented (Young et al. 2008a). Determining how many individual plants exist when each has multiple stems is difficult in dense populations, and there is an obvious tendency to underestimate abundances. Additionally, smaller plants growing in marginal habitats tend to be overlooked (Young et al. 2008b). Finally, plants growing along the borders of the plots may have been mistakenly assigned to the incorrect plot, and presumably account for the overestimates documented when no plants were actually present (Table 2).

Studies that have addressed the question of whether observer error represents random variation or a systematic bias have frequently found no evidence of a systematic bias (e.g. Smith 1944; Lepš and Hadincová 1992; Klimeš 2003). In a study of five different variables, Archaux et al. (2007) reported that the magnitude of random variation was twice as high as that of a systematic bias. Thus, our finding of a strong systematic bias is somewhat unusual.

Many studies have reported that some observers (or teams) tended consistently to record under- or overestimates compared to other observers (or teams) (Sykes et al. 1983; Tonteri 1990; Bråkenhielm and Qinghong 1995; Kercher et al. 2003; Carlsson et al. 2005). Frequently, a single observer (or team) has been documented to be relatively far from the group mean. Goodall (1952) reported that while the cover estimates of two observers agreed consistently, a third frequently made estimates that exceeded or fell below those of the other two. Tonteri (1990) reported that one observer tended to consistently overestimate cover compared to 10 others.

In a study using indices, Gorrod and Keith (2009) reported that the total site scores of one observer were consistently different from the group mean (10 observers total). Similar findings are evident from studies of species composition: McCune et al. (1997) reported that 1 observer out of 11 found very few species and inflated the between-crew variance, and Oredsson (2000) found that 1 observer out of 6 recorded significantly fewer species. Thus, in contrast to the relative uniqueness of the observed systematic bias, the inherent variability we documented among observers seems to be quite common.

Effect of experience

Perhaps surprisingly, there was no clear effect of experience on accuracy of estimates. The mechanism underlying the marginally significant negative relationship between per cent agreement of estimates with counts and number of estimates among observers is unclear, although it is the opposite of that expected given the hypothesis that experience improves accuracy. Observer fatigue may have played some role; observer 1, for example, estimated all

plots in 2 years (2011 and 2012). Additionally, less training would have been conducted in these 2 years, as no new observers were involved.

The effect of prior experience has been evaluated in numerous studies of vegetation sampling. The results have been mixed, as some studies have found an effect of experience (Hall and Okali 1978; McCune et al. 1997; Oredsson 2000; Scott and Hallam 2002; Ringvall et al. 2005; Vittoz and Guisan 2007; Bergstedt et al. 2009) whereas others have not (Sykes et al. 1983; Kéry and Gregg 2003; Cheal 2008; Chen et al. 2009; Moore et al. 2011; Burg et al. 2015). It has been suggested that the equivocal effect of experience may be due to some degree to wide ranges in the amount of relevant experience among observers, which differed greatly among studies (Morrison 2016). Given that all experience in estimating abundance of *P. filiformis* in this study was relatively uniform, this could not be the explanation for the findings presented here. Rather, the available evidence suggests that experience does not increase the accuracy of estimates, or at least not beyond a relatively low threshold.

The effect of training must also be considered, however, as training for a particular sampling regime is a type of experience. The literature is less equivocal about the effect of training, as numerous studies have reported that training increased the precision or accuracy of estimates to some degree (Smith 1944; Kennedy and Addison 1987; Stapanian et al. 1997; Campbell and Arnold 1973; Murphy and Lodge 2002; Symstad et al. 2008; but see Archaux et al. 2009 for an exception). All observers in this study of *P. filiformis* sampling received training in each year prior to sampling. Once a certain amount of training has been done, the additional effect of previous-related experience may be trivial.

Subjective estimation ability

Estimation error appears to be inevitable to some extent, reflecting an inherent limitation of our subjective ability to quantify objects accurately. Categories are frequently employed for cover estimation due to the inability to visually estimate cover precisely. Hahn and Scheuring (2003), using computer simulations to test estimation of per cent cover, found that estimation error was minimal when the range of cover was divided into 10 equal categories. Most test subjects divided the cover range into 10–20 intervals in their minds, even when they were given the opportunity to make more precise estimates.

Plots with actual values near the boundaries of the categories should be the most difficult to categorise. We evaluated how much of the overall error such plots accounted for by applying a 10% error factor to densities occurring near density class boundaries, as in Young et al. (2008b). For example, an estimate of density class 4 (100–499 individuals) would be considered correct if the actual count was between 90 and 549 plants. Of the 208 total misestimates in which plants were actually present, only

12.5% represented cases in which actual counts were within 10% of the boundaries (15.7% excluding cases in which plants were present but 0 plants were estimated). Thus, relatively few cases of incorrect classification involved cases in which actual abundances were close to density class boundaries.

What can be done?

Sykes et al. (1983) suggested that if individual observers either consistently underestimated or overestimated cover, a correction factor could be used to reduce inter-observer bias. In community studies, a separate correction factor would be necessary for each species, and this approach, to our knowledge, has not been attempted. For diverse communities, such an approach using multiple correction factors may be impractical. For a single species, however, application of a correction factor to individual observer estimates would be practical. The error rates documented could be used to assign category-dependent correction factors as suggested by Young et al. (2008b). Given the documented inter-observer variability, however, such correction factors may need to be both density class-dependent and observer-specific (e.g. a separate correction factor for each observer-density class combination).

Sykes et al. (1983) also suggested ‘screening’ observers and rejecting individuals who were unable to produce ‘acceptably consistent results’ after training. Given that many studies have found evidence that single observers or teams produced estimates that were outliers relative to the group mean (Goodall 1952; Tonteri 1990; McCune et al. 1997; Oredsson 2000; Gorrod and Keith 2009), the existence of such ‘extreme’ observers seems to be common and rejecting such individuals could increase overall precision.

Multiple observers are frequently recommended (Klimeš et al. 2001; Klimeš 2003; Symstad et al. 2008; Archaux 2009; Archaux et al. 2009; Gorrod and Keith 2009; Vittoz et al. 2010). The advantage of multiple observers is that extreme estimates may be adjusted (e.g. Klimeš 2003), and average estimates may be closer to true values as errors associated with individual estimates are cancelled out (Wintle et al. 2013). The main disadvantage of additional observers would be greater cost.

Other factors may also affect the precision or accuracy of estimates in vegetation sampling (see review in Morrison 2016). Field workers should be vigilant regarding these sources of error and include error rates as standard components of reports (Kercher et al. 2003). Although it will not usually be possible to eliminate observer error, ignoring such error is not a responsible practice.

Acknowledgements

Michael DeBacker provided helpful comments on an earlier draft. Laszlo Nagy and an anonymous reviewer contributed many helpful suggestions to 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, US Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Lloyd Morrison is the quantitative ecologist for the Heartland Network of the National Park Service Inventory and Monitoring Program (NPS I&M) and adjunct faculty at Missouri State University. His research interests include biogeography, macroecology and conservation. His publications can be found at <http://people.missouristate.edu/LloydMorrison/>.

Craig Young is an ecologist and exotic plant management team leader with the Heartland Network of the NPS I&M, with an interest in the use of empirical data in land management decisions.

References

- Al-Shehbaz IA. 2010. Brassicaceae. In: Editorial Committee, editors. *Flora of North America*. New York (NY): Oxford University Press. p. 224–746.
- Archaux F. 2009. Could we obtain better estimates of plot species richness from multiple-observer plant censuses? *Journal of Vegetation Science* 20:603–611.
- Archaux F, Bergès L, Chevalier R. 2007. Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecology* 188:179–190.
- Archaux F, Camaret S, Dupouey J-L, Ulrich E, Corcket E, Bourjot L, Brêthes A, Chevalier R, Dobremez J-F, Dumas Y, et al. 2009. Can we reliably estimate species richness with large plots? An assessment through calibration training. *Plant Ecology* 203:303–315.
- Bergstedt J, Westerberg L, Milberg P. 2009. In the eye of the beholder: bias and stochastic variation in cover estimates. *Plant Ecology* 204:271–283.
- Bråkenhielm S, Qinghong L. 1995. Comparison of field methods in vegetation monitoring. *Water, Air, & Soil Pollution* 79:75–87.
- Burg S, Rixen C, Stöckli V, Wipf S. 2015. Observation bias and its causes in botanical surveys on high-alpine summits. *Journal of Vegetation Science* 26:191–200.
- Campbell NA, Arnold GW. 1973. The visual assessment of pasture yield. *Australian Journal of Experimental Agriculture and Animal Husbandry* 13:263–267.
- Carlsson ALM, Bergfur J, Milberg P. 2005. Comparison of data from two vegetation monitoring methods in semi-natural grasslands. *Environmental Monitoring and Assessment* 100:235–248.
- Cheal D. 2008. Repeatability of cover estimates? *Ecological Management & Restoration* 9:67–68.
- Chen G, Kéry M, Zhang J, Ma K. 2009. Factors affecting detection probability in plant distribution studies. *Journal of Ecology* 97:1383–1389.
- Elzinga CL, Salazar DW, Willoughby JW, Gibbs JP. 2001. *Monitoring plant and animal populations*. Oxford (UK): Blackwell Science.
- Goodall DW. 1952. Some considerations in the use of point quadrats for the analysis of vegetation. *Australian Journal of Biological Science* 5:1–41.
- Gorrod EJ, Keith DA. 2009. Observer variation in field assessments of vegetation condition: implications for biodiversity conservation. *Ecological Management & Restoration* 10:31–40.
- Gray AN, Azuma DL. 2005. Repeatability and implementation of a forest vegetation indicator. *Ecological Indicators* 5:57–71.
- Hahn I, Scheuring I. 2003. The effect of measurement scales on estimating vegetation cover: a computer-assisted experiment. *Community Ecology* 4:29–33.
- Hall JB, Okali DUU. 1978. Observer bias in a floristic survey of complex tropical vegetation. *The Journal of Ecology* 66:241–250.
- Hope-Simpson JF. 1940. On the errors in the ordinary use of subjective frequency estimations in grassland. *The Journal of Ecology* 28:193–209.
- IBM Corp. 2011. *IBM SPSS statistics, version 20.0*. Armonk (NY): IBM Corp.
- Kennedy KA, Addison PA. 1987. Some considerations for the use of visual estimates of plant cover in biomonitoring. *The Journal of Ecology* 75:151–158.
- 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.
- Kéry M, Gregg KB. 2003. Effects of life-state on detectability in a demographic study of the terrestrial orchid *Cleistes bifaria*. *Journal of Ecology* 91:265–273.
- Klimeš L. 2003. Scale-dependent variation in visual estimates of grassland plant cover. *Journal of Vegetation Science* 14:815–821.
- Klimeš L, Dančák M, Hájek M, Jongepierová I, Kučera T. 2001. Scale-dependent biases in species counts in a grassland. *Journal of Vegetation Science* 12:699–704.
- Legg CJ, Nagy L. 2006. Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management* 78:194–199.
- Lepš J, Hadincová V. 1992. How reliable are our vegetation analyses? *Journal of Vegetation Science* 3:119–124.
- McCune B, Dey JP, Peck JE, Cassell D, Heiman K, Will-Wolf S, Neitlich PN. 1997. Repeatability of community data: species richness versus gradient scores in large-scale lichen studies. *The Bryologist* 100:40–46.
- Moore JL, Hauser CE, Bear JL, Williams NSG, McCarthy MA. 2011. Estimating detection-effort curves for plants using search experiments. *Ecological Applications* 21:601–607.
- Morrison LW. 2016. Observer error in vegetation surveys: a review. *Journal of Plant Ecology* 9:367–379.
- Morrison LW, Smith DR, Young CC, Nichols DW. 2008. Evaluating sampling designs by computer simulation: a case study with the Missouri bladderpod. *Population Ecology* 50:417–425.
- Murphy SR, Lodge GM. 2002. Ground cover in temperate native perennial grass pastures. I. A comparison of four estimation methods. *The Rangeland Journal* 24:288–300.
- Oredsson A. 2000. Choice of surveyor is vital to the reliability of floristic change studies. *Watsonia* 23:287–291.
- 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.
- Scott WA, Hallam CJ. 2002. Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology* 165:101–115.
- Smith AD. 1944. A study of the reliability of range vegetation estimates. *Ecology* 25:441–448.
- Stapanian MA, Cline SP, Cassell DL. 1997. Evaluation of a measurement method for forest vegetation in a large-scale ecological survey. *Environmental Monitoring and Assessment* 45:237–257.
- Sykes JM, Horrill AD, Mountford MD. 1983. Use of visual cover assessments as quantitative estimators of some British woodland taxa. *The Journal of Ecology* 71:437–450.
- Symstad AJ, Wienk CL, Thorstenson AD. 2008. Precision, repeatability, and efficiency of two canopy-cover estimate

- methods in northern Great Plains vegetation. *Rangeland Ecology & Management* 61:419–429.
- Thomas LP, Willson GD. 1992. Effects of experimental trampling on the federally endangered species, *Lesquerella filiformis* Rollins, at Wilson's Creek National Battlefield, Missouri. *Natural Areas Journal* 12:101–105.
- Tonteri T. 1990. Inter-observer variation in forest vegetation cover assessments. *Silva Fennica* 24:189–196.
- Vittoz P, Bayfield N, Brooker R, Elston DA, Duff EI, Theurillat J-P, Guisan A. 2010. Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *Journal of Vegetation Science* 21:1035–1047.
- 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.
- Wintle BC, Fidler F, Vesk PA, Moore JL. 2013. Improving visual estimation through active feedback. *Methods in Ecology and Evolution* 4:53–62.
- Young CC, Kelrick MI, Morrison LW, DeBacker MD, Haack JL, Rowell GA. 2008a. Missouri bladderpod monitoring protocol for Wilson's creek national battlefield. Natural Resource Report NPS/MWR/HTLN/NRR—2008/043. Fort Collins (CO): National Park Service.
- Young CC, Morrison LW, Haack JL. 2009. Habitat relationships and management implications for *Lesquerella filiformis* Rollins (Missouri bladderpod) on a xeric limestone prairie. *The Journal of the Torrey Botanical Society* 136:233–241.
- Young CC, Morrison LW, Kelrick MI, DeBacker MD. 2008b. Monitoring *Lesquerella filiformis* Rollins (Missouri bladderpod): application and evaluation of a grid-based survey approach. *Natural Areas Journal* 28:370–378.