

# Evaluating Livestock Grazing Use With Streambank Alteration Protocols: Challenges and Solutions

Jeremiah D. Heitke,<sup>1</sup> Richard C. Henderson,<sup>2</sup> Brett B. Roper,<sup>3</sup> and Eric K. Archer<sup>2</sup>

Authors are <sup>1</sup>Fisheries Biologist and <sup>2</sup>Biological Scientist, PACFISH/INFISH Biological Opinion Effectiveness Monitoring Program, National Forest Systems, USDA Forest Service, Logan, UT 84321, USA; and <sup>3</sup>National Aquatic Monitoring Coordinator, Fish and Aquatic Ecology Unit, National Forest Systems, USDA Forest Service, Logan, UT 84321, USA.

## Abstract

Appropriate management of livestock in riparian areas can help ensure that these ecosystems are maintained. We evaluated how one indicator of livestock grazing in riparian areas, streambank alteration, was affected by choices related to protocols and personnel used for these assessments. We found that although streambank alteration protocols were generally repeatable among observers, results were affected by factors not directly related to grazing intensity, including 1) training, 2) professional background, 3) location and intensity of measurements, and 4) the protocol used. Training reduced estimates of alteration and observer variability. Rangeland professionals had higher estimates of streambank alteration than seasonal technicians. Rapid assessments of alteration were correlated with more intensive estimates; however, the relationship was not 1:1. Different protocols resulted in different alterations estimates when alterations at the same locations were estimated. Given the large number of monitoring programs, personnel, and methods used to assess streambank alteration, we suggest more thought be given on how to standardize monitoring efforts so results consistently reflect the true amount of alteration at a site. We also remind managers that no protocol can be implemented without some error. Managers should therefore be careful when taking action based on a single evaluation—especially when the result is near a management standard or threshold. When these concerns are addressed, indicators such as streambank alteration can help ensure management decisions maintain both sustainable allotments and landscapes.

## Resumen

El manejo apropiado del ganado en áreas riverseñas puede favorecer el mantenimiento de estos ecosistemas. El objetivo fue evaluar la alteración del banco del arroyo, cómo un indicador del pastoreo del ganado en áreas riverseñas y cómo las decisiones tomadas y el personal utilizado para estas evaluaciones afectaron este indicador. Se encontró que mientras que las propuestas de alteración del banco del arroyo fueron generalmente repetibles entre observadores, los resultados fueron afectados por factores no directamente relacionados con la intensidad del pastoreo, incluyendo; 1) entrenamiento, 2) experiencia profesional, 3) localización e intensidad de las medidas, y 4) la propuesta que utilizaron. El entrenamiento redujo estimaciones de la alteración y de la variabilidad del observador. Los expertos en pastizales hicieron estimaciones más altas de la alteración del banco del arroyo que la que hicieron los técnicos temporales. Las evaluaciones rápidas de la alteración fueron correlacionadas con estimaciones más intensivas; sin embargo la relación no fue de 1:1. Diversas propuestas dieron lugar a distintas estimaciones de la alteración en las mismas localidades. Dado el gran número de programas de monitoreo, de personal, y métodos utilizados para determinar la alteración del banco del arroyo sugerimos que debe tenerse más cuidado para estandarizar esfuerzos para supervisión. Asimismo, los resultados reflejan consistentemente el grado de alteración en un sitio. También sugerimos a los encargados que ninguna propuesta se puede aplicar sin cierto error. Los encargados debieran por lo tanto tener precaución al tomar medidas basadas en una sola evaluación, especialmente cuando el resultado está cerca de un estándar o de un umbral de manejo. Cuando se consideran estos factores, los indicadores tales como alteración del banco del arroyo ayudan a tomar decisiones de manejo para mantener una asignación y paisajes sostenibles.

**Key Words:** allowable use standards, annual monitoring indicators, observer variability, riparian grazing, training

## INTRODUCTION

Unmanaged livestock grazing can negatively affect conditions and processes in riparian areas (Platts 1991). Negative impacts associated with riparian grazing are often the focus of lawsuits seeking to question grazing practices on federally managed

lands (Shepard 2007). Appropriate allotment planning and administration of livestock use in riparian areas not only minimizes these negative effects (Bengeyfield 2006), but can also be used to maintain or improve rangeland conditions (Blackmore and Vitousek 2000; Marty 2005). Therefore, it is important that indicators of riparian conditions be identified and standards set for those indicators so managers can quickly assess whether management was appropriate and take action if it was not (Clary and Leininger 2000).

Stream and riparian characteristics, such as streambank angle, prevalence and depths of undercut banks, and riparian community composition, can be good indicators of grazing intensity (Platts et al. 1987; Bauer and Burton 1993; Green and

Research was funded by the USDA Forest Service Fish and Aquatic Ecology Unit and USDA Forest Service and USDI Bureau of Land Management PACFISH/INFISH Biological Opinion Effectiveness Monitoring Program.

Correspondence: Brett B. Roper, USDA Forest Service, Fish and Aquatic Ecology Unit, Forest Sciences Laboratory, Logan, UT 84321, USA. Email: broper@fs.fed.us

Manuscript received 13 February 2008; manuscript accepted 19 August 2008.

**Table 1.** Definitions of the types of alteration used to define streambank alteration in all four protocols evaluated in this study. A streambank was defined as altered when the presence of any of these three alterations was determined to have occurred in the current year.

Types of alteration	Definition
Shearing	Removal of a portion of the streambank by ungulate hooves leaving a smooth vertical surface and an indentation of a hoofprint at the bottom or along the sides.
Trampling	Indentation of a hoofprint and exposed roots or soil, resulting in a depression at least 13 mm deep or soil displacement at least 13 mm upwards.
Trailing	Trails and other severe trampling were counted as alteration if there were signs of current-year use. Because of the compacted soils, trailing was counted even if hoofprints did not result in 13-mm displacement of soil.
Current-year alteration	Discernible from previous years' alteration because of weathering effects of freeze/thaw cycles, rain events, and erosion by stream flow or vegetative regrowth.

Kauffman 1995; Clary 1999; Winward 2000). The conditions of these attributes, however, change slowly over the time frame of several grazing seasons (Green and Kauffman 1995), making them most useful as long-term indicators/goals of riparian grazing management plans. To provide annual accountability for riparian grazing programs, managers instead rely on indicators that reflect within-season disturbance and promote the attainment of long-term indicators.

Examples of annual indicators include measures of streambank alteration by livestock, height of forage remaining following grazing (stubble height), and amount of woody species utilization (Clary and Leininger 2000; Turner and Clary 2001; Bengeyfield 2006). Each of these indicators has been used as end-of-season compliance standards or indicators for riparian areas, is quick and cost effective to measure (Cowley and Burton 2005), and, if maintained above an annual threshold, facilitates long-term riparian health.

Although not explicitly stated, use of annual indicators for management decisions is predicated on the assumptions that different evaluators will get similar results and these results are related to true grazing intensity. If monitoring results differ substantially among observers, decisions based on these assessments could be dependent upon who completed the assessment rather than on environmental impacts (MacDonald et al. 1991; Elzinga et al. 1998; Bauer and Ralph 2001). Therefore, poorly designed and implemented protocols threaten the credibility and trust of federal management agencies while increasing economic risk to permittees (Conley et al. 2007).

Previous studies evaluating repeatability among observers in evaluating annual indicators have had mixed results, with some protocols deemed repeatable (Turner and Clary 2001) and others unrepeatably (Platts et al. 1983; Hall and Max 1999). Factors shown to affect repeatability of a protocol among observers include methodology and operational clarity of the protocol, the length and complexity of the evaluated stream reach, and the amount of training evaluators received (Whitacre et al. 2007).

Although replication of results among observers is important, it is also important for results to be related to the disturbance of interest. For example, results of a protocol that evaluates ground disturbance by cattle should be related to the true amount of ground disturbance caused by cattle. One concern is that in order to develop a cost-effective and easy-to-use protocol, methodologies could be simplified to a point where they may no longer be related to the disturbance they seek to evaluate (Anderson 2003). So, although environmental monitoring programs often evaluate indices and surrogates of an attribute of interest, it is important that we not blindly

assume values produced by these protocols are directly related to that attribute (Thompson et al. 1998; Anderson 2001).

In this article we evaluate the repeatability and accuracy of four protocols used to measure one annual indicator of grazing intensity in riparian areas: streambank alteration. Streambank alteration is defined by the presence of current-year shearing, trampling, and trailing by livestock in the evaluated area (Table 1). We chose this indicator because limiting streambank alteration has been shown to maintain or improve riparian conditions (Bengeyfield 2006), is widely used (Burton et al. 2007), simple to learn, and can be easily related to a more precise measurement of livestock disturbance. In assessing streambank alteration, we evaluated 1) differences in mean values for different streambank alteration protocols, 2) variation in protocol results among and within observers, 3) the effect of training on results, 4) the effect of experience/background on the results, and 5) how rapid protocols were related to more intensive measurements of streambank alteration.

## METHODS

We evaluated streambank alteration within 10 fall-grazing allotments located on lands managed by the US Forest Service in western Montana in 2003 and 2004 (Table 2). Forest Service personnel selected these sites to assure a broad representation of grazing intensities, stream sizes, channel types, and vegetation communities. All four protocols used paces to determine spacing of the samples, and then used the tip of the

**Table 2.** Stream width and vegetation community of 10 sampling sites.

Stream name	Bankfull width (m)	Riparian vegetation
Beaverhead–Deerlodge National Forests		
Bowles Creek	6.0	Willow/sedge/grass
Meadow, east branch	2.5	Willow/sedge
Meadow, west branch	3.8	Forb
Middle Fork Rock Creek	18.0	Conifer/sedge/grass
Sand Basin Creek	3.0	Conifer/sedge/grass
Lewis and Clark National Forest		
Allen Gulch	0.8	Wet sedge meadow
Cabin Creek	1.3	Alder/mixed shrub
Calf Creek	4.5	Conifer/mixed shrub
Daniels Creek	1.0	Douglas fir/grass
Newlan Creek	3.5	Willow dominated

boot as the reference point for the sample (see Wolman 1954 for genesis of this sampling approach). Data were collected 23–29 September in 2003 and 21–26 September in 2004.

The 2003 tests focused on how training affected means and variability associated with two streambank alteration protocols. The 2004 tests assessed whether the same observer could replicate his or her results, the effect of professional experience, and whether intensive and rapid streambank alteration protocols produced similar results. Evaluated protocols differed between years because the primary goal of this study was to examine methods that would improve streambank alteration assessment techniques.

Because of the number of people evaluating each site each year, the observers were reminded to minimize their impacts to the site. Throughout both years of the study observers worked independently, were asked not to discuss interpretation or application of the sampling method with other observers, and were not given individual feedback by investigators.

### 2003 Evaluation

To evaluate the role training played in differences among observers, we asked eight summer seasonal employees (hereafter called technicians) with little experience evaluating grazing to estimate streambank alteration at the 10 sites twice, once before training and again following training. We evaluated two streambank alteration protocols and randomly assigned four technicians to each protocol. Prior to the first visit (untrained), each of the technicians were given a 30-min introduction to livestock alteration approaches, a written copy of the assigned protocol, and time to familiarize him- or herself with the method at a site not used for the study.

Following the first visit to sites, technicians received a half-day training session (4 h) for the protocol they were applying. Training involved detailed operational instructions on where to take measurements and how to identify and distinguish different forms of current-year streambank alteration. Following training, observers resampled the same 10 sites. To minimize the likelihood of crews memorizing results from their first visit (untrained), observers were not told of the second part of the study until they had completed their first visit. All data were entered on data recorders so results of the first visits were not directly available to the technicians.

One of the evaluated protocols, the top of streambank (TS), measured disturbance along a visualized line that was perpendicular to the stream and extended from the water's edge through the tip of the observers shoe to a point 46 cm beyond the TS (Fig. 1A; TS). This protocol evaluated alteration along a line that did not have a fixed length but varied with the size and shape of the streambank. With the use of the TS protocol, the visualized line was considered altered if any part of the line was altered by livestock (each step was either a "0" for no disturbance or "1" for disturbed). Total streambank alteration for the monitoring site with this protocol was the percent of the surveyed lines that had disturbance.

The second protocol, the bankfull (BF), measured disturbance with the use of a 61 × 30.5 cm sampling frame, the length of which was centered at BF elevation and parallel to the stream (Fig. 1B; BF). The BF protocol divided the sampling frame into 10 equally spaced lines. At each placement of the frame, the number of the lines (0–10, with intervals of 1) that

had been disturbed by cattle was counted. A "0" meant none of the 10 lines were altered by cattle and a "10" meant all the lines were altered by cattle. Total streambank alteration for the monitoring site measured with the use of this approach was the average of all the frames times 10, resulting in a disturbance rating between 0 and 100 (a percentage scale).

Both protocols evaluated disturbance on both streambanks, used paces as the interval to space measurements (every step for TS and every second step for BF), had the same definition of alteration categories (Table 1), and started their surveys at the same locations. The protocols did differ in the length of streambank evaluated, 30.5 m for TS vs. 110 m for BF. Both protocols were considered rapid techniques because they used visual techniques to assess alteration.

### 2004 Evaluation

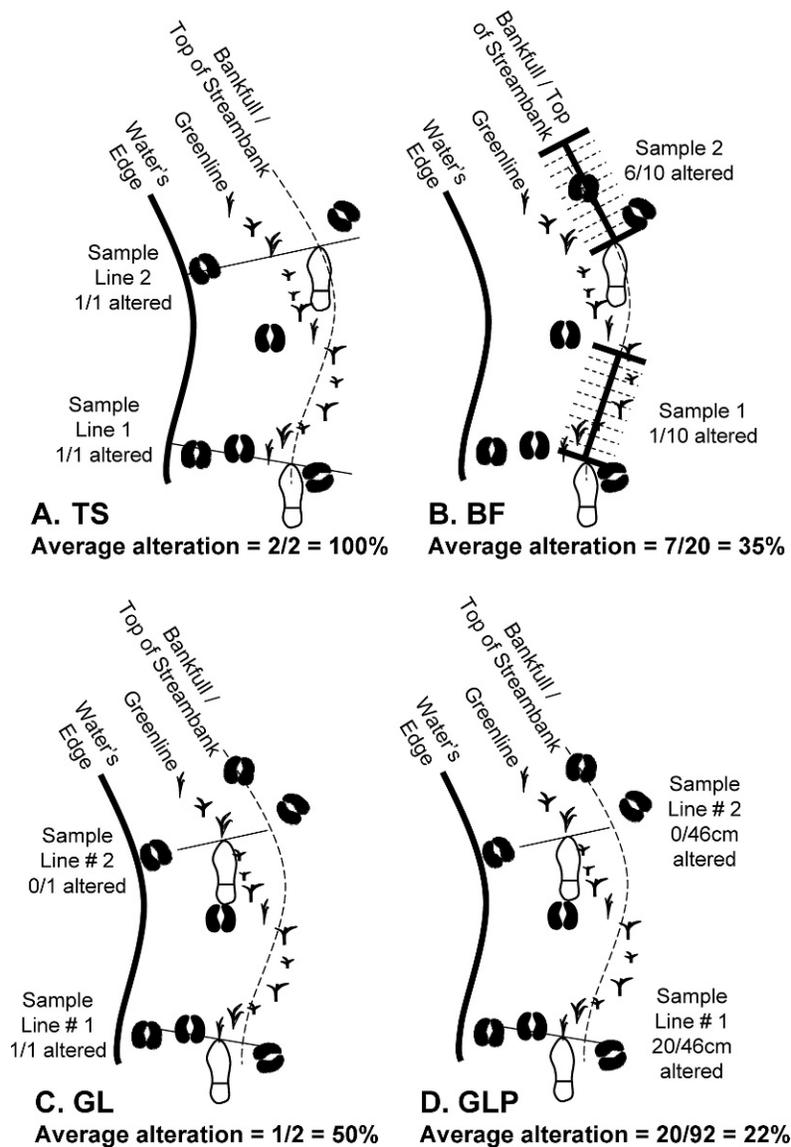
Rather than using BF elevation or TS, the methods evaluated in 2004 followed the first line of perennial vegetation, commonly referred to as the greenline (Winward 2000). This change in protocol was made because unpublished data suggested that greenline could be identified by observers more consistently than BF elevation or TS. The two evaluated protocols, greenline (GL; Fig. 1C) and greenline precise (GLP; Fig. 1D), measured disturbance along a line centered on the GL and perpendicular to the stream channel. The line extended through the toe of the observer 46 cm toward the stream and 46 cm up the streambank (92 cm in total length).

When GL was used, the sample line was considered altered if current-year disturbance was observed at any point along the line and unaltered if alteration was from the prior year or no alteration was observed (Fig. 1C; Table 1). The percent streambank alteration with the use of GL was the number of sample lines altered divided by the total number of sample lines. This protocol was considered a rapid technique (like those evaluated in 2003) because it used visual techniques to assess alteration at a site.

The GLP evaluated streambank alteration with the exact same approach as the GL protocol, except the observers measured the length of the line (in centimeters) that had been altered (Fig. 1D). Because GLP was a measurement of the total length of a line affected by disturbance, this protocol was a good estimate of the true amount of alteration at the monitoring site (line-intercept approach; Elzinga et al. 1998). An estimate of the true amount of streambank alteration with the use of GLP was the percent of the cumulative length of the lines evaluated that had been altered. In contrast to the other protocols, this technique was considered intensive because it required the physical measurement of alteration at each site.

Four technicians assessed disturbance with the GL and three technicians used GLP. Because of a concern that some of the difference between protocols in 2003 could have been due to different reach lengths (Whitacre et al. 2007), all 2004 evaluations were performed on the same 75-m stream segment within each of the 10 allotments. Measurements were taken on both stream banks. At each site the presence or measurement of alteration on the line was taken every second step (87–102 samples per site, depending upon the individual's pace length).

We asked three of the four technicians that applied the GL to reevaluate the 10 sites in order to assess the ability of observers



**Figure 1.** An illustration of the four streambank alteration protocols we evaluated. Footprints illustrate how sample lines were oriented relative to the observer's foot while the observer was walking along the top of streambank (TS, **A**), bankfull (BF, **B**), or greenline (GL, **C**; greenline precise [GLP], **D**). All evaluated lines were perpendicular to stream flow. Lengths of the lines evaluated for alteration were from the stream to 46 cm beyond the top of streambank for TS, at 10 equally spaced lines, with a length of 30.5 cm within a 61-cm quadrat for BF, and 46 cm both toward and away (total assessed length 92 cm) from the stream for GL and GLP. The hoofprints represent current-year livestock alterations. The average alterations calculated for two paces are meant to display how different protocols might result in different average alteration percentages even when the same locations are being measured.

to repeat their results. The technicians were not told they would resurvey sites until after the first visit, to reduce the likelihood the observers would remember results from their first visit. Data were entered in data loggers, so results were not directly available to the technicians.

We evaluated the effect of professional background by using a second group of observers consisting of seven Forest Service employees with between 5 yr and 30 yr of experience in rangeland management (hereafter called professionals). These professionals were asked to evaluate the same 10 sites evaluated by the technicians using the GL protocol. Although each technician sampled all 10 sites, two of the professionals evaluated 5 sites and the five other professionals sampled the remaining sites.

All technicians and professionals received 4 h of training prior to completing the site evaluations. The trainer focused on how to identify the GL (illustrations were provided to clarify the GL in different situations; Cowley and Burton 2005), how to identify streambank alteration, and how to distinguish current-year from previous-year alteration. Training was conducted at locations not used for the study.

## DATA ANALYSIS

### Protocol Differences

Protocol differences were determined by comparing data collected for the two protocols assessed in 2004 (GL and

GLP). We used these protocols because they differed in their general measurement approach (rapid versus intensive) but evaluated the exact same stream reaches. We used the data from technicians in analysis of variance (ANOVA) techniques, with streams as a block effect, to determine if the protocols produced significantly different ( $P \leq 0.10$ ) estimates of streambank alteration.

### Among-Observer Variability

We used the posttraining results of the technicians from both 2003 and 2004 to describe the observer variability associated with the four different protocols (TS, BF, GL, GLP). Variability was partitioned between that associated with differences among streams and residual error (Kaufmann et al. 1999; Larsen et al. 2001; Roper et al. 2002). With the use of this approach, if all observers arrived at the same estimate of alteration for each of the streams, then all the variation would be due to stream differences and there would be no residual error. Observer variability within a method was derived from the residual error estimate and described with the use of standard deviation (SD) and coefficient of variation (CV; Kaufmann et al. 1999). We used a Hartley's F-Max test to determine whether observer variance estimates among the four methods were different (Milliken and Johnson 1997).

### Within-Observer Variability

We used the three observers from the 2004 tests who sampled each site twice by the same method (GL) to determine variability within an individual. We took the results from the first visit and subtracted the second-visit results (paired  $t$  tests) to determine if the observers differed in their assessments (significantly different from 0). Within-observer variability was determined by calculating the SD of the difference between the two visits; if an observer perfectly repeated the evaluation there would be no within-observer variation.

### Training Effect

We used the 2003 pre- and posttraining data for both protocols (TS and BF) to evaluate the effect of training. The analysis was conducted separately for each protocol. To evaluate whether training affected mean values or observer variability we estimated the means and SD independently for each visit (untrained/trained). We used a paired  $t$  test and evaluated whether the difference between the two visits was significantly different from 0. We used a Hartley's F-Max test to determine whether observer variance differed between the untrained and trained observations (Milliken and Johnson 1997).

### Professional Background

We used the 2004 GL data and an ANOVA to compare differences in means and variances between technicians and professionals. We used professional background as a main effect and blocked for differences among streams. We estimated unique variances for each group (Milliken and Johnson 1997). Observer variability was described with the use of SD and CV (Kaufmann et al. 1999). We used a Hartley's F-Max test to determine whether observer variance estimates between the two groups were equal (Milliken and Johnson 1997).

**Table 3.** Overall mean percent of streambank altered, standard deviation (SD), and coefficient of variation (CV) of the four streambank alteration protocols for the 10 allotments evaluated. Protocols: GL = greenline, GLP = greenline precise, BF = bankfull, TS = top of streambank. Data were summarized in a manner that represented the protocol's measurement approach (see Fig. 1 and text for specifics). Abbreviations in parentheses represent personnel used in the comparison; Tech = technician, Pro = experienced range conservationist, Tr = trained technician, Unt = untrained technicians. Because not all protocols were used to evaluate all objectives, discussion of statistical significance is found in the text.

	2004			2003			
	GL (Tech) <i>n</i> = 4	GL (Pro) <i>n</i> = 7	GLP <i>n</i> = 3	BF (Tr) <i>n</i> = 4	BF (Unt) <i>n</i> = 4	TS (Tr) <i>n</i> = 4	TS (Unt) <i>n</i> = 4
Mean	31.4	41.7	8.3	22.9	32.6	38.6	45.7
SD	6.3	8.8	4.7	8.1	13.7	9.7	15.6
CV	20	21	56	35	42	25	34

### Relationship Between Rapid and Intensive Protocols

We tested whether results generated with the use of a rapid protocol was related to a protocol that more intensively and accurately measured streambank alteration. We used linear regression to relate the mean site disturbance of three technicians applying GLP ("truth," dependent variable) to the mean site disturbance of four technicians applying GL (independent variable). We used GL as the independent variable because it was the most rapid and therefore the cheaper of the evaluated protocols.

## RESULTS

### Protocol Difference

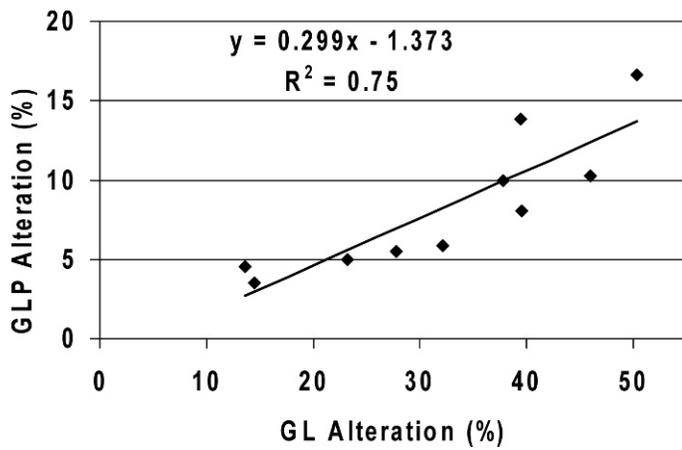
Streambank alteration as determined by the two protocols evaluated in 2004 (GL and GLP) differed significantly (Table 3;  $P < 0.1$ ). The overall average alteration for all 10 streams was 31.4% for GL (the rapid protocol) and 8.3% for GLP (the intensive protocol).

### Among-Observer Variability

Variability in trained observers differed among the four protocols (Table 3). Only one pairwise comparison of the three rapid protocols differed significantly (GL vs. TS), whereas all the rapid protocols had significantly higher variability than the intensive protocol (GLP). The GLP had the lowest overall variability as measured by the SD (4.7), followed by GL (SD = 6.3), BF (SD = 8.1), and TS (SD = 9.7). The CVs were similar among the three rapid protocols (20 for GL, 35 for BF, and 25 for TS). Conversely, the GLP protocol had the largest CV (56).

### Within-Observer Variability

Paired  $t$  tests suggest no significant differences in the mean estimates between visits by the same technician (confidence intervals of the difference included 0). GL variance estimates for multiple visits by the same observer (within observer; SD = 5.2) were not significantly different than the variability associated with visits among different observers (SD = 6.3).



**Figure 2.** The linear relationship between the mean values of two protocols, greenline (GL; independent variable) and greenline precise (GLP; dependent variable), for the 10 assessed allotments. Estimates for each of the 10 allotments for GL are the means of the four technicians and GLP are the means of the three technicians.

### Training Effect

Training significantly reduced estimates of mean streambank alteration for both TS and BF (Table 3). The overall average estimate of streambank alteration for untrained technicians was 32.6% for BF and 45.7% for TS, whereas posttraining results were 22.9% and 38.6%. Training also lowered observer variability by at least a third (Table 3).

### Professional Background

Technicians and professionals had significantly different mean estimates of streambank alteration when using GL, but had similar among-observer variability (Table 3). The mean estimate of overall alteration for the technicians was 31.4%, whereas professionals averaged 41.7% ( $P < 0.1$ ). The technicians had a SD of 6.3 and a CV of 20, whereas the professionals had a SD of 8.8 and a CV of 21 (these were not significantly different).

### Relationship Between Rapid and Intensive Protocols

Average GL values, as measured by the technicians, were linearly ( $r^2 = 0.75$ ,  $P < 0.1$ ) related to the more intensively measured GLP values (Fig. 2). The slope of the regression equation was 0.299 (90% confidence interval; 0.184–0.413), suggesting a 1-unit increase under the GLP protocol was approximately equivalent to a 3.3-unit increase in streambank alterations as determined by using GL. These results indicate that more rapid protocols are related to the true amount of disturbance, but that this relationship was not 1:1.

## DISCUSSION

In this article we evaluated many aspects thought to influence the assessment of streambank alteration by livestock grazing. We found that although measurements of streambank alteration are relatively repeatable by different observers, results were affected by several factors not directly related to grazing intensity. These factors include 1) training, 2) professional

background, 3) location and intensity of measurements, and 4) the protocol used.

Observer consistency in site evaluations was significantly improved by training. Variability among observers applying both the TS and BF protocols was cut by more than a third following a 4-h training session. Training has consistently been shown to reduce variability among observers when measuring stream habitat (Roper and Scarnecchia 1995; Hannaford et al. 1997) and vegetative attributes (Smith 1944; Thorne et al. 2002). However, until recently there has been little effort expended to develop standardized training for assessing livestock disturbance in riparian areas (but see Cowley and Burton 2005 for an exception). Given the higher means and variability associated with untrained observers, assessing streambank alteration with untrained observers could seriously undermine the credibility of land management decisions based on those assessments. Further studies should be conducted to determine if additional training (beyond 4 h) will reduce uncertainty associated with individual assessments of streambank alteration.

A question related to training that is often discussed (Allen and Hoekstra 1992; Hummel 1994) but infrequently evaluated is this: does professional background influence monitoring results? We found that trained Forest Service technicians and rangeland specialists differed significantly in their mean estimates, but had similar variability when using a rapid approach (GL) to assess streambank alteration. Even though both groups applied the same protocol, the professionals' estimates of streambank alteration were 30% higher than the technicians'. Because both groups received the same training and surveyed the same reaches, differences between technicians and professionals had to be the result of differences associated with the placement of sampled lines and/or how alteration was assessed along the sample line (see Galloway et al. 2006 for similar findings in a different assessment setting). Regardless of the specific reasons for this difference, our data suggest that professional background can influence survey results. A simple (but not the only) approach to address this concern is the use of independent and consistently trained technicians to evaluate grazing allotments. The use of independent observers would not only standardize the application of protocols, but could reduce criticisms that imply local politics influence monitoring outcomes.

One potential reason why protocols yield different estimates of streambank alteration could be that measurements are taken in slightly different locations such as BF, TS, and GL. The inability of different individuals to identify these stream channel features consistently could affect variability associated with the protocol. For example, BF elevation has been shown to be notoriously difficult for different observers to locate consistently (Williams 1978; Harrelson et al. 1994; Whitacre et al. 2007). Although unpublished data we collected in association with this study did find observers more consistently identified GL than either BF or the TS (median disagreement among observers; GL = 0.13 m, BF = 0.43 m, and TS = 0.35 m), this increased consistency led to GL having only slightly (but significantly) less variability than the other presence/absence protocol, TS. The small difference between these two protocols likely reflects the proximity of these stream channel indicators, the considerable overlap in the area evaluated (Figs. 1A and 1C) and the coarseness (presence/

absence) of the protocol. We suggest that if enough measurements are taken to capture the spatial variation in alteration (Dowdy and Wearden 1983; Elzinga et al. 1998) within the near-stream area, the choice of sample location will have only a minor effect on the overall variation. Based on the experience gained in this test, the primary advantage of GL is not that it reduces variability, but that it is the simplest to teach, because it does not rely on an understanding of hydrology (as does measurement at BF).

Differences in protocol approaches (rapid versus intensive) resulted in substantially different mean estimates of streambank alteration. Because GL and GLP had sample lines with the same length, placed in the same way, in the same stream reaches by similarly trained observers (technicians), differences reflect how alterations were measured and summarized rather than differences in actual alteration. The GLP is a good estimate of the true amount of streambank alteration (similar to Heady et al. 1959), whereas GL is an index of stream alteration. Because GL treats a line that barely intercepted a hoofprint as if it were 100% altered (presence/absence), whereas GLP might indicate only 1% of the line had actually been altered, index values derived with GL will always equal or exceed GLP. In this study the index method (GL) was on average 3.3 times larger than the true amount of alteration (GLP). Our comparisons suggest that as measurement precision increases, all other things being equal, estimates of streambank alteration generated by a protocol will be lowered. Even in cases where all things were not exactly equal (different reach lengths and measurement configurations), in 2003 the more precise protocol (11 possible levels of disturbance; BF) had lower estimates of alteration than the protocol that used presence/absence (TS). The lower the measurement precision of a protocol (e.g., altered/not altered), the greater the divergence results will be from protocols more precisely measuring streambank alteration (e.g., measured to the nearest centimeter).

Regardless of the protocol being used, it is imperative that the outcome of a protocol be informative relative to the resource condition/attribute the protocol is being used to estimate (Anderson 2001; Whitacre et al. 2007). For example, consider an allotment where it had been determined that when the near-stream area altered by hoofprints exceeds 10%, then the long-term outcome is a stream with increasing streambank angles and decreasing biotic production. In this situation, a protocol must be able to identify when this 10% threshold is crossed (Anderson 2003). Our evaluations of these four protocols suggest that if the observer receives training, the three fixed-length approaches (BF, GL, GLP) could consistently indicate whether this threshold had been crossed. However, it is important to note that the exact value of the threshold would be dependent on the protocol; 10% for GLP, approximately 35% for GL, and a value somewhere between these two for BF. Although not directly assessed, deduction suggests it is unlikely that TS could be used to identify a single threshold criterion consistently. Although this protocol was repeatable among observers at a site, the variable length of the sampled line would result in estimates of streambank alteration not only being a function of livestock alteration but also being affected by stream size (the bigger the stream or longer the line, the more likely the evaluated line will pass through at least one

hoofprint). This suggests that if TS was used, there would be a need to have different standards for different-size streams.

The rationale given to support the use of rapid protocols, which result in indices of alteration rather than direct measures of the attribute of interest, is that doing so lowers the cost or ease of monitoring. But as we have shown, how things are measured affects results and how those results can be interpreted (also see Anderson 2003; Wolman 2006). So, in deciding which protocol to use it is important to balance the time it takes to complete a survey against the precision and accuracy of the estimate (Krebs 1989). Technicians were able to assess a site with GL in approximately 30 min, whereas the GLP generally took about an hour. When the limited additional time it takes to complete the more precise protocol is weighed against the possible concerns associated with using indices of alterations, there seems little rationale not to use the more precise and accurate GLP assessment techniques.

We found that streambank alteration protocols were as repeatable as or better than many protocols used to assess physical stream habitat and riparian attributes (see Kaufmann et al. 1999; Roper et al. 2002). But even if protocols are repeatable, managers must use caution when basing a decision on a single observer's assessment of a site (Olsen et al. 2005). This is because even well-trained observers will not get the exact same estimate at a site. We found that individuals repeating a survey had only slightly (and not significantly) lower variability than different observers. This suggests that even if the same observer samples through time there is going to be error associated with estimates of streambank alteration. As a result, any single estimate is simply the best available indication of the amount of streambank alteration at that site—not a perfect estimate.

Because of the uncertainty regarding single estimates of streambank alteration, we suggest observer variability associated with a protocol be incorporated into management decisions. For example, any site sampled with the use of the GL method ( $SD = 6.3$ ) should acknowledge uncertainty around the estimate of  $\pm 10\%$  (approximately 90% confidence interval of individual estimates). Managers can feel relatively confident when making allotment decisions if assessment values plus or minus this uncertainty do not include the threshold value. In contrast, when the estimate plus or minus this uncertainty includes the threshold value, then we suggest repeating the assessment (perhaps with multiple observers) or completing complementary assessments (Cowley and Burton 2005; Burton et al. 2007). This is especially important when the decision has the potential to have high economic cost to a permittee, high ecological cost to the riparian area, greatly affects the presence of rare species, or fails to meet legal obligations.

## MANAGEMENT IMPLICATIONS

Both the public and permittees expect federal land management and regulatory agencies to use the best available science to inform decisions relative to grazing (for a general discussion of science and the law, see Jasanoff 1995). The use of standards in combination with monitoring is, and should remain, an important management tool to minimize the impacts of grazing in riparian areas (Bengetyfield and Svoboda 1998; Clary and

Leininger 2000). But it is also important that bank-alteration standards not only indicate the desired environmental conditions at a site, but also identify the protocol to be used to assess streambank alteration. Given that there are at least nine different streambank alteration methods currently used by land-management agencies in the western United States (Bengeyfield and Svoboda 1998; Cowley and Burton 2005), specialists need to understand how the choice of protocol, level of training, and professional background affect these estimates. A clear understanding of these issues is necessary to interpret results accurately and to articulate clearly the implications of streambank alteration assessments to land management agencies, regulators, permittees, and the public.

Finally, it is important to remember that no protocol can be implemented without measurement error (Krebs 1989; Ramsey et al. 1992; Roper et al. 2002). Managers should therefore be careful when taking action based on a single evaluation—especially when the result is near a management standard or threshold. When these concerns are addressed, indicators such as streambank alteration can help ensure that management decisions maintain sustainable allotments and landscapes.

## ACKNOWLEDGMENTS

The authors acknowledge the field technicians who collected the data for this study. We extend special thanks to Mike Enk, Bo Stuart, and Dan Svoboda for providing input on the study design and for training the participants. Finally, this article was greatly improved by reviews from Pete Bengeyfield, Marc Coles-Ritchie, Ervin Cowley, Mike Enk, Allison Kelly, Jeffrey Kershner, Mark Nienow, Janice Statz, Deanna J. Stouder, Dan Svoboda, and Marina Whitacre. We are grateful for their comments.

## LITERATURE CITED

- ALLEN, T. F. H., AND T. HOEKSTRA. 1992. *Toward a unified ecology*. New York, NY, USA: Columbia University Press. 384 p.
- ANDERSON, D. R. 2001. The need to get things right in wildlife field studies. *Wildlife Society Bulletin* 29:1294–1297.
- ANDERSON, D. R. 2003. Response to Engeman: index values rarely constitute reliable information. *Wildlife Society Bulletin* 31:288–291.
- BAUER, S. B., AND T. A. BURTON. 1993. Monitoring protocols to evaluate water quality effects of grazing management of western rangeland streams. Seattle, WA, USA: US Environmental Protection Agency (EPA), Region 10, Water Division. Report EPA 910/R-93-017. 166 p.
- BAUER, S. B., AND S. C. RALPH. 2001. Strengthening the use of aquatic habitat indicators in the Clean Water Act programs. *Fisheries* 26(6):14–25.
- BENGEYFIELD, P. 2006. Managing streams with cows in mind. *Rangelands* 28:3–6.
- BENGEYFIELD, P., AND D. SVOBODA. 1998. Determining allowable use levels for livestock movement in riparian areas. Proceedings of the Specialty Conference on Rangeland Management and Water Resources; 27–29 May 1998. Middleburg, VA, USA: American Water Resources Association.
- BLACKMORE, M., AND P. M. VITOUSEK. 2000. Cattle grazing, forest loss and fuel loading in dry forest ecosystem at Pu'u Wa'aWa'a ranch Hawai'i. *Biotropica* 32:625–632.
- BURTON, T. A., E. R. COWLEY, AND S. J. SMITH. 2007. Monitoring stream channels and riparian vegetation – multiple indicators. Boise, ID, USA: Idaho State office of Bureau of Land Management and Rocky Mountain Research Station, Forest Service BLM/ID/GI-07. 41 p.
- CLARY, W. P. 1999. Stream grazing and vegetation responses to late spring cattle grazing. *Journal of Range Management* 52:218–227.
- CLARY, W. P., AND W. C. LEININGER. 2000. Stubble height as a tool for management of riparian areas. *Journal of Range Management* 53:562–573.
- CONLEY, J. L., M. E. FERNANDEZ-GIMENEZ, G. B. RUYLE, AND M. BRUNSON. 2007. Forest service grazing permittee perception of the Endangered Species Act in Southeastern Arizona. *Rangeland Ecology and Management* 60:136–145.
- COWLEY, E. R., AND T. A. BURTON. 2005. Monitoring streambanks and riparian vegetation—multiple indicators. Boise, ID, USA: US Department of Interior, Bureau of Land Management. Technical Bulletin No. 2005-2. 29 p.
- DOWDY, S., AND S. WEARDEN. 1983. *Statistics for research*. New York, NY, USA: John Wiley and Sons. 537 p.
- ELZINGA, C. L., D. W. SALZER, AND J. W. WILLOUGHBY. 1998. Measuring and monitoring plant populations. Denver, CO, USA: US Department of Interior, Bureau of Land Management. Technical Reference 1730-1. 477 p.
- GALLOWAY, A. W. E., M. T. TUDOR, AND W. M. VANDER HAEGEN. 2006. The reliability of citizen science: a case study of Oregon white oak stand surveys. *Wildlife Society Bulletin* 34:1425–1429.
- GREEN, D. G., AND J. B. KAUFFMAN. 1995. Succession and livestock grazing in a northeastern Oregon riparian ecosystem. *Journal of Rangeland Management* 48:307–313.
- HALL, F. C., AND T. MAX. 1999. Test of observer variability in measuring riparian shrub twig length. *Journal of Range Management* 52:633–636.
- HANNAFORD, M. J., M. T. BARBOUR, AND V. H. RESH. 1997. Training reduces observer variability in visual-based assessments of stream habitat. *Journal of the North American Benthological Society* 16:853–860.
- HARRELSON, C. C., C. L. RAWLINS, AND J. P. POTYONDY. 1994. Stream channel reference sites: an illustrated guide to field techniques. Fort Collins, CO, USA: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. General Technical Report RMRS-245. 61 p.
- HEADY, H. F., R. P. GIBBENS, AND R. W. POWELL. 1959. A comparison of charting, line intercept, and line point methods of sampling shrub types of vegetation. *Journal of Range Management* 12:180–188.
- HUMMEL, R. P. 1994. *The bureaucratic experience: a critique of life in the modern organization*. New York, NY, USA: St. Martin's Press. 286 p.
- JASANOFF, S. 1995. *Science at the bar: law, science, and technology in America*. Cambridge, MA, USA: Harvard University Press. 285 p.
- KAUFMANN, P. R., P. LEVINE, E. G. ROBISON, C. SEELIGER, AND D. V. PECK. 1999. Quantifying physical habitat in Wadeable streams. Washington, DC, USA: US Environmental Protection Agency. Report EPA/620/R-99/003. 99 p.
- KREBS, C. J. 1989. *Ecological methodology*. New York, NY, USA: HarperCollins. 654 p.
- LARSEN, D. P., T. M. KINCAID, S. E. JACOBS, AND N. S. URQUHART. 2001. Designs for evaluating local and regional scale trends. *Bioscience* 51:1069–1078.
- MACDONALD, L. H., A. SMART, AND R. C. WISSMAR. 1991. Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. Seattle, WA, USA: US Environmental Protection Agency, Region 10, Water Division. Report EPA/910/9-91-001. 166 p.
- MARTY, J. 2005. Effects of cattle grazing on diversity in ephemeral wetlands. *Conservation Biology* 19:1626–1632.
- MILLIKEN, G. A., AND D. E. JOHNSON. 1997. *Analysis of messy data. Volume 1. Designed experiments*. London, United Kingdom: Chapman and Hall. 473 p.
- OLSEN, D. S., B. B. ROPER, J. L. KERSHNER, R. HENDERSON, AND E. ARCHER. 2005. Sources of variability in conducting pebble counts: their potential influence on the results of stream monitoring programs. *Journal of the American Water Resources Association* 41:1225–1236.
- PLATTS, W. S. 1991. Livestock grazing. *American Fisheries Society Special Publication* 19:389–424.
- PLATTS, W. S., C. ARMOUR, G. D. BOOTH, M. BRYANT, J. L. BUFFORD, P. CUPLIN, S. JENSEN, G. W. LIENKAEMPER, G. W. MINSHALL, S. P. MONSEN, R. L. NELSON, J. R. SEDELL, AND J. S. TUHY. 1987. Methods for evaluating riparian habitats with applications to management. Ogden, UT, USA: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. General Technical Report INT-221. 177 p.
- PLATTS, W. S., W. F. MEGAHAN, AND G. W. MINSHALL. 1983. Methods for evaluating stream, riparian, and biotic conditions. Ogden, UT, USA: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. General Technical Report INT-138. 70 p.
- RAMSEY, M. H., M. THOMPSON, AND M. HALE. 1992. Objective evaluation of precision requirements for geochemistry analysis of variance. *Journal of Geochemical Exploration* 44:23–36.

- ROPER, B. B., J. L. KERSHNER, E. K. ARCHER, R. C. HENDERSON, AND N. BOUWES. 2002. An evaluation of physical stream habitat attributes used to monitor streams. *Journal of the American Water Resources Association* 38:1637–1646.
- ROPER, B. B., AND D. L. SCARNECCHIA. 1995. Observer variability in classifying habitat types in stream surveys. *North American Journal Fisheries Management* 15:49–53.
- SHEPARD, H. S. 2007. The future of livestock grazing and the Endangered Species Act. *Journal of Environmental Law and Litigation* 21:383–444.
- SMITH, A. D. 1944. A study of the reliability of range vegetation estimates. *Ecology* 25:441–448.
- THOMPSON, W. L., G. C. WHITE, AND C. GOWAN. 1998. Monitoring vertebrate populations. San Diego, CA, USA: Academic Press. 365 p.
- THORNE, M. S., Q. D. SKINNER, M. A. SMITH, J. D. RODGERS, W. A. LAYCOCK, AND S. A. CEREKCI. 2002. Evaluation of a technique for measuring canopy volume of shrubs. *Journal of Range Management* 55:235–241.
- TURNER, D. L., AND W. P. CLARY. 2001. Sequential sampling protocol for monitoring pasture utilization using stubble height criteria. *Journal of Range Management* 54:132–137.
- WHITACRE, H. W., B. B. ROPER, AND J. L. KERSHNER. 2007. A comparison of protocols and observer precision for measuring stream attributes. *Journal of the American Water Resources Association* 43:923–937.
- WILLIAMS, G. P. 1978. Bank-full discharge of rivers. *Water Resources Research* 14:1141–1154.
- WINWARD, A. H. 2000. Monitoring the vegetation resources in riparian areas. Ogden, UT, USA: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. General Technical Report RMRS-47. 49 p.
- WOLMAN, A. G. 2006. Measurement and meaningfulness in conservation science. *Conservation Biology* 20:1626–1634.
- WOLMAN, M. G. 1954. A method of sampling coarse river-bed material. *Transactions of the American Geophysical Union* 35:951–956.