The Statistical Power Of Rangeland Monitoring Data

A range consultant shares his thoughts on when numbers are really justified – and when they aren't.

By Peter Sundt

Demands for monitoring of all types of natural resources have steadily increased in recent years. Monitoring of rangeland condition and grazing utilization are now required by court order on federal grazing leases, and agencies are scrambling to comply.

"Monitoring" has joined "ecosystem" as a catchword of progressive resource management without which no proposal is complete. But in the clamor for more quantitative information little attention has been paid to the methods of monitoring, and even less to the statistical power of those methods to detect change.

This study evaluates power in a set of monitoring data in which the point-line method was used to estimate foliar and basal cover of rangeland vegetation. Statistically valid quantitative monitoring is more difficult than people seem to realize. I conclude that some of the rangeland monitoring being called for could likely be accomplished by qualitative evaluations by experienced professionals, and numbers should be reserved for properly-designed projects in which statistical power is adequate.

Methods For The Study

The data set consists of cover estimates for 48 study plots comprising an ongoing research project of the US Forest Service Rocky Mountain Research Station, in which Chihuahuan desert scrub vegetation was subjected to roller-chopping and grass-seeding in hopes of converting scrub to grass.

At each of four separate locations in the vicinity of Douglas, Arizona, twelve 500 ft x 500 ft plots were arranged contiguously in a large grid. At each site, eight of the twelve study plots were treated by pulling a large water-filled drum with sharp edges over the vegetation. Half of the treated plots were broadcast-seeded with species of native perennial grasses. Four plots were left as untreated controls.

Before and after treatment the cover of rock, bare soil, litter, the foliar cover of all plant species and the basal cover of perennial grasses were estimated by the point-line method, primarily to evaluate the success of the treatments in replacing shrubs with grass. Ten temporary transects per plot were systematically located and at 100 points along the transects a sharp steel pin was used to identify the plant or soil category occurring at the point, for a total of 100 points per transect, 1000 points per plot. Transect data was kept separate to estimate within-plot variance. The data set analyzed for this study consists of estimates from one pre-treatment and two post-treatment years.

Statistical Power And Minimum Detectable Change

Statistical power is a way to measure how effective monitoring methods are in detecting change. Power can be defined as the probability that a real change from one sampling time to the next is detected and not mistaken for sampling error or random variation.

Related to power is minimum detectable change (MDC): the smaller the minimum detectable change the more powerful the methods. Given the number of samples and the variance of the data (i.e. the variation of cover values among the 10 transects in a plot), one specifies the acceptable probabilities of false-change and missed-change errors (alpha and beta) and can calculate the minimum change that can be detected (see formula in sidebar).

For this analysis I specified alpha=beta=0.20; in other words I want 80% confidence that an observed difference in mean cover value from one year to another is not due to random events (such as variation in point placement). I also want to be at least 80% sure that a real change will be detected. The more confidence one requires the more difficult it is to detect change or to distinguish real change from random effects. All else equal, the more con-

Formula for MDC
From Zar (1984):

$$MDC = [\sqrt{(2s^2/n)}]\,(t_{\alpha,v} + t_{\beta,v})$$

where MDC=minimum detectable change, s^2 = pooled sample variance, n = sample size, t_α = t value for alpha (two-tailed), t_β = t value for beta (one-tailed), v = degrees of freedom (= n-1).

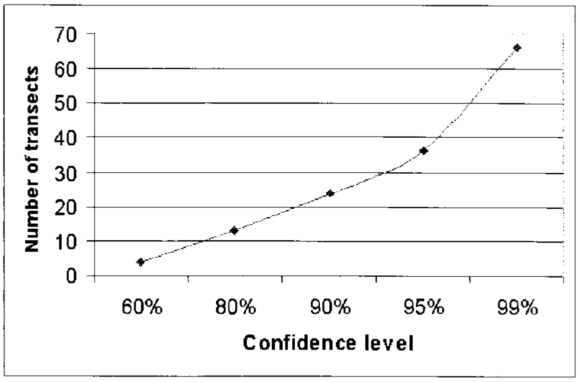

Fig. 1. *Sample sizes (number of transects) needed to detect a 20% change in the bare ground cover at plot GW1, at various levels of confidence. For 60% confidence alpha=beta=0.4, for 80% confidence alpha=beta=0.2, etc. Calculated from the formulae in Elzinga (1998), p. 346.*

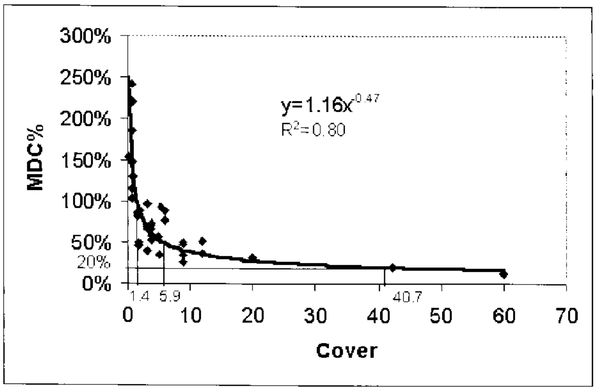

Fig. 2. *Minimum detectable change (expressed as a percentage of initial cover) vs. initial cover for plot GW6. The best-fit curve is illustrated and cover values at MDC%=100%, 50% and 20% calculated from the equation of the curve are indicated. R^2 is a measure of the accuracy with which the curve fits the data points (values range from 0 to 1.0).*

fidence one demands, the more samples are needed (Figure 1). I have settled on 80% confidence as a reasonable level for most rangeland monitoring. For most biological research the required level of confidence is 95%.

Evaluating Power

Minimum detectable change is a function of alpha (the acceptable probability of false change error), beta (the acceptable probability of missed change error), sample size, and pooled sample variance. Consider the data in Table 1 from one of the 48 study plots. A change between years 1 and 2 from the initial cover of 60 to at least 60+7=67 or to less than 60-7=53 could have been detected – any smaller change would have to be attributed to random variation.

In percentage terms, the methods were sufficiently powerful to detect changes of 12% or more of the initial cover value between years 1 and 2. In later years the methods were less powerful largely because of the increased pooled variance—a minimum change of 13% was necessary between years 1 and 3, and of 19% between years 2 and 3.

Table 1. Data for bare ground cover in plot GW6 for 3 sample times (1996, 1999, 2000). The upper half of the table compares the mean cover values and variances among the 3 years. The lower half indicates the pooled sample variance, the minimum detectable change (MDC) and minimum detectable change as a percent of the mean initial cover value (MDC%) for each between-year comparison.

YEAR	1	2	3
MEAN COVER	60	42	67
VARIANCE	43	51	73

YEARS COMPARED	1 to 2	1 to 3	2 to 3
POOLED VAR.	47	58	62
MDC	7	8	8
MDC as % of cover	12%	13%	19%

These calculations were made for all the species and soil categories in all 48 plots for all possible between-year comparisons, about 1575 comparisons in all. To distill the enormous amount of data into a readily evaluated form, graphs were produced of MDC% versus cover for each plot (MDC% is the minimum detectable change expressed as a percent of the initial cover). One such graph appears as Figure 2. The power of the method is greatest for species with high initial cover. Most species in this plot have cover values less than 10% and any change in their cover must be 20-250% to be detected.

The data in Figure 2 can be fitted with a curve, the equation of which is a negative exponential function. While some species' points lie above or below this curve, the equation is a useful generalization of the method's power integrated across all species. For the average species in Figure 2 with initial cover less than 1.4% a change must be 100% or more to be detected; for species with initial cover between 1.4% and 5.9%, the change must generally be 50%-100%; and for those species with initial cover greater than 40.7%

Table 2. Cover values for MDC%=100%, 50% and 20% calculated from equations of best-fit curves for all the data pooled at each of the 4 study sites, and for all data pooled in the entire dataset. In each column the range of cover values calculated from the curves from each of the 12 plots at each site is in parentheses. In the right hand column the results are from the graph made by pooling all the data in the dataset (Figure 3).

MDC%	Site WG	Site GW	Site ER	Site SB	All data pooled
100%	2 (1-2)	2 (1-2)	2 (1-4)	2 (1-3)	2
50%	8 (4-15)	7 (4-14)	7 (5-9)	13 (6-32)	8
20%	49 (20->100)	48 (23->100)	58 (33->100)	132 (41->100)	61

a change of only 20% could be detected. Such graphs and their equations were created for each plot in the dataset, and provide a useful way to compare the power of the methods among plots (Table 2).

For Most Species Power Is Low

In Figure 3 all of the data in the study has been combined into a single graph, and the curve provides an overall assessment of the power of the monitoring methods integrated across all species in all plots at all sites in the study.

Generalizing from the equation of the curve one can say that for the average species with less than 2% initial cover a change must be 100% or more of the initial cover to be detected with 80% confidence of making neither a missed change nor a false change error. For the average species with cover between 8-61%, a 50% change can be detected, while for those species with initial cover ≥ 61% a change need be only 20% of the initial value to be detected.

Figure 4 shows the distribution of all year-to-year comparisons in the dataset, with corresponding ranges of minimum detectable change. For almost one third of the species in the data set the change must have been

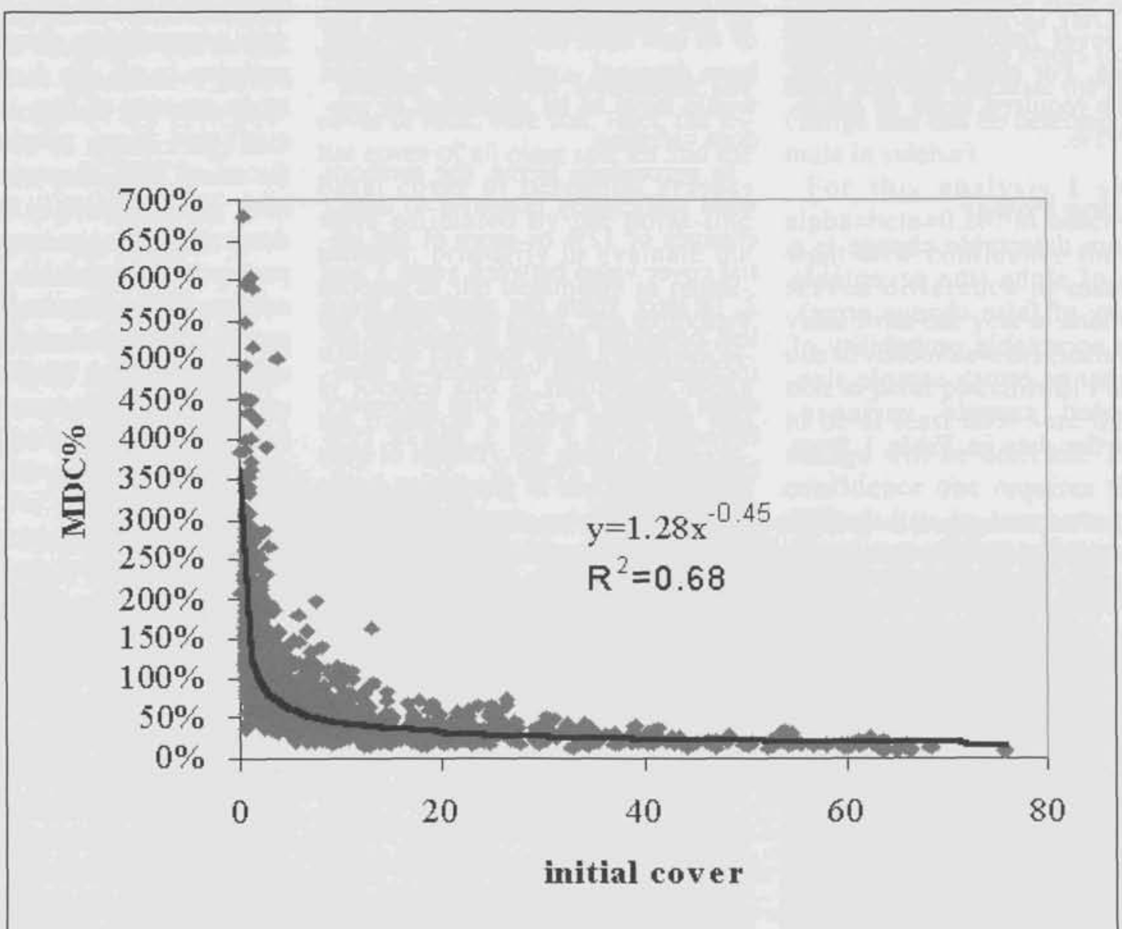

Fig. 3. Minimum detectable change as a percent of initial cover (MDC%) graphed as a function of initial cover for all 1575 between-year comparisons in the dataset. The equation of the best-fit curve and its R^2 value are indicated. R^2 is a measure of the accuracy with which the curve fits the data points (values range from 0 to 1.0).

100% to 700% of the initial cover value to be detected--rather low power of the methods to detect change for many species.

Fortunately, most conclusions in this project and in rangeland monitoring generally are based on changes in common species and in bare ground cover, for which the methods have reasonable power. For bare ground, which in this data set is generally 50–75% cover, the methods are quite sensitive. The top 2 or 3 species in a plot generally exceed 2% cover, so that a 50-100% change in their initial cover values can be detected. These methods are rarely able to detect a change in basal cover of perennial grass species, however, simply because basal cover values are generally <1% and rarely exceed 2% in this study.

How To Boost Statistical Power

Much rangeland monitoring has been conducted with sample sizes considerably smaller than the 1,000 points per plot used in this study. Cover data will often be taken in conjunction with frequency data, using a single point indicator on the frequency frame (or a mark on a boot), and typically resulting in cover estimates based on 100 or 200 points.

In many rangeland situations these sample sizes are far too small for adequate statistical power, even for common species. Furthermore, many monitoring projects do not incorporate replication (i.e. by keeping data separate by transect) and so no evaluation of sample variance and confidence can be made.

Statistical power can be increased by

a) increasing sample sizes,

b) reducing within-plot and between-year variances, and

c) lumping species into groups to attain larger cover values.

More samples are almost always desirable, but the cost in time and effort is directly related to sample size. Reducing variance can often increase power with less effort than raising the sample size. Variance can be minimized by carefully selecting homogeneous vegetation for study plots, by permanently marking sample points or transects, and in some cases by using a rectangular rather than a square sampling frame. Lumping uncommon species into functional groups (i.e. annual forbs, perennial grasses) can boost power because, as this study shows, power increases with cover values. It is

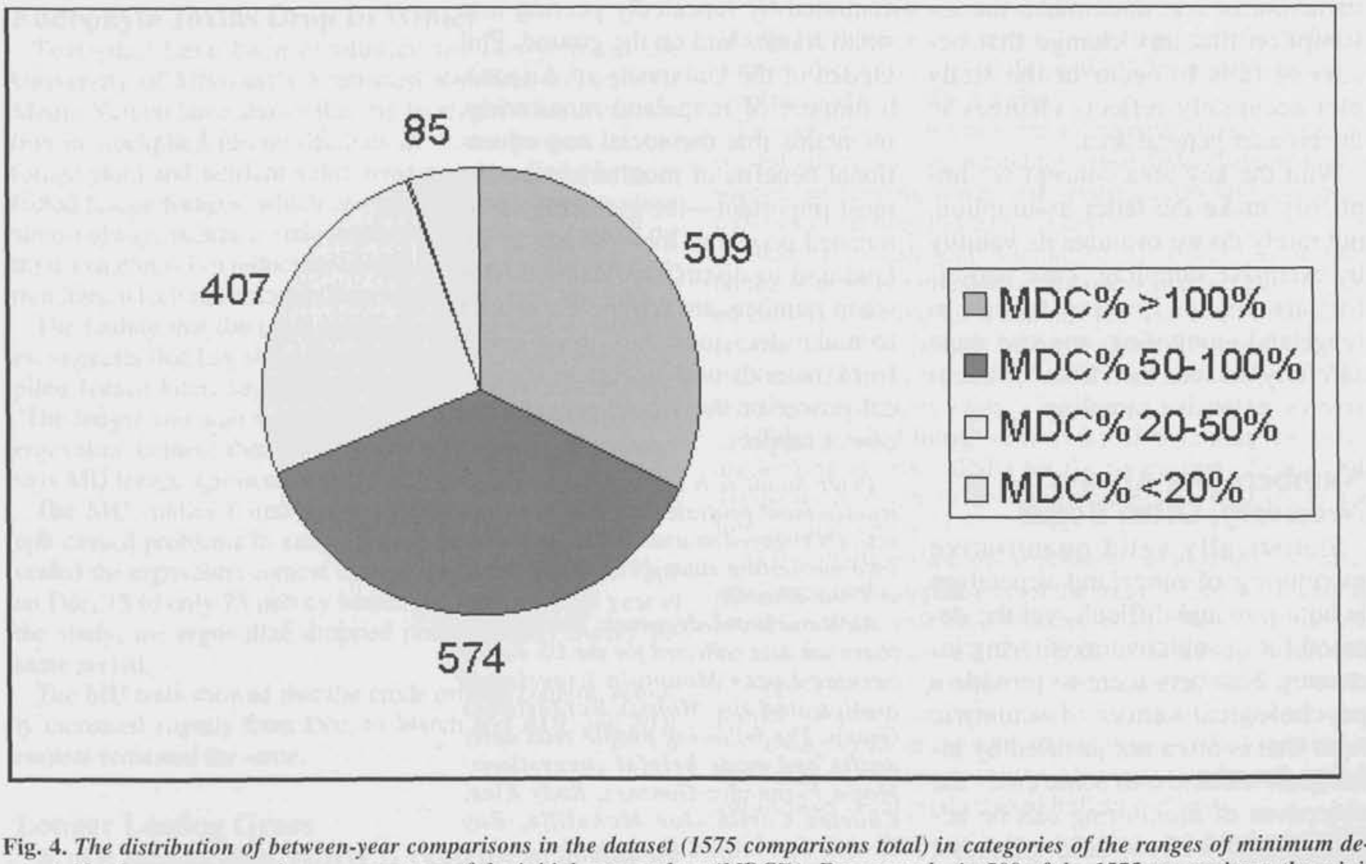

Fig. 4. ***The distribution of between-year comparisons in the dataset (1575 comparisons total) in categories of the ranges of minimum detectable change expressed as a percentage of the initial cover values (MDC%). For example, in 509 of the 1575 comparisons the minimum detectable change must have exceeded 100% of the initial cover value to be detected with 80% confidence.***

easier to detect change in the group "annual forbs" than in any individual species of annual forb, and oftentimes it is a change in the functional group, rather than the individual species, that is of most interest. In my data set, lumping the basal covers of all perennial grasses into a group would make changes in this important vegetation component much easier to detect.

There is an unavoidable tradeoff between statistical power, which comes from large sample sizes and minimized variance, and adequate representation of the heterogeneity of real rangeland vegetation. With a small plot, permanently-marked transects and large numbers of unbiased samples one can attain considerable power to detect change within the study plot. But small plots encompass less of the real variation that exists, so there must be several such plots located to represent the variation, or one must make the assumption that any change that occurs or fails to occur in the study plot accurately reflects changes in the broader general area.

With the key area concept we implicitly make the latter assumption, but rarely do we evaluate its validity by extensive sampling. Time and effort are always limiting factors in rangeland monitoring, and one must carefully balance the values of intensive vs. extensive sampling.

Numbers Not Always Necessary, Often Bogus

Statistically valid quantitative monitoring of rangeland vegetation is both rare and difficult, yet the demand for quantitative monitoring increases. Numbers seem to provide a psychological veneer of scientific rigor that is often not justified by inadequate methods. In some cases the objectives of monitoring can be accomplished without hard numbers, based on professional opinion guided by semi-quantitative methods.

For example, an experienced observer can learn to estimate cover in broad categories (e.g. 10–25%, 25–50%) by calibration with small, intensively sampled plots or single transects, and such estimates may often provide sufficient precision for the objectives of the project. Qualitative assessments of species composition, erosion, and ground cover supplemented by photos are adequate for many projects and are always preferable to bogus numbers (reported with exaggerated significant digits and no estimates of uncertainty) resulting from inadequate methods.

There is a place for rigorous quantitative methods, such as when early detection of change is important and cannot be accomplished by visual inspection, or when, for legal reasons, an objective, quantitative record is demanded. Furthermore, sampling of vegetation has collateral benefits—one's attention is wonderfully concentrated by repeatedly peering into small frames laid on the ground. Phil Ogden of the University of Arizona, a pioneer of rangeland monitoring, maintains that the social and educational benefits of monitoring are the most important—the gathering of interested people to look closely at the land and to discuss its changes. But when numbers are reported and used to make decisions they must result from methods with adequate statistical power or the whole process becomes suspect.

Peter Sundt is a SRM certified range management professional who has been actively engaged as a consultant in rangeland monitoring since 1990. He is based in Pima, Arizona.

Author acknowledgements: This study is based on data collected for the US Forest Service Rocky Mountain Experiment Station and the Malpai Borderlands Group. The following people read early drafts and made helpful suggestions: Maria Fernandez-Gimenez, Rudy King, Charles Curtin, Joe McAuliffe, Ray Turner, Larry Allen, Gary Frasier and an anonymous reviewer.

References

Bonham, CD. 1989. Measurements for terrestrial vegetation. John Wiley & Sons. New York.

Elzinga, CL, DW Salzer and JW Willoughby. 1998. Measuring and monitoring plant populations. BLM Technical Reference 1730-1. This excellent reference book can be obtained free from BLM National Business Center, BC-650B, P.O. Box 25047, Denver, Colo., 80225-0047.

Zar, JH. 1984. Biostatistical analysis. Prentice-Hall. Englewood Cliffs, New Jersey.