



A COMPARISON OF PROTOCOLS AND OBSERVER PRECISION FOR MEASURING PHYSICAL STREAM ATTRIBUTES¹

Heath W. Whitacre, Brett B. Roper, and Jeffrey L. Kershner²

ABSTRACT: Stream monitoring programs commonly measure physical attributes to assess the effect of land management on stream habitat. Variability associated with the measurement of these attributes has been linked to a number of factors, but few studies have evaluated variability due to differences in protocols. We compared six protocols, five used by the U.S. Department of Agriculture Forest Service and one by the U.S. Environmental Protection Agency, on six streams in Oregon and Idaho to determine whether differences in protocol affect values for 10 physical stream attributes. Results from Oregon and Idaho were combined for groups participating in both states, with significant differences in attribute means for 9 out of the 10 stream attributes. Significant differences occurred in 5 of 10 in Idaho, and 10 of 10 in Oregon. Coefficients of variation, signal-to-noise ratio, and root mean square error were used to evaluate measurement precision. There were differences among protocols for all attributes when states were analyzed separately and as a combined dataset. Measurement differences were influenced by choice of instruments, measurement method, measurement location, attribute definitions, and training approach. Comparison of data gathered by observers using different protocols will be difficult unless a core set of protocols for commonly measured stream attributes can be standardized among monitoring programs.

(KEY TERMS: sampling protocols; aquatic habitat; attribute measurement; precision; stream monitoring; quality control.)

Whitacre, Heath W., Brett B. Roper, and Jeffrey L. Kershner, 2007. A Comparison of Protocols and Observer Precision for Measuring Physical Stream Attributes. *Journal of the American Water Resources Association* (JAWRA) 43(4):923-937. DOI: 10.1111/j.1752-1688.2007.00074.x

INTRODUCTION

Aquatic monitoring programs have been developed throughout the country to quantify the status and trends in physical instream fish habitat and channel condition and to evaluate the response to resource

management activities (Conquest *et al.*, 1994; Larsen *et al.*, 2001). Information from these programs has commonly been used to justify changes in resource management, restore degraded resources, and determine compliance with laws and regulations (Platts *et al.*, 1987; USEPA, 1991; MacDonald, 1994). Habitat monitoring programs have become the basis

¹Paper No. J05093 of the *Journal of the American Water Resources Association* (JAWRA). Received July 8, 2005; accepted October 26, 2006. © 2007 American Water Resources Association. No claim to original U.S. government works. **Discussions are open until February 1, 2008.**

²Respectively, Hydrologist, Aquatic, Watershed, and Earth Resources Department, Utah State University, 5210 University Hill, Logan, Utah 84322-5210 [Whitacre now at U.S. Forest Service, Tongass National Forest, Petersburg Ranger District, 12 Nordic Dr. Petersburg, Alaska 99833]; Aquatic Monitoring Center Program Leader, Fish and Aquatic Ecology Unit, U.S. Forest Service, Logan Forestry Science Laboratory, 860 North 1200 East, Logan, Utah 84321; and Center Director, Fish and Aquatic Ecology Unit, U.S. Forest Service, Aquatic, Watershed, and Earth Resources Department, Utah State University, 5210 University Hill, Logan, Utah 84322-5210 [Kershner now at U.S. Geological Survey, Northern Rocky Mountain Science Center, 211 AJM Johnson Hall, Montana State University, Bozeman, Montana 59717-3492] (E-Mail/Whitacre: hwhitacre@fs.fed.us).

of many aquatic impact assessments, resource inventories, species management plans, mitigation planning, and environmental regulation (Bain and Stevenson, 1999; Bain *et al.*, 1999).

Since the advent of the Federal Water Quality Act in 1965 (PL 89-234), scientists and land managers have attempted to develop and refine parameters and techniques to monitor trends in water quality and instream habitat. This process has helped distill the key steps to successful monitoring plans and improved understanding of the sources of variability within monitoring designs (Larsen *et al.*, 2001; Kershner *et al.*, 2004). Variability can arise from the sampling design: the process of specifying how and where to select the population units of interest, and the response design: the process of deciding what and how to measure (Conquest *et al.*, 1994; Urquhart *et al.*, 1998; Kaufmann *et al.*, 1999; Stevens and Urquhart, 2000; Larsen *et al.*, 2001). Variability in the response design can affect the ability to draw conclusions relative to the habitat variable being measured (Roper *et al.*, 2002). Proper choice of habitat variables (MacDonald *et al.*, 1991; Kaufmann *et al.*, 1999; Roper *et al.*, 2002), choice of measurement instruments (Isaak *et al.*, 1999; Ryan and Porth, 1999), and observer bias (Hogle *et al.*, 1993; Marcus *et al.*, 1995; Roper and Scarnecchia, 1995; Wang *et al.*, 1996; Wohl *et al.*, 1996; Poole *et al.*, 1997; Bunte and Abt, 2001) are all components of the response design that have been evaluated. However, few studies have attempted to quantify variability associated with the use of different protocols for measurement of the same habitat variables (a notable exception are protocols used to measure streambed substrate; see Wolman, 1954; Kondolf and Li, 1992; Wohl *et al.*, 1996; Bunte and Abt, 2001).

The objective of this article was to evaluate whether the use of different protocols leads to differences in reported values for commonly measured physical stream attributes. This study compares mean values, measurement precision, and sources of variability among different protocols used by the U.S. Forest Service (USFS) and U.S. Environmental Protection Agency (USEPA). Measurement precision in this study is defined as the ability of crews using the same protocol to produce the same value for a particular stream attribute. Accuracy in the context of stream habitat assessment is the conformity of a measured attribute value to the physical reality of the attribute. Accuracy has been assessed in previous studies by assuming values measured by precision instruments represent the "true" value (Isaak *et al.*, 1999), by comparing estimates of substrate composition with values obtained from digital photographs (Wang *et al.*, 1996), and with high sample sizes producing a strong "signal" for the attribute of interest (Kaufmann *et al.*, 1999). Since collecting a "true" value for all attributes

on each stream was beyond the scope of this study, accuracy was not evaluated. Nonetheless, documenting differences in means and precision among protocols used for stream habitat data collection is critical for determining whether the results among stream monitoring programs can be compared and how they should be used to inform management decisions.

METHODS

Agency Monitoring Protocols

Five stream habitat monitoring protocols currently used by the USFS and one used by the USEPA were evaluated in this study. The U.S. Forest Service's Aquatic and Riparian Effectiveness Monitoring Program (AREMP) is a large-scale multi-federal agency program developed to monitor aquatic and riparian ecosystems on federal lands in the Pacific Northwest managed under the Northwest Forest Plan (Reeves *et al.*, 2004). The PACFISH/INFISH effectiveness monitoring program (PIBO) was developed to respond to monitoring needs identified in the biological opinions for bull trout and steelhead (Kershner *et al.*, 2004). PIBO conducts large-scale monitoring of aquatic and riparian resources on Forest Service and Bureau of Land Management Lands within the Interior Columbia River Basin. Forest Service monitoring groups from the Pacific Northwest Region (R6) (USDA, 2000), the Intermountain and Rocky Mountain Regions (R1/R4) (Overton *et al.*, 1997), and the Alaska Region (R10) (USDA, 2001), also participated. All three of the regional Forest Service protocols are habitat-based and derived from the basin-wide fisheries assessment developed by Hankin and Reeves (1988). The USEPA's Environmental Monitoring and Assessment Program (EMAP) protocols were developed for use across the United States and are intended to be applied by state agencies (Kaufmann *et al.*, 1999). Crews from the Department of Environmental Quality (DEQ) of both Oregon and Idaho used the EMAP protocols for this study.

Study Design

We evaluated three stream reaches located in the Boise National Forest in Idaho and three stream reaches on the west side of the Cascade Mountains in the Mount Hood National Forest in Oregon. Stream surveys were conducted in the summer of 2002 during the low-flow sampling season from July 3 to July 25 in Idaho, and July 28 to August 20 in

Oregon. Stream stage was recorded to ensure data collected was under comparable conditions. Streams were chosen in both states to avoid bias caused by regional differences in geology and climate and to reflect a range of physical characteristics found under normal sampling conditions in each region (Table 1). Stream reaches were chosen to have relative uniformity in gradient, habitat width and depth, and to avoid tributary junctions. Two of the six streams surveyed, one in Idaho and one in Oregon were sand-bed channels while the rest were gravel-bed streams. The lower boundary of each sample reach was flagged to indicate the starting point for each field crew.

TABLE 1. General Characteristics of Streams in This Study.

State	Stream	Geology	Elevation (m)	Gradient (%)	Bankfull Width (m)
Oregon	Linney	Volcanic	802	1.62	10.47
Oregon	Oak Grove	Volcanic	988	0.95	10.22
Oregon	Still	Volcanic	1088	1.93	14.37
Idaho	Anderson	Granitic	1094	1.21	8.02
Idaho	Pine	Granitic	1155	3.23	4.96
Idaho	Tripod	Granitic	1553	0.88	1.69

Note: Values determined as averages from all protocols.

Four monitoring groups were involved in the Idaho sample—AREMP, PIBO, EMAP and R1/R4. Both AREMP and PIBO provided three crews while EMAP and R1/R4 provided two crews. In Oregon five monitoring groups were involved—AREMP, EMAP, PIBO, R10, and R6. Each of these five groups provided three crews. A crew was made up of two or three crew members and each crew made observations independent of other crews.

All crews used the same reach starting point. Reach length was determined differently by each monitoring group and was dependent on individual protocol instructions. For example, reach length for AREMP and PIBO crews was a length 20 times bankfull width or a distance not less than 150 m for AREMP and 80 m for PIBO. The reach length for EMAP protocols was 40 times wetted width but not less than 150 m. Since R10, R6 and R1/R4 crews typically survey longer reaches (100s to 1000s m) we flagged both the downstream and upstream ends of the reaches to be surveyed. We chose lengths for these groups based on relative homogeneity of the entire reach and to ensure the distance could be surveyed in a day; depending upon the stream these distances ranged between approximately 500 and 1000 m. In both states, crews from each monitoring group had 3 days to complete all streams (one crew per stream per day) before the next monitoring group arrived. Groups were assigned a 3-day sampling slot on a first come, first served basis within logistics and

scheduling constraints. All crews measured stream attributes following protocols for their specific groups.

Analysis

We summarized data at the reach level for each crew for 10 stream attributes common to many of the groups. Attribute means for each crew using the same protocol were averaged for an overall protocol mean. The attributes included in the analysis were reach length (m), gradient (%), sinuosity, percent pools, residual pool depth (m), bankfull width (m), bankfull width-to-depth ratio, median particle size (D_{50}) (mm), percent fines (< 6 mm), and large woody debris (LWD) pieces (100/m) (Table 2). For detailed protocols see Kaufmann *et al.* (1999) (EMAP); USDA (2000) (R6); USDA (2001) (R10); Overton *et al.* (1997) (R1/R4); Gallo (2002) (AREMP); and Henderson *et al.* (2002) (PIBO).

A two-way analysis of variance (ANOVA) using stream and monitoring group as the main effects was used to evaluate differences in means among the monitoring groups for each of the 10 attributes. Visits by different crews of the same monitoring group to the same stream served as replicates for this study. When significant differences among monitoring groups were found for a specific attribute we made pairwise comparisons using a Tukey adjustment ($\alpha = 0.05$). Because we had different monitoring groups as well as different crews (i.e., for EMAP we had Oregon DEQ in Oregon and Idaho DEQ in Idaho) in the two states, we conducted three different analyses: (1) a combined analysis for the three monitoring groups, AREMP, EMAP, and PIBO, which participated in evaluating all six streams in both states; (2) an analysis of the five groups in the Oregon streams; and (3) an analysis of the four groups in the Idaho streams.

In order to estimate both the signal and noise associated with measuring the habitat variables, we partitioned variance associated with streams and crews using a random effects analysis of variance model, with streams treated as a random effect (Littell *et al.*, 1996). Separate analyses of attributes were run for each of the monitoring groups within each state. Parametric statistical methods were used for our analysis because inspection of the residuals provided little evidence that the error distribution of different crews evaluating the same attribute within the same stream were not normally distributed and estimates of variances are difficult to interpret when not using raw data. In estimating variance, all error in the model not associated with variation among streams was attributed to crews (Kaufmann *et al.*, 1999; Roper *et al.*, 2002). Estimates of means and variances were also used to calculate coefficients of variation

TABLE 2. Comparisons of Protocols Used by the Six Monitoring Groups.

Reach length – total length surveyed by the crews	
AREMP	20 times bankfull width, minimum length 150 m, nearest 0.1 m
EMAP	40 times wetted width, minimum length 150 m, nearest 0.1 m
PIBO	20 times bankfull width class (<4,6,8...) minimum length 80 m, nearest 2.0 m
R10	Fixed length, each habitat unit measured then distances summed, nearest 0.1 m
R6	Fixed length, each habitat unit measured then distances summed, nearest 0.1 m
R1/R4	Fixed length, each habitat unit measured then distances summed, nearest 0.1 m
Gradient	
AREMP	Laser rangefinder, line of site, single evaluation, nearest 0.01%
EMAP	Clinometers, average slope of 11 equally spaced transects within the reach, nearest 0.01%
PIBO	Surveyors level, line of site, average of two whole reach measurements; if they differ by more than 10% a third estimate is made and averaged, nearest 0.01%
R10	Surveyors level, line of site, one measurement 20 channel widths long or one distinct channel feature to another similar feature over shorter distance
R6	Determined from USGS 1:24,000 topographic map, nearest 0.01%
R1/R4	Hand level, representative segment (200 m long), nearest 0.01 m
Sinuosity	
AREMP	Distance along thalweg and between top to bottom of reach determined by laser (0.1 m)
EMAP	Sum distances between 11 transects/measured straight line top to bottom of reach (0.1 m)
PIBO	Measured stream length at thalweg/measured straight line top to bottom of reach (0.1 m)
R10	Not determined for this study
R6	Not determined for this study
R1/R4	Not determined for this study
Percent pool- percent reach length in pool habitat	
AREMP	Pools have slower flow, reduced turbulence, and zone of scour. Must be longer than wide
EMAP	Pools have still water, low velocity, glassy surface, and deep compared to other parts of channel. Must be as long as channel is wide
PIBO	Pools must be bounded by pool head and tail crest, 1.5 times deeper than pool tail crest, within main channel. Must be longer than wide
R10	Pools must have noticeable change in bed elevation caused by pool forming agent, macro pools meet residual depth requirements. Length or width $\geq 10\%$ of the channel bed width
R6	Pools have slow flow, deeper than riffles, bowl or tub appearance, have hydrologic control. Longer than wetted width
R1/R4	Pools have slow flow, deeper than riffles, have hydrologic control
Residual pool depth (using the pools as defined above)	
AREMP	Not determined for this study
EMAP	Graphically represented residual surface along a longitudinal depth profile, with corrections made for reach slope
PIBO	Average for all pools in reach of maximum depth minus depth at pool tail crest
R10	Average for all pools in reach of maximum depth minus depth at pool tail crest
R6	Average for all pools in reach of maximum depth minus depth at pool tail crest
R1/R4	Average for all pools in reach of maximum depth minus depth at pool tail crest
Median particle size – determined from the following collection approach	
AREMP	10 pebbles collected at systematic intervals at 11 bankfull transects. Measured 1 mm
EMAP 5	Pebbles collected at systematic intervals at 21 transects (wetted). Visually estimated and placed into generic size classes (fines, sand, fine gravel, coarse gravel, cobbles, boulders)
PIBO 25	Pebbles collected systematically in first four riffles (bankfull). Measured –1 mm
R10	Collected from five transects in at least one representative riffle (bankfull), 100 pebbles. Measured (gravel template), $\frac{1}{2}$ phi classes
R6	Collected from five transects in at least one representative riffle, 100 pebbles (wetted). Measured (gravel template), $\frac{1}{2}$ phi classes
R1/R4	Not determined for this study
Percent fines	
AREMP	Determined from data used for median particle size – percent ≤ 6 mm
EMAP	Determined from data used for median particle size – percent ≤ 6 mm
PIBO	Determined from data used for median particle size – percent ≤ 6 mm
R10	Determined from data used for median particle size – percent ≤ 6 mm
R6	Determined from data used for median particle size – percent ≤ 6 mm
R1/R4	Determined in pool tail out. Number of particles ≤ 6 mm on a (size) grid with 49 intersections

TABLE 2. (Continued)

Bankfull width – channel width at bankfull discharge	
AREMP	Unconstrained reaches; average 11 equally spaced transects. Constrained reaches; average six equally spaced transects (0.01 m)
EMAP	Average 11 equally spaced transects (0.01 m)
PIBO	Average 20 equally spaced transects (0.01 m)
R10	One representative transect (0.01 m)
R6	One representative transect (0.01 m)
R1/R4	Not determined for this study
Bankfull width-to-depth ratio	
AREMP	Bankfull width divided by average depth (10 systematic depths per transect). Average of 11 transects
EMAP	Not determined for this study
PIBO	Bankfull width at divided by average depth (10 systematic depths per transect). Average of four transects evaluated in riffles
R10	Bankfull width divided by cross-sectional area. One representative transect
R6	Bankfull width divided by cross-sectional area at systematic riffles
R1/R4	Not determined for this study
Large woody debris (number of pieces per 100 m)	
AREMP	Minimum size; 0.3 m diameter × 3 m long. Count if in, partially in, or above bankfull channel
EMAP	Minimum size; 0.1 m diameter × 1.5 m long. Count if in, partially in, or above bankfull channel
PIBO	Minimum size; 0.1 m diameter × 1 m long or 2/3 wetted width. Count if in, partially in, or above bankfull channel
R10	Minimum size; 0.1 m diameter × 1 m long. Counted if in bankfull channel
R6	Minimum size; 0.3 m diameter × 3 m long, or 2 × average bankfull width. Count if part of the wood interacts with stream channel at bankfull flow
R1/R4	Minimum size; 0.1 m diameter × 3 m long or 2/3 wetted width. Count if in, partially in, or above bankfull channel

This table focuses on describing the differences among the groups. For more specifics related to protocols see each of the groups published monitoring protocols.

[CV = (crew variance)^{0.5}/mean*100], signal-to-noise ratios (S:N=stream variance/crew variance) and root mean square errors [RMSE = (crew variance)^{0.5}].

Coefficients of variation, S:N, and RMSE were used to compare measurement precision of each crew as well as provide estimates of overall protocol precision. CV provide a dimensionless measure of variability in which scaling is relative to the mean, with values ≤ 20 suggested for acceptable measurement precision (Ramsey *et al.*, 1992). Signal-to-noise ratio provides an estimate of precision relative to the inherent variation among streams, with values from 2-10 reflecting moderate to high precision (Kaufmann *et al.*, 1999). Values of S:N reflect the ability of a protocol to distinguish differences among streams for the particular attribute measured. RMSE represents the pooled standard deviation among crews using the same protocol. Lower RMSE for an attribute suggests the attribute is more consistently evaluated by the different crews, especially when the measurements come from the same set of streams.

RESULTS

Combined Sites

In our analysis which combined Idaho and Oregon data, 10 attributes—reach length, gradient, sinuosity,

percent pools, residual pool depth, bankfull width, bankfull width-to-depth ratio, D₅₀, percent fines, and LWD were evaluated by at least two of the monitoring groups. Insufficient data were collected by AREMP and EMAP crews for residual pool depth comparisons on Idaho streams; therefore this attribute was not included in the combined site analysis. Significant differences among protocols (*p* ≤ 0.05) occurred in 9 of the 10 attributes compared (Table 3).

Measurement precision varied among crews using different protocols for each stream attribute. Reach length, gradient, sinuosity, and percent fines were measured with relatively high precision (CV ≤ 20) by all crews. Precision among protocols varied for

TABLE 3. Summary of Statistical Analysis of Attribute Means, Described as Being the “Same” or “Different.”

Attribute	Combined Data	Idaho	Oregon
Reach length (m)	Different	Different	Different
Gradient (%)	Different	Same	Different
Sinuosity	Different	Different	Different
Percent pools	Different	Different	Different
Residual pool depth (m)	Different	Same	Different
Bankfull width (m)	Different	Same	Different
BF width:depth ratio	Same	Same	Different
D ₅₀ (mm)	Different	Different	Different
Percent fines (< 6 mm)	Different	Same	Different
Large woody debris (100/m)	Different	Different	Different

Different = Significant difference (*p* ≤ 0.5) in attribute means between at least two protocols.

TABLE 4. Attribute Means and Precision Estimates [Root Mean Square Error (RMSE), Coefficient of Variation (CV), Signal-to-Noise (S:N)] by Protocol for the Combined Site Dataset.

Stream Attribute	Protocol	Mean	Significant Difference	RMSE	CV	S:N
Reach length (m)	AREMP	205.17	A	36.35	17.72	2.83
	EMAP	246.67	B	31.74	12.87	9.16
	PIBO	175.17	C	25.13	14.35	8.37
Gradient (%)	AREMP	1.72	A	0.4	23.39	7.02
	EMAP	2.04	B	0.42	20.82	7.27
	PIBO	1.45	C	0.14	9.69	66.62
Sinuosity ¹	AREMP	1.48	A	0.17	11.38	1.28
	PIBO	1.28	B	0.1	8.22	2.32
Percent pools	AREMP	21.58	A	12.22	56.62	0.32
	EMAP	25.16	A	11.05	43.93	1.35
	PIBO	51.7	B	8.22	15.91	6.27
Residual pool depth (m) ¹	PIBO	0.53		0.05	10.07	37.31
Bankfull width (m)	AREMP	11.36	A	4.31	37.94	1.2
	EMAP	8.72	B	3.05	34.96	1.93
	PIBO	7.81	B	0.35	10.42	30.32
BF width:depth ratio ¹	AREMP	25.9	NS	10.97	42.36	0.53
	PIBO	26.94	NS	5.14	19.07	4.01
D ₅₀ (mm)	AREMP	22.88	A	11.85	51.77	5.1
	EMAP	31.47	AB	21.06	66.9	2.39
	PIBO	36.5	B	12.85	35.2	5.13
Percent fines (<6 mm) ¹	AREMP	51.25	A	6.38	12.45	21.72
	EMAP	44.93	B	3.72	8.28	69.94
	PIBO	36.06	C	7.26	20.14	21.24
LWD (100/m)	AREMP	7.82	A	3.29	42.03	6.64
	EMAP	34.61	B	13.44	38.82	0.74
	PIBO	42.62	B	17.86	41.92	1.19

Note: Protocol means with the same letters are not significantly different ($p \leq 0.05$).

¹Protocols missing in comparisons do not typically measure these attributes, or there were insufficient data gathered for comparison.

bankfull width and percent pools suggesting some protocols are more repeatable than others (Table 4). Precision was low among all crews for measurement of D₅₀ and large woody debris (100 m⁻¹). Signal-to-noise and RMSE estimates reflect the same general precision pattern as CV, with the exception of D₅₀, which had low to moderate precision (S:N 2-5).

Idaho

There were significant differences ($p \leq 0.05$) in attribute means among protocols for 5 of 10 commonly collected stream attributes in Idaho (Table 3). Attributes with similar means among protocols included gradient, residual pool depth, bankfull width, width-to-depth ratio, and percent fines. Percent pools and LWD were the only attributes with more than one significant pairwise difference among protocols (Table 5).

Estimates of precision varied among protocols for each stream attribute measured. Reach length, gradient, sinuosity, and residual pool depth were measured with relatively high precision by all crews.

Five stream attributes – gradient, percent pools, bankfull width, percent fines, and D₅₀—exhibited a wide range of precision, indicating considerable differences in the consistency of crews among protocols (Table 5).

Oregon

All stream attributes (10 of 10) had statistical differences in means among protocols in Oregon (Table 3). Crews using the AREMP protocol had means for bankfull width that were different from all others except EMAP (Table 6).

Precision also varied among protocols for each stream attribute measured in Oregon. Most crews measured reach length, sinuosity, and mean residual pool depth with relatively high precision, while precision was low for percent pools, D₅₀, and LWD (Table 6). Three stream attributes; gradient, bankfull width, and percent fines had the same wide range of precision among protocols as occurred in Idaho (Table 6). The attributes measured with the most variability were percent fines and percent pools.

TABLE 5. Attribute Means and Precision Estimates (RMSE, CV, S:N) by Protocol in Idaho.

Stream Attribute	Protocol	Mean	Significant Difference	RMSE	CV	S:N
Reach length (m)	AREMP	153.22	A	5.24	3.42	0.8
	EMAP	190.33	B	29.9	15.71	2.61
	PIBO	166.22	AB	15.35	12.72	0.79
	R1/R4	520.92		38.14	7.32	0.81
Gradient (%)	AREMP	1.97	NS	0.24	12.14	45.44
	EMAP	2	NS	0.16	8.16	117.81
	PIBO	1.59	NS	0.09	5.4	412.26
	R1/R4	1.62	NS	0.29	17.78	0
Sinuosity ¹	AREMP	1.55	A	0.17	11.09	1.99
	PIBO	1.38	B	0.11	7.85	3.21
Percent pools	AREMP	26.54	A	11.17	42.1	0.72
	EMAP	30.61	A	12.71	41.52	2.38
	PIBO	59.51	B	4.81	8.08	29.9
	R1/R4	26.5	A	16.65	62.82	0
Residual pool depth (m) ¹	PIBO	0.35	NS	0.04	11.36	22.89
	R1/R4	0.35	NS	0.02	5.38	67.33
Bankfull width (m) ¹	AREMP	6.83	NS	2.86	41.93	0
	EMAP	5.3	NS	1.73	32.66	3.66
	PIBO	4.37	NS	0.36	8.22	61.23
BF width:depth ratio ¹	AREMP	17.42	NS	6.72	38.59	0.01
	PIBO	21.11	NS	4.77	22.61	3.91
D ₅₀ (mm) ¹	AREMP	13.58	A	6.85	50.44	7.39
	EMAP	16.01	AB	1.04	6.47	332.71
	PIBO	20.56	B	7.13	34.7	9.72
Percent fines (<6 mm)	AREMP	63.24	NS	2.48	3.92	170.59
	EMAP	60.86	NS	1.53	2.51	487.63
	PIBO	54	NS	4.07	7.53	100.06
	R1/R4	59.27	NS	37.93	64	0.24
LWD (100/m)	AREMP	1.51	A	0.51	33.69	6.88
	EMAP	28.82	B	9.02	31.32	2.82
	PIBO	31.42	B	11.6	36.92	2.81
	R1/R4	9.82	A	8.62	87.82	0

Note: Protocol means with the same letters are not significantly different ($p \leq 0.05$).

¹Protocols missing in comparisons do not typically measure these attributes, or there were insufficient data gathered for comparison.

DISCUSSION

Variability associated with field measurement of physical stream attributes has been linked to a number of factors including habitat complexity, instrumentation, inconsistent training, and inconsistent application of protocols among observers (Ralph *et al.*, 1994; Hannaford and Resh, 1995; Roper and Scarnecchia, 1995; Wang *et al.*, 1996; Hannaford *et al.*, 1997; Kondolf, 1997; Isaak *et al.*, 1999; Ryan and Porth, 1999). This study indicates that the variability can also be the result of specific protocol used to measure an attribute. Group differences were due to differences in how and where stream attributes were measured, how they defined the attribute, and training approach.

The stream attributes compared in this study were chosen because they are commonly measured by most monitoring groups, although they are often defined differently. Attributes, such as percent pools and LWD best illustrate how different definitions can

influence reported values and precision. We discuss the influence of how and where stream attributes are measured on mean values and measurement precision followed by the differences which cross-cut all protocols, such as attribute definitions and training approaches.

Gradient

Differences in instrumentation, the number of measurements, and sampling location resulted in variable estimates of gradient among protocols. Although gradient was measured by most crews with high precision, mean values were significantly different among protocols in both states; varying up to 64% in Oregon (Tables 5 and 6). Differences in this magnitude are of concern to land managers due to the common use of gradient for determining channel classification and the subsequent management direction (USDA, 1997). Instruments used to evaluate

TABLE 6. Attribute Means and Precision Estimates (RMSE, CV, S:N) by Protocol in Oregon.

Stream Attribute	Protocol	Mean	Significant Difference	RMSE	CV	S:N
Reach length (m)	AREMP	257.11	AB	51.13	19.89	0.55
	EMAP	302.22	A	32.66	10.81	10.72
	PIBO	229.69	B	31.85	13.87	1.33
	R10	800.76		22.41	2.8	NA
	R6	856.01		102.43	11.97	NA
Gradient (%) ¹	AREMP	1.48	A	0.52	35.02	0.38
	EMAP	2.02	B	0.48	23.78	1.27
	PIBO	1.3	AC	0.18	13.71	5.44
	R10	1.23	AC	0.39	31.78	1.37
Sinuosity ¹	AREMP	1.41	A	0.16	11.69	0.75
	EMAP	1.26	AB	0.1	7.67	5.18
	PIBO	1.17	B	0.09	7.65	0
Percent pools	AREMP	16.63	A	12.62	75.89	0
	EMAP	19.47	A	9.44	48.47	0
	PIBO	43.89	B	10.59	24.13	1.75
	R10	50.01	B	20.94	41.86	0.1
	R6	49.16	B	22.37	45.5	0.35
Residual pool depth (m) ¹	EMAP	0.19	A	0.05	23.14	0.69
	PIBO	0.7	B	0.06	9.04	33.16
	R10	0.59	B	0.06	9.81	0.58
	R6	0.7	B	0.15	22.09	0
Bankfull width (m)	AREMP	15.88	A	5.18	32.6	0
	EMAP	12.28	AB	3.52	28.68	0.15
	PIBO	11.26	B	0.35	9.71	8.67
	R10	11.54	B	3.39	29.35	0.57
	R6	8.5	B	0.76	8.98	18.93
BF width:depth ratio ¹	AREMP	34.39	A	12.98	37.74	0
	PIBO	32.78	A	5.9	16.71	2.75
	R10	24.45	AB	6.41	26.23	1.51
	R6	19.97	B	5.12	25.63	0.55
D ₅₀ (mm)	AREMP	32.19	A	15.29	47.51	5.16
	EMAP	46.23	A	25.89	56.01	2.6
	PIBO	52.44	AB	16.71	31.87	3.18
	R10	49.75	AB	11.79	23.69	3.11
	R6	69.5	B	9.15	13.17	35.16
Percent fines (<6 mm)	AREMP	39.26	A	8.67	22.09	9.79
	EMAP	29.25	AB	4.42	15.13	27.34
	PIBO	18.11	C	9.43	52.06	2.09
	R10	10.93	C	7.3	66.73	3.49
	R6	19.05	BC	6.32	33.16	7.24
LWD (100/m)	AREMP	14.12	A	4.62	32.71	2.81
	EMAP	41.14	B	15.11	36.72	0.33
	PIBO	53.82	B	22.44	41.7	0.5
	R10	49.94	B	27.26	54.6	0.79
	R6	4.95	A	0.92	18.69	13.87

Note: Protocol means with the same letters are not significantly different ($p \leq 0.05$).

¹Protocols missing in comparisons do not typically measure these attributes, or there were insufficient data gathered for comparison.

gradient in this study included laser rangefinders (AREMP), surveyor's levels (PIBO and R10), hand-held clinometers (EMAP), and hand levels (R1/R4). Isaak *et al.* (1999) found clinometers and hand levels had low precision when compared with other common instruments used to measure gradient, yet EMAP, which uses clinometers, reported gradient with higher precision than AREMP in both states and R10 in Oregon (Tables 5 and 6). Higher EMAP precision despite the use of an imprecise instrument was influ-

enced by protocol differences in how gradient is measured. The AREMP and R10 crews use a single observation to estimate gradient while EMAP crews estimate gradient between 11 transect locations and average these values. Averaging values takes advantage of the central limit theorem as well as minimizing the effects of small errors in each of the individual measurements. In contrast, a single measurement is not able to account for errors made by the observer and any errors will directly affect any

estimate of precision. Measurement location may have also influenced R10 and R1/R4 estimates of gradient since crews using these protocols often measure gradient through a particular habitat type in a location deemed representative of the overall reach (Table 2). Repeatability among crews using this approach can be difficult because gradient estimates are likely to vary when crews choose different locations. Gradient was measured with the highest precision by crews using the PIBO protocol (Table 6). This protocol combines a precision instrument with redundancy; requiring repeated measurements of gradient throughout the entire reach. If the values are not within 10% of one another, a third measurement is taken and the values averaged (Table 2).

Gradient measurement precision may be improved by using precision instrumentation, increasing the number of measurements, and choosing the same location for measurement. Additionally, measuring the elevation of a known benchmark at the beginning and end of a gradient survey would ensure data are within a specified tolerance, increasing both accuracy and repeatability among crews (Harrelson *et al.*, 1994).

Stream Substrate

Monitoring the size and composition of stream substrate through attributes, such as D_{50} and percent fines is a common way for water resource managers to assess water and habitat quality for fisheries. These attributes can also be used to characterize channel roughness, sediment budgets, habitat descriptions, and when used in conjunction with flow data can describe how flow and sediment interact to create and maintain channel form (Wohl *et al.*, 1996; Ryan and Emmett, 2002).

Many different methods were used to describe stream substrate in this study. The strongest trends for differences in D_{50} means and measurement precision emerged when considering how and where these attributes were measured. Measurement methods included the use of gravel templates (R10, R6), direct measurement of each particle (AREMP, PIBO), and visual estimation (EMAP). EMAP crews using visual estimations had the lowest measurement precision in Oregon and the combined site dataset (Tables 4 and 6). This pattern is confounded, however, by the high EMAP precision in Idaho. Additionally, EMAP had the highest measurement precision for determining percent fines in both states and the combined dataset (Tables 4-6). An important distinction of the EMAP protocol is that when particle sizes are visually estimated, they are

placed into size categories up to 4 phi sizes in interval width (Kaufmann *et al.*, 1999). Higher measurement precision for the EMAP visually estimated data may result from observers placing the data into a limited number (7) of size classes. A potential drawback of this method is the data may not be as accurate as other methods where observers measure each particle to the nearest mm or $\frac{1}{2}$ phi interval. Kaufmann *et al.* (1999) found visually-assessed substrate metrics to be reasonably precise and suggested that carefully designed visual estimates made at multiple locations within a reach can be nearly as precise as quantitative measurements. Wang *et al.* (1996) also found visual estimates of substrate composition were sufficiently accurate for many fisheries applications. Others suggest that quantitative measurement of at least 100 pebbles decreases random measurement error and is a better alternative than visual techniques in fisheries and instream flow studies (Wolman, 1954; Hey and Thorne, 1983; Kondolf and Li, 1992). The use of gravel templates or other mechanical devices has been suggested as a means to improve the accuracy and precision of pebble counts (Wohl *et al.*, 1996; Kondolf, 1997; Bunte and Abt, 2001). Using gravel templates does not add much time to a protocol taking visual estimates at multiple locations and appeared to improve D_{50} precision for R10 and R6 crews in this study (Table 6).

When comparing relationships between D_{50} and percent fines in Tables 4-6 some values may seem counterintuitive. For example, substrate data for AREMP in Table 4 shows a D_{50} of 22.88 mm while the percent fines (< 6 mm) value is 51.25%. Since fines are greater than 50%, one might assume that the D_{50} has to be less than 6 mm; and this assumption would be true if only a single stream was evaluated. Results like this occur because streams, such as Tripod Creek with streambeds made up primarily of fine material (100%, D_{50} < 1 mm; AREMP results) were averaged with streams, such as Still Creek which have streambeds with low fines (13%) and large D_{50} s (71 mm). The average of these two streams is 56.5% fines with a D_{50} of 36 mm; an even larger contradiction than presented in the table.

The multiple approaches to measuring D_{50} and percent fines illustrate the difficulty of comparing protocols that, while quantifying the same stream reach, are not intended to target the same habitat. For example, AREMP and EMAP measured substrate at equally spaced transects with no regard to whether transects were in riffle or pool habitat, and generally reported smaller values of D_{50} in both states than crews evaluating substrate only in riffles (PIBO, R10, R6) (Tables 5 and 6).

Reach Length

Within any given protocol, reach length is one of the most critical stream variables to measure with consistency. Results from this study illustrate the potential benefits and pitfalls of different measurement methods. Criteria for determining a sufficient length of stream to survey vary among monitoring programs and are often related to study objectives (Kershner *et al.*, 1992; Simonson *et al.*, 1994; Kaufmann *et al.*, 1999). To ensure representative estimates of habitat, monitoring programs use reach lengths long enough to incorporate repeating patterns of variation associated with riffle-pool sequences and meander bend morphology (Kaufmann *et al.*, 1999). Crews in this study measured reach length as 40 times the low flow wetted width (EMAP), 20 times bankfull width (AREMP, PIBO), or from predetermined start and end points (R10, R6, R1/R4). Surveyed reach lengths varied among crews using the same protocol as well as among crews using different protocols. In Oregon, reach lengths among AREMP crews differed by 128 m and 90 m on Linney and Still Creek, respectively, and 244 m among EMAP and PIBO crews on Still Creek. Distances determined by crews measuring reach lengths with fixed start and end points also differed. Reach lengths differed by 72 m among R6 crews on Oak Grove in Oregon and by 79 m among R6 and R10 crews on Linney Creek. These differences resulted from variable estimates of bankfull and wetted stream widths among crews, and the ability of successive crews to repeat distance measurements along the thalweg, center, or side of the stream channel.

Variability in reach length can influence the means and measurement precision of other habitat attributes. For example, ending a survey 10 m downstream of a log jam can result in a different mean value for LWD metrics than ending 10 m above. In this study, bankfull width estimates by AREMP and EMAP crews on Oak Grove in Oregon were influenced by differences in reach length. Approximately 220 m from the start on Oak Grove the stream channel was less confined and meandered through a low gradient meadow complex. Some crews that measured longer reach lengths had difficulty distinguishing bankfull width in this portion of the stream, with reported values ranging from 6.5 to 65.0 m. The high values inflated bankfull width means, resulting in AREMP having the highest protocol mean and lowest measurement precision among protocols (Table 6).

Variability due to differences in evaluated reach length can be reduced in a number of ways. The use of permanent markers at the beginning and end of

the stream reach, the use of Global Position System locations, map descriptions, and drawings would ensure more precise location and length of the sites. Roper *et al.* (2003) found that sites relocated to distances < 10 m from the original site required smaller sample sizes to detect a 20% change in a stream variable. As site relocation distances increased, a higher sample size was needed to detect the same amount of change.

Variability may also be reduced by increased training for measuring bankfull width and by adopting width categories for field calculation of reach length. Many crews in this study determined reach length by multiplying an estimated bankfull or wetted width by a predetermined number. This results in reach lengths ± 20 m for each 1 m difference in bankfull width estimation for AREMP crews, and ± 40 m for each 1 m difference in wetted width estimation for EMAP crews (Kaufmann *et al.*, 1999; Gallo, 2002). Alternately, PIBO crews place measured bankfull width into 2 m width categories which are then multiplied by 20 to determine minimum reach length, effectively giving crews a 2 m buffer for measuring bankfull width. While the same problems exist on the margins of these width categories, a 2 m measurement buffer decreases the opportunity for greater differences in final reach length.

Pool Metrics

Pool metrics are commonly summarized to characterize the quality of stream habitat, which is then evaluated to estimate the potential effects of management on aquatic resources (Hankin and Reeves, 1988; Kaufmann *et al.*, 1999; Buffington *et al.*, 2002). Additionally, pool metrics are often used for regional target values in channel assessments (INFS, 1995; Buffington *et al.*, 2002). In this study, percent pools was measured with a wide range in mean values and low measurement precision for all but one monitoring group (Tables 4-6).

The results of this study are similar to those of previous studies describing the difficulty of consistently quantifying pool habitat (Ralph *et al.*, 1994; Roper and Scarnecchia, 1995; Woodsmith and Buffington, 1996; Poole *et al.*, 1997; Kaufmann *et al.*, 1999; Archer *et al.*, 2004). The variable results exhibited in this and previous studies can have management consequences, particularly if different protocols are adopted between years. For example, in the case of Oak Grove in Oregon the attainment of a common forest standard of 35% pool habitat for this channel type would be dependent on whether pools were measured using a transect-based approach (AREMP and EMAP) or a habitat-unit based approach (PIBO,

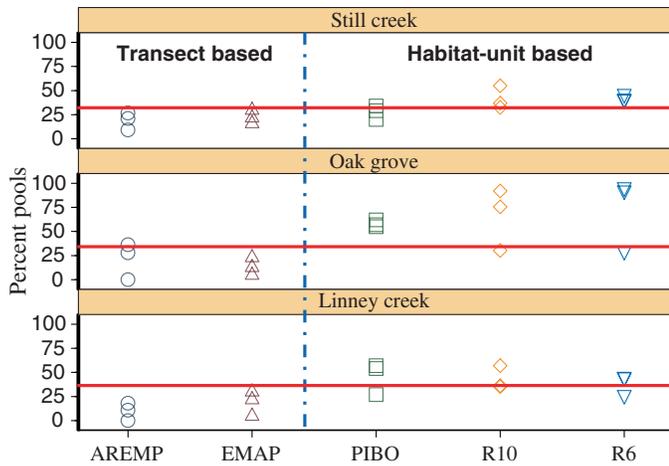


FIGURE 1. Management Implications of a Hypothetical Regional Forest Standard of 35% on Reported Percent Pool Values in Oregon.

R10, R6, R1/R4). This section of Oak Grove would meet the standard if measured by most PIBO, R10 and R6 crews, but would not if measured by AREMP and EMAP crews (Figure 1). The lack of consistency for measuring pool habitat can limit the ability to detect change, and may limit the utility of the percent pools metric for estimating the effects of management.

Cross-Cutting Issues Affecting Means and Precision

Differences in attribute definitions contributed to variable estimates in means and precision among monitoring groups. Previous studies have indicated precision and accuracy of habitat measurement is often related to how thoroughly and clearly attributes are defined in a protocol, with quantitative definitions reducing subjectivity and measurement variability (Hogle *et al.*, 1993; Ralph *et al.*, 1994; Wang *et al.*, 1996). Definitions for determining pool habitat were inconsistent among groups in this study, ranging from qualitative descriptions of channel bed form and flow patterns, to quantitative measurements, such as determining if the pool occupies greater than half the wetted channel, if length is greater than width, and if maximum pool depth is at least 1.5 times the pool tail depth (Table 2). Most groups had varying degrees of both, but the PIBO protocol had the most specific pool definition and was the only protocol with high measurement precision (Tables 4-6). Most crews in this study measured maximum pool depth and pool tail depth for calculations of residual pool depth. Residual pool depth is typically calculated as the difference between the two values and defined as the hypothetical depth of the pool if water ceases to flow

and the stream becomes a series of standing, disconnected pools (USDA, 2001). Measurements for this habitat variable are generally much more precise than those used for determining percent pools (Tables 4-6). Assessment of pool habitat through use of qualitative descriptors, in combination with quantitative measurements as described above, would improve repeatability among observers.

Different size categories for LWD pieces counted in a survey affected mean count estimates and variability among protocols. Large wood is quantified in most stream surveys because it plays an important role in physical and biological stream processes by influencing channel width and meander patterns, trapping organic matter, providing storage for sediment and bed load, and forming pools used by fish and aquatic insects for cover (Bilby, 1984; Ralph *et al.*, 1994; Beechie and Sibley, 1997). Minimum length and width requirements of wood are influenced by differences in region, ecosystem, and predominant tree species. AREMP and R6 sampling effort is focused in the Pacific Northwest, therefore these protocols have larger minimum size criteria than protocols designed for sampling in drier regions with fewer and smaller trees. All protocols in this study except AREMP and R6 used different size categories to evaluate LWD, resulting in count differences in both states (Tables 5 and 6). Counts of LWD vary by an order of magnitude among some protocols as a result of size category differences, with values for AREMP and R6 significantly lower than EMAP, PIBO, and R10. Count differences between AREMP and R6 crews are further exacerbated by how they determine which pieces are to be counted. In contrast to all other protocols in the study, R6 only includes trees that come in contact with water at bankfull discharge and not trees spanning the channel above bankfull discharge, while others include trees that span the channel whether or not they come in contact with water at bankfull discharge. These differences led to R6 having the lowest LWD counts among protocols in Oregon (Table 6). The adoption of common, overlapping length and width size categories, in addition to reconciliation of differences in philosophy concerning which pieces to measure (e.g., estimation or measurement in log jams, inclusion of pieces spanning the bankfull channel) is needed to improve comparison potential of LWD metrics among monitoring programs and subsequently its utility as a long-term monitoring tool.

Similarly, fundamental differences in philosophy concerning how and where substrate was measured impacted values of D₅₀ and percent fines (< 6 mm) in this study. The decision to use gravel templates, visual estimates, or measuring each piece with a ruler can affect precision, accuracy, time spent on

measurement, and calculation method. For example, D_{50} values for protocols using gravel templates or visual estimates are commonly interpolated using cumulative frequency distribution curves, whereas percentile functions in Microsoft Excel are often used for particles measured individually. EMAP calculates D_{50} by assigning particles in each size class an integer value from 6 (bedrock, concrete, hardpan) to 1 (clay/silt), then uses one of two equations based on particle size to calculate the \log_{10} of the geometric mean (Kaufmann *et al.*, 1999). Locations for measuring stream particles included within the bankfull channel, within the wetted channel, and between the left and right streambed; either along habitat specific (riffle) transects or systematic transects placed in both pools and riffles, depending on the protocol. Because substrate size can be expected to vary according to where particles are measured, and calculations are made depending on how they are measured, these differences in philosophy must be reconciled before substrate data can be shared among monitoring groups.

Training and experience have been shown to influence accuracy and precision of habitat measurement, as well as being an important component of monitoring programs with high personnel turnover (Hogle *et al.*, 1993; Roper and Scarnecchia, 1995; Penrose and Call, 1995; Wang *et al.*, 1996; Wohl *et al.*, 1996; Hannaford *et al.*, 1997; Thorne *et al.*, 2002; for an exception see Smith, 1944). In this study, training and experience were assessed for all crews via questionnaire (Table 7). Results indicated that measurement precision was likely linked to the amount of time spent training new employees and not necessarily to overall experience. For example, PIBO crews received the most training and also had the highest overall measurement precision in both states, even though observer experience was relatively low. There were exceptions however, as illustrated by bankfull width measurement among crews in Oregon. This metric was most precisely measured by R6 crews with the most experience, although PIBO crews with little experience but extensive training also measured with precision (Tables 6 and 7).

TABLE 7. Average Cumulative Stream Assessment Experience and Training Level of Crews Using the Six Protocols.

Protocol	Number of Crews	Assessment Experience (months)	Field Experience With Protocol (months)	Typical Training For New Field Techs (days)
PIBO	6	8	3	10
EMAP	5	23	5	5
AREMP	6	6	3	5
R10	3	145	13	4
R6	3	36	34	4
R1/R4	2	8	5	3

Many crews had difficulty in distinguishing bankfull width on particular streams, perhaps resulting from a lack of training in a variety of channel types. Another recent study described considerable observer variation for determining bankfull elevation in floodplain channels (Woodsmith *et al.*, 2005). Hannaford *et al.* (1997) found that training on one habitat type did not necessarily prepare observers to assess the same attribute in another. Many crews had difficulty recognizing bankfull indicators on the low gradient meandering channels of Tripod Creek in Idaho and Oak Grove in Oregon. Exit surveys of all crews indicated that many had not received adequate training on these types of channels.

Although difficult to quantify, quality of training undoubtedly impacted overall measurement precision. Differences in training quality among monitoring groups were evident through communication with crews and participation in several training sessions. Quality ranged from daily and weekly training camps with multiple experienced instructors to newly trained employees responsible for conducting 2 days of on-the-job training for field partners. Some programs in this study conduct multiple quality assurance tests on practice stream reaches prior to collecting field data while others begin collecting data from the start. Differences in approach to training reflect priorities related to sampling objectives, budget, and quality of data collected. Qualified instructors, consistency among instructors, exposure to a variety of habitat types, and commitment to quality assurance testing are all components improving the collection of stream habitat data.

The development of multiple protocols for measuring, analyzing, and reporting habitat conditions has exacerbated the challenges associated with comparison of data gathered from different monitoring efforts or through different years. Different monitoring objectives and data requirements influence study design and implementation, leading to inherent differences in methods. Subsequent differences in data accuracy, precision, and effort required confound the ability of government agencies and private interests to share and synthesize information (Bain and Stevenson, 1999; Johnson *et al.*, 2001). A recent study by Johnson *et al.* (2001) reviewed documents describing 429 protocols for measuring salmonid habitat in the Pacific Northwest. Similarly, the American Fisheries Society and the U.S. Fish and Wildlife Service reviewed 52 methods, identified 705 different habitat variables used in assessment and monitoring programs, and concluded that the large variation in habitat measurements precluded any meaningful synthesis across regions, provinces, states, and even through time within single agencies (Bain and Stevenson, 1999).

Reducing the number of approaches and types of data used for stream habitat assessment is essential. This may be done by standardizing protocols among monitoring programs to the extent possible. Standardized assessment protocols would result in higher sample sizes through the sharing and combining of data among monitoring programs, thereby increasing the statistical power to describe spatial and temporal trends. Financially, the ability to use data collected and paid for by other monitoring programs reduces the need for large and redundant expenditures for aquatic habitat monitoring. The advantages of standardization have yet to be fully realized, primarily because monitoring programs are numerous, varied, and have acquired historical data that may be difficult to assimilate into a different protocol. Although these difficulties exist, there is a potential for standardizing a core set of physical stream attributes because most monitoring programs within a particular geographic area already measure a common set of attributes.

CONCLUSION

While field measurement of stream attributes has continued to be refined, the results of this study suggest that differences within and among USFS and USEPA protocols affect means and measurement precision for many commonly evaluated attributes. Statistical differences in means occurred for every stream attribute despite small sample sizes. Sources of variability among monitoring groups included how and where stream attributes were measured, attribute definitions, and training approach.

That measurement precision differs among protocols evaluated in this study suggests some protocols may be better than others (Table 8). The PIBO proto-

col emphasized repeated measurements for quality control, use of precision instruments, and consistency in measurement location and performed better when assessing reach length, gradient, and sinuosity (Table 8). Transect-based protocols (AREMP, EMAP) generally measured substrate attributes with more precision than habitat-based protocols (PIBO, R10, and R1/R4 – for an exception see R6 in Oregon).

Differences in the relative influence of cross-cutting issues, such as attribute definitions and training approach were also apparent. PIBO crews using quantitative attribute definitions performed better than those relying on more qualitative descriptions for determining pool habitat, and the monitoring group investing the most time training field technicians (PIBO) performed better measuring attributes known to cause consistency problems, such as percent pools, bankfull width, and bankfull width to depth ratio (Tables 4-8).

To further the comparability and synthesis of data among monitoring groups we recommend the adoption of standardized protocols for measuring a core set of habitat attributes. The results of this study suggest that of the 10 habitat variables compared, gradient, sinuosity, bankfull width, and bankfull width-to-depth ratio may be the easiest to integrate into a standardized protocol because definitions and measurement approaches were similar. Attributes requiring more effort to standardize due to larger discrepancies among protocols include reach length, percent pools, residual pool depth, D₅₀, percent fines, and LWD. Standardizing measurement protocols would reduce redundancy in collection efforts, lower aquatic monitoring costs through the sharing of data, increase defensibility of data as a result of higher statistical power, and increase the power to detect negative trends in time to mitigate their effects. In addition, the adoption of minimum accuracy and precision standards, as well as a commitment to quality control and quality assurance, would further ensure and strengthen the benefits of a standardized approach to monitoring stream habitat.

TABLE 8. Protocols With Highest Measurement Precision. High Precision, for These Comparisons, Were Based on the Lowest Root Mean Square Error for a Given Attribute.

Attribute	Combined ¹	Idaho	Oregon
Reach length (m)	PIBO	AREMP	PIBO
Gradient (%)	PIBO	PIBO	PIBO
Sinuosity	PIBO	PIBO	PIBO
Percent pools	PIBO	PIBO	EMAP
Residual pool depth (m)	N/A	R1/R4	EMAP
Bankfull width (m)	PIBO	PIBO	PIBO
BF width:depth ratio	PIBO	PIBO	R6
D ₅₀ (mm)	AREMP	EMAP	R6
Percent fines (< 6 mm)	EMAP	EMAP	EMAP
Large woody debris (100/m)	AREMP	AREMP	R6

¹AREMP, EMAP, and PIBO.

ACKNOWLEDGMENTS

We thank the many monitoring programs who contributed crews, time, and expertise to make this comparison possible, and the 65 members of the sampling crews who collected and/or summarized the data used in the analysis. We also thank Jack Schmidt, Fred Provenza, Burk Warner, and 3 anonymous reviewers for their insight and suggestions for improving this manuscript. Funding for this project was provided by the USDA Forest Service, USDI Bureau of Land Management, U.S. Environmental Protection Agency, and the states of Idaho and Oregon Department of Environmental Quality.

LITERATURE CITED

- Archer, E.K., B.B. Roper, R.C. Henderson, N. Bouwes, S.C. Mellison, and J.L. Kershner, 2004. *Testing Common Stream Sampling Methods for Broad-Scale, Long-Term Monitoring. General Technical Report, RMRS-GTR-122*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, 15 pp.
- Bain, M.B., T.C. Hughes, and K.K. Arend, 1999. Trends in Methods for Assessing Freshwater Habitats. *Fisheries* 24(4):16-21.
- Bain, M.B., and N.J. Stevenson, 1999. *Aquatic Habitat Assessment: Common Methods*. American Fisheries Society, Bethesda, Maryland, 215 pp.
- Beechie, T.J. and T.H. Sibley, 1997. Relationships Between Channel Characteristics, Woody Debris, and Fish Habitat in Northwestern Washington Streams. *Transactions of the American Fisheries Society* 126:217-229.
- Bilby, R.E., 1984. Removal of Woody Debris May Affect Stream Channel Stability. *Journal of Forestry* 82(10):609-613.
- Buffington, J.M., T.E. Lisle, R.D. Woodsmith, and S. Hilton, 2002. Controls on the Size and Occurrence of Pools in Coarse-Grained Forest Rivers. *River Research and Applications* 18:507-531.
- Bunte, K. and S.R. Abt, 2001. Sampling Frame for Improving Pebble Count Accuracy in Coarse Gravel-Bed Streams. *Journal of the American Water Resources Association* 37(4):1001-1014.
- Conquest, L.L., S.C. Ralph, and R.J. Naiman, 1994. Implementation of Large-Scale Stream Monitoring Efforts: Sampling Design and Data Analysis Issues. In: *Biological Monitoring of Aquatic Systems*, S.L. Loeb, and A. Spacie (Editors). Lewis Publishers, Boca Raton, Florida, 381 pp.
- Gallo, K., 2002. *Aquatic and Riparian Effectiveness Monitoring Program for the Northwest Forest Plan*. U.S. Forest Service, Corvallis, Oregon, 51 pp.
- Hankin, D.G., and G.H. Reeves, 1988. Estimating Total Fish Abundance and Total Habitat Area in Small Streams Based on Visual Estimation Methods. *Canadian Journal of Fisheries and Aquatic Science* 45:834-844.
- 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(4):853-860.
- Hannaford, M.J. and V.H. Resh, 1995. Variability in Macroinvertebrate Rapid-Bioassessment Surveys and Habitat Assessments in a Northern California Stream. *Journal of the North American Benthological Society* 14:430-439.
- Harrelson, C.C., C.L. Rawlins, and J.P. Potyondy, 1994. *Stream Channel Reference Sites: An Illustrated Guide to Field Technique. General Technical Report RM-245*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, 61 pp.
- Henderson, R.C., E.K. Archer, and J.L. Kershner, 2002. *Effectiveness Monitoring for Streams and Riparian Areas, Sampling Protocol for Integrator Reaches: Stream Channel Parameters*. U.S. Forest Service, Fish and Aquatic Ecology Unit, Logan, Utah, 45 pp.
- Hey, R.D. and C.R. Thorne, 1983. Accuracy of Surface Samples From Gravel Bed Material. *Journal of Hydrologic Engineering* 109:842-851.
- Hogle, J.S., T.A. Wesch, and W.A. Hubert, 1993. A Test of the Precision of the Habitat Quality Index Model II. *North American Journal of Fisheries Management* 13:640-643.
- INFS (Inland Native Fish Strategy), Environmental Assessment, 1995. *Decision Notice and Finding of No Significant Impact*. U.S. Department of Agriculture, Forest Service, Intermountain, Northern, and Pacific Northwest Regions. <http://www.fs.fed.us/r6/fish/9506-infish.pdf>, accessed February 6, 2004.
- Isaak, D.J., W.A. Hubert, and K.L. Krueger, 1999. Accuracy and Precision of Stream Reach Water Surface Slopes Estimated in the Field and From Maps. *North American Journal of Fisheries Management* 19:141-148.
- Johnson, D.H., N. Pittman, E. Wilder, J.A. Silver, R.W. Plotnikoff, B.C. Mason, K.K. Jones, P. Roger, T.A. O'Neil, and C. Barrett, 2001. *Inventory and Monitoring of Salmon Habitat in the Pacific Northwest - Directory and Synthesis of Protocols for Management/Research and Volunteers in Washington, Oregon, Idaho, Montana, and British Columbia*. Washington Department of Fish and Wildlife, Olympia Washington, 212 pp.
- Kaufmann, P.R., E.G. Levine, E.G. Robison, C. Seeliger, and D.V. Peck, 1999. *Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003*. U.S. Environmental Protection Agency, Washington, D.C., 102 pp.
- Kershner, J.L., M. Coles-Ritchie, E. Cowely, R.C. Henderson, K. Kratz, C. Quimby, D.M. Turner, L.C. Ulmer, and M.R. Vinson, 2004. *A Plan to Monitor the Aquatic and Riparian Resources in the Area of PACFISH/INFISH and the Biological Opinions for Bull Trout Salmon and Steelhead. General Technical Report*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, Utah. http://www.fs.fed.us/biology/resources/pubs/feu/pibo_final_011003.pdf, accessed February 6, 2004.
- Kershner, J.L., W.M. Snider, D.M. Turner, and P.B. Moyle, 1992. Distribution and Sequencing of Mesohabitats: Are There Differences at the Reach Scale? *Rivers* 3:179-190.
- Kondolf, G.M., 1997. Application of the Pebble Count: Notes on Purpose, Method, and Variants. *Journal of the American Water Resources Association* 33:79-87.
- Kondolf, G.M. and S. Li, 1992. The Pebble Count Technique for Quantifying Surface Bed Material Size in Instream Flow Studies. *Rivers* 3:80-87.
- Larsen, D.P., T.M. Kincaid, S.E. Jacobs, and N.S. Urquhart, 2001. Designs for Evaluating Local and Regional Scale Trends. *Bioscience* 51(12):1069-1078.
- Littell, R.C., G.A. Miliken, W.W. Stroup, and R.D. Wolfinger, 1996. *SAS® System for Mixed Models*. SAS Institute Inc., Cary, North Carolina, 633 pp.
- MacDonald, L.H., 1994. Developing a Monitoring Project. *Journal of Soil and Water Conservation* 49(3):221-227.
- MacDonald, L.H., A.W. Smart, and R.C. Wissmar, 1991. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Stream in the Pacific Northwest and Alaska*. U.S. Environmental Protection Agency, Region 10, Seattle, Washington, 166 pp.
- Marcus, W.A., S.C. Ladd, J.A. Stoughton, and J.W. Stock, 1995. Pebble Counts and the Role of User-Dependent Bias in Documenting Sediment Size Distributions. *Water Resources Research* 31(10):2625-2631.
- Overton, C.K., S.P. Wollrab, B.C. Roberts, and M.A. Radko, 1997. *R1/R4 (Northern/Intermountain Regions) Fish and Fish Habitat Standard Inventory Procedures Handbook. General Technical Report INT-GTR-346*. U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, Utah, 74 pp.
- Penrose, D. and S.M. Call, 1995. Volunteer Monitoring of Benthic Macroinvertebrates: Regulatory Biologists' Perspectives. *Journal of the North American Benthological Society* 14:203-209.
- Platts, W.S., C. Armour, G.D. Booth, M. Bryant, J.L. Bufford, P. Cuplin, S. Jensen, G.W. Lienkaimper, G.W. Minshall, S.B. Monsen, R.L. Nelson, J.R. Sedell, and J.S. Tuhy, 1987. *Methods for Evaluating Riparian Habitats With Applications to Management. General Technical Report INT-221*, U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, Utah, 17 pp.

- Poole, G.C., C.A. Frissell, and S.C. Ralph, 1997. In-Stream Habitat Unit Classification: Inadequacies for Monitoring and Some Consequences for Management. *Journal of the American Water Resources Association* 33:879-896.
- Ralph, S.C., G.C. Poole, L.L. Conquest, and R.J. Naiman, 1994. Stream Channel Morphology and Woody Debris in Logged and Unlogged Basins of Western Washington. *Canadian Journal of Fisheries and Aquatic Science* 51:37-51.
- Ramsey, M.H., M. Thompson, and M. Hale, 1992. Objective Evaluation of Precision Requirements for Geochemical Analysis Using Robust Analysis of Variance. *Journal of Geochemical Exploration* 44:23-36.
- Reeves, G.H., D.B. Hohler, and D.P. Larsen, 2004. *Effectiveness Monitoring for the Aquatic and Riparian Component of the Northwest Forest Plan: Conceptual Framework and Options. Gen. Tech. Rep. PNW-GTR-577*. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, 71 pp.
- 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* 38:1-10.
- Roper, B.B., J.L. Kershner, and R.C. Henderson, 2003. The Value of Using Permanent Sites When Evaluating Stream Attributes at the Reach Scale. *Journal of Freshwater Ecology* 18(4):585-592.
- Roper, B.B. and D.L. Scarnecchia, 1995. Observer Variability in Classifying Habitat Types in Stream Surveys. *North American Journal of Fisheries Management* 15:49-53.
- Ryan, S.E. and W.W. Emmett, 2002. *The Nature of Sediment Movement in Little Granite Creek Near Bondurant, Wyoming. RMRS-GTR-90*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, Utah, 48 pp.
- Ryan, S.E. and L.S. Porth, 1999. A Field Comparison of Three Pressure-Difference Bedload Samplers. *Geomorphology* 30:307-322.
- Simonson, T.D., J. Lyons, and P.D. Kanehl, 1994. Quantifying Fish Habitat in Streams: Transect Spacing, Sample Size, and a Proposed Framework. *North American Journal of Fisheries Management* 14:607-615.
- Smith, A.D., 1944. A Study of the Reliability of Range Vegetation Estimates. *Ecology* 25(4):441-448.
- Stevens, D.L. Jr and N.S. Urquhart, 2000. Response Designs and Support Regions in Sampling Continuous Domains. *Environmetrics* 11:13-41.
- 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.
- Urquhart, N.S., S.G. Paulsen, and D.P. Larsen, 1998. Monitoring for Policy-Relevant Regional Trends Over Time. *Ecological Applications* 8(2):246-257.
- USDA (U.S. Department of Agriculture), 1997. *U.S. Department of Agriculture, Forest Service. Tongass Land Management Plan Revision*. Alaska Region R10-MB-338a, Juneau, Alaska.
- USDA (U.S. Department of Agriculture), 2000. *U.S. Department of Agriculture, Forest Service. Stream Inventory Handbook, Level I & II, Version 2.0*. Pacific Northwest Region R6, Portland, Oregon, 84 pp.
- USDA (U.S. Department of Agriculture), 2001. *U.S. Department of Agriculture, Forest Service. Aquatic Habitat Management Handbook, Chapter 20 – Fish and Aquatic Stream Habitat Survey*. Alaska Region R10, Juneau, Alaska, 55 pp.
- USEPA (U.S. Environmental Protection Agency), 1991. *Guidance for Water Quality-Based Decisions: The TMDL Process. EPA 440/4-91-001*. Office of Water, U.S. Environmental Protection Agency, Washington, D.C., 41 pp.
- Wang, L., T.D. Simonson, and J. Lyons, 1996. Accuracy and Precision of Selected Stream Habitat Estimates. *North American Journal of Fisheries Management* 16:340-347.
- Wohl, E.E., D.J. Anthony, S.W. Madsen, and D.M. Thompson, 1996. A Comparison of Surface Sampling Methods for Coarse Fluvial Sediments. *Water Resources Research* 32(10):3219-3226.
- Wolman, M.G., 1954. A Method of Sampling Coarse River-Bed Material. *Transactions of the American Geophysical Union* 36(6):951-956.
- Woodsmith, R.D. and J.M. Buffington, 1996. Multivariate Geomorphic Analysis of Forest Streams: Implications for Assessment of Land Use Impacts on Channel Condition. *Earth Surface Processes and Landforms* 21:377-393.
- Woodsmith, R.D., J.R. Noel and M.L. Dilger, 2005. An Approach to Effectiveness Monitoring of Floodplain Channel Aquatic Habitat: Channel Condition Assessment. *Landscape and Urban Planning* 72:177-204.