

A procedure for classifying textural facies in gravel-bed rivers

John M. Buffington¹ and David R. Montgomery

Department of Geological Sciences, University of Washington, Seattle

Abstract. Textural patches (i.e., grain-size facies) are commonly observed in gravel-bed channels and are of significance for both physical and biological processes at subreach scales. We present a general framework for classifying textural patches that allows modification for particular study goals, while maintaining a basic degree of standardization. Textures are classified using a two-tier system of ternary diagrams that identifies the relative abundance of major size classes and subcategories of the dominant size. An iterative procedure of visual identification and quantitative grain-size measurement is used. A field test of our classification indicates that it affords reasonable statistical discrimination of median grain size and variance of bed-surface textures. We also explore the compromise between classification simplicity and accuracy. We find that statistically meaningful textural discrimination requires use of both tiers of our classification. Furthermore, we find that simplified variants of the two-tier scheme are less accurate but may be more practical for field studies which do not require a high level of textural discrimination or detailed description of grain-size distributions. Facies maps provide a natural template for stratifying other physical and biological measurements and produce a retrievable and versatile database that can be used as a component of channel monitoring efforts.

1. Introduction

Bed surfaces of natural and laboratory gravel channels are frequently organized into distinct textural patches (i.e., facies) that are distinguished from one another by differences in grain size and sorting (Figure 1) [Iseya and Ikeda, 1987; Collins, 1988, 1990; Collins and Dietrich, 1988; Dietrich et al., 1989, 1993; Ferguson et al., 1989; Kinerson, 1990; Wolcott and Church, 1990; Lisle and Madej, 1992; Lisle et al., 1993; Paola and Seal, 1995; Powell and Ashworth, 1995; Kondolf, 1997; J. M. Buffington and D. R. Montgomery, Effects of hydraulic roughness on surface textures of gravel-bed rivers, submitted to *Water Resources Research*, 1999; hereinafter referred to as submitted paper]. Textural patches likely result from size-selective deposition or entrainment caused by spatial variations in shear stress, sediment supply, and lateral bed slope [Dietrich and Smith, 1984; Parker and Andrews, 1985; Dietrich et al., 1993]. Grain interactions, such as kinematic waves [Langbein and Leopold, 1968], intergranular friction angles [Miller and Byrne, 1966; Buffington et al., 1992; Johnston et al., 1998], relative grain protrusion [Kirchner et al., 1990], and grain wake effects [Iseya and Ikeda, 1987; Naden and Brayshaw, 1987; Whiting et al., 1988], also may play a role in textural patch development.

Patchy bed surfaces affect physical and biological processes within a stream reach. For example, textural patches affect bed load transport rates by creating patch-specific mobility thresholds that give rise to spatially nonsynchronous sediment motion and the appearance of size-selective transport [Lisle and

Madej, 1992; Paola and Seal, 1995]. Textural patches also influence bed load transport rates by providing spatially variable surface roughness [Dietrich et al., 1989]; fine patches with low surface roughness are expressways for coarse grains, while coarse patches with high surface roughness trap fine particles in downstream grain wakes [Iseya and Ikeda, 1987; Whiting et al., 1988] or in deep intergranular pockets [Buffington et al., 1992]. Differential roughness of textural patches also influences local boundary shear stress [Naot, 1984]. Patchy bed surfaces and spatial variability of physical environments have direct biological implications [Townsend, 1989], as many aquatic animals prefer specific substrate sizes [Cummins and Lauff, 1969; Reice, 1980; Kondolf and Wolman, 1993] and particular hydraulic regimes for different life stages [Sullivan, 1986].

In addition to their physical and biological significance, textural patches provide a natural template for stratifying sediment sample sites. In facies-stratified sampling, grain-size characteristics of each textural patch are quantified through random or systematic (i.e., grid/transect) sampling within patches (Figures 2a and 2b), and then weighted by patch area to determine reach-average grain-size statistics [Kinerson, 1990; Wolcott and Church, 1990; Lisle and Madej, 1992; Kondolf, 1997; J. M. Buffington and D. R. Montgomery, submitted paper, 1999]. Provided that each textural type is adequately sampled (both in terms of the number of observations per patch and the number of patches sampled per facies type), facies-stratified sampling can be a very accurate means of determining reach-average grain-size statistics, as it correctly weights particle sizes by their areal extent. Facies stratification also allows statistical comparison of between-patch differences in grain-size distribution and physical environment [Krumbein, 1953, 1960]. However, stratified sampling is a two-step process that requires a method of classifying bed-surface facies. Many

¹Now at the U.S. Geological Survey, Water Resources Division, Boulder, Colorado.

This paper is not subject to U.S. copyright. Published in 1999 by the American Geophysical Union.

Paper number 1999WR900041.

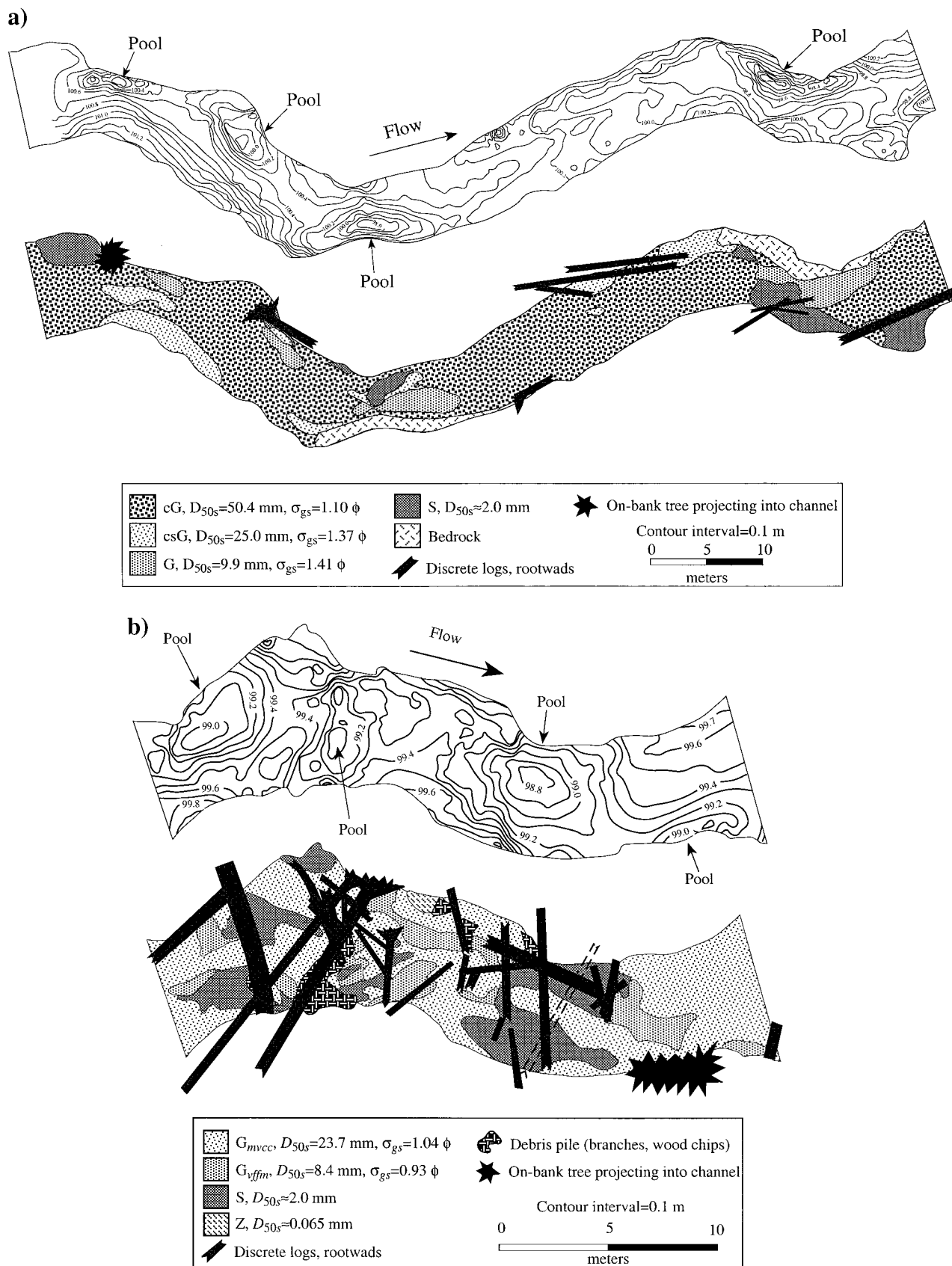


Figure 1. Textural and topographic maps of (a) Skunk Creek and (b) Mill Creek, forest pool-riffle channels of the Olympic Peninsula, western Washington. Each facies type is named according to our textural classification (discussed in text). D_{50s} is the patch, median, bed-surface grain size, and σ_{gs} is Folk's [1974] graphic standard deviation ($[\phi_{84} - \phi_{16}]/2$, where ϕ_{84} and ϕ_{16} are the $\log_2$ grain sizes [Krumbein, 1936] for which 16% and 84% of the surface grain sizes are finer).

investigators choose a simpler approach of unstratified random sampling (Figure 2c) or unstratified systematic sampling (Figure 2d). The ability of unstratified sampling to measure the true underlying sediment distribution depends on the areal coverage and density of sample sites, as well as the spatial distribution of textural patches (i.e., their size and frequency) [Smartt and Grainger, 1974; McCammon, 1975; Wolcott and Church, 1990]; too few observations and too large a sampling interval will produce a poor representation of the true underlying sediment population. For example, *Bevenger and King* [1995] recently proposed a pebble count procedure [Wolman, 1954] in which grains are selected from a transect that zig-zags bank-to-bank downstream through a channel reach to produce a reach-average grain-size distribution. While well intended, the proposed sampling coverage of this procedure would be insufficient to give the correct areal weighting of textures and their grain sizes in channels with complex spatial arrangements of textural patches, such as those of Figure 1b [Kondolf, 1997].

Although the occurrence of textural patches is fairly common in gravel-bed channels and is of significance to both physical and biological processes, there is no standard procedure for identifying and classifying bed-surface facies; a variety of approaches exist that vary in subjectivity and utility. Fisheries biologists commonly differentiate bed-surface textures visually [Platts *et al.*, 1983; Shirazi and Seim, 1981; Shirazi *et al.*, 1981] but do not have a formal textural classification scheme and rarely conduct accompanying grain-size measurements of the bed surface. Although visual differentiation is the primary means of classifying bed-surface textures, accompanying grain-size measurements are required to verify visual classification and reduce subjectivity [Kondolf and Li, 1992]. In contrast to visual identification, *Crowder and Diplas* [1997] recently proposed a grid sampling technique that uses a moving window comparison of mean grain size to locate textural boundaries. While more rigorous, the success of their approach depends on grid spacing (i.e., density of sample sites) and may be unnecessarily laborious in channels that have distinct textural boundaries that can be located more simply by visual inspection. Other methods of classifying textural patches are frequently site- or study-specific [Lisle and Madej, 1992; Paola and Seal, 1995]. In this paper, we present a standard procedure for classifying textural patches that combines visual identification and quantitative grain-size measurement. Our intent is to provide a general classification framework that can be modified, as needed, for particular study goals. We also examine the relative accuracy of different classification schemes and explore the trade-off between classification simplicity and accuracy.

2. Classifying Textural Patches

No single textural classification can satisfy all study goals. However, some degree of standardization for classifying bed-surface textures is appealing, particularly due to the growing number of studies dealing with textural patches [Collins, 1988, 1990; Collins and Dietrich, 1988; Dietrich *et al.*, 1989, 1993; Ferguson *et al.*, 1989; Kinerson, 1990; Lisle and Madej, 1992; Lisle *et al.*, 1993, 1997; Paola and Seal, 1995; Powell and Ashworth, 1995; Kondolf, 1997; J. M. Buffington and D. R. Montgomery, submitted paper, 1999]. Here, we present a classification procedure that is purposefully general, allowing study-specific adaptation, while maintaining a basic degree of standardization.

It is common practice in the earth sciences to use ternary

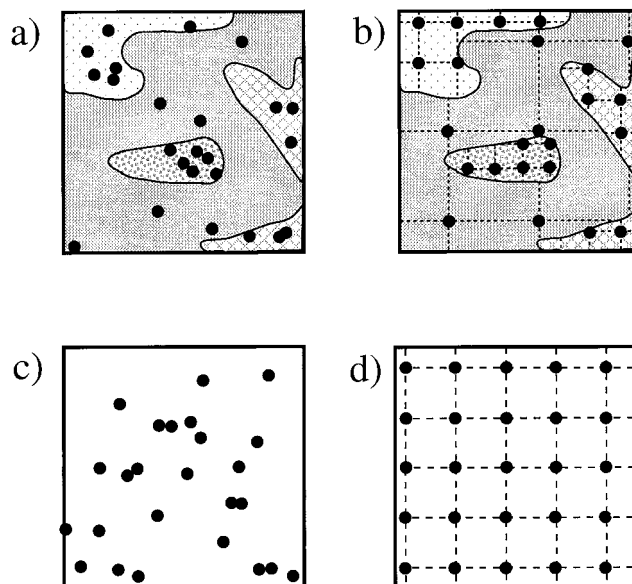


Figure 2. Cartoon illustrating basic strategies for sampling bed-surface material in a channel reach: (a) facies-stratified random, (b) facies-stratified systematic (grid/transect), (c) unstratified random, and (d) unstratified systematic. Textural patches are indicated by different fill patterns, and sample locations are shown by black dots. In Figure 2b the interval of sample sites within patches is scaled to individual patch area, with each patch having the same number of samples.

diagrams for classifying chemical composition of minerals [Deer *et al.*, 1982], mineral composition of rocks [Dietrich and Skinner, 1979; Williams *et al.*, 1982], and grain-size composition of sediments and soils [Krumbein and Pettijohn, 1938; Folk, 1954; Plumley and Davis, 1956; U.S. Department of Agriculture (USDA), 1962; Ritter, 1967]. Continuing this tradition, we propose a two-level ternary classification for textural patches (Figure 3) that uses standard grain-size divisions and names (Table 1). Ternary diagrams classify objects according to the relative abundance of three primary components. Each of our ternary diagrams is divided into 15 categories: six equal-area central fields that represent relative abundances of three primary grain sizes, and nine edge classes that represent end-member conditions with one or two dominant grain sizes, effectively reducing the system to a binary or unary classification. In constructing the ternary diagrams, we assumed equal probability of occurrence of different grain-size mixtures (i.e., no inherent mixture bias). Consequently, the central six fields within each ternary diagram are equal area and trisymmetrical.

Level I of the classification is used to classify the relative abundance of the three primary grain-size classes of a texture (i.e., silt, sand, gravel, cobble, or boulder) (Figure 3). For example, if a textural patch is predominantly composed of sand, gravel, and cobble, the fourth ternary of Figure 3 is used to classify the texture based on the relative abundance of these three primary size classes. To classify a texture, rank the three primary size classes from least to most abundant and name the texture according to that ranking (e.g., a texture with sand < gravel < cobble is classified as sgC, a sandy, gravelly cobble facies; the order of the adjectives (lower case letters of the classification) denotes the relative abundance of each of the subordinate size classes [Wentworth, 1922; Folk, 1954; Washburn *et al.*, 1963]). If a subordinate size class comprises less

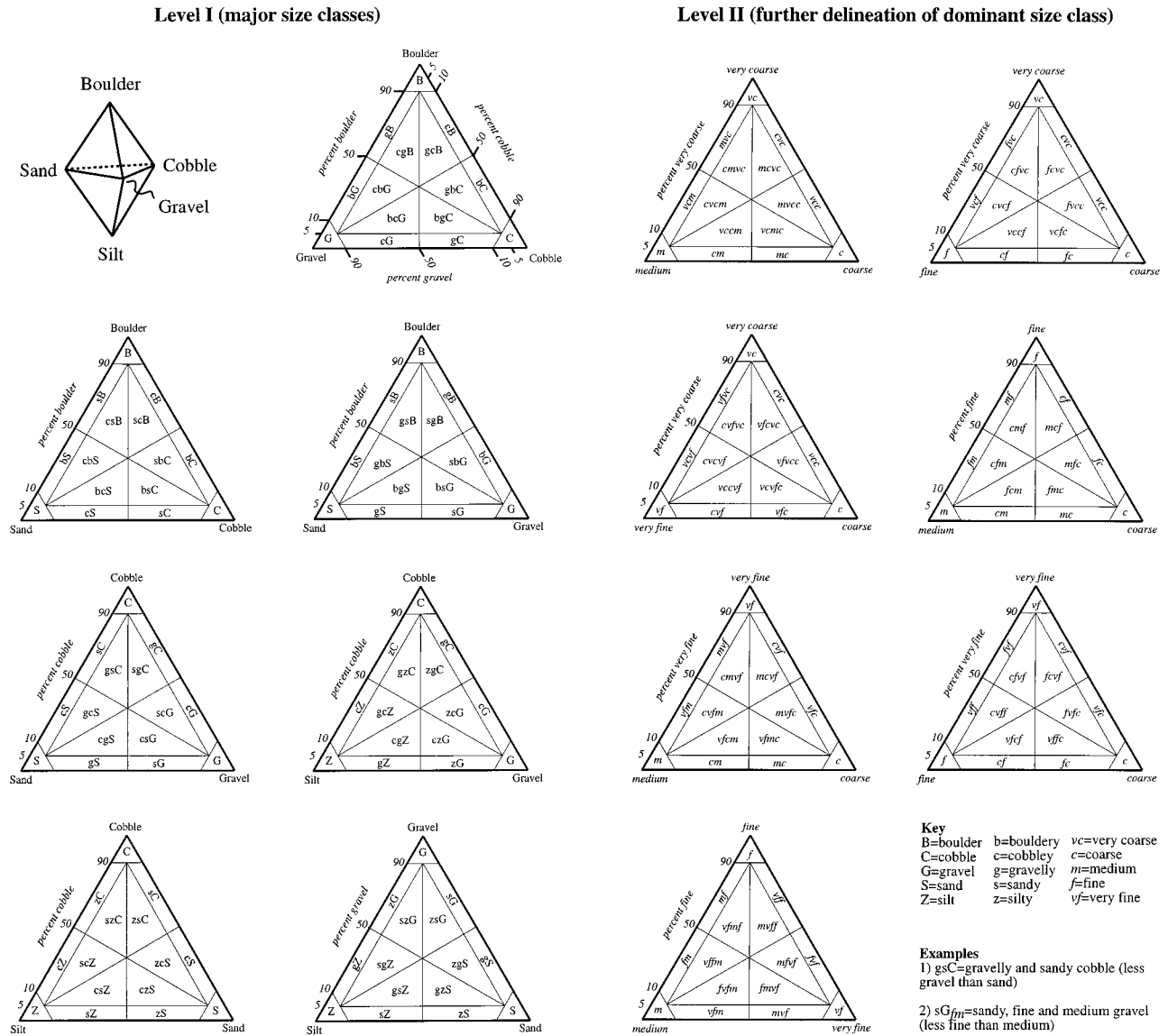


Figure 3. Two-level ternary classification for bed-surface facies. Level I decomposes the basic hexahedron of major size classes into seven ternary diagrams for classifying textures according to their relative abundance of three major size classes; standard grain-size names and divisions are used (Table 1). Level II further delineates the grain-size composition of the dominant size class (see figure text for symbol key and examples). Note that the figures are not drawn to scale.

than 5% of the relative abundance, it is considered negligible and is not used in naming the texture (in the above example, if there is <5% sand, then the texture is classified as gC, a gravelly cobble facies). Similarly, if both subordinate size classes together comprise less than 10% relative abundance, they are both omitted from the texture name (the above texture would then be classified as C, a cobble facies). These rules for classification are defined graphically by the ternary diagrams (Figure 3). However, once the above rules are learned, textures are easily classified without recourse to the diagrams.

The relative percentages that define the unary and binary fields (edge classes) within each ternary diagram are somewhat arbitrary. Here we use 5% relative abundance and 10% relative abundance as levels of significance for including subordinate size classes in textural names and defining binary and unary categories, as discussed above. Other textural classifica-

tions use limits of significance that range from 0.01 to 60% relative abundance for inclusion of subordinate size classes [USDA, 1904, 1962; Wentworth, 1922; Folk, 1954; Washburn *et al.*, 1963]. As there is no physical rationale or compelling historical precedent, the particular limits for defining unary and binary categories is a matter of personal preference.

Using similar ternary diagrams, level II of the classification further delineates the grain-size composition of the dominant size class (Figure 3). The purpose of the second level of the classification is to distinguish visually distinct textures that have the same level I name. For example, two gravel textures (G, >90% relative abundance, level I) might be distinguished as G_{fm} and G_c (fine to medium gravel versus uniformly coarse gravel) using the second level of the classification. Level II delineates boulder, cobble, and gravel subsizes only; fine-scale divisions of sand and silt classes are not differentiated here

because most surface-based sampling techniques (e.g., pebble counts [Wolman, 1954]) cannot realistically resolve those fine-scale divisions, nor are such divisions readily discernible by visual inspection. To minimize the number of level II ternary diagrams, the usual terms of “small” and “large” used to describe cobble and boulder subsizes are replaced by the terms “fine” and “coarse” in Table 1.

The above classification scheme is applied as follows to identify textural patches within a study area.

1. Conduct a preliminary reconnaissance of the stream reach, visually identifying textural patches according to Figure 3.

2. Map surface textural patches according to the names given in step 1. Discrete patches of a given textural type may occur several times within a reach, as illustrated in Figure 1.

3. Determine grain-size distributions for several patches of each facies type (see discussion below) and compare them with the visually estimated grain-size components of step 1. If there is a discrepancy, reclassify according to the measured grain-size distributions.

4. Group visually similar, but technically dissimilar, textures as equivalent for practical purposes (Figure 4). The imposed classification boundaries should not artificially separate textures that are visually and functionally similar.

The method outlined here is an iterative procedure of visual identification, quantitative grain-size measurement, and reclassification as needed.

In step 3 above, grain-size analysis of a given textural patch should encompass the entire surface area of the patch. The number of measurements per patch and the number of patches measured per facies type depend on the desired level of precision [Mosely and Tindale, 1985; Church et al., 1987; Rice and Church, 1996]. Choice of surface sampling technique (i.e., pebble count [Wolman, 1954], areal adhesion [Fripp and Diplas, 1993], photo-sieving [Ibbeken and Schleyer, 1986], etc.), as well as the strategy by which the technique is applied to a patch (i.e., random versus systematic sampling, Figures 2a and 2b),

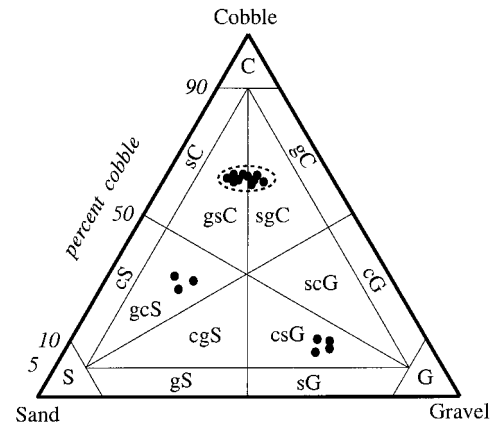


Figure 4. Technically dissimilar textures (sgC and gsC) grouped as functionally similar (circled points), but distinctly different from gcS and csG textures in the same reach. Figure not to scale.

may also influence sample precision and requisite size. Systematic (i.e., grid/transect) sampling is generally more accurate than random sampling [Smartt and Grainger, 1974; Wolcott and Church, 1990]; however, the performance of each sampling strategy is influenced by the size and frequency of the objects being sampled and the areal coverage of the sampling effort (i.e., density of observations) [Smartt and Grainger, 1974; McCammon, 1975; Wolcott and Church, 1990].

While there are a variety of surface sampling techniques to choose from, not all yield equivalent results [Leopold, 1970; Kellerhals and Bray, 1971; Potter, 1979; Church et al., 1987; Diplas and Sutherland, 1988]. Of the methods available, systematic (i.e., grid/transect) pebble counts [Wolman, 1954] are particularly attractive because they are easy to perform, relatively cheap (not much time and labor investment), and can be directly compared with subsurface samples sieved by weight [Kellerhals and Bray, 1971; Church et al., 1987; Diplas and Sutherland, 1988]. To minimize methodological differences between pebble counts and sieved samples, a gravelometer (square grain-size template [e.g., Hey and Thorne, 1983; Church et al., 1987]) should be used when conducting pebble counts.

Some bed surfaces exhibit continuous spatial gradients of grain size and sorting that make it difficult to identify discrete textural boundaries. Consequently, it may be necessary, in some cases, to classify bed surfaces as gradational from one textural composition to another. Regardless of whether bed surfaces are composed of punctuated (i.e., discrete) facies or gradational textures, our procedure for classifying bed surfaces will reduce subjectivity and observer bias.

3. Field Test

Demonstration of textural similarity within classification categories, as well as textural difference between categories, is required for acceptance of facies classification as a reliable means of quantifying textural variation within and between channel reaches. Our proposed textural classification distinguishes bed-surface facies based on differences in both grain size and sorting. Consequently, we test our classification by comparing grain-size distribution medians and variances within and between textural categories. In particular, we test for within-group similarity of both grain-size median and variance, and

Table 1. Standard Grain-Size Divisions and Names

Name	ϕ	Size, mm
Boulder		
Very coarse	−12 to −11	2048–4096
Coarse	−11 to −10	1024–2048
Medium	−10 to −9	512–1024
Fine	−9 to −8	256–512
Cobble		
Coarse	−8 to −7	128–256
Fine	−7 to −6	64–128
Gravel		
Very coarse	−6 to −5	32–64
Coarse	−5 to −4	16–32
Medium	−4 to −3	8–16
Fine	−3 to −2	4–8
Very fine	−2 to −1	2–4
Sand	−1 to 4	0.0625–2
Silt	4 to 8	0.0039–0.0625
Clay	≥ 8	≤ 0.0039

Udden [1898, 1914]–Wentworth [1922] grain-size scale adapted from Lane et al. [1947] and Church et al. [1987]. Note that we replace the usual cobble and boulder terms “small” and “large” with “fine” and “coarse” to minimize the number of ternary diagrams required for level II of our classification (Figure 3). Here ϕ is the Wentworth exponent, the standard log₂ unit of grain-size measurement [Krumbein, 1936]; $\phi = -\log_2 D$, where D is grain size in millimeters.

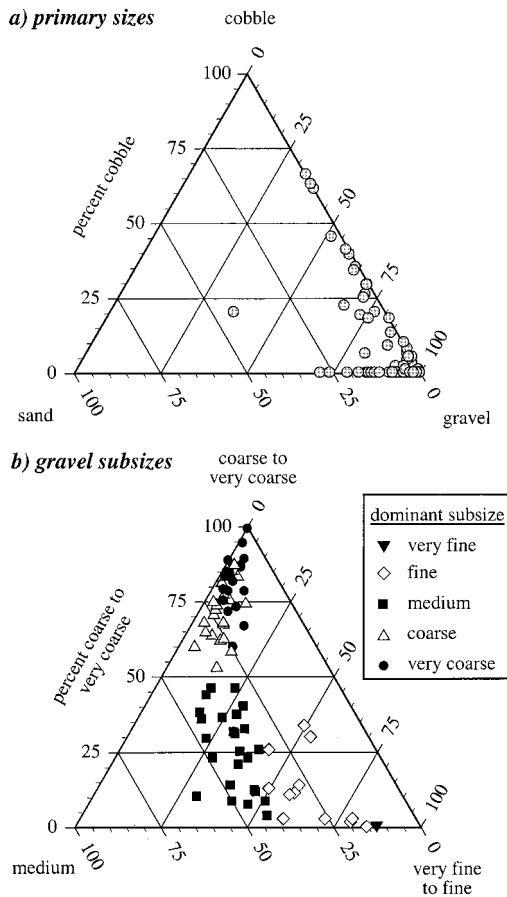


Figure 5. Grain-size composition of textures sampled at our study sites: (a) composition of primary size classes (level I) for all but two samples (see text), and (b) composition of gravel subsize (level II) (gravel is the dominant size class for 78 out of 83 textures sampled). Here, the seven, level II, gravel ternaries (Figure 3) are simplified to a single ternary diagram (Figure 5b), with the data stratified by the dominant subsize class.

between-group difference of one of these quantities. Because our classification distinguishes textures based on differences in either grain size or sorting, a difference of one of those quantities will suffice to demonstrate between-texture difference. For example, textures are distinguishable if they have different variances despite similar medians, or vice versa. We choose to compare grain-size distribution medians, rather than means, because the median value is a more robust measure of the central tendency of a distribution, particularly for coarse-grained fluvial sediments which rarely have either normal (Gaussian) or lognormal size distributions [Church and Kellerhals, 1978; Church et al., 1987; Rice and Church, 1998]. Similarly, we examine the variance about the median, rather than about the mean.

We also test whether more detailed classification schemes improve one's ability to discriminate statistically different textures. Specifically, we explore the potential compromise between classification simplicity and accuracy. To address this issue, we compare the statistical accuracy of our proposed classification to that of a simplified version (less categories per ternary), as well as to that of a complicated one (more categories per ternary). We further investigate the issue of simplic-

ity versus accuracy by examining whether a level I + II classification significantly improves one's ability to discriminate statistically different textures compared to that of a level I classification alone.

3.1. Field Sites

We examined the above questions using 83 bed-surface textures sampled in 17 gravel-bed rivers in northwestern Washington and southeastern Alaska; the study sites are located in forested mountain drainage basins and are further described by J. M. Buffington and D. R. Montgomery (submitted paper, 1999). Textural patches were classified in the field using a prototype procedure similar to that proposed here (Figure 3). Surface grain-size distributions of each texture were determined from patch-spanning, random pebble counts that sampled 100+ grains [Wolman, 1954]. All but two of the sampled textures plot in the sand-gravel-cobble ternary, with gravel being the most common size class (Figure 5a); the remaining two textures plot in the gravel-cobble-boulder ternary. While most of the sampled facies are gravel textures (78 out of 83), the subsize composition of the gravel is quite variable and defines a roughly arcuate band of occurrence (Figure 5b), the physical explanation for which is uncertain. Median grain sizes of the sampled textures have a flat-lying (platykurtic) distribution, predominantly composed of fine to very coarse gravel (Figure 6). In contrast, the distribution of grain-size variance is peaked (leptokurtic), with most patches composed of moderately well sorted to poorly sorted sediment (Figure 6). For this analysis, we are interested in the relative accuracy of different classification schemes and assume that the pebble counts, themselves, are sufficiently accurate.

3.2. Statistical Analysis

To assess the accuracy of different classification schemes, we reclassified the sampled bed-surface textures using 12, successively more complex approaches: three level I classifications, with ternaries divided into 15, 27, and 39 categories, respectively, (Figure 7); three level I + II classifications using the above level I ternaries combined with a unary level II classification (i.e., one that describes only the primary component of the dominant size class; e.g., a G_{fmc} texture would be classified as G_c in a unary level I + II scheme); three level I + II classifications like the previous, but with a binary level II classification (i.e., one that identifies the primary and secondary

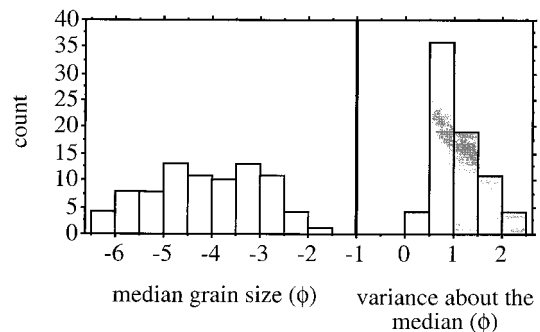


Figure 6. Distributions of grain-size median and variance for textures sampled at our study sites. Here we define sample variance about the median as $s'^2 = \sum_{j=1}^n (x_j - x')^2 / (n - 1)$ where x_j is the j th observation in a sample of size n and x' is the sample median (i.e., D_{50}).

components of the dominant size class; G_{mc} for the above example); and three complete level I + II classifications, using the 15-category, 27-category, and 39-category ternaries for both levels. We use a short hand representation for each of these classifications by indicating in parentheses the number of textural categories per ternary (e.g., I (15) + II (1) is a 15-category level I classification combined with a unary level II classification).

Our bed-surface textures tend to be positively skewed (Figure 8), indicating distributions shifted toward coarser sizes and having long fine tails, similar to surface textures observed in gravel-bed rivers of British Columbia [Rice and Church, 1998]. Because our grain-size distributions are asymmetrical, we use nonparametric statistical tests for comparing sample medians and variances, avoiding any formal assumption of distribution shape or parameter specification.

Median grain sizes were compared using a sign test [Conover, 1971; Ferguson and Takane, 1989]. To conduct the test, the samples of interest are combined to determine a grand median (X'). A contingency table is then constructed tallying the number of observations in each sample that are less than or equal to X' versus the number of observations that are greater than X' . The null hypothesis of equal medians is evaluated via a χ^2 statistic:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

where O_{ij} and E_{ij} are the observed and expected frequencies of grain sizes $\leq X'$ versus grain sizes $> X'$ for a contingency table composed of r rows and c columns. E_{ij} is defined as

$$E_{ij} = \frac{1}{n} \sum_{j=1}^c O_{ij} \sum_{i=1}^r O_{ij} \quad (2)$$

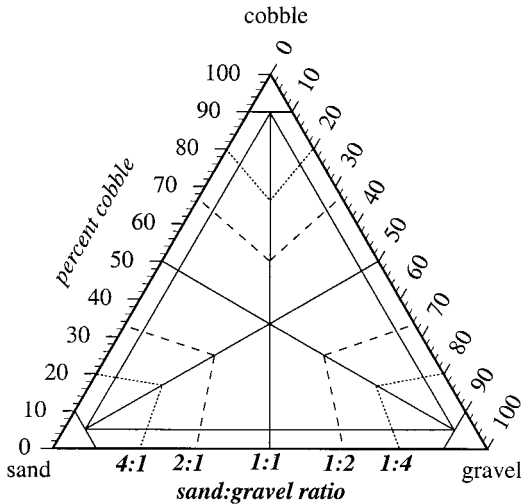


Figure 7. Sand-gravel-cobble ternary illustrating classification schemes using 15 categories (solid lines), 27 categories (solid lines plus long-dashed lines), and 39 categories (all lines shown). The 27-category classification subdivides the two dominant size classes at relative ratios of 1:2 and 2:1 (long dashes), while the 39-category classification additionally subdivides the two dominant size classes at relative ratios of 1:4 and 4:1 (short dashes).

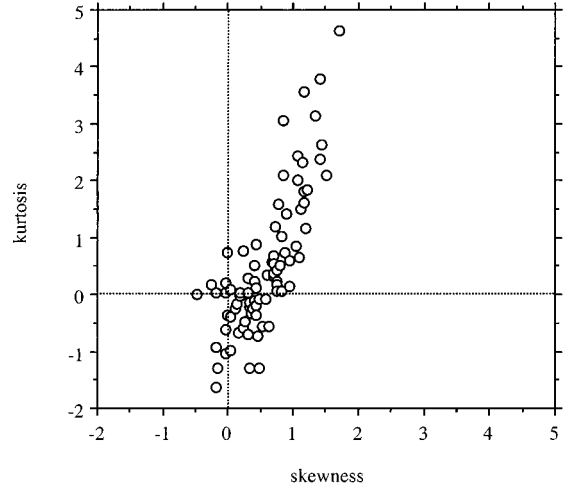


Figure 8. Skewness ($Sk = m_3/m_2^{3/2}$) and kurtosis ($K = (m_4/m_2^2) - 3$) of our sampled textures compared to a normal distribution ($Sk = K = 0$), where $m_2 = \sum_{j=1}^n (x_j - \bar{x})^2/n$, $m_3 = \sum_{j=1}^n (x_j - \bar{x})^3/n$, $m_4 = \sum_{j=1}^n (x_j - \bar{x})^4/n$, x_j is the j th observation in a sample of size n , and $\bar{x}$ is the mean grain size.

where n is the total number of observations for the combined samples. We conducted two-tailed tests of the null hypothesis (equal medians) at a significance level of $\alpha = 0.05$.

Sample variances were compared using Levene's [1960] test as modified by Brown and Forsythe [1974]. The procedure is a robust test for the equality of variances when sample distributions are nonnormal. The test statistic is a single-factor analysis of variance expressed in terms of sample medians:

$$W_0 = \frac{\sum_{i=1}^g n_i (\bar{z}_i - \bar{z}_{..})^2 / (g - 1)}{\sum_{i=1}^g \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2 / \sum_{i=1}^g (n_i - 1)} \quad (3a)$$

such that

$$z_{ij} = |x_{ij} - x'_i| \quad (3b)$$

$$\bar{z}_i = \sum_{j=1}^{n_i} z_{ij} / n_i \quad (3c)$$

$$\bar{z}_{..} = \sum_{i=1}^g \sum_{j=1}^{n_i} z_{ij} / \sum_{i=1}^g n_i = \sum_{i=1}^g \bar{z}_i / g \quad (3d)$$

where x_{ij} is the j th observation ($j = 1, \dots, n_i$) in the i th group (grain-size distribution) ($i = 1, \dots, g$) and x'_i is the median of the i th group. The null hypothesis of equal variances is evaluated by comparing W_0 to upper tail values of Fisher's [1928] F-distribution (f_α) with $g - 1$ and $\sum_{i=1}^g (n_i - 1)$ degrees of freedom; the null hypothesis is rejected when $W_0 > f_\alpha$ (here we choose a significance of $\alpha = 0.05$).

3.3. Results

For each of the 12 candidate classifications, we conducted 3403 tests for equality of grain-size medians and variances, comparing each sample to all of the others. Results are sum-

Table 2. Statistical Accuracy of Textural Classification

Classification Type	Median Grain Size*		Median Grain Size and Variance†	
	Percent of Within-Group Comparisons That Are Statistically Similar	Percent of Between-Group Comparisons That Are Statistically Different	Percent of Within-Group Comparisons That Are Statistically Similar	Percent of Between-Group Comparisons That Are Statistically Different
Level I (15 categories)	28	85	18	98
Level I (27)	27	84	17	97
Level I (39)	27	84	17	97
Level I (15) + II (1)	63	86	39	97
Level I (27) + II (1)	64	86	39	96
Level I (39) + II (1)	64	86	41	96
Level I (15) + II (2)	81	85	53	96
Level I (27) + II (2)	83	85	54	95
Level I (39) + II (2)	83	84	53	95
Level I (15) + II (15)	88	84	64	95
Level I (27) + II (27)	93	83	71	95
Level I (39) + II (39)	93	83	70	94

*Median grain sizes compared via two-sample sign tests evaluated using a χ^2 statistic. Two-tailed tests of the null hypothesis (equal medians) were conducted at a significance level of $\alpha = 0.05$. Note that even if all samples in a group had identical medians, one could only expect about 95% of the within-group comparisons to be “statistically similar” because the type I error rate (rejection of the null hypothesis when it is true) is 5% for $\alpha = 0.05$.

†Median grain sizes compared as per the footnote above. Texture variances compared via two-sample Brown-Forsythe test. One-tailed tests of the null hypothesis (equal variances) were conducted at a significance level of $\alpha = 0.05$. Here within-group similarity requires acceptance of the null hypotheses for both the sign test (equal medians) and the Brown-Forsythe test (equal variances), while between-group difference requires rejection of either null hypothesis. Because two tests are used for assessing within-group similarity, each evaluated at a significance level of $\alpha = 0.05$, the type I error rate for the combined tests is $1 - (1 - \alpha)^2 \approx 0.10$. Consequently, one could only expect about 90% of the within-group comparisons to be statistically similar.

marized in Table 2, with the first two columns presenting results for sample medians alone and the second two columns presenting results for sample medians and variances together. Our analysis demonstrates that level I classifications do not perform as well as level I + II classifications in terms of grouping statistically similar textures. For a level I classification of our data, only 27–28% of within-group sample medians are similar, and only 17–18% of both sample medians and variances are similar. However, within-group similarity doubles when level I classifications are supplemented with a unary level II classification and nearly triples when a binary level II classification is used (Table 2). A full, 15- to 39-category level I + II classification produces further, but less dramatic, improvement (last three entries, Table 2). Although statistical similarity of within-group textures increases from level I to level I + II, there is little difference among the 15- through 39-category classification schemes. For example, there is only a 5% difference in the within-group similarity of medians between the 15-category and 39-category level I + II classifications (88% versus 93%, Table 2). Our analysis also demonstrates that the classification schemes examined here discriminate median grain size of textures better than median and variance together. For example, 88% of within-group medians are statistically similar for a 15-category level I + II classification of our data, while only 64% of both within-group medians and variances are similar (Table 2). In contrast to the within-group comparisons, between-group comparisons show a slight decrease in accuracy from level I to level I + II but are relatively insensitive to the type of textural classification used (Table 2).

We compared the combined results of the sign tests and Brown-Forsythe tests of each classification (last two columns of Table 2) to one another to assess which classification schemes provide statistically significant improvements over the

others. The results of the sign tests and the Brown-Forsythe tests describe binomial distributions of significant versus non-significant observations, allowing comparison of the reported percentiles (Table 2) via one-tailed z-tests:

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\{p(1-p)[(1/n_1) + (1/n_2)]\}^{1/2}} \quad (4)$$

where z_0 is the test statistic, $\hat{p}$ is the proportion of interest, and p is the pooled proportion defined as $(x_1 + x_2)/(n_1 + n_2)$, where x is the number of observations out of n that define $\hat{p}$. Results of these comparisons are presented in Table 3, with $P \leq 0.05$ representing a significant change in percent accuracy of classification. The first 12 entries of Table 3 assess the significance of increasing the number of textural categories per ternary (i.e., 15, 27, 39) for a given type of classification (i.e., I, I + II(1), I + II(2), I + II), and demonstrate that, for our data, there is no statistical difference between the 15-category, 27-category, and 39-category classifications. The last nine entries of Table 3 assess the significance of increasing the number of levels used per classification (e.g., I versus I + II for a given number of textural categories per ternary). These data demonstrate that within-group similarity is significantly improved by greater classification detail, while between-group difference is either unchanged (Table 3, entries 13–15 and entries 19–21) or significantly degraded (Table 3, entries 16–18; see also Table 2).

It is important to note that because our statistical analyses are based on multiple tests the probability for type I errors (rejection of the null hypothesis when it is true) is greater than the nominal value of $\alpha = 0.05$, our chosen level of significance. In general, the greater the number of simultaneous comparisons, the greater the chance of type I errors. For example, the experiment-wise type I error rate for the 42 comparisons made

in Table 3 is $1 - (1 - \alpha)^{42} = 0.88$ for a nominal error rate of $\alpha = 0.05$. In other words, there is an 88% chance that a type I error will occur for a family of 42 tests each conducted at a significance level of $\alpha = 0.05$. Similarly, type I errors are a near certainty in Table 2, which reports the results of thousands of simultaneous comparisons between grain-size median and variance. The *Bonferroni* [1935, 1936] correction can be used to adjust α values for multiple comparisons and reduce the probability of type I errors. However, we choose not to use the Bonferroni correction because it reduces the power of the tests, increasing the potential for type II errors (acceptance of the null hypothesis when it is not true). Despite the high probability for type I errors in our analysis, it is the differences between reported proportions (Table 3), rather than the absolute significance of the tests or the absolute proportions reported, that are important for comparing different classification schemes.

3.4. Simplicity Versus Accuracy

Our analysis shows that level I classification is a poor discriminator of statistically significant differences in median grain size and variance of bed-surface textures. However, discriminatory power is significantly improved with a level I + II classification. Of the level I + II classifications examined, the full 15- through 39-category classifications are more accurate than either the unary level I + II classifications or the binary level I + II classifications (Table 2). But there is no statistical difference in classification accuracy among the 15- through 39-category schemes (Table 2). Consequently, the simpler, 15-category, level I + II classification (as proposed) is recommended over the more complicated 27-category and 39-category approaches. We emphasize, however, that the results of this analysis are specific to our particular data set, the generality of which remains to be tested.

While the complete level I + II classification is the most accurate, its complexity makes it difficult to apply in the field. In general, as the degree of finer-scale categorizing increases, the more opportunity one has for visually misclassifying a texture. Visual assessment of the relative proportions of level II size classes often is not very accurate, resulting in extra time spent in reclassifying visually identified textures after completion of surface sampling. Moreover, the fine-scale textural distinctions generated from the full level I + II classifications are not necessary for all studies. Consequently, we recommend the simpler, unary or binary, 15-category, level I + II classifications [I(15) + II(1) or I(15) + II(2)] for general field application, while reserving the full, 15-category, level I + II classification [I(15) + II(15)] for field studies that demand a high level of textural discrimination. The unary level I + II classification will likely involve the least amount of reclassifying but does not discriminate statistical differences among bed-surface textures as well as the binary level I + II classification (Table 2).

4. Discussion and Conclusion

Textural patches represent spatial differences in physical environments within a stream reach and provide a natural, easily discernible, stratification for sampling both physical and biological conditions. For example, surface and subsurface grain-size percentiles of textural patches are roughly correlated with one another in forest channels of western Washington (Figure 9a); coarser surface textures have correspondingly coarser subsurfaces (see figure caption for methodology). Con-

Table 3. Statistical Significance of Greater Classification Detail

Classification Comparison	P Value*	
	Within Groups	Between Groups
<i>More Divisions per Ternary</i>		
I (15) vs. I (27)	0.341	
I (15) vs. I (39)	0.288	(0.009)
I (27) vs. I (39)	0.439	0.179
I (15) + II (1) vs. I (27) + II (1)	0.469	0.127
I (15) + II (1) vs. I (39) + II (1)	0.270	0.074
I (27) + II (1) vs. I (39) + II (1)	0.300	0.380
I (15) + II (2) vs. I (27) + II (2)	0.467	0.238
I (15) + II (2) vs. I (39) + II (2)	0.482	0.178
I (27) + II (2) vs. I (39) + II (2)	0.451	0.416
I (15) + II (15) vs. I (27) + II (27)	0.133	0.094
I (15) + II (15) vs. I (39) + II (39)	0.166	0.079
I (27) + II (27) vs. I (39) + II (39)	0.449	0.460
<i>More Levels per Classification</i>		
I (15) vs. I (15) + II (1)	0.000	0.052
I (27) vs. I (27) + II (1)	0.000	0.125
I (39) vs. I (39) + II (1)	0.000	0.322
I (15) + II (1) vs. I (15) + II (2)	0.000	(0.002)
I (27) + II (1) vs. I (27) + II (2)	0.001	(0.009)
I (39) + II (1) vs. I (39) + II (2)	0.003	(0.011)
I (15) + II (2) vs. I (15) + II (15)	0.014	0.271
I (27) + II (2) vs. I (27) + II (27)	0.001	0.114
I (39) + II (2) vs. I (39) + II (39)	0.002	0.137

Here vs., versus.

*The *P* values reported here are for one-tailed z-tests (4) of proportions reported in the last two columns of Table 2. *P* values ≤ 0.05 and not in parentheses indicate that the latter classification is a significant improvement over the former in its ability to discriminate statistical differences in median grain size and variance of bed-surface textures. *P* values ≤ 0.05 and in parentheses indicate the latter classification significantly worsens discriminatory power relative to that of the former.

sequently, classification and mapping of surface facies produce a useful template for locating and stratifying subsurface sample sites. However, using surface grain-size composition to infer subsurface sizes is not recommended because surface textures may be draped by ephemeral, low-flow deposition of fine sediment that is not representative of the underlying subsurface material. The relationship between surface and subsurface grain sizes improves when the fine-grained, suspendable particles are removed from the grain-size distributions (Figure 9b; see figure caption for methodology).

Between-patch differences in grain size, sorting, shear stress, and sediment supply also create a practical template for stratifying biological measurements, such as species preference for spawning, feeding, and resting sites. For example, the 24 mm patches of Figure 1b should appeal most to spawning steelhead or chum salmon, while the 8 mm patches should appeal most to spawning brook trout [Kondolf and Wolman, 1993]. However, facies stratification is only one of many ways to stratify sample sites within a stream reach. Depending on the study goals and the hypotheses to be tested, it may be more desirable to stratify physical and biological measurements by factors such as flow depth, velocity, or channel unit morphology (i.e.,

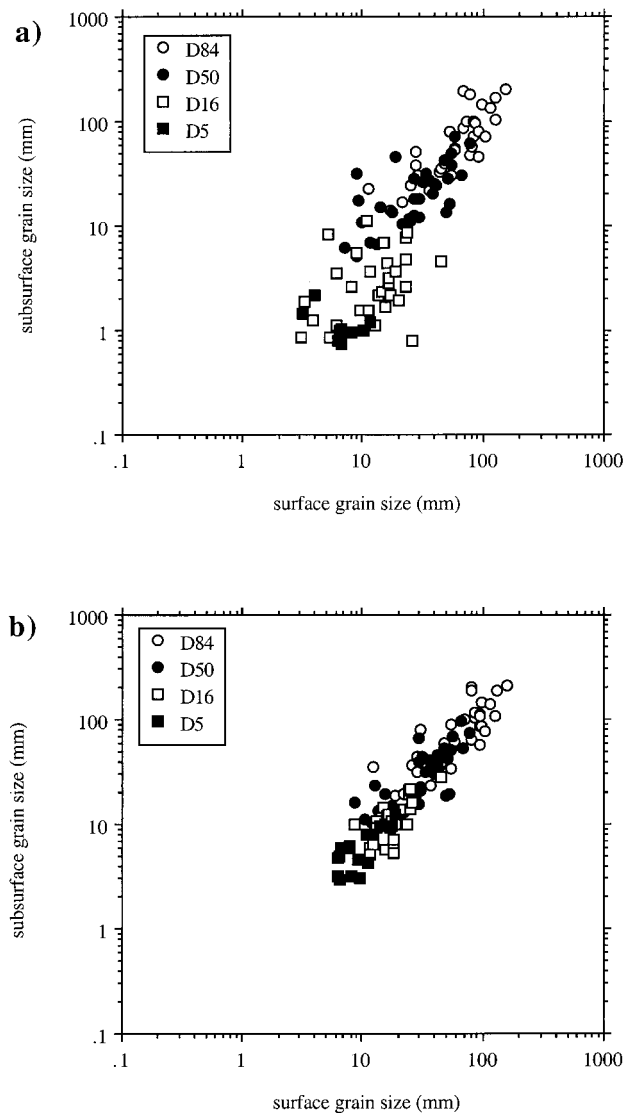


Figure 9. Comparison of surface and subsurface grain-size percentiles for textural patches in forest channels of the Olympic Peninsula, western Washington [data from Buffington, 1995]. D_{84} , D_{50} , D_{16} , and D_5 are the grain-size percentiles for which 84%, 50%, 16%, and 5% of the sizes are finer. Surface grain-size distributions were determined from patch-spanning, random, pebble counts [Wolman, 1954] of 100+ grains. Sub-surface grain-size distributions were determined from sieved bulk samples, following the Church *et al.* [1987] sampling criterion (i.e., the largest grain is $\leq 1\%$ of the total sample weight). In Figure 9b, particle sizes that are suspendable at bank-full stage are removed from the grain-size distributions to separate the bed load distribution from the suspended load distribution. The maximum suspendable size was calculated from Dietrich's [1982] settling velocity curves, assuming a Corey shape factor of 0.7, a Power's roundness of 3.5, and a settling velocity equal to the bank-full shear velocity.

types of pools, bars, riffles, and steps [Church, 1992; Wood-Smith and Buffington, 1996]).

Classification and mapping of textural patches provide an important and versatile database, in itself, regardless of whether such maps are used to structure further facies-stratified sampling. Textural mapping yields an easily under-

stood visual record of channel conditions (particularly when combined with topographic and morphologic maps (Figure 1)) from which a variety of data can be derived, such as subreach spawning-habitat availability (a function of grain size and sorting [Kondolf and Wolman, 1993]), patterns of sediment transport and dispersal [Dietrich and Smith, 1984], and textural response to sediment supply [Dietrich *et al.*, 1993] and hydraulic roughness, i.e., bars, wood (J. M. Buffington and D. R. Montgomery, submitted paper, 1999). Combined textural, topographic, and morphologic mapping produces a retrievable database that allows one to associate channel processes and morphologic response and is a defensible means of monitoring channel characteristics.

Our classification provides a standard method for identifying textural patches and conducting facies-stratified sediment sampling. While this method of sampling can be quite time consuming and laborious in channels with complex textural distributions (Figure 1b), it produces an accurate areal weighting of grain sizes. Unstratified sampling strategies (Figures 2c and 2d) are less time consuming and less costly alternatives for sampling sediments within a stream reach but are abstractions of the underlying, natural facies and may yield inaccurate results if there is insufficient areal coverage and density of sample sites. Furthermore, the underlying facies stratification and all of its uses, as discussed above, are, in some cases, unretrievable from unstratified sampling. Consequently, important channel characteristics and process insights may be hidden by economical, but abstract, unstratified sampling strategies. For example, textural mapping conducted in forest channels of western Washington demonstrates that the frequency and diversity of textural patches (and therefore potential diversity of aquatic habitat) is well correlated with the frequency of in-channel wood and its consequent form drag and forced shear stress divergence (J. M. Buffington and D. R. Montgomery, submitted paper, 1999). This insight would not be evident without textural mapping and facies-stratified sampling.

The results of our field test indicate that statistically meaningful textural classification requires a level I + II analysis (Tables 2 and 3); statistical differences in median grain size and variance of bed-surface textures are poorly represented by a level I classification alone. This result is not surprising given that the level I size classes (silt, sand, gravel, cobble, and boulder) each include a broad range of possible grain sizes and sortings (0.2–0.5 logarithmic orders of magnitude each, Table 1) and are therefore best used for large-scale, regional quantification of bed-surface texture. However, mechanistic studies of subreach-scale physical and biological processes require quantification of patch-scale variations of bed-surface texture and therefore use of a level I + II type classification. While a level I + II classification is more laborious than a level I classification, our analysis demonstrates that there is no statistical difference among the 15-category, 27-category, or 39-category textural classifications (Table 3, first 12 entries). Consequently, the simpler, 15-category scheme (as proposed) can be used without significant loss of classification accuracy.

Although a full level I + II classification produces better statistical discrimination of bed-surface textures, it is cumbersome and less practical than the simpler, unary and binary, level I + II schemes [I(15) + II(1) or I(15) + II(2)], both of which yield reasonably accurate results (Table 2). Consequently, we recommend these simpler approaches for general field application and suggest reserving the full 15-category level I + II classification [I(15) + II(15)] for studies that

require more detailed textural discrimination or more complete description of grain-size distributions. For example, two textural patches classified as medium gravel in a unary level I + II scheme may have very different fine- and coarse-gravel contents, the distinction and quantification of which may be of prime interest to fisheries biologists charged with assessing spawning gravel quality. The complete 15-category level I + II classification provides more detail about the relative abundance of component size classes in gravel textures than is available from level I and unary level I + II type classifications currently used by fisheries biologists [e.g., Shirazi and Seim, 1981; Platts *et al.*, 1983]. Consequently, it may offer better quantification of the physical differences driving habitat quality and use of the channel by aquatic animals.

While our data set samples 83 bed-surface textures from 17 rivers, the sampled textures are dominantly different subcategories of gravel (Figure 5a). Furthermore, the samples tend to plot along the edges of the sand-gravel-cobble ternary, indicating that our textures are dominated by one to two major grain sizes, rather than three; only one texture plots in the central portion of the ternary. The data also demonstrate that when there are two dominant size classes, they tend to be closely related (i.e., sand and gravel, or gravel and cobble, but not sand and cobble), suggesting a hydrologic control on the grain-size composition of patches. A similar hydrologic control on grain size is commonly observed at reach scales; there is a general downstream sequence of boulder-bed, cobble-bed, gravel-bed, and sand-bed morphologies that covaries with the downstream decline of channel slope and shear stress [e.g., Montgomery and Buffington, 1997]. If our data are generally representative, then most bed-surface textures may plot along the edges of the level I grain-size ternaries, making the number of commonly observed textural types less than the full suite of possible textures proposed in our classification.

We have not conducted an exhaustive test of textural classifications, nor do we attempt to present the optimal scheme for classifying bed-surface facies. Rather, our approach is a minimalistic one that uses as few textural categories as possible without compromising statistical accuracy of the classification, thereby producing an economical classification procedure that is more likely to be used. Moreover, we present a purposefully general classification framework that can be modified for particular study interests, while maintaining a basic degree of standardization.

Acknowledgments. This manuscript was inspired by facies maps and studies of bed-surface texture made by Laurel Collins. Financial support was provided by the Washington State Timber, Fish and Wildlife agreement (TFW-SH10-FY93-004, FY95-156) and the Pacific Northwest Research Station of the U.S. Department of Agriculture Forest Service (cooperative agreement PNW 94-0617). We appreciate the insightful reviews provided by Kristin Bunte, Jack Lewis, Tom Lisle, and Steve Rice.

References

- Bevenger, G. S., and R. M. King, A pebble count procedure for assessing watershed cumulative effects, *For. Serv. Res. Pap. RM-RP-319*, 17 pp., U.S. Dep. of Agric., Fort Collins, Colo., 1995.
- Bonferroni, C. E., Il calcolo delle assicurazioni su gruppi di teste, in *Studi in Onore del Professore Salvatore Ortu Carboni*, pp. 13–60, Tip. del Senato, Rome, 1935.
- Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità, *Pubbl. R. Ist. Super. Sci. Econ. Commer. Firenze*, 8, 3–62, 1936.
- Brown, M. B., and A. B. Forsythe, Robust tests for the equality of variances, *J. Am. Stat. Assoc.*, 69, 364–367, 1974.
- Buffington, J. M., Effects of hydraulic roughness and sediment supply on surface textures of gravel-bedded rivers, M.S. thesis, 184 pp., Univ. of Wash., Seattle, 1995.
- Buffington, J. M., W. E. Dietrich, and J. W. Kirchner, Friction angle measurements on a naturally formed gravel streambed: Implications for critical boundary shear stress, *Water Resour. Res.*, 28, 411–425, 1992.
- Church, M., Channel morphology and typology, in *The Rivers Handbook*, edited by P. Carlow and G. E. Petts, pp. 126–143, Blackwell, Cambridge, Mass., 1992.
- Church, M., and R. Kellerhals, On the statistics of grain size variation along a gravel river, *Can. J. Earth Sci.*, 15, 1151–1160, 1978.
- Church, M. A., D. G. McLean, and J. F. Wolcott, River bed gravels: Sampling and analysis, in *Sediment Transport in Gravel-Bed Rivers*, edited by C. R. Thorne, J. C. Bathurst, and R. D. Hey, pp. 43–88, John Wiley, New York, 1987.
- Collins, L. M., The shape of Wildcat Creek, park log, East Bay Reg. Park Dist., Oakland, Calif., 1988.
- Collins, L. M., Maps, documents and exhibits for U.S. Forest Service reserved water rights case, *Case W-8439-76*, Colo. Water Div. 1, Greeley, Colo., 1990.
- Collins, L. M., and W. E. Dietrich, The influence of hillslope-channel interaction on fish habitat in Wildcat Creek, California (abstract), *Eos Trans. AGU*, 69, 1225, 1988.
- Conover, W. J., *Practical Nonparametric Statistics*, 462 pp., John Wiley, New York, 1971.
- Crowder, D. W., and P. Diplas, Sampling heterogeneous deposits in gravel-bed streams, *J. Hydrol. Eng.*, 123, 1106–1117, 1997.
- Cummins, K. W., and G. H. Lauff, The influence of substrate particle size on the microdistribution of stream macrobenthos, *Hydrobiologia*, 34, 145–181, 1969.
- Deer, W. A., R. A. Howie, and J. Zussman, *An Introduction to the Rock-Forming Minerals*, 528 pp., Longman, White Plains, N. Y., 1982.
- Dietrich, R. V., and B. J. Skinner, *Rocks and Rock Minerals*, 319 pp., John Wiley, New York, 1979.
- Dietrich, W. E., Settling velocity of natural particles, *Water Resour. Res.*, 18, 1615–1626, 1982.
- Dietrich, W. E., and J. D. Smith, Bed load transport in a river meander, *Water Resour. Res.*, 20, 1355–1380, 1984.
- Dietrich, W. E., J. W. Kirchner, H. Ikeda, and F. Iseya, Sediment supply and the development of the coarse surface layer in gravel-bedded rivers, *Nature*, 340, 215–217, 1989.
- Dietrich, W. E., D. Kinerson, and L. Collins, Interpretation of relative sediment supply from bed surface texture in gravel bed rivers (abstract), *Eos Trans. AGU*, 74, 151, 1993.
- Diplas, P., and A. J. Sutherland, Sampling techniques for gravel sized sediments, *J. Hydrol. Eng.*, 114, 484–501, 1988.
- Ferguson, G. A., and Y. Takane, *Statistical Analysis in Psychology and Education*, 587 pp., McGraw-Hill, New York, 1989.
- Ferguson, R. I., K. L. Prestegard, and P. J. Ashworth, Influence of sand on hydraulics and gravel transport in a braided gravel bed river, *Water Resour. Res.*, 25, 635–643, 1989.
- Fisher, R. A., On a distribution yielding the error functions of several well known statistics, in *Proceedings International Mathematics Congress*, pp. 805–813, Univ. of Toronto Press, Toronto, 1928.
- Folk, R. L., The distinction between grain size and mineral composition in sedimentary-rock nomenclature, *J. Geol.*, 62, 344–359, 1954.
- Folk, R. L., *Petrology of Sedimentary Rocks*, 182 pp., Hemphill, Austin, Tex., 1974.
- Fripp, J. B., and P. Diplas, Surface sampling in gravel streams, *J. Hydrol. Eng.*, 119, 473–490, 1993.
- Hey, R. D., and C. R. Thorne, Accuracy of surface samples from gravel bed material, *J. Hydrol. Eng.*, 109, 842–851, 1983.
- Ibbeken, H., and R. Schleyer, Photo-sieving: A method for grain-size analysis of coarse-grained, unconsolidated bedding surfaces, *Earth Surf. Processes Landforms*, 11, 59–77, 1986.
- Iseya, F., and H. Ikeda, Pulsations in bedload transport rates induced by a longitudinal sediment sorting: A flume study using sand and gravel mixtures, *Geogr. Ann.*, 69A, 15–27, 1987.
- Johnston, C. E., E. D. Andrews, and J. Pitlick, In situ determination of particle friction angles of fluvial gravels, *Water Resour. Res.*, 34, 2017–2030, 1998.

- Kellerhals, R., and D. I. Bray, Sampling procedures for coarse fluvial sediments, *J. Hydrol. Eng.*, 97, 1165–1180, 1971.
- Kinerson, D., Bed surface response to sediment supply, M.S. thesis, 420 pp., Univ. of Calif., Berkeley, 1990.
- Kirchner, J. W., W. E. Dietrich, F. Iseya, and H. Ikeda, The variability of critical shear stress, friction angle, and grain protrusion in water worked sediments, *Sedimentology*, 37, 647–672, 1990.
- Kondolf, G. M., Application of the pebble count: Notes on purpose, method, and variants, *J. Am. Water Resour. Assoc.*, 33, 79–87, 1997.
- Kondolf, G. M., and S. Li, The pebble count technique for quantifying surface bed material size in instream flow studies, *Rivers*, 3, 80–87, 1992.
- Kondolf, G. M., and M. G. Wolman, The sizes of salmonid spawning gravels, *Water Resour. Res.*, 29, 2275–2285, 1993.
- Krumbein, W. C., Application of logarithmic moments to size frequency distribution of sediments, *J. Sediment. Petrol.*, 6, 35–47, 1936.
- Krumbein, W. C., Statistical designs for sampling beach sand, *Eos Trans. AGU*, 34, 857–868, 1953.
- Krumbein, W. C., The “geological population” as a framework for analysing numerical data in geology, *Liverpool Manchester Geol. J.*, 2, 341–368, 1960.
- Krumbein, W. C., and F. J. Pettijohn, *Manual of Sedimentary Petrography*, 549 pp., D. Appleton-Century, New York, 1938.
- Lane, E. W., C. Brown, G. C. Gibson, C. S. Howard, W. C. Krumbein, G. H. Matthes, W. W. Rubey, A. C. Trowbridge, and L. G. Straub, Report of the subcommittee on sediment terminology, *Eos Trans. AGU*, 28, 936–938, 1947.
- Langbein, W. B., and L. B. Leopold, River channel bars and dunes—Theory of kinematic waves, *U.S. Geol. Surv. Prof. Pap.*, 422-L, 20 pp., 1968.
- Leopold, L., An improved method for size distribution of stream bed gravel, *Water Resour. Res.*, 6, 1357–1366, 1970.
- Levene, H., Robust tests for equality of variances, in *Contributions to Probability and Statistics*, edited by I. Olkin, pp. 278–292, Stanford Univ. Press, Stanford, Calif., 1960.
- Lisle, T. E., and M. A. Madej, Spatial variation in armouring in a channel with high sediment supply, in *Dynamics of Gravel-Bed Rivers*, edited by P. Billi et al., pp. 277–293, John Wiley, New York, 1992.
- Lisle, T. E., F. Iseya, and H. Ikeda, Response of a channel with alternate bars to a decrease in supply of mixed-size bed load: A flume experiment, *Water Resour. Res.*, 29, 3623–3629, 1993.
- Lisle, T. E., J. M. Nelson, and M. A. Madej, Poor correlation between boundary shear stress and surface particle size in four natural, gravel-bed channels (abstract), *Eos Trans. AGU*, 78(46), Fall Meet. Suppl., 278, 1997.
- McCammon, R. B., On the efficiency of systematic point-sampling in mapping facies, *J. Sediment. Petrol.*, 45, 217–229, 1975.
- Miller, R. T., and R. J. Byrne, The angle of repose for a single grain on a fixed rough bed, *Sedimentology*, 6, 303–314, 1966.
- Montgomery, D. R., and J. M. Buffington, Channel-reach morphology in mountain drainage basins, *Geol. Soc. Am. Bull.*, 109, 596–611, 1997.
- Mosely, M. P., and D. S. Tindale, Sediment variability and bed material sampling in gravel-bed rivers, *Earth Surf. Processes Landforms*, 10, 465–482, 1985.
- Naden, P. S., and A. C. Brayshaw, Small- and medium-sized bedforms in gravel-bed rivers, in *River Channels: Environment and Process*, edited by K. S. Richards, pp. 249–271, Blackwell, Cambridge, Mass., 1987.
- Naot, D., Response of channel flow to roughness heterogeneity, *J. Hydrol. Eng.*, 110, 1568–1587, 1984.
- Paola, C., and R. Seal, Grain size patchiness as a cause of selective deposition and downstream fining, *Water Resour. Res.*, 31, 1395–1407, 1995.
- Parker, G., and E. D. Andrews, Sorting of bed load sediment by flow in meander bends, *Water Resour. Res.*, 21, 1361–1373, 1985.
- Platts, W. S., W. F. Megahan, and G. W. Minshall, Methods for evaluating stream, riparian, and biotic conditions, *For. Serv. Gen. Technol. Rep. GTR-INT-138*, 70 pp., U.S. Dep. of Agric., Ogden, Utah, 1983.
- Plumley, W. J., and D. H. Davis, Estimation of recent sediment size parameters from a triangle diagram, *J. Sediment. Petrol.*, 26, 140–155, 1956.
- Potter, K. W., Derivation of the probability density function of certain biased samples of coarse riverbed material, *Water Resour. Res.*, 15, 21–22, 1979.
- Powell, D. M., and P. J. Ashworth, Spatial pattern of flow competence and bed load transport in a divided gravel bed river, *Water Resour. Res.*, 31, 741–752, 1995.
- Reice, S. R., The role of substratum in benthic macroinvertebrate microdistribution and litter decomposition in a woodland stream, *Ecology*, 61, 580–590, 1980.
- Rice, S., and M. Church, Sampling surficial fluvial gravels: The precision of size distribution percentile estimates, *J. Sediment. Res., Sect. A*, 66, 654–665, 1996.
- Rice, S., and M. Church, Grain size along two gravel-bed rivers: Statistical variation, spatial pattern and sedimentary links, *Earth Surf. Processes Landforms*, 23, 345–363, 1998.
- Ritter, J. R., Bed-material movement, Middle Fork Eel River, California, *U.S. Geol. Surv. Prof. Pap.*, 575-C, 3 pp., 1967.
- Shirazi, M. A., and W. K. Seim, Stream system evaluation with emphasis on spawning habitat for salmonids, *Water Resour. Res.*, 17, 592–594, 1981.
- Shirazi, M. A., W. K. Seim, and D. H. Lewis, Characterization of spawning gravel and stream evaluation, in *Salmon-Spawning Gravel: A Renewable Resource in the Pacific Northwest?*, Rep. 39, pp. 227–278, State of Wash. Water Res. Cent., Pullman, 1981.
- Smartt, P. F. M., and J. E. A. Grainger, Sampling for vegetation survey: Some aspects of the behaviour of unrestricted, restricted, and stratified techniques, *J. Biogeogr.*, 1, 193–206, 1974.
- Sullivan, K., Hydraulics and fish habitat in relation to channel morphology, Ph.D. dissertation, 407 pp., Johns Hopkins Univ., Baltimore, Md., 1986.
- Townsend, C. R., The patch dynamics concept of stream community ecology, *J. N. Am. Benthological Soc.*, 8, 36–50, 1989.
- Udden, J. A., The mechanical composition of wind deposits, *Augustana Libr. Publ.* 1, 69 pp., Augustana Coll. and Theological Seminary, Rock Island, Ill., 1898.
- Udden, J. A., Mechanical composition of clastic sediments, *Bull. Geol. Soc. Am.*, 25, 655–744, 1914.
- U.S. Department of Agriculture (USDA), Instructions to field parties and descriptions of soil types, 198 pp., Bur. of Soils, Washington, D. C., 1904.
- U.S. Department of Agriculture (USDA), Soil survey manual, *U.S. Dep. of Agric. Handbook* 18, 437 pp., Washington, D. C., 1962.
- Washburn, A. L., J. E. Sanders, and R. F. Flint, A convenient nomenclature for poorly sorted sediments, *J. Sediment. Petrol.*, 33, 478–580, 1963.
- Wentworth, C. K., A scale of grade and class terms for clastic sediments, *J. Geol.*, 30, 377–392, 1922.
- Whiting, P. J., W. E. Dietrich, L. B. Leopold, T. G. Drake, and R. L. Shreve, Bedload sheets in heterogeneous sediment, *Geology*, 16, 105–108, 1988.
- Williams, H., F. J. Turner, and C. M. Gilbert, *Petrography*, 626 pp., W. H. Freeman, New York, 1982.
- Wolcott, J., and M. Church, Strategies for sampling spatially heterogeneous phenomena: The example of river gravels, *J. Sediment. Petrol.*, 61, 534–543, 1990.
- Wolman, M. G., A method of sampling coarse bed material, *Eos Trans. AGU*, 35, 951–956, 1954.
- Wood-Smith, R. D., and J. M. Buffington, Multivariate geomorphic analysis of forest streams: Implications for assessment of land use impact on channel condition, *Earth Surf. Processes Landforms*, 21, 377–393, 1996.

J. M. Buffington, U.S. Geological Survey, Building RL6, 3215 Marine Street, Boulder, CO 80303. (jbuffing@usgs.gov)
D. R. Montgomery, Department of Geological Sciences, University of Washington, Seattle, WA 98195.

(Received July 24, 1998; revised February 5, 1999; accepted February 8, 1999.)