



United States
Department of
Agriculture

Forest Service

Rocky Mountain
Forest and Range
Experiment Station

Fort Collins
Colorado 80526

Research Paper
RM-RP-319



A Pebble Count Procedure for Assessing Watershed Cumulative Effects

Gregory S. Bevenger
Rudy M. King



Bevenger, Gregory S.; King, Rudy M. A pebble count procedure for assessing watershed cumulative effects. Res. Pap. RM-RP-319. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 17 p.

Abstract

Land management activities can result in the delivery of fine sediment to streams. Over time, such delivery can lead to cumulative impacts to the aquatic ecosystem. Because numerous laws require Federal land managers to analyze watershed cumulative effects, field personnel need simple monitoring procedures that can be used directly and consistently. One approach to such monitoring is described. The approach involves sampling a longitudinal reach of stream channel several hundred feet long using a zig-zag pebble count procedure that crosses all habitat features within a stream channel. The approach accommodates reference (unimpacted) and study (impacted) reaches so that impact comparisons can be made. Case studies show how the procedure is applied.

Keywords: Watershed cumulative effects, monitoring, watershed field techniques, pebble counts, sample sizes

The Authors

Gregory S. Bevenger is a hydrologist with the Shoshone National Forest, Cody, Wyoming. He graduated from Utah State University in 1978 with a B.S. in Forestry. Greg provides technical consultation on land management projects that affect soil and water resources.

Rudy M. King is Station Biometrician, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado. He has been a member of the Station's Biometrics Unit since 1971 and holds a M.S. in Mathematical Statistics from Colorado State University. Rudy provides statistical consultation for Station scientists and cooperators.

A Pebble Count Procedure for Assessing Watershed Cumulative Effects

**Gregory S. Bevenger, Hydrologist
Shoshone National Forest¹**

**Rudy M. King, Station Biometrician
Rocky Mountain Forest and Range Experiment Station²**

¹*Forest Supervisor's Office is in Cody, Wyoming.*

²*Headquarters is in Fort Collins, in cooperation with Colorado State University.*

Contents

	Page
Introduction	1
Preliminary Sampling and Analysis	1
Validation Efforts	3
How to Use the Zig-Zag Pebble Count Procedure	6
Using the Procedure in a Watershed Assessment	10
Conclusions	13
Literature Cited	13
Appendix 1. Statistical Framework	15
Appendix 2. Sample Size Estimation	16

A Pebble Count Procedure for Assessing Watershed Cumulative Effects

Gregory S. Bevenger and Rudy M. King

Introduction

Of concern in many National Forest watersheds is the cumulative effect of land management. One aspect of this concern is the delivery of fine sediment to streams and its effect on aquatic habitat, particularly within pool and riffle complexes. To this end, the Clean Water Act requires the control of non-point sources of pollution through application of Best Management Practices. To evaluate the effectiveness of such controls, stream monitoring needs to occur. The most desirable form of monitoring is one that is relatively simple to implement, can be used directly and consistently by field personnel, and is sensitive enough to provide a measure of impact. Ideally, indices of sediment in streams can be obtained by measuring channel substrate material in comparison to reference streams, where little or no management has occurred.

We have developed a field procedure for characterizing particle size distributions of reference and study streams, where we define reference streams as "natural" or "unimpacted" and study streams as "disturbed" or "impacted." These distributions can be used for comparative purposes to determine, with statistical reliability, if there has been a shift toward finer size material (fine gravels to sands) in the study stream. Knowledge of such a shift is important because increases in this size material can negatively impact the aquatic ecosystem (Meehan 1991; MacDonald et al. 1991). If monitoring demonstrates that a shift toward finer size material has or is taking place, corrective actions can be taken within either (1) the watershed to reduce the amount of sediment being delivered to the stream or (2) the stream itself to remove the introduced sediment.

Wolman (1954) described a method for sampling coarse riverbed material that is rapid and produces statistically reliable information on the median bed particle size. His procedure involves randomly sampling bed material within a grid system and then plotting these data as a cumulative frequency distribution. His method is commonly modified by establishing transects across the channel (cross-sections

perpendicular to the flow) and sampling on a toe-to-heel spacing until 100 pebbles within habitat features of interest (pools, riffles, or combinations thereof) are collected (Leopold 1970).

Our interest is in shifts in the fine gravel and smaller portion of the distribution, not the median. Therefore, through a longitudinal stream reach (most commonly a pool and riffle complex) we modified the Wolman procedure into a zig-zag pattern such that a stream reach is sampled along a continuum instead of an individual cross-section. By doing this, numerous meander bends and all associated habitat features can be sampled as an integrated unit rather than as individual cross-sections.

Preliminary Sampling and Analysis

In 1992, we conducted preliminary sampling and analysis to determine if the zig-zag procedure was worth pursuing in detail. Many reference and study reaches within the headwaters of the Clarks Fork of the Yellowstone River drainage were sampled. We located four stream classes within these watersheds that have characteristics common to both reference and study reaches of immediate interest. The classes are based on a stream classification hierarchy developed for the Shoshone National Forest (Hoskins 1979) that integrates gradient, watershed area, landform, lithology, and channel pattern as differentiating characteristics. Of the 37 classes identified by Hoskins, we limited sampling to classes 3, 5, 10, and 16. Class 5 reference reaches were sampled in anticipation of future needs even though no study reaches having this class were present in the sampled watersheds. Table 1 displays the characteristics of these particular stream classes.

From topographic maps that showed the stream class delineations, individual reaches to be sampled were located in the field. Then we followed a zig-zag pattern of going upstream from bankfull to bankfull (as illustrated in figure 1 and explained in more detail later) and measured pebbles at approximately 7-foot intervals. One hundred pebbles were mea-

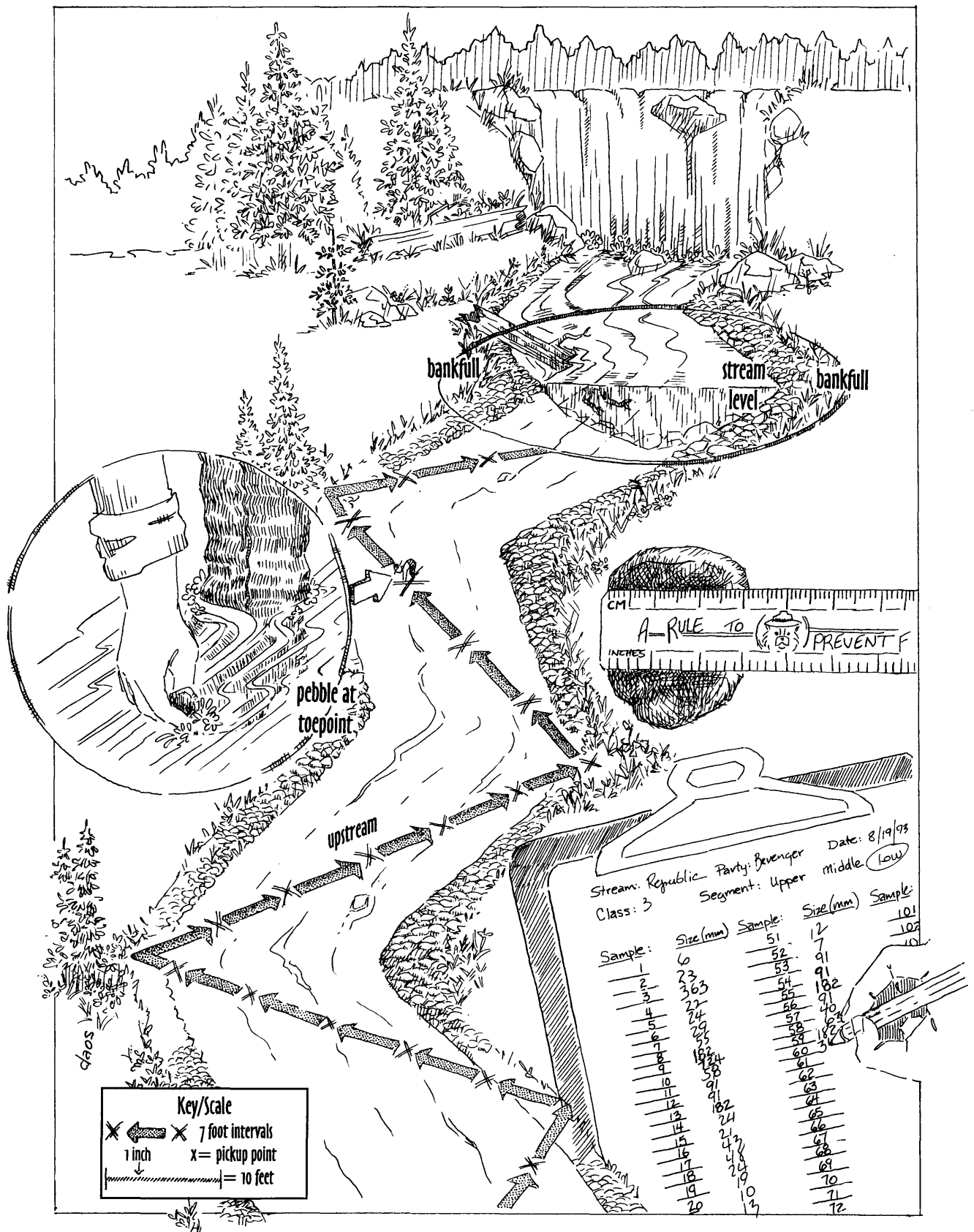


Figure 1.—Zig-zag pebble count procedure.

Table 1.— Stream classes used in this study and their characteristics.

Stream class	Stream gradient	Watershed area	Landform	Lithology	Channel pattern
3	6–12%	< 4 mi ²	Not defined	Volcanic	Not defined
5	3–6%	< 4 mi ²	Not defined	Volcanic	Not defined
10	3–6%	4–20 mi ²	Alluvium	Volcanic	Not defined
16	2–3%	4–20 mi ²	Alluvium	Volcanic	Not defined

sured within each reach, which translated to sampling a longitudinal profile several hundred feet long. Pebbles were randomly selected by reaching over the toe of the wader with an extended forefinger and picking up the first rock that was touched. The intermediate axis of each selected pebble was measured and tallied using standard Wentworth size classes (particles less than 2 mm; 2–4 mm; 4–8 mm; 8–16 mm; 16–32 mm; 32–64 mm; 64–128 mm; 128–256 mm; 256–512 mm; 512–1,024 mm; 1,024–2,048 mm; and 2,048–4,096 mm). Figure 2 displays particle size distributions for reference streams sampled in 1992, while figure 3 displays distributions for study streams.

These data were analyzed using contingency tables (gross Wentworth size classes of sands, gravels, cobbles, and boulders as columns and different reaches as rows) and the Pearson chi-squared statistic. Comparing reaches within classes, particle size distributions were homogeneous for class 5 and larger ($p > 0.13$). Within class 3, particle size distributions were heterogeneous ($p < 0.001$), but large differences were present only for cobble and boulder size classes. When particle sizes were summarized as proportion of fine particles and proportion of larger particles, no differences were present among reference reaches across all classes and using a variety of cut-off points <8mm to define fine particles ($p > 0.68$). We concluded that reference particle size distributions are relatively homogeneous for smaller size classes (fine gravels and smaller) and that study reach particle size distributions are significantly different from reference distributions ($p < 0.001$). This finding encouraged us to pursue a more rigorous analysis of the proposed procedure.

Validation Efforts

Using the results of the previously described preliminary effort in concert with several specific pragmatic assumptions, a statistical framework was de-

signed so the zig-zag pebble count procedure could be validated. This framework is described in detail in Appendix 1. Within this framework, additional field data were collected in 1993 and subjected to comprehensive statistical analysis.

The generic null hypothesis for the validation effort is the population of pebbles from the reference condition and the population of pebbles from the study condition are the same. This hypothesis requires testing several assumptions: (1) individual pebbles in a sample are independent members of the appropriate target population; (2) location of sample points within a stream reach is unimportant; (3) data from multiple reference stream reaches can be combined to appropriately represent a reference population (condition); and (4) sample results are independent of the field observer.

Each of these assumptions was analyzed by performing additional pebble counts in 16 reference reaches within 2 of the same streams sampled in 1992, Republic and Closed Creeks. As noted earlier, these reaches have similar lithology (volcanic) and a specified range of drainage area and gradient. Each reach was sampled by the senior author and one of three other observers. Each observer sampled 300 pebbles in each reach in Republic Creek and 200 pebbles in each reach in Closed Creek. Distance between samples was approximately 3.5 feet in all cases. Pebble size was recorded to the nearest millimeter to permit detailed analysis of the particle size distribution.

Assumption 1: Independence of Pebbles

To analyze independence of pebbles within a sample, autocorrelations of up to 20 lags were estimated for the senior author's data (to control for any variation among observers) for each reach. Autocorrelation is defined as the correlation between pairs of sequentially observed points (Chatfield 1989). Lag 1 autocorrelation (or serial correlation, r_1) is the correlation between each pair of adjacent points; lag 2

autocorrelation (r_2) is the correlation between every other point; lag 3 autocorrelation (r_3) is the correlation between every third point; and so on. Lag 1 autocorrelations on 2 of the 16 reaches were 0.26 and 0.43, respectively, and these were significant at $\alpha=0.05$. Typical values for r_1 ranged from 0.10 to 0.20. Autocorrelations at longer lags, r_2 to r_{20} , were not significant for any reach. In summary, there was occasionally a worrisome level of correlation between pebbles sampled 3.5 feet apart; there was never significant correlation between pebbles sampled at least 7 feet apart.

Assumption 2: Sampling Location Unimportant

To analyze the importance of sampling location within a reach, the senior author's data (again, to control for any observer variation) were divided into 2 (Closed Creek) or 3 (Republic Creek) 100-point sub-reaches. Sub-reaches within each reach were

compared using contingency tables structured to: (1) divide particle sizes into gross Wentworth size classes of sands, gravels, cobbles, and boulders and (2) divide particles into ≤ 4 mm vs. >4 mm; ≤ 6 mm vs. >6 mm; ≤ 8 mm vs. >8 mm; and ≤ 10 mm vs. >10 mm. These latter divisions were made to represent various definitions of sediment that might be defined as biologically significant. Only 2 reaches exhibited significant ($\alpha=0.05$) differences between the sub-reaches; one reach for the gross Wentworth size classes and one reach for the 10 mm cutoff point. In general, within-reach variability was minor.

Assumption 3: Multiple Stream Reaches Can Be Combined to Define a Reference

The reference condition must be well-defined to assume that a study reach would be the same as the reference condition if not impacted. Otherwise a null

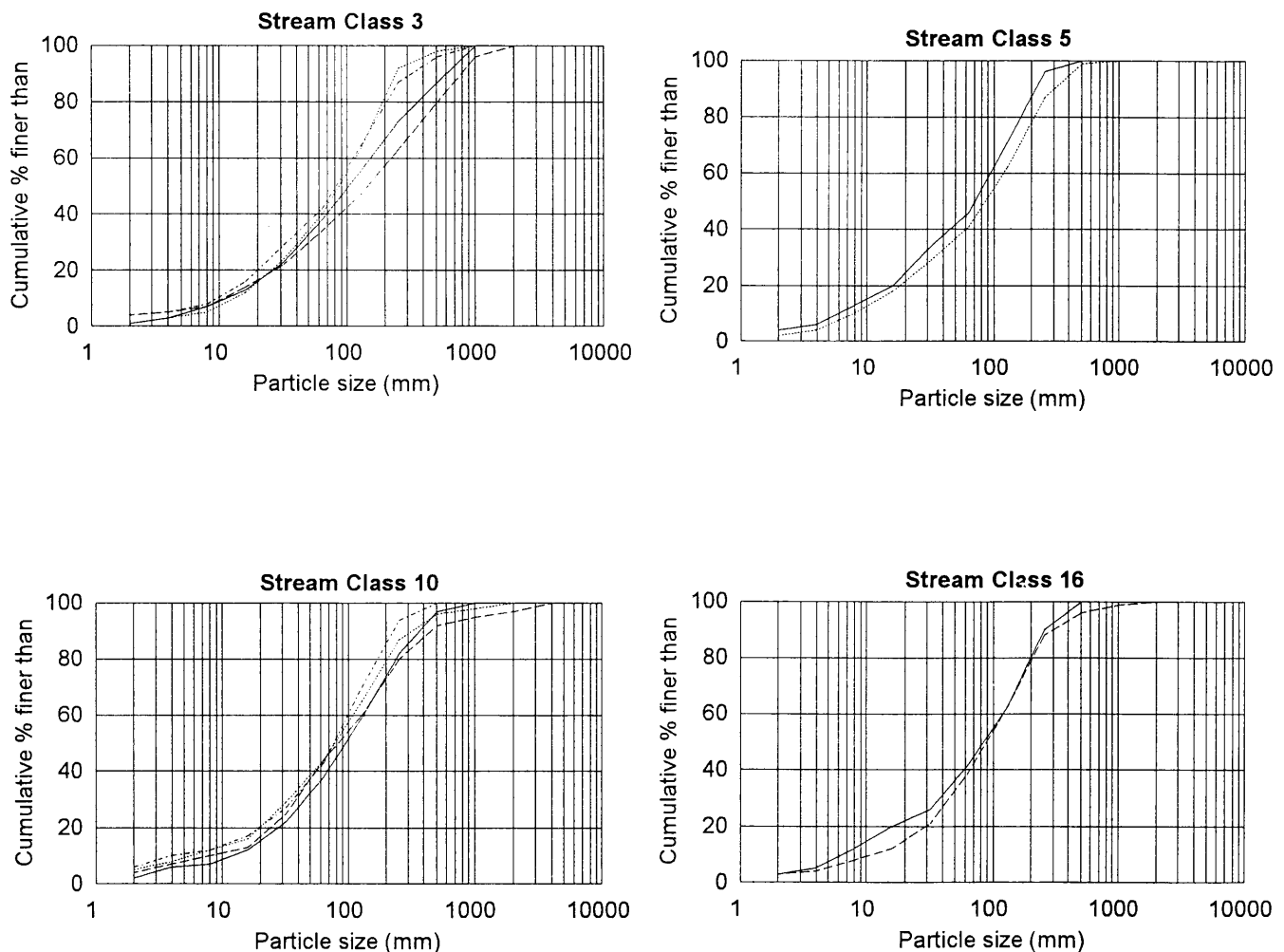


Figure 2.— Particle size distributions of reference streams sampled in 1992. Different lines refer to different reaches.

hypothesis that reference and study streams are not different would not be tenable. This suggests one should define a reference population using as many reference reaches as possible and use this as the basis for comparing impacted reaches.

To analyze this assumption, we explored the question, "To what degree are particle size distributions homogeneous among reference reaches?" The senior author's data was (1) compared among reference reaches within and between creeks (Republic and Closed Creeks) for 4 individual stream classes (Shoshone stream classes 3, 5, 10, and 16) using contingency tables and (2) compared between creeks and among stream classes using log linear models. Log linear models allow for frequency analysis of more than the two factors analyzed in a contingency table (see Agresti 1990). Particle size distributions were characterized as before using both gross Wentworth

size classes (sands, gravels, cobbles, and boulders) and "small" and "large" particles as defined by various cutoff points (4, 6, 8, and 10 mm).

For individual stream classes, comparisons of standard size fractions between creeks were significant for stream classes 10 and 16 (significance levels are displayed in table 2a), with differences apparent only for gravel and larger material. Similar comparisons assuming the various cutoff points of 4, 6, 8, and 10 mm were significant only for stream class 3 (table 2a). When multiple stream classes were added to the analysis using log linear models, heterogeneity among stream classes was evident. Additionally, stream class 3 exhibited evidence of internal heterogeneity (table 2b).

In summary, for the cutoff points used, reaches in individual stream classes 5, 10, and 16, both within and between Republic and Closed Creek, were homogeneous. Class 3 pooling attempts, both within

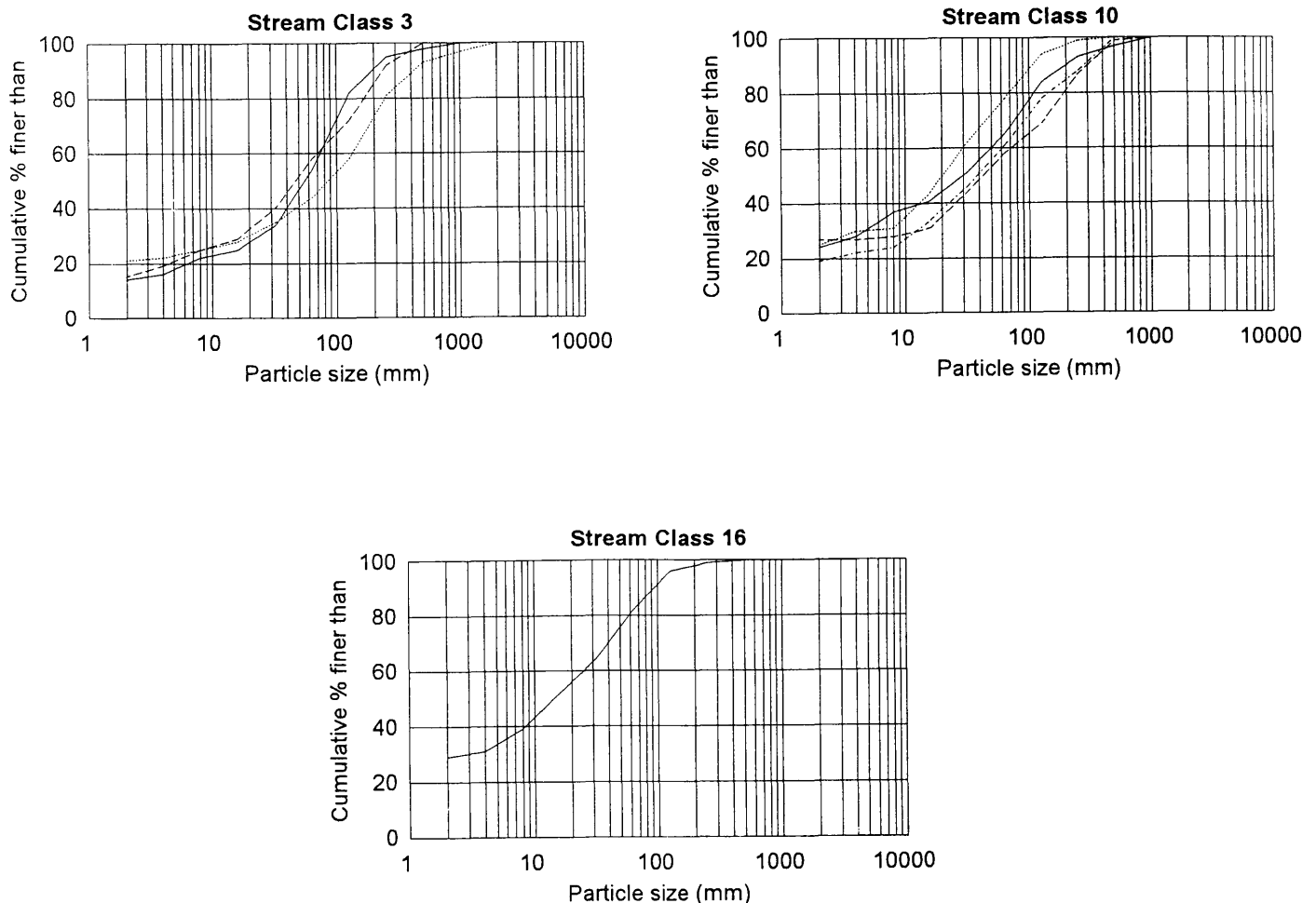


Figure 3.—Particle size distributions of study streams sampled in 1992. Different lines refer to different reaches.

and between Republic and Closed Creek, exhibited heterogeneity. This was considered in the comparison of class 3 reference conditions to a study reach (e.g., Case Study 2 later in this paper). Pooling reaches among stream classes was not successful.

Assumption 4: Independent of Field Observer

The fourth assumption tested was whether sample results are independent of field observer. Data from the two observers for each reach were compared using contingency tables structured similar to those used for assessing sampling location. When the particle size distribution was characterized using gross Wentworth size classes (sands, gravels, cobbles, and boulders), 7 of 10 reaches for Republic Creek and 5 of 6 reaches for Closed Creek showed significant differences ($\alpha=0.05$) between the data collected by the senior author and the second observer. When particles were divided into "small" and "large" classes as defined by the various cutoff points of 4, 6, 8, and 10 mm used in previous analyses, 3 of 10 reaches for Republic Creek showed significant differences; no significant differences were present for Closed Creek. The improved agreement between observers for Closed Creek may be due to increased training and monitoring of the second observer while sampling. However, second observers were different between Republic and Closed Creek, so a conclusive statement cannot be made from these data. Regardless, where observer variability was present, it was more severe for larger particle sizes (cobbles and boulders). This may be because cobbles and boulders are difficult to measure to the nearest millimeter or because there is inherent wide variability in cobble and boulder size material. In conclusion, data should be analyzed using cutoff points of interest rather than by gross Wentworth size classes.

How to Use the Zig-Zag Pebble Count Procedure

Step 1. Think about how to delineate reference and study reaches. We used the Shoshone stream classification system. Other stream classification systems may work just as well. Classifying reaches by watershed area, gradient, and lithology may be sufficient. Whatever delineation technique is used, you should conduct preliminary sampling and analysis to determine if the effort is worth pursuing. It is particularly important to stratify reference reaches into homogeneous groups. When this cannot be accomplished (e.g., as was the case for class 3 reaches on the Shoshone National Forest), the analysis should reflect the heterogeneity of the reference condition (e.g., Case Study 2 later in this paper).

Before initiating field work, you should ask, "How many pebble counts are needed to detect a real difference between reference and study conditions?" To answer this question, the nature of the reference condition, the size of change that might occur from a management action, and acceptable levels of risk must be specified. Discussion of this analysis is postponed until Step 4 below to allow a direct link to be made with the statistical analysis, but it should be completed before Step 2.

Step 2. Delineate reference and study reaches on topographic maps, and then locate the reaches in the field. Figure 1 illustrates the basic zig-zag sampling technique. For each reach, choose a random location on one bank at bankfull stage. Once you locate this point, identify an upstream target point on the opposite bank that can be used to hold a straight line while sampling. Then pace off 7 feet (3 to 4 steps depending on your pace), reach over the toe of the

Table 2a.—Significance levels of contingency table comparisons between Republic and Closed creeks for reference data collected by Bevenger.

Stream class	Standard fractions	Small vs. large particles greater than:				
		10 mm	8 mm	6 mm	4 mm	2 mm
3	0.469	0.018	0.004	0.021	0.088	0.124
5	0.096	0.292	0.573	0.760	0.932	0.685
10	<0.001	0.239	0.142	0.172	0.485	0.999
16	0.019	0.444	0.481	0.416	0.958	0.999

Table 2b.—Significance levels of log linear model components¹ for comparisons between Republic and Closed Creeks and among stream classes for reference data collected by Bevenger.

Model components	Standard fractions	Small vs. large particles greater than:				
		10 mm	8 mm	6 mm	4 mm	2 mm
1st order	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2nd order	<0.001	<0.001	<0.001	<0.001	0.001	0.003
3rd order	0.039	0.075	0.031	0.071	0.367	0.366

¹ 1st order factors are particle size class, stream class, and creek. 2nd order components are interactions between 2 of the factors. The 3rd order component is interaction among all 3 factors.

wader with the forefinger without looking down, pick up the first pebble touched, and measure it in millimeters. If it is less than 64 mm, record the actual size on the data form (figure 4 contains a sample form). If it is greater than or equal to 64 mm, assign the sample to the appropriate Wentworth size class by recording the class midpoint based on a log scale:

- 91 for pebbles 64–127 mm;
- 182 for pebbles 128–255 mm;
- 363 for pebbles 256–511 mm;
- 724 for pebbles 512–1,023 mm;
- 1,445 for pebbles 1,024–2,047 mm; and
- 2,884 for pebbles 2,048–4,095 mm.

This recording method allows for analysis of cut-off points such as 4, 6, 8, and 10 mm as well as other types of pebble count applications as mentioned in the introduction. Once you record the pebble, discard it, focus on the target, pace 7 feet, and sample another pebble. Continue working across and up the stream until you reach the far bankfull. Then locate another target, this time on the bank on which the survey line began, and work across the creek. Continue this technique until the desired number of pebbles are sampled.

The result should be a zig-zag line that traverses through a stream reach such that somewhere along the zig-zag the inside and outside of riffles, runs, pools, and meander bends are sampled. The angle of the zig-zag depends upon the meander pattern of the stream reach being sampled. We cannot give precise instructions on what the angle of the zig-zags should be except that low sinuosity reaches require less sharp angles than high sinuosity reaches. If channel length is defined as the distance along the thalweg (i.e., deepest point of the channel), and if the sampling design called for 100 pebbles at 7-foot intervals, then 700 feet of stream reach would be sampled

if all samples were taken from the thalweg. Since the zig-zag pattern follows a bankfull to bankfull line across and up the channel, less than 700 feet of thalweg length will actually be sampled. Our field experience indicates that the ratio of zig-zag length to thalweg length ranges from 0.85 to 0.95. In other words, our final reach lengths were between 595 and 665 feet, instead of 700 feet. Of course, these ratios hold only for the streams we sampled. Streams of different sinuosity would have different ratios.

Individuals should be trained before starting field sampling and regularly monitored while sampling to ensure they are properly applying the technique. We found that poorly or improperly trained people collected data significantly different from that collected by trained people. Additional training remedied this. Important factors to consider and monitor during training are that (1) spacing between pebble samples is correct, (2) pebbles are being selected properly, (3) pebbles are being measured properly, and (4) the zig-zag pattern is being properly laid out.

Step 3. Randomly choose several streams (and reaches within streams) from those available in a class to define the variability within the reference condition. Reaches within each hypothesized reference condition should be analyzed for homogeneity, and subsequent analyses should reflect any heterogeneity found (e.g., class 3 reference reaches and Case Study 2 below). If a rationale can be found and sufficient data exists, it may be possible to further stratify heterogeneous reference conditions into homogeneous subsets. However, sufficient data should be obtained to precisely describe each subset for subsequent comparison to study reaches. The intensity of sampling needed to define each reference condition and study reach is discussed in Step 4.

Stream:

Party:

Date:

Class:

Segment:

Upper Middle Lower

Size: 1 <2mm; 2mm<actual>63mm; 64<91>127; 128<182>255; 256<363>511; 512<724>1023; 1024<1445>2047; 2048<2884>4095.

#	Size	#	Size	#	Size	#	Size	#	Size	#	Size	#	Size	#	Size
1		46		91		136		181		226		271			
2		47		92		137		182		227		272			
3		48		93		138		183		228		273			
4		49		94		139		184		229		274			
5		50		95		140		185		230		275			
6		51		96		141		186		231		276			
7		52		97		142		187		232		277			
8		53		98		143		188		233		278			
9		54		99		144		189		234		279			
10		55		100		145		190		235		280			
11		56		101		146		191		236		281			
12		57		102		147		192		237		282			
13		58		103		148		193		238		283			
14		59		104		149		194		239		284			
15		60		105		150		195		240		285			
16		61		106		151		196		241		286			
17		62		107		152		197		242		287			
18		63		108		153		198		243		288			
19		64		109		154		199		244		289			
20		65		110		155		200		245		290			
21		66		111		156		201		246		291			
22		67		112		157		202		247		292			
23		68		113		158		203		248		293			
24		69		114		159		204		249		294			
25		70		115		160		205		250		295			
26		71		116		161		206		251		296			
27		72		117		162		207		252		297			
28		73		118		163		208		253		298			
29		74		119		164		209		254		299			
30		75		120		165		210		255		300			
31		76		121		166		211		256		301			
32		77		122		167		212		257		302			
33		78		123		168		213		258		303			
34		79		124		169		214		259		304			
35		80		125		170		215		260		305			
36		81		126		171		216		261		306			
37		82		127		172		217		262		307			
38		83		128		173		218		263		308			
39		84		129		174		219		264		309			
40		85		130		175		220		265		310			
41		86		131		176		221		266		311			
42		87		132		177		222		267		312			
43		88		133		178		223		268		313			
44		89		134		179		224		269		314			
45		90		135		180		225		270		315			

Figure 4.—Field data form.

Based on the analyses that we performed and considering that only two streams were available for comparison, reaches of similar Shoshone stream class could be pooled both within and between streams to describe a reference condition involving small particle sizes. Reaches cannot be pooled among different stream classes, and describing a reference condition for larger particle sizes seems problematic. For similar Shoshone stream classes, the number of reaches used to describe a reference condition should increase as the spatial scale of the reference condition increases.

Step 4. The hypothesis recommended for testing is

$$H_0 : p_s - p_r = 0 \text{ versus } H_a : p_s - p_r > 0$$

where p_r is the reference proportion (for instance, the proportion of fines <8mm) and p_s is the respective proportion for a study condition. This is a test where detection of a significant *increase* on the study condition is all that is considered important (that is, in statistical terms a one-tailed test is being used). As noted in Step 1, you should answer the question “How many pebble counts are needed to detect a real difference between reference and study conditions?” before initiating field work. Fleiss (1981) shows how to estimate needed sample sizes, both when equal sample sizes are used for the two conditions (equations 3.14 and 3.15) and when unequal sample sizes are used (equations 3.18 and 3.19). Fleiss’s equations are provided in Appendix 2 for use in situations that require exact sample size calculations. Figures 5 and 6 plot various combinations of proportions and sample sizes for situations in which coarser estimates of needed sample sizes are sufficient.

To demonstrate use of the sample size graphs, assume there is a timber sale and associated road building planned in a municipal water supply. The local sawmill may suffer economically if the sale does not occur, and the municipality will be impacted if the sale results in increased sedimentation in the water supply. A decision is made to use the zig-zag pebble count procedure as one tool in the analysis of the sale.

The potential costs associated with both Type I (α) and Type II (β) errors should be considered when estimating sample sizes. A Type I error in the present example is the risk of falsely concluding that there is a difference between the study stream and its reference, which might lead to a decision to postpone the sale. A Type II error in the present example is the

probability of falsely concluding there is no difference between the study stream and its reference, which might lead to a decision to proceed with the sale as proposed, potentially resulting in increased water treatment costs. Typically, sample sizes are estimated assuming $\beta=4\alpha$, most commonly $\alpha=0.05$ and $\beta=0.20$ (Fleiss 1981).

The sample size question can be broken into two parts: (1) what sample size is needed to define the reference condition and (2) what sample size is needed to define the study reach? To answer these questions, the nature of the reference condition, the size of change that might occur from a management action, and the risk levels (α and β) that are acceptable must be specified. Suppose no existing reference data are available, but you hypothesize that the proportion of particles <8 mm for the reference condition is about $p_r=0.10$ and a change on the study reach to $p_s=0.20$ or larger (i.e., a change of 0.10 or larger) is considered biologically significant.

For standard values of $\alpha=0.05$ and $\beta=0.20$, the upper right graph in figure 5 shows sample sizes needed to detect a difference of $p_s - p_r > 0.10$. First, locate the difference point of 0.10 on the vertical axis and scan across the plot horizontally to the curve representing the reference sample size. Then, read down to the horizontal axis to obtain the sample size needed for the study reach. For example, if a sample size of $n_r=450$ or larger were used to characterize the reference condition, then a sample size (n_s) of 100 or less would be needed for the study condition. If $n_r=300$ were used to define the reference condition, $n_s=115$ would be needed for the study condition, while an $n_r=150$ would require about $n_s=215$. Sample sizes substantially larger than 300 would be required for the study condition if $n_r=100$ were used for the reference condition. Alternatively, if the minimum difference considered biologically significant were relaxed to $p_s - p_r = 0.15$, then a sample size of 100 would be more than adequate for both reference and study conditions.

In some situations, such as if a threatened species were present, it might be appropriate to strike a balance between both errors rather than worrying more about the risk of falsely declaring a difference. Figure 6 contains plots of sample sizes needed when both α and β are set equal to 0.05. When $p_r=0.10$, detecting a difference of $p_s - p_r = 0.10$ would require about $n_s=290$ for $n_r=300$. Increasing the sample size for the reference condition to $n_r=450$ or 600 would decrease the sample size needed for the study condition to about $n_s=215$ or 190, respectively.

After a point, increasing the sample size for the reference condition does not result in an equal reduction in the sample size needed for the study condition. This can be seen in both figures 5 and 6, where distance between reference lines decreases at larger reference condition sample sizes. The trade-off between this and minimizing sample size for both conditions is that additional sampling will result in a better description of the reference condition that could be used for more than one study reach, each of which could be sampled with fewer samples. In general, we recommend sampling several reaches to characterize the range of variability of the reference condition with the side benefit that less sampling will be needed for comparing several study conditions to the same reference condition.

Using the Procedure in a Watershed Assessment

A coarse screening process was recently developed on the Shoshone National Forest to analyze watershed cumulative effects. This screening process revealed numerous watersheds of concern (Shoshone National Forest 1993). A watershed of concern is defined as a fourth-order drainage where inherent soil and water sensitivity plus land disturbing activities may be causing significant degradation of aquatic resources. In such watersheds any additional surface disturbance may only add to the level of concern. Since the effort was programmatic and the screens were coarse, managers required that a watershed of concern be field verified to determine if the watershed and streams within it are actually in a degraded

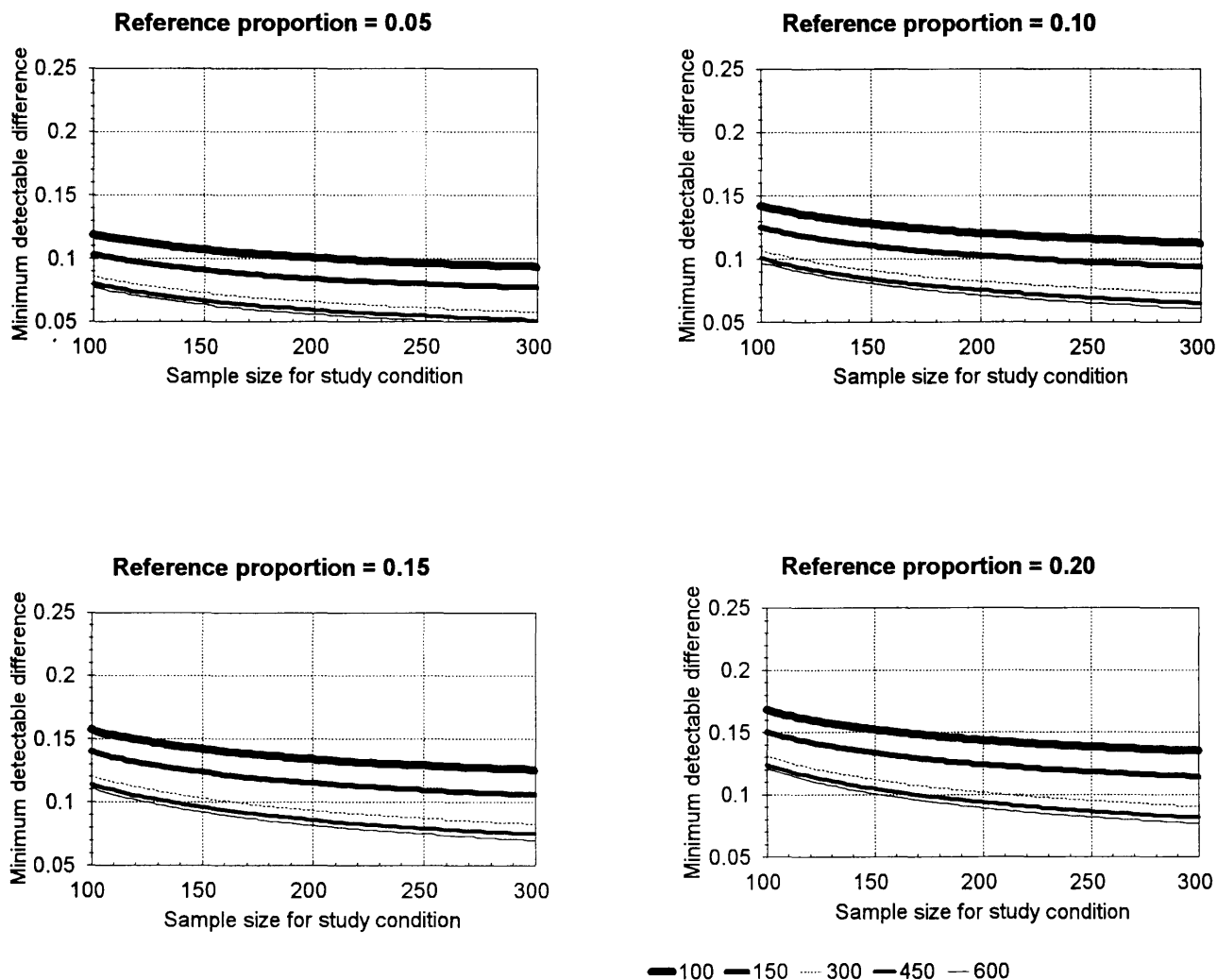


Figure 5.—Minimum detectable difference and study condition sample size estimates assuming $\alpha = 0.05$ and $\beta = 0.20$.

condition. After field verification, appropriate site-specific mitigation requirements could then be developed and implemented. Mitigation requirements may be in the form of “standard operating procedures” if a watershed is not one of concern or in the form of “extraordinary mitigation” if the watershed is one of concern.

Since pebble count data were available for both reference and study reaches, we used these data to demonstrate one part of the process of verifying the level of concern and amount of mitigation that would be needed if additional management were to take place. More specifically, contingency tables were established to compare small and large particle size fractions between reference and study reaches. We used the data collected in 1993 from Republic and

Closed Creeks to represent reference reaches. For study reaches we used data collected from our initial reconnaissance work conducted in 1992. MacDonald et al. (1991) summarize the literature by reporting that particles up to 6.4 mm are of most concern to the fishery resource. Since the 1992 data were collected by Wentworth size classes, we used <8 mm as the cutoff point because this is the size class closest to, but yet greater than or equal to, 6.4 mm. We are confident with our approach because we compared reaches that are similar in lithology, watershed area, and gradient as incorporated by the Shoshone stream classification system. The following four case studies briefly describe the results of these comparisons.

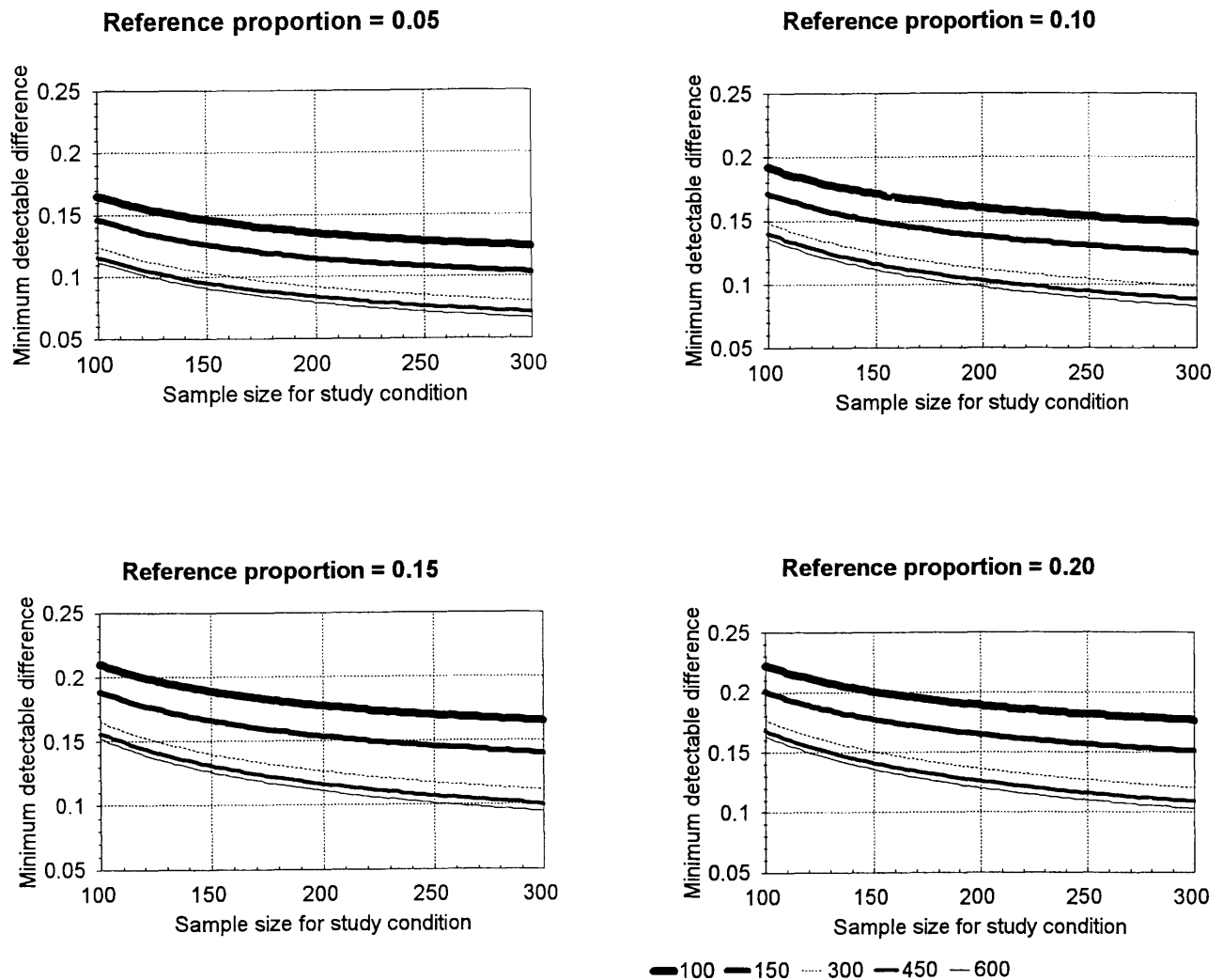


Figure 6.—Minimum detectable difference and study condition sample size estimates assuming $\alpha = 0.05$ and $\beta = 0.05$.

Case Study 1: Recent Wildfire

Lodgepole and Pilot Creeks were heavily burned by the Clover-Mist Fire in 1988. Recovery of surface-protecting vegetation within both watersheds has been very slow and was practically nonexistent the first 2 years after the fire. The fire caused a considerable water yield increase that has resulted in major adjustments in stream channel morphology (Troendle and Bevenger 1993). Additionally, Lodgepole Creek received a significant rainfall event the summer after the fire that resulted in exceptionally large amounts of overland flow, substantial changes in channel morphology, and considerable inner gorge mass movement activity. Figure 7 displays post-fire particle size distributions for a class 10 reach in both Lodgepole and Pilot Creeks versus the distribution for the class 10 reference. These 2 reaches are significantly different from the reference (significance level

$p < 0.001$ for both reaches) for particles less than 8 mm ($p_r = 0.10$ versus $p_s = 0.37$ and $p_s = 0.28$, respectively).

Case Study 2: Recent Timber Harvest

Jim Smith Creek flows through an area that was logged in the late 1980's. Trees were harvested along the banks of the creek. Due to the inherent instability of the channel, many of the root wads have since sloughed into the channel. Fine material surrounding the root wads is being scoured by the streamflow and deposited downstream. Figure 7 displays the post-harvest particle size distribution for a class 3 study reach versus the distributions for the class 3 reference reaches. For particles less than 8 mm, this reach is significantly different from 2 of the reference reaches ($p_s = 0.22$ versus $p_r = 0.06$ [$p < 0.001$] and $p_r = 0.11$ [$p = 0.009$]), but not the third ($p_r = 0.14$ [$p = 0.068$]). Therefore, the small size fraction on Jim Smith Creek

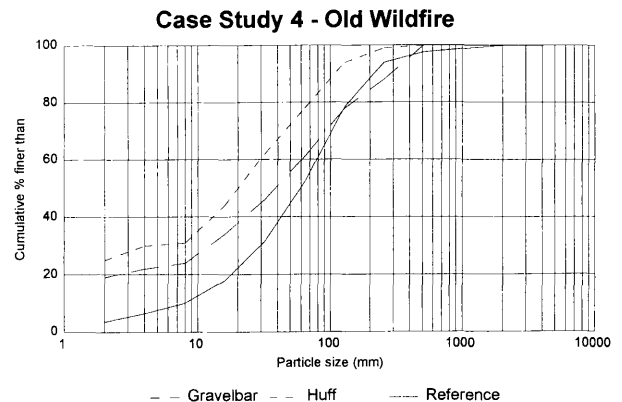
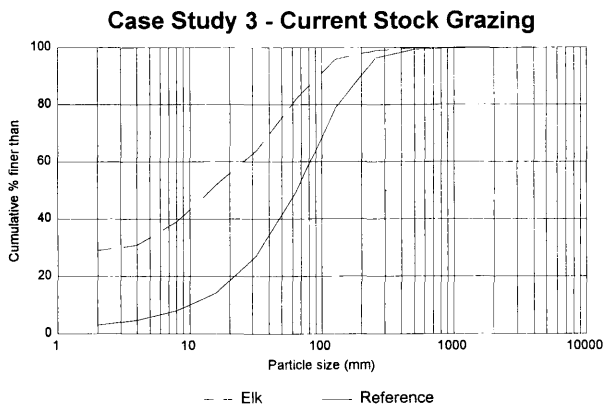
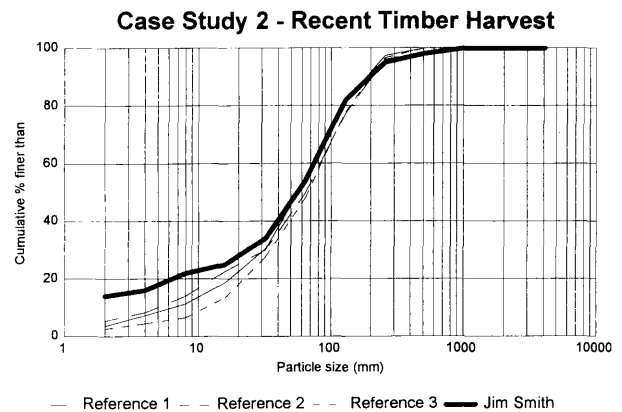
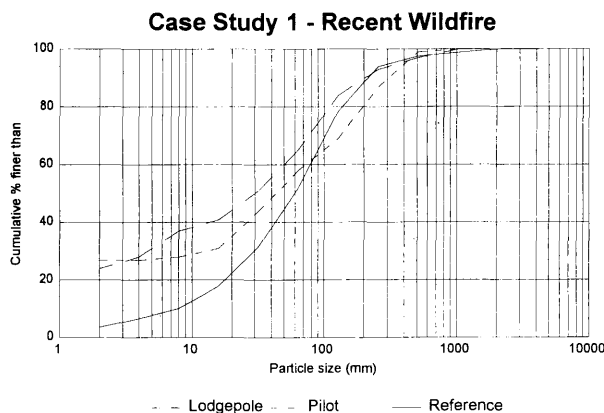


Figure 7.—Particle size distributions of case study streams.

is comparable to the upper end of the variability observed in the reference condition, but quite different from the lower end of the reference variability. This suggests that before making mitigation recommendations in this situation further analysis is needed. Additional class 3 reference reaches could be sampled to establish a better estimate of the "norm" and thereby provide for a conclusive interpretation of the fine sediment load in Jim Smith Creek. Also, the logging effects analysis could be expanded by additional on-site assessment of other indicators of watershed condition and stream health.

Case Study 3: Current Stock Grazing

Elk Creek flows through an active cattle grazing allotment. Cattle congregate within the riparian zone, resulting in considerable trailing near the channel and sloughing of the banks. Figure 7 displays the current particle size distribution for a class 16 study reach versus the distribution for the class 16 reference. The study reach is significantly different from the reference ($p < 0.001$) for particles less than 8 mm ($p_r = 0.08$ versus $p_s = 0.39$).

Case Study 4: Old Wildfire

Gravelbar and Huff Gulch Creeks flow through an area burned 50 years ago by the Gravelbar Fire. Slopes less than 50% have recovered quite well to grasses on south aspects and lodgepole pine on north aspects. Slopes steeper than 50% show minimal recovery, probably due to continued loss of the very thin soil during summer thunderstorm events. Figure 7 displays the current particle size distribution for class 10 reaches in Gravelbar and Huff Gulch Creeks versus the distribution for the class 10 reference. The study reaches are significantly different from the reference ($p < 0.001$ for both reaches) for particles less than 8 mm ($p_r = 0.10$ versus $p_s = 0.24$ and $p_s = 0.31$, respectively). This result suggests stream health recovery in this particular geographic setting lags considerably behind upslope recovery, which may not be typical. For example, Potyondy and Hardy (1994) and Wells et al. (1979) showed statistically significant increases in fine sediment in streams appear to last only a few years after the fires before returning to pre-fire conditions.

Conclusions

The zig-zag pebble count procedure we have developed appears to be an effective and defensible field monitoring technique that can be used as part of a monitoring program designed to assess land

management activities and watershed cumulative effects. Currently, the Shoshone National Forest uses the procedure as part of a watershed cumulative effects monitoring program. A range of stream types and management conditions have been and are being assessed. Additionally, the procedure is being tested in other physiographic settings by other users. The procedure will not be adaptable to every situation (for example, sand-bed streams) but may have merit in many National Forest monitoring programs.

Literature Cited

- Agresti, A. 1990. Categorical data analysis. New York, NY: Wiley. 558 p.
- Chatfield, C. 1989. The analysis of time series: an introduction, 4th ed. London, Eng.: Chapman and Hall. 241 p.
- Clean Water Act, as amended by the Water Quality Act of 1987, Public Law 100-4. 1988.
- Fleiss, J.L. 1981. Statistical methods for rates and proportions, 2nd ed. New York, NY: Wiley. 321 p.
- Haber, M. 1980. A comparison of some continuity corrections for the chi-squared test on 2x2 tables. *Journal of the American Statistical Association*. 70:510-515.
- Hoskins, W. 1979. Shoshone National Forest Stream Classification System. USDA Forest Service, Shoshone National Forest.
- Leopold, L.B. 1970. An improved method for size distribution of stream-bed gravel. *Water Resources Research*. 6(5):1357-1366.
- MacDonald, L.H.; Smart, A.; Wissmar, R.C. 1991. Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. U.S. Environmental protection Agency, Region 10, NPS Section, EPA/910/9-91-001. Seattle, Washington. 166 p.
- Meehan, W.R. 1991. Influence of forest and rangeland management on salmonid fishes and their habitats. Bethesda, Maryland: American Fisheries Society; Special Publication 19. 751 p.
- Potyondy, J.P.; Hardy, T. 1994. Use of pebble counts to evaluate fine sediment increase in stream channels. *Water Resources Bulletin*. 30(3):509-520.
- Shoshone National Forest. 1993. Oil and Gas Leasing Final Environmental Impact Statement. USDA Forest Service, Shoshone National Forest.
- Troendle, C.A.; Bevenger, G.S. 1993. Effect of fire on streamflow and sediment transport, Shoshone National Forest, Wyoming. *In: proceedings, The*

Ecological Implications of Fire in Greater Yellowstone (in print).
Wells, C.G.; Campbell, R.E.; DeBano, L.F.; Lewis, C.E.;
Fredriksen, R.L.; Franklin, E.C.; Froelich, R.C.;
Dunn, P.H. 1979. Effects of fire on soil: a state-of-

knowledge review. Washington, D.C.: USDA Forest Service, General Technical Report WO-7. 34 p.
Wolman, M.G. 1954. A method of sampling coarse river-bed material. Transactions American Geophysical Union. Volume 35. Number 6. Pp. 951–956.

Appendix 1. Statistical Framework

Fleiss (1981, chapter 2) is a comprehensive reference used in developing this statistical framework. Data for comparison of reference and study conditions are obtained through purposive sampling (Fleiss's Method II). A predetermined number of pebbles are obtained to represent the reference (n_r) and study (n_s) conditions. The sample of pebbles for each condition characterizes the particle size distribution for the reach(s) sampled. One way to compare these particle size distributions between reference and study conditions, and to focus on a particular question or issue, requires division of the data into two categories (for example, smaller versus larger particles). For instance, let n_{r1} and n_{s1} be the number of particles < 8 mm observed in the reference and study samples, respectively. Estimates of the proportion of sediment for each condition are $p_r = n_{r1}/n_r$ and $p_s = n_{s1}/n_s$, respectively.

The statistical significance of the difference between p_r and p_s can be tested using:

(1) the Pearson chi-squared statistic in a 2x2 contingency table (table A1) (available in most general-purpose statistical software systems)

or

2) computation of a statistic

$$z = \frac{|p_r - p_s| - \frac{1}{2}(1/n_r + 1/n_s)}{\sqrt{pq}(1/n_r + 1/n_s)}$$

where $\bar{p} = (n_{r1} + n_{s1}) / (n_r + n_s) = n_{.1} / n_{..}$, $\bar{q} = 1 - \bar{p}$, and z refers to the standard normal distribution. $-1/2 (1/n_r + 1/n_s)$ is Yate's correction for continuity to account for using a continuous distribution (the chi-squared or normal, respectively) to represent discrete frequencies (Fleiss 1981). Some statistical software systems provide this correction (e.g., SPSS CROSSTABS) or optional exact methods that do not require the correction (e.g., SAS PROC FREQ). Haber (1980) shows that more complicated continuity corrections (not provided in standard statistical systems) provide better estimates of true significance levels, but also that Yate's correction is conservative in the

sense of erring in the direction of overestimating significance levels. Haber (1980) also shows that the uncorrected Pearson chi-squared statistic is susceptible to under-estimating significance levels, and therefore may not be providing Type I error protection at the desired level. Most statistical systems also estimate the likelihood ratio chi-squared statistic for contingency tables, but Agresti (1990) indicates that this statistic provides inconsistent results, especially for small sample sizes. Therefore, exact methods should be used to evaluate significance in contingency tables whenever possible; otherwise, use the Pearson chi-squared statistic corrected for continuity.

Significance of either of these tests indicates p_s differs from p_r , but not whether it is smaller or larger. In some situations such as cumulative effects analysis, it is appropriate to formulate a uni-directional (or one-tailed) test for whether $p_s > p_r$. That is, did a management action increase the fine sediment proportion of the particle size distribution? To evaluate a contingency table or z-statistic for a difference in only one direction, first observe whether $p_s > p_r$. If it is, then compute test statistics as before, but assess significance against tabled values for 2α (where α is the predetermined level of Type I error) or divide computed significance levels in half (e.g., computed $p=0.06$ becomes $p=0.03$). For the same sampling effort, uni-directional tests are more sensitive than their two-directional counterparts but do carry the danger of ignoring significant change in the opposite direction.

Table A1.—Notation for 2x2 contingency table.

Row	Column		Total
	Smaller size	Larger size	
Reference	n_{r1}	$n_{r2} = n_{r.} - n_{r1}$	$n_{r.}$
Study	n_{s1}	$n_{s2} = n_{s.} - n_{s1}$	$n_{s.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Appendix 2. Sample Size Estimation

Study condition sample size requirements for selected combinations of reference proportions (p_r) and sample sizes (n_r) are presented in figures 5 and 6. Estimation equations applicable to any situation are presented here. A one-sided hypothesis testing framework is assumed:

$$H_0 : p_s = p_r \text{ versus } H_a : p_s > p_r$$

or

$$H_0 : p_s - p_r = 0 \text{ versus } H_a : p_s - p_r > 0$$

where p_r is the reference proportion (for instance, the proportion of fines <8mm) and p_s is the respective proportion for a study condition. If equal sample sizes are used to sample both conditions, then the estimated sample size (n) for a one-tailed test is given by (Fleiss 1981, Section 3.2):

$$n' = \frac{\left(z_\alpha \sqrt{2pq} - z_{1-\beta} \sqrt{p_r q_r + p_s q_s} \right)^2}{(p_s - p_r)^2} \quad \text{and} \quad [1]$$

$$n = \frac{n'}{4} \left(1 + \sqrt{1 + \frac{4}{n' |p_s - p_r|}} \right)^2$$

where $\bar{p} = (p_r + p_s)/2$, $\bar{q} = 1 - \bar{p}$, $q_r = 1 - p_r$, $q_s = 1 - p_s$, n is the estimated sample size for each condition, and z_α and $z_{1-\beta}$ refer to ordinates of the standard normal distribution. α refers to the probability of a Type I error, which is the assumed probability of declaring a false difference; β refers to the probability of a Type II error, which is the assumed probability of not detecting a true difference.

To estimate the number of samples (pebbles) needed when reference and study conditions have unequal sample sizes, you must first specify a factor (f) relating sampling effort between the two conditions, $n_s = f n_r$. For instance, setting $f=0.5$ would result in a sample size for the study condition (n_s) half the size of the sample size for the reference condition (n_r). In this case, sample size estimation equations are (Fleiss 1981, Section 3.4):

$$m' = \frac{\left(z_\alpha \sqrt{(f+1)pq} - z_{1-\beta} \sqrt{f p_r q_r + p_s q_s} \right)^2}{f(p_s - p_r)^2} \quad \text{and} \quad [2]$$

$$m = \frac{m'}{4} \left(1 + \sqrt{1 + \frac{2(f+1)}{m' f |p_s - p_r|}} \right)^2$$

where $\bar{p} = (p_r + f p_s)/(f+1)$, $\bar{q} = 1 - \bar{p}$, the estimated reference sample size is $n_r = m$, and estimated study sample size is $n_s = f m$. All of these equations can readily be incorporated into a computer spreadsheet.

To demonstrate use of the sample size equations, we continue the timber sale example started on page 9. Suppose no existing reference data is available and there is no reason to think that the two conditions should be sampled differently. If it is hypothesized that the proportion of particles <8 mm for the reference condition is about $p_r = 0.10$, and it is desirable to be able to detect a change on the study reach to $p_s = 0.20$ or greater, then equation [1] becomes

$$n' = \frac{\left(z_\alpha \sqrt{2 \times .15 \times .85} - z_{1-\beta} \sqrt{.1 \times .9 + .2 \times .8} \right)^2}{(.2 - .1)^2} \quad \text{and} \quad [3]$$

$$n = \frac{n'}{4} \left(1 + \sqrt{1 + \frac{4}{n' [.2 - .1]}} \right)^2$$

where typically-used values for α and β , and corresponding Z_α and $Z_{1-\beta}$ are:

α or β	Z_α	$Z_{1-\beta}$
0.01	-2.327	2.327
0.05	-1.645	1.645
0.10	-1.282	1.282
0.20		0.842

Table A2 contains sample size estimates from equation [3] for these levels of Type I and Type II error. Tables A3 and A4 contain unequal sample size estimates using equation [2], and $f=1/2$ and $f=1/3$, respectively. In this example, standard values of $\alpha=0.05$

Table A2.—Equal sample sizes for $p_r = 0.10$ and $p_s = 0.20$.

β	α		
	0.01	0.05	0.10
0.01	566	417	347
0.05	419	293	236
0.10	349	236	185
0.20	275	177	134

Table A3.—Unequal sample size estimates ($f = 1/2$) for $p_r = 0.10$ and $p_s = 0.20$.

β	α		
	0.01	0.05	0.10
0.01	$n_r = 848, n_s = 424$	$n_r = 635, n_s = 318$	$n_r = 534, n_s = 267$
0.05	$n_r = 617, n_s = 309$	$n_r = 439, n_s = 220$	$n_r = 357, n_s = 179$
0.10	$n_r = 510, n_s = 255$	$n_r = 350, n_s = 175$	$n_r = 278, n_s = 139$
0.20	$n_r = 394, n_s = 197$	$n_r = 257, n_s = 129$	$n_r = 197, n_s = 99$

and $\beta=0.20$ were initially assumed for Type I and Type II error levels. Sample size estimates for these error levels are read from the middle cell/bottom row of tables A2–A4. For equal sampling of reference and study conditions, estimated $n_r=n_s=177$. If reference and study conditions are sampled unequally, then alternative sample sizes to consider are (1) $n_r=257$ and $n_s=129$, or (2) $n_r=336$ and $n_s=112$. If it were more appropriate to assume equal risk for Type I and Type II errors, then estimated sample sizes would be read from the diagonal cells of Tables A2–A4. For instance, assuming $\alpha=\beta=0.05$, alternative sample sizes to consider are (1) $n_r=n_s=293$, (2) $n_r=439$ and $n_s=220$, or (3) $n_r=583$ and $n_s=191$.

Table A4.—Unequal sample size estimates ($f = 1/3$) for $p_r = 0.10$ and $p_s = 0.20$.

β	α		
	0.01	0.05	0.10
0.01	$n_r = 1127, n_s = 376$	$n_r = 849, n_s = 283$	$n_r = 718, n_s = 240$
0.05	$n_r = 813, n_s = 271$	$n_r = 583, n_s = 191$	$n_r = 476, n_s = 159$
0.10	$n_r = 668, n_s = 223$	$n_r = 462, n_s = 154$	$n_r = 368, n_s = 123$
0.20	$n_r = 513, n_s = 171$	$n_r = 336, n_s = 112$	$n_r = 259, n_s = 87$

Acknowledgments

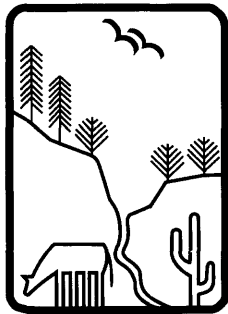
Corky Ohlander motivated us to pursue our pebble count procedure. Corky's exhaustive review and knowledge of the Clean Water Act gave us a firm foundation for "going to the stream." We also greatly appreciate the pebble-counting efforts of Mark Hight, Robert Rush, and Brent Jenkins, and the hospitality of horse packer Gary Carver. Working in remote locations occupied by grizzly bears can be unnerving. Without the critical manuscript reviews offered by Pete Hawkins, Corky Ohlander, John Potyondy, Lee MacDonald, Larry Schmidt, Charles Troendle, and David Turner our document would not be past the starting gate. It is much improved because of their suggestions. Donna Sullenger drew the zig-zag procedure illustration.

The United States Department of Agriculture (USDA) prohibits discrimination in its programs on the basis of race, color, national origin, sex, religion, age, disability, political beliefs and marital or familial status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (braille, large print, audiotape, etc.) should contact the USDA Office of Communications at (202) 720-5881 (voice) or (202) 720-7808 (TDD).

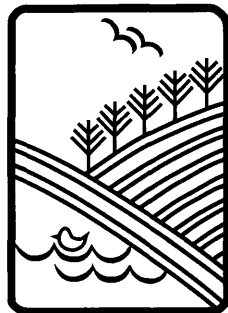
To file a complaint, write the Secretary of Agriculture, U.S. Department of Agriculture, Washington, DC 20250, or call (202) 720-7327 (voice) or (202) 720-1127 (TDD). USDA is an equal employment opportunity employer.



Rocky
Mountains



Southwest



Great
Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526