JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 44, No. 2 AMERICAN WATER RESOURCES ASSOCIATION April 2008

THE ROLE OF OBSERVER VARIATION IN DETERMINING ROSGEN STREAM TYPES IN NORTHEASTERN OREGON MOUNTAIN STREAMS[1]

Brett B. Roper, John M. Buffington, Eric Archer, Chris Moyer, and Mike Ward[2]

ABSTRACT: Consistency in determining Rosgen stream types was evaluated in 12 streams within the John Day Basin, northeastern Oregon. The Rosgen classification system is commonly used in the western United States and is based on the measurement of five stream attributes: entrenchment ratio, width-to-depth ratio, sinuosity, slope, and substrate size. Streams were classified from measurements made by three monitoring groups, with each group fielding multiple crews that conducted two to three independent surveys of each stream. In only four streams (33%) did measurements from all crews in all monitoring groups yield the same stream type. Most differences found among field crews and monitoring groups could be attributed to differences in estimates of the entrenchment ratio. Differences in entrenchment ratio were likely due to small discrepancies in determination of maximum bankfull depth, leading to potentially large differences in determination of Rosgen's flood-prone width and consequent values of entrenchment. The result was considerable measurement variability among crews within a monitoring group, and because entrenchment ratio is the first discriminator in the Rosgen classification, differences in the assessment of this value often resulted in different determination of primary stream types. In contrast, we found that consistently evaluated attributes, such as channel slope, rarely resulted in any differences in classification. We also found that the Rosgen method can yield nonunique solutions (multiple channel types), with no clear guidance for resolving these situations, and we found that some assigned stream types did not match the appearance of the evaluated stream. Based on these observations we caution the use of Rosgen stream classes for communicating conditions of a single stream or as strata when analyzing many streams due to the reliance of the Rosgen approach on bankfull estimates which are inherently uncertain.

(KEY TERMS: monitoring; quality control/quality assurance; restoration; Rosgen; stream classification; bankfull.)

Roper, Brett B., John M. Buffington, Eric Archer, Chris Moyer, and Mike Ward, 2008. The Role of Observer Variation in Determining Rosgen Stream Types in Northeastern Oregon Mountain Streams. *Journal of the American Water Resources Association* (JAWRA) 44(2):417-427. DOI: 10.1111/j.1752-1688.2008.00171.x

[1]Paper No. JAWRA-07-0009-P of the *Journal of the American Water Resources Association* (JAWRA). Received January 18, 2007; accepted August 2, 2007. **Discussions are open until October 1, 2008**.

[2]Respectively, U.S. Forest Service, Fish and Aquatic Ecology Unit, Utah State University, 860 N. 1200 E., Logan, Utah 84321; U.S. Forest Service, Rocky Mountain Research Station, Idaho Water Center, 322 E. Front St., Boise, Idaho 83702; U.S. Forest Service, PacFish InFish Biological Opinion Monitoring Program, 860 N. 1200 E, Logan, Utah 84321; Bureau of Land Management, Aquatic Riparian Effectiveness Monitoring Program, 4077 S.W. Research Way, Corvallis, Oregon 97333; and Terraqua Inc., PO Box 85, Wauconda, Washington 98859 (E-Mail/Roper: broper@fs.fed.us).

INTRODUCTION

Rivers reflect the local physiographic setting and disturbance regime in which they are found (Leopold *et al.*, 1964; Ebersole *et al.*, 1997; Buffington *et al.*, 2003). Within and between these settings, streams typically have similar suites of channel morphologies, with repeatable patterns of occurrence, that have resulted in numerous classification efforts (see reviews by Kondolf, 1995; Rosgen, 1996; Montgomery and Buffington, 1998; Juracek and Fitzpatrick, 2003; Downs and Gregory, 2004; Simon *et al.*, 2007). The rational for these classification systems spans a broad spectrum of goals and objectives, including the need to meet legal requirements for environmental standards, to improve communication, and to provide a better understanding of fluvial processes (Kondolf, 1995; Kondolf *et al.*, 2003; Downs and Gregory, 2004; Brierley and Fryirs, 2005; Simon *et al.*, 2007). One classification system that has found widespread application, especially in the mountainous river basins of the western United States (U.S.), was developed by Rosgen (1994). In formulating his classification system, Rosgen (1994) postulated it would meet four objectives: (1) predict a river's behavior from its appearance, (2) allow the development of specific hydraulic geometry and sediment transport relationships for different channel types, (3) permit extrapolation of site-specific data to reaches of similar character, and (4) provide a consistent frame of reference for communication amongst those working with river systems.

In determining Rosgen (1994, 1996) stream types, three aspects of the stream's appearance (entrenchment ratio, bankfull width-to-depth ratio, and sinuosity) are used to divide channels into eight primary stream types denoted by the capital letters – A, B, C, D, DA, E, F, and G. These primary stream types are further divided into secondary types based on stream slope and substrate size. The result is 42 major and 94 total stream types.

While the Rosgen stream classification system has been widely applied, it has also been widely criticized (Malakoff, 2004). Critics argue that the relationship between Rosgen stream types and fluvial processes is poorly demonstrated and that the approach provides little mechanistic insight regarding channel processes and response potential to natural and anthropogenic disturbance (Miller and Ritter, 1996; Montgomery and Buffington, 1997; Doyle and Harbor, 2000; Juracek and Fitzpatrick, 2003; Simon *et al.*, 2007). The disconnect between Rosgen stream types and channel processes has led several authors to suggest that this classification system has the potential to be applied inappropriately (Juracek and Fitzpatrick, 2003; Kondolf *et al.*, 2003) as demonstrated in several recent case studies of failed stream restoration efforts based on the Rosgen system (Kondolf *et al.*, 2001; Downs and Kondolf, 2002; Juracek and Fitzpatrick, 2003; Smith and Prestegaard, 2005). As a result, there are serious questions whether this classification system meets the first three objectives described by Rosgen (1994).

Despite these criticisms, many state and federal management agencies continue to rely on the Rosgen system for conducting stream inventories, designing channel restoration, and monitoring aquatic habitat (Savery *et al.*, 2001; Juracek and Fitzpatrick, 2003; Environmental Protection Agency 2006; Simon *et al.*, 2007). Given the shortcomings of the Rosgen system to represent mechanistic fluvial processes, its remaining strength is likely to be in enabling communication among professionals in aquatic fields (Miller and Ritter, 1996; Juracek and Fitzpatrick, 2003). But for a classification system to improve communication, it must assure that different observers provide equivalent identifications of stream type (Kondolf *et al.*, 2003). This assumption, however, has yet to be rigorously evaluated for the Rosgen classification system. This paper therefore seeks to determine whether measurements made by different observers yield consistent classification of Rosgen stream types and, if these classifications differ, to determine the reasons for these differences.

STUDY AREA AND METHODS

Consistency in determining Rosgen stream types was evaluated in 12 study reaches within the John Day Basin, northeastern Oregon (Figure 1). All of the stream reaches examined in this study were derived from random sampling strategies used by the Oregon Department of Fish and Wildlife (n = 7 of the 12 study sites), Environmental Protection Agency (n = 3), or U.S. Forest Service (n = 2) for monitoring physical characteristics of fish-bearing streams. Study sites were selected from candidate lists to represent three stream types defined by Montgomery and Buffington (1997): step-pool, plane-bed, and pool-riffle channels, with four channels of each stream type, and with each set representing a range of channel complexity [simple, free-formed channels *vs.* complex wood-forced ones (e.g., Buffington and Montgomery, 1999)]. The result was a set of stream reaches with variable physical characteristics that could be used to evaluate the consistency of Rosgen classification determined from measurements reported by different observers (Table 1).

During the summer of 2005 (July 16 to September 12), each of the 12 stream reaches was evaluated by

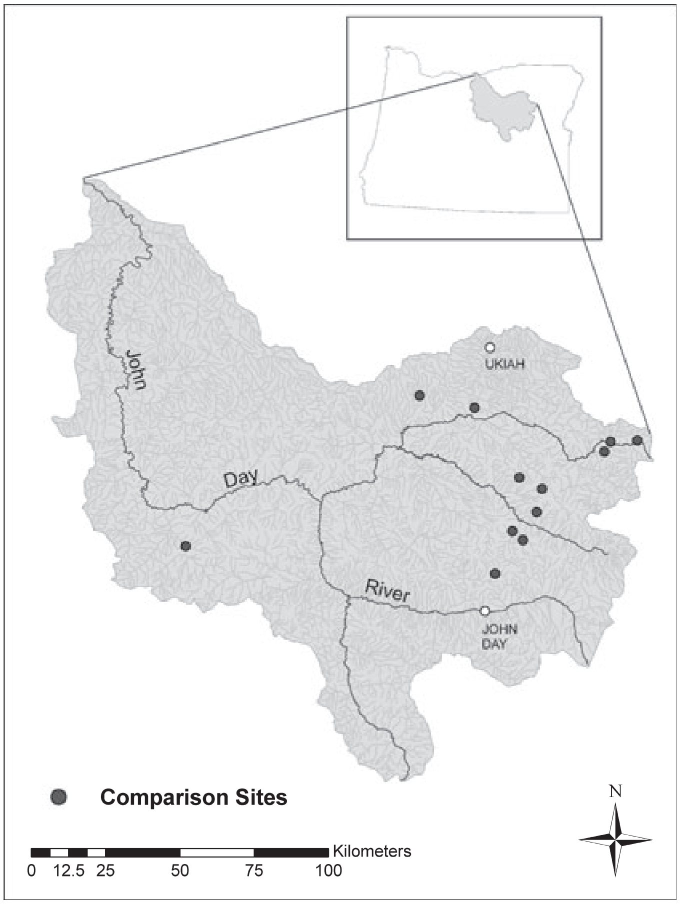

FIGURE 1. Locations of Study Sites Within the John Day River Basin, OR.

TABLE 1. Stream Reach Characteristics of the 12 Study Sites.

Stream	BFW	Ent	W/D	Sin	Slope	D_{50} (mm)
Big (pr)	3.15	2.33	13.8	1.44	0.0113	5
Bridge (pb)	4.35	1.57	18.7	1.28	0.0099	23
Camus (pb)	13.83	1.61	29.4	1.04	0.0116	97
Crane (pr)	4.10	2.72	21.1	1.47	0.0110	7
Crawfish (sp)	6.68	1.35	17.5	1.15	0.0503	82
Indian (sp)	4.78	1.75	23.0	1.17	0.0582	17
Myrtle (sp)	3.02	1.61	17.6	1.13	0.0935	29
Potamus (pb)	8.11	1.63	36.6	1.11	0.0242	75
Tinker (pb)	2.21	3.23	14.7	1.18	0.0272	18
Trail (pr)	5.52	3.20	22.5	1.39	0.0176	47
West Fork Lick (pr)	2.82	1.69	15.6	1.28	0.0330	26
Whiskey (sp)	2.75	1.73	16.8	1.11	0.0688	41

Notes: Values are crew averages across all monitoring groups. BFW, bankfull width; Ent, entrenchment ratio; W/D, width-to-depth ratio; Sin, sinuosity; D_{50}, median surface grain size. Montgomery and Buffington (1997) stream types are given in parentheses (pr, pool-riffle; pb, plane-bed; sp, step-pool).

seven state, tribal, and federal monitoring groups as part of a comparison of the repeatability and equivalence of different protocols used for measuring physical stream attributes (Lanigan *et al.*, 2006). These monitoring groups conduct extensive stream surveys each year throughout the western U.S., fielding hundreds of personnel, as part of legally mandated state and federal environmental assessment programs. Of these seven monitoring groups, three collected information on the five attributes necessary for Rosgen (1994) stream classification; entrenchment ratio, bankfull with-to-depth ratio, sinuosity, slope and substrate size. The three groups were the Aquatic Riparian Effectiveness Monitoring Program (AREMP; Reeves *et al.*, 2004), the PacFish InFish Biological Opinion Monitoring Program (PIBO; Kershner *et al.*, 2004a), and the Upper Columbia Monitoring Program (UC; Hillman, 2004).

In most cases, three independent surveys of each stream were conducted by each monitoring group, but because of data omissions by the PIBO group, several streams (8 of 12) only had data for two independent observations. Evaluations were conducted by a total of six different crews for AREMP, five for PIBO, and three for UC. Each group used their own protocols to evaluate the five attributes necessary to classify Rosgen (1994) stream type (Table 2). Two of the groups, AREMP and PIBO, had identical operational definitions for these stream attributes, but differed in training, instruments, and locations within a reach where attributes were evaluated.

We used Rosgen's (1994) Level II classification key (as modified by Rosgen, 1996) to determine stream types based on the summarized reach data collected by each of the crews (see Figure 2 for stream types most likely in this study; see Rosgen (1996) for all stream types); the classification was not performed by the field crews, but rather from their measurements. In determining Rosgen stream type, we used the classification parameters listed in the key, as well as their suggested possible variation (Rosgen, 1994, 1996). For example, the entrenchment ratio corresponding with a Rosgen A channel type is less than 1.4, but because the suggested variation is ±0.2 units, for classification purposes we permitted entrenchment ratios up to 1.6 for Rosgen A channel types. In using the Rosgen classification key in conjunction with this study, an effort was made to interpret data so that all crews within a monitoring group arrived at the same stream type for each site; different stream types were reported only when it was not possible to assign a common stream type using the allowed attribute variation (Rosgen, 1994, 1996). We refer to this as our "consistency rule," which provides a conservative assessment of classification differences within each monitoring group.

TABLE 2. Protocols Used by the Three Monitoring Groups for Evaluating the Stream Attributes Used in Rosgen's (1994, 1996) Classification System.

Attribute	Group	Definition
Entrenchment ratio	Rosgen	Ratio of the flood-prone width to the bankfull width (flood-prone width = width at elevation of twice maximum bankfull depth)
	AREMP	Same as Rosgen; measured in first riffle
	PIBO	Same as Rosgen; average, measured at first four riffles
	UC	Same as Rosgen; average, measured at three equally spaced transects
Width-to-depth ratio	Rosgen	The ratio of bankfull channel width to mean bankfull depth
	AREMP	Same as Rosgen; average, measured at 11 equally spaced transects
	PIBO	Same as Rosgen; average, measured at first four riffles
	UC	Average width measured at 11 equally spaced transects, mean depth is average depth of thalweg
Sinuosity	Rosgen	Stream length (thalweg)/valley length
	AREMP	Same as Rosgen
	PIBO	Same as Rosgen
	UC	Same as Rosgen
Slope	Rosgen	Reach-average water-surface slope
	AREMP	Same as Rosgen; measured with laser level
	PIBO	Same as Rosgen; measured with hand level
	UC	Same as Rosgen; measured with hand level
Substrate	Rosgen	Wolman (1954) pebble counts including streambanks
	AREMP	Pebble counts at 21 equally spaced transects – active channel only.
	PIBO	Pebble counts at 11 equally spaced transects – active channel only
	UC	Pebble counts at 21 equally spaced transects – active channel only
Reach length	Rosgen	Tens of meters to kilometers
	AREMP	20 times bankfull width; minimum of 150 m, maximum of 500 m
	PIBO	Same as above
	UC	150 m

In applying the classification key, we kept track of observations where no stream type was possible even with the variation of channel attributes allowed by Rosgen (Figure 2). We also noted stream types that would have resulted from a single measurement in the absence of both the suggested variation and our consistency rule, thereby representing the broadest variation amongst observers. Finally, we noted cases where the absence of allowable variation in classification parameters would have resulted in no possible determination of stream type in the Rosgen system. An example would be sinuosity less than 1.2 when the entrenchment ratio is greater than 1.4.

RESULTS

Rosgen stream types for each monitoring group and for crews within a group are shown in Table 3. We found that all field crews in all monitoring groups agreed on the Rosgen stream type in 4 of the 12 streams (33% of the sites). Agreement increases to 50% (6 of the 12 sites) for crews in the two monitoring groups that used the same operational definitions for physical attributes (AREMP and PIBO). Differences among crews were primarily due to differences in values of the entrenchment ratio. Entrenchment is defined as the ratio of the flood-prone width to the bankfull width, where the flood-prone width is measured across the river valley at an elevation twice the maximum bankfull depth (Rosgen, 1994). Crews had a difficult time consistently evaluating this attribute (Figure 3). For example in Trail Creek, estimates of the entrenchment ratio ranged from slightly above 1 to nearly 10 depending upon the crew.

Consistency among crews within monitoring groups also differed; AREMP crews were different in five, PIBO four, and UC six of the 12 streams (Table 3). Differences among crews using the same protocols occurred for multiple reasons, but differences in entrenchment ratio accounted for 60% of the classification differences among crews in the AREMP group, 100% of the cases for PIBO and 50% of the cases for UC. Sediment size accounted for the next largest classification difference within monitoring groups (40% of AREMP and 75% of UC), followed by width-to-depth ratio (25% of AREMP) and gradient (25% of UC).

Although we determined stream types for all crew evaluations, there was one set of measurements that did not fit Rosgen's classification. In this case, an AREMP crew determined West Fork Lick Creek to be moderately entrenched (1.85) with a low width-to-depth ratio (6.2). This stream was labeled as a B type

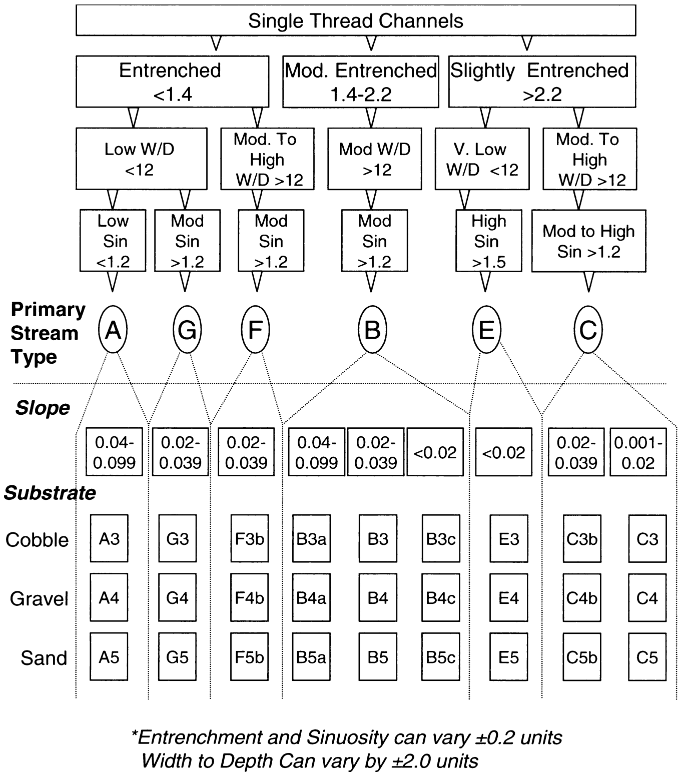

FIGURE 2. The Rosgen Stream Types Which Were Most Likely for the 12 Stream Reaches Evaluated in This Study. Entrenchment is defined as the ratio of the flood-prone width to the bankfull width, where the flood-prone width is measured across the river valley at an elevation twice the maximum bankfull depth (Rosgen, 1994). Sin is sinuosity (stream length/valley length), and W/D is the bankfull width-to-depth ratio. Substrate is the median surface grain size (D_{50}), with cobble, gravel, and sand defined as D_{50} = 64-256, 2-64, and <2 mm, respectively. Channel types not shown were not encountered in this study (e.g., Aa+ and Cc– channels). Modified from Rosgen (1996).

due to the observed entrenchment ratio, but clearly the low width-to-depth ratio is outside the range expected for this stream type and better fits streams with lower or higher entrenchment ratios (Figure 2).

We also found that 50% or more of the stream evaluations for each monitoring group had attribute values outside the defined limits, but within the expected variation, for a given stream type. This happened primarily when moderately or slightly entrenched streams (>1.6) had sinuosity less than 1.2 (Table 1). As a result, there were a large number of sites that could not have been classified without the allowable variation in classification parameters (Table 3, NC values).

The allowed attribute variation led to an increase in consistency in determining Rosgen stream type at each site. For example, one AREMP crew found the following characteristics for Myrtle Creek: entrenchment ratio 1.26, width-to-depth ratio 17.6, sinuosity 1.12, and slope 0.0945. Based on entrenchment ratio, sinuosity, gradient, and the allowable variation of these first two parameters (±0.2 units), this stream could be either a Rosgen A or B channel type. The width-to-depth value, however, forced assignment into the B stream type. This was not true of another AREMP crew, which described Myrtle Creek as having an entrenchment ratio of 1.23, width-to-depth ratio of 13.8, sinuosity of 1.09, and slope of 0.0942. Again, this channel could be a Rosgen A or B stream type, but because of the lower width-to-depth ratio, it is likely that if this crew had been the sole evaluator of this reach, it would have been assigned a Rosgen A

TABLE 3. Rosgen Stream Types Determined for Each Crew Within the Three Monitoring Groups at Each of the 12 Streams.

		Monitoring Group		
Creek	**Crew**	**AREMP**	**PIBO**	**UC**
Big	1	E4 (NC)	B4c (B4c)	B4c (B4c)
	2	E4 (NC)	E4 (E4)	B4c (B4c)
	3	B4c (B4c)		B4c (B4c)
Bridge	1	B4c (F4)	B4c (B4c)	B4c (B4c)
	2	B4c (NC)	B4c (B4c)	B4c (B4c)
	3	B4c (NC)		B4c (B4c)
Camas	1	B3c (NC)	B3c (NC)	B3c (NC)
	2	B3c (NC)	B3c (NC)	B3c (NC)
	3	B3c (NC)	B3c (NC)	B3c (NC)
Crane	1	C4 (C4)	C4 (C4)	C4 (C4)
	2	C4 (C4)	B4c (B4c)	B4c (B4c)
	3	B4c (B4c)		B5c (B5c)
Crawfish	1	A3 (NC)	B3a (NC)	B4a (B4a)
	2	A3 (NC)	B3a (NC)	B4a (B4a)
	3	A4 (NC)		B3a (NC)
Indian	1	B4a (B4a)	B4a (B4a)	B4a (NC)
	2	B4a (NC)	B4a (NC)	B4a (NC)
	3	B4a (NC)	B4a (NC)	B5a (NC)
Myrtle	1	B4a (NC)	B4a (NC)	B4a (NC)
	2	B4a (NC)	B4a (NC)	B4a (NC)
	3	B4a (NC)		B4a (NC)
Potamus	1	F3b (F3b)	B3 (NC)	B3 (NC)
	2	F3b (NC)	B3 (NC)	B4 (NC)
	3	F3b (NC)		C4b (NC)
Tinker	1	C4b (NC)	C4b (C4b)	C4b (NC)
	2	C4b (NC)	C4b (C4b)	C4b (NC)
	3	C4b (NC)	C4b (E4b)	B4 (NC)
Trail	1	C4 (C4)	C4 (C4)	B4c (B4c)
	2	C4 (NC)	F4 (F4)	B4c (B4c)
	3	C3 (C3)		B4 (B4)
WF Lick	1	G4 (G4)	B4 (B4)	B4 (B4)
	2	G4 (NC)	C4b (C4b)	B4 (F4b)
	3	B4 (NP)		B4 (F4b)
Whiskey	1	B4a (NC)	B4a (NC)	B4a (NC)
	2	B4a (NC)	B4a (NC)	B4a (NC)
	3	B4a (NC)		B4a (NC)

Notes: The first value is the stream type based on applying our rule set for consistency, as described in the text. The value in parenthesis is the stream type without allowing for variation of classification parameters specified by Rosgen (Figure 2) and without applying our consistency rule. NC means no stream class could have been determined without allowed variation of classification parameters. NP means not possible to classify even with allowed variation of parameters.

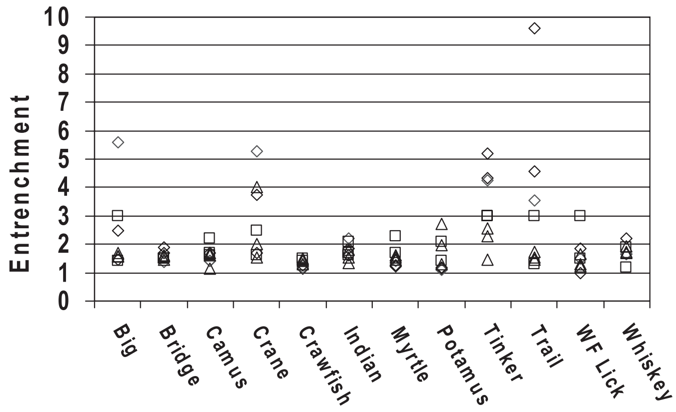

FIGURE 3. Entrenchment Ratios for Each of the 12 Streams Determined by Each of the Three Monitoring Groups. The diamonds are AREMP observations. The squares are PIBO observations. The triangles are UC observations.

stream type. However, because of our rule set for consistency, both observations were determined to be B stream types for this study.

Channel types determined without the allowed attribute variation and without the use of our consistency rule are also reported in Table 3 (values in parentheses). Only 41% of the observations could be classified without the allowed attribute variation. Of these, less than 10% differed from the channel types determined with our consistency rule.

DISCUSSION

Observer Differences

We found that monitoring groups and field crews within groups often differed in their determination of Rosgen stream type. In only 33% of the streams evaluated did all monitoring groups and all crews within a group agree on the stream type. In each of these cases, consistency was only possible because of the permissible variation in the primary classification attributes and the use of our rule for maximizing consistency.

Complete agreement in stream type among field crews increased to 50% for the two groups that used similar definitions of measured stream attributes (AREMP and PIBO). Within a monitoring group, consistent determination of stream type was higher still, with all crews agreeing on the primary stream type (A-G) in 75% of the evaluated streams. This suggests that if all crews used similar protocols for evaluating attributes and received similar training, variability in classification among crews would likely decrease. Although consistent protocols and training may be desirable, the large number of aquatic monitoring programs and their affiliation with different state and federal agencies (Johnson *et al.*, 2001) make implementing this option a challenge at a regional or national scale.

While requiring similar training and protocols would increase consistency, this step alone may not be enough to ensure similar identification of Rosgen stream type. Because many of the observations in this study (>50%) have channel attributes that span multiple Rosgen stream types given the allowable variation of classification parameters, differences would have been greater if each monitoring group had only a single evaluation of stream type at each site, or if our consistency rule had not been applied. The data indicate that at least 36% of AREMP, 42% of PIBO, and 31% of UC determinations could have been placed in another stream type (i.e., nonunique solutions).

The primary cause for differences in classification of Rosgen stream type was variation among field crews in estimating entrenchment ratio. The average deviation of each crew's entrenchment ratio from their monitoring group's mean value for that stream, averaged over the 12 streams and three monitoring groups, was 0.78 (overall mean = 2.04; coefficient of variation = 38%). This indicates that assessing whether the entrenchment ratio is less than or greater than 1.4 or 2.2 (critical values in Rosgen's classification) is more dependent on the observer than the site. The average observer variability in determining this attribute was nearly four times greater than the allowable variation (0.2) suggested by Rosgen (1994) for classification of channel types.

One possible explanation for this large variation in the assessment of entrenchment ratios was that these monitoring groups do a poor job of consistently evaluating stream characteristics in general. While this problem can not be ruled out, these crews were consistent in their evaluation of other attributes used in the Rosgen classification system, such as slope (average variation among observers of 0.0027; coefficient of variation = 8%) and sinuosity (average variation among observers of 0.083; coefficient of variation = 7%). In addition, these crews receive more training, have experience surveying, and have better defined protocols than the vast majority of federal and state personnel used to conduct stream surveys (Whitacre *et al.*, 2007). Source of training could also be a factor. Although all crews were trained in measuring Rosgen classification parameters, not all crew members received training from Rosgen. However, scientific procedures should allow for replication by any competent investigator, regardless of who trained them.

We suggest that the large amount of variability associated with estimating entrenchment ratios results from differences among field crews in determining the elevation of the bankfull floodplain and consequent values of bankfull depth. Recall that the entrenchment ratio is the flood-prone width (measured across the valley at an elevation twice the maximum bankfull depth) normalized by the bankfull width. As such, slight differences in one's estimate of the bankfull depth will literally be multiplied by two, potentially resulting in large differences in the flood-prone elevation, and even larger differences in the flood-prone width, particularly in unconfined alluvial channels. For example at Big Creek, one AREMP crew chose a somewhat lower location for the bankfull floodplain compared to a second AREMP crew (Figure 4), resulting in similar bankfull widths (3.15 *vs.* 3.33 m, 6% difference), but different values of both average bankfull depth (0.279 *vs.* 0.371 m, 33% difference) and maximum bankfull depth (0.521 *vs.* 0.623 m, 20% difference). These modest differences in depth led to substantially different estimates of the flood-prone width in this unconfined alluvial channel (7.88 *vs.* 18.58 m, 136% difference), resulting in very different assessments of channel entrenchment (2.50 *vs.* 5.58, 123% difference). While the above example was one of the more extreme in these data, even minor differences among observer estimates of the entrenchment ratio can easily result in different primary stream types (A-G) since entrenchment ratio is the first step of Rosgen's classification, and the allowable variation of this attribute separating different channel types is small (0.2) (Figure 2).

Differences in the number and location of cross-sectional measurements may also explain some of the variation between field crews (Table 2). Natural variability of channel characteristics along a reach may result in different estimates of Rosgen classification parameters and potentially different channel types, depending on how cross sections are arrayed, particularly if the number of cross sections is too small or if their locations are not "characteristic."

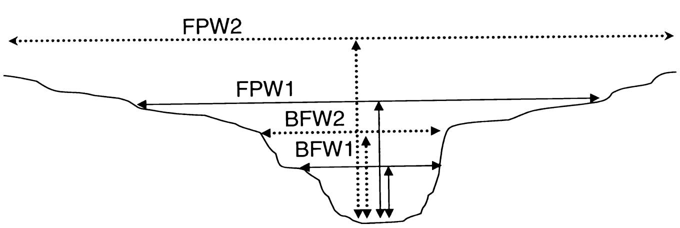

FIGURE 4. A Theoretical Cross-Section Analyzed by Two Independent Observers. The first observer identifies the lower terrace as bankfull and measures width there (BFW1), while the second observer uses a slightly higher terrace (BFW2); maximum bankfull depths for each are shown by shorter vertical arrows. Doubling the maximum bankfull depth (longer vertical arrows) gives the elevation of Rosgen's flood-prone width. The flood-prone width of the first observer (FPW1) is approximately 2.5 times his/her bankfull width, while the flood-prone width of the second observer (FPW2) is not contained within this cartoon and could be many times greater than the observer's bankfull width, depending upon the extent of the floodplain.

Another cause for observer differences may have to do with the use of ratios. Two of Rosgen's classification parameters are ratios of measured values (entrenchment and width-to-depth). Ratios can either reduce or magnify differences between observers when the differences in the numerator and denominator of the ratio are disproportionate. In the above example for Big Creek, the large difference in flood-prone widths (136%) is reduced slightly (123%) when these values are normalized by bankfull width to calculate entrenchment ratio. This is due to the disproportionate and relatively smaller difference in bankfull widths compared to flood-prone widths between the two field crews (6% *vs.* 136%). Similarly, a 6% difference in observed bankfull width at Big Creek (3.15 *vs.* 3.33 m) is magnified to a 21% difference when these values are normalized by disproportionate differences in average bankfull depth (0.279 *vs.* 0.371 m; width-to-depth values of 11.29 *vs.* 8.98, respectively). Although ratios are commonly used in geomorphology for scaling processes and physical characteristics of landforms (e.g., Richards, 1982), they can distort observer differences in the underlying parameters, which may mask true differences, or exaggerate minor ones, as illustrated above. The use of ratios for the first two tiers of Rosgen's classification (entrenchment and width-to-depth, Figure 2) may facilitate and partially explain observer differences in identification of channel type.

Our findings suggest that measurement of bankfull channel geometry and classification parameters derived from it may be a primary source for observer differences. Identifying bankfull elevation from field indicators can be difficult due to a variety of reasons, such as irregularity of the floodplain surface, poorly defined banks, and uncertainty in differentiating between terraces and the active floodplain surface. Furthermore, there are numerous methods for defining bankfull, each of which may yield somewhat different results (Williams, 1978; Johnson and Heil, 1996). Tools are available to assist in field identification of bankfull geometry (USDA 1995, 2003, 2005), but bankfull measurements can have large uncertainties associated with them and in some cases may be subjective (Johnson and Heil, 1996). In addition, bankfull flow is typically defined as that which begins to spill out of the channel onto the floodplain, which by definition makes it applicable only to floodplain rivers. In practice, however, bankfull geometry is also measured in nonfloodplain rivers (e.g., confined step-pool and cascade channels, or Rosgen's A and G stream types) using bankfull-like indicators (discontinuous and irregular floodplain surfaces, vegetation limits, cut banks, flow staining on boulders and bedrock, etc.).

Within a monitoring group, observer variability in determining bankfull dimensions has been shown to be ±15% (Roper *et al.*, 2002). Furthermore, different monitoring groups which receive different training – even if they use the same protocol – will often consistently differ in the characteristics they use to identify the bankfull surface (Whitacre *et al.*, 2007), increasing the potential for between-group differences in both bankfull widths and depth. Even if bankfull depth was consistently estimated, differences among field crews can still occur in determining the bankfull and flood-prone widths due to observer variability in cross-section location and orientation; small angular differences in the trend of cross sections (orientation in the horizontal plane) can lead to large differences in width for wide channels and broad floodplains.

In contrast, we found only one minor difference in stream type due to field crew differences in the estimate of channel slope, an attribute which tends to be more consistently evaluated (Isaak *et al.*, 1999; Roper *et al.*, 2002). In this case, UC Crew 3 measured slope to be 0.0202 in Trail Creek, while Crews 1 and 2 measured slope to be 0.0195 and 0.0189, respectively. Although this is a small difference, the Rosgen (1994, 1996) classification system has no allowed variation for slope. The final outcome was that data from two of the crews yielded B4c channel types while data from the final crew indicated a B4 channel type.

The fact that this study constantly needed to incorporate the expected variation of classification parameters to ensure consistent identification of stream type indicates two potential problems for application of Rosgen's approach. First, without the allowable variation of classification parameters, the overall mean values of 5 of the 12 evaluated streams (>40%) did not fall into a primary stream type because they had entrenchment ratios between 1.4 and 2.2 and sinuosity less than 1.2 (Table 2). Although our sample of streams was fairly small and not strictly randomized, this calls into question the suitability of the parameter ranges used for classifying Rosgen stream types, and whether those ranges adequately describe typical values for a given stream type as intended (Rosgen, 1994, 1996). Rosgen's use of allowable parameter variation recognizes that all classification systems necessarily impose artificial boundaries that may not fully capture the range of natural variability for a given channel morphology and was intended to provide a continuum of channel types by allowing "fuzzy" boundaries between channel types (Rosgen, 1996). However, because more than 40% of our sites fell in "the gray zone" between channel types, the specified parameter ranges do not seem representative, at least in northeastern Oregon. Another troubling aspect of this classification system is that the sinuosity criterion for stream types B, C, G, and F is 1.2 with an allowable variation of 0.2. The result is a criterion that will be met by all streams (sinuosity ≥1); not an insightful trait for an attribute used for classification.

Communication

It is clear from the widespread use of the Rosgen (1994, 1996) classification system that the general descriptions of the stream types used in this approach resonate with field practitioners. For example, most practitioners can quickly visualize a stream which fits the description of C channels; "low gradient, meandering, point-bar, riffle/pool, alluvial channels with broad well-defined floodplains" (Rosgen, 1994). Yet, we found that how one operationally defines attributes in a classification system not only affects repeatability among observers, but can also alter classification in a manner so as to be incongruent with the visualized ideal.

This was readily observable within the four streams with slope greater than 0.05 (Crawfish, Indian, Myrtle, and Whiskey). In 30 of the 33 observations for these channels, they were classified as B streams (moderate-gradient, riffle-dominated channels according to Rosgen (1994) Table 2), rather than as A streams (steep, cascading, step-pool channels according to Rosgen). In most cases, this was because the observed entrenchment ratio was close to or greater than 1.4 and width-to-depth ratios were greater than 12, which forced classification as B channels, despite slopes steep enough for classification as A channels. Inspection of these sites clearly shows that they are steep, step-pool channels (Figure 5), more akin to what is described for Rosgen's A channel type, than the B channel type that results from the measured attributes. This suggests that even when reported data consistently yield the same major stream type, it might differ from what practitioners are visualizing when they talk about that stream type. Although it is unclear whether this disconnect is due to problems with the classification system or its application (observer training and field methods), it is clear that the intended communication of channel condition is not being achieved in this test of the approach.

The difference between classified channel types and their observed appearance may partially reflect our rule set, which sought to maximize consistency

FIGURE 5. Crawfish Creek, a Steep Step-Pool Channel (Rosgen's A stream type) That Five of Eight Times Was Misclassified by Observer Measurements as a B Channel Type, Which Rosgen (1994) Describes as a Moderate-Gradient, Riffle-Dominated Channel.

among field crews within a monitoring group. Using this rule set it is possible that misclassification relative to the observed appearance increased if one or more of the crews in a monitoring group made measurements that did not reflect the stream's true condition. For example, the three UC crews found the entrenchment ratio in West Fork Lick Creek to be 1.64, 1.31, and 1.21, with a width-to-depth ratio greater than 12 and sinuosity greater than 1.2, resulting in two F channels and one B. Because the critical entrenchment ratio separating B and F channel types is 1.4, and allowable variation is 0.2, the only way to get agreement among all crews is to add 0.2 to the pair of lower estimates rather than subtract 0.2 from the high observation (1.64-0.2 is still greater than 1.4), yielding classification as a B stream type for consistency among crews. If the real stream type had been an F, this rule set for consistency would result in three incorrect determinations (all B), when there may have been only one misclassification without applying our consistency rule. While we acknowledge that our consistency rule may create errors of this sort, some type of rule set is required to decide between nonunique solutions that result from Rosgen's allowed variation of classification parameters (Figure 2). Visual assessment of the reach morphology is likely the best way to decide between nonunique solutions. However, the purpose of measuring channel characteristics for classifying stream type is undermined if in the end stream type is determined by visual assessment.

Performing the classification in the field might also reduce misidentification of stream type by providing visual verification of the assigned stream type. For example, if the resultant field classification of stream type differed from the observed morphology, the field crew would have the option of re-evaluating their measurements. However, this would make visual identification of channel type the primary classification tool, which is not how the Rosgen method has been presented; it is a parameter-based classification tree, rather than a visual method.

Summary

Based on the above analysis, it is evident that Rosgen's (1994, 1996) use of entrenchment ratio and width-to-depth ratio as primary attributes in classification undermines consistent application and interpretation of this classification system because of the sensitivity of those parameters to identification of bankfull. Small differences among field crews in determining bankfull depth can have a large effect on the resultant classification of stream type, regardless of how well those determinations fit Rosgen's more generalized description of those stream types. So while it may be helpful if individuals in different aquatic professions are able to communicate stream type quickly using this classification system, we need to be mindful of observer variation and the compromises implicit in any classification system (for further discussion of this issue see Kondolf, 1995; Juracek and Fitzpatrick, 2003; Kondolf *et al.*, 2003; and Downs and Gregory, 2004).

CONCLUSION

We found that application of the Rosgen (1994, 1996) classification system at our study sites (1) resulted in inconsistent determination of stream type among observers, (2) presented no clear guidance for determining stream type when more than one was possible, and (3) often ended up assigning streams to types that did not fit the generalized appearance of the evaluated stream. The Rosgen (1994, 1996) classification system, therefore, appears to do little to improve communication among practitioners beyond what the raw measures of channel attributes would have done. If the objective of collecting stream data is to evaluate patterns among a number of stream reaches (as is the intent in many large-scale aquatic monitoring programs), we suggest analysis might be better served by using the raw data and statistical techniques to model general stream processes (Kershner *et al.*, 2004b), rather than stratify based on Rosgen stream type. In contrast, if the evaluation is to be conducted for a single stream reach, reliance on a potentially variable determination of Rosgen stream type (due to observer and methodological biases) may not provide the necessary information, either in time or space, to make sound recommendations concerning the state of a specific stream reach (Kondolf *et al.*, 2003).

Reducing observer differences in bankfull estimates through increased training and use of consistent protocols across monitoring groups would likely decrease observer bias in determining Rosgen stream types. However, other difficulties identified in our study [the repeated need to incorporate allowable parameter variation in order to classify streams, differences between classified stream type and observed morphology, and lack of guidance for cases where measurements yield nonunique solutions (multiple stream classes)] have less clear solutions for successful application of the Rosgen classification. No classification is perfect or infallible, but it is important to quantify observer variability and associated uncertainty in the Rosgen approach given its widespread use and acceptance despite few formal tests of the method.

ACKNOWLEDGMENTS

We thank all the monitoring groups involved with collecting data within the John Day Basin, as well as the Pacific Northwest Aquatic Monitoring Partnership, which provided the energy necessary to make this project possible. We also thank four anonymous reviewers for comments that substantially increased the quality of this manuscript. Funding for this project was provided by the Bonneville Power Administration, NOAA Fisheries, Forest Service, Bureau of Land Management, the Environmental Protection Agency, Northwest Indian Fisheries Commission, and the states of Oregon, Washington and California.

LITERATURE CITED

Brierley, G.J. and K.A. Fryirs, 2005. Geomorphology and River Management. Blackwell, Oxford, UK.

Buffington, J.M. and D.R. Montgomery, 1999. Effects of Hydraulic Roughness on Surface Textures of Gravel-bed Rivers. Water Resources Research 35:3507-3522.

Buffington, J.M., R.D. Woodsmith, D.B. Booth, and D.R. Montgomery, 2003. Fluvial Processes in Puget Sound Rivers and the Pacific Northwest. In: Restoration of Puget Sound Rivers, D.R. Montgomery, S. Bolton, D.B. Booth, and L. Wall (Editors). University of Washington Press, Seattle, Washington, pp. 46-78.

Downs, P.W. and K.J. Gregory, 2004. River Channel Management. Oxford University Press, New York, New York.

Downs, P.W. and G.M. Kondolf, 2002. Post-Project Appraisal in Adaptive Management of River Channel Restoration. Environmental Management 29:477-496.

Doyle, M.W. and J.M. Harbor, 2000. Discussion of "Evaluation of Rosgen's Streambank Erosion Potential Assessment in Northeast Oklahoma". Journal of the American Water Resources Association 36:1191-1192.

Ebersole, J.L., W.J. Liss, and C.A. Frissell, 1997. Restoration of Stream Habitats in the Western United States: Restoration as Reexpression of Habitat Capacity. Environmental Management 21:1-14.

Environmental Protection Agency, 2006. Fundamentals of Rosgen Stream Classification; Watershed Academy web. http://www.epa.gov/watertrain/stream_class/, *accessed* February 5, 2008.

Hillman, T.W., 2004. Monitoring Strategy for the Upper Columbia Basin. BioAnalysts, Inc. Eagle, Idaho. Prepared for Upper Columbia Regional Technical Team; Upper Columbia Salmon Recovery Board, Wenatchee, Washington, 103 pp.

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, P.A. and T.M. Heil, 1996. Uncertainty in Estimating Bankfull Conditions. Water Resource Bulletin 32:1283-1290.

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.

Juracek, K.E. and F.A. Fitzpatrick, 2003. Limitations and Implications of Stream Classification. Journal of the American Water Resources Association 39:659-670.

Kershner, J.L., E.K. Archer, M. Coles-Ritchie, E.R. Cowley, R.C. Henderson, K. Kratz, C.M. Quimby, D.L. Turner, L.C. Ulmer, and M.R. Vinson, 2004a. Guide to Effective Monitoring of Aquatic and Riparian Resources. Gen. Tech. Rep. RMRS-GTR-121, U.S. Department of Agriculture, Rocky Mountain Research Station, Fort Collins, Colorado, 57 pp.

Kershner, J.L., B.B. Roper, N. Bouwes, R.C. Henderson, and E.A. Archer, 2004b. A Comparison of Stream Reach Habitat Characteristics in Managed and Reference Watersheds on Federal Lands Within the Columbia Basin. North American Journal of Fisheries Management 24:1353-1365.

Kondolf, G.M., 1995. Geomorphological Stream Channel Classification in Aquatic Habitat Restoration: Uses and Limitations. Aquatic Conservation. Marine and Freshwater Ecosystems 5:127-141.

Kondolf, G.M., D.R. Montgomery, H. Piégay, and L. Schmitt, 2003. Geomorphic Classification of Rivers and Streams. In: Tools in Fluvial Geomorphology, G.M. Kondolf and H. Piegay (Editors). pp. 171-204. John Wiley and Sons, New York, New York, pp. 171-204.

Kondolf, G.M., M.W. Smeltzer, and S. Railsback, 2001. Design and Performance of a Channel Reconstruction Project in a Coastal California Gravel-bed Stream. Environmental Management 28:761-776.

Lanigan, S., B. Roper, J.M. Buffington, E. Archer, S. Downie, J. Faustini, S. Hubler, K. Jones, G. Merritt, D. Konnoff, A. Pleus, M. Ward, and K. Wolf, 2006. Pacific Northwest Side-by-Side Protocol Comparison Test. 5th National Water Quality Monitoring Conference, Monitoring Networks: Connecting for Clean Water, San Jose, California.

Leopold, L.B., M.G. Wolman, and J.P. Miller, 1964. Fluvial Processes in Geomorphology. Freeman, San Francisco, California.

Malakoff, D., 2004. The River Doctor. Science 305:937-939.

Miller, J.R. and J.B. Ritter, 1996. Discussion. An Examination of the Rosgen Classification of Natural Rivers. Catena 27:295-299.

Montgomery, D.R. and J.M. Buffington, 1997. Channel-Reach Morphology in Mountain Drainage Basins. Geological Society of America Bulletin 109:596-611.

Montgomery, D.R. and J.M. Buffington, 1998. Channel Processes, Classification, and Response. In: River Ecology and Management, R.J. Naiman and R.E. Bilby (Editors). Springer-Verlag, New York, New York, pp. 13-42.

Reeves, G.H., D.B. Hohler, D.P. Larsen, D.P. Busch, K. Kratz, K. Reynolds, K.F. Stein, P. Hays, and M. Tehan, 2004. Effectiveness Monitoring for the Aquatic and Riparian Component of the Northwest Forest Plan: Conceptual Framework and Options. General Technical Report PNW-GTR-577. U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, Portland, Oregon, 71 pp.

Richards, K., 1982. Rivers: Form and Process in Alluvial Channels. Methuen & Co., New York, New York.

Roper, B.B., J.L. Kershner, E. Archer, R. Henderson, and N. Bouwes, 2002. An Evaluation of Physical Habitat Used to Monitor Streams. Journal of the American Water Resources Association 38:1637-1646.

Rosgen, D.L., 1994. A Classification of Natural Rivers. Catena 22:169-199.

Rosgen, D.L., 1996. Applied River Morphology (Second Edition). Wildland Hydrology, Pagosa Springs, Colorado.

Savery, T.S., G.H. Belt, and D.A. Higgins, 2001. Evaluation of the Rosgen Stream Classification System in Chequamengon-Nicolet National Forest, Wisconsin. Journal of the American Water Resources Association 37:641-654.

Simon, A., M. Doyle, M. Kondolf, F.D. Sheilds, B. Rhoads, and M. McPhillips, 2007. Critical Evaluation of how the Rosgen Classification and Associated "Natural Channel Design" Methods Fail to Integrate and Quantify Fluvial Processes and Channel Response. Journal of the American Water Resources Association 43:1117-1131.

Smith, S.M. and K.L. Prestegaard, 2005. Hydraulic Performance of a Morphology-Based Stream Channel Design. Water Resources Research 41:W11413, doi: 10.1029/2004WR003926.

USDA, 1995. A Guide for Field Identification of Bankfull Stage in the Western United States. Rocky Mountain Research Station, Stream Systems Technology Center, video.

USDA, 2003. Identifying Bankfull Stage in Forested Streams in the Eastern United States. Rocky Mountain Research Station, Stream Systems Technology Center, video.

USDA, 2005. Guide to Identification of Bankfull Stage in the Northeastern United States. Rocky Mountain Research Station, Stream Systems Technology Center, CD.

Whitacre, H.W., B.B. Roper, and J.L. Kershner, 2007. A Comparison of Protocols and Observer Precision for Measuring Physical. Stream Attributes. Journal of the American Water Resources Association 43:923-937.

Williams, G.P., 1978. Bank-Full Discharge of Rivers. Water Resources Research 14:1141-1154.

Wolman, M.G., 1954. A Method of Sampling Coarse River-Bed Material. Transactions of the American Geophysical Union 35(6):951-956.