AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Earth Surface

RESEARCH ARTICLE

10.1002/2014JF003191

This article is a companion to *Buscombe et al.* [2014] doi:10.1002/2014JF003189.

Key Points:
- Examines relationship between backscatter spectral properties and sediment types
- Slope, intercept, and integral of power law spectral form sensitive to sediments
- Uses this to successfully classify Colorado River bed sediment in Grand Canyon

Correspondence to:
D. Buscombe,
dbuscombe@usgs.gov

Citation:
Buscombe, D., P. E. Grams, and M. A. Kaplinski (2014), Characterizing riverbed sediment using high-frequency acoustics 2: Scattering signatures of Colorado River bed sediment in Marble and Grand Canyons, *J. Geophys. Res. Earth Surf.*, *119*, doi:10.1002/2014JF003191.

Received 23 APR 2014
Accepted 24 OCT 2014
Accepted article online 29 OCT 2014

Characterizing riverbed sediment using high-frequency acoustics: 2. Scattering signatures of Colorado River bed sediment in Marble and Grand Canyons

D. Buscombe[1], P. E. Grams[1], and M. A. Kaplinski[2]

[1]Grand Canyon Monitoring and Research Center, Southwest Biological Science Center, U.S. Geological Survey, Flagstaff, Arizona, USA, [2]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona, USA

Abstract In this, the second of a pair of papers on the statistical signatures of riverbed sediment in high-frequency acoustic backscatter, spatially explicit maps of the stochastic geometries (length and amplitude scales) of backscatter are related to patches of riverbed surfaces composed of known sediment types, as determined by georeferenced underwater video observations. Statistics of backscatter magnitudes alone are found to be poor discriminators between sediment types. However, the variance of the power spectrum and the intercept and slope from a power law spectral form (termed the spectral strength and exponent, respectively) successfully discriminate between sediment types. A decision tree approach was able to classify spatially heterogeneous patches of homogeneous sands, gravels (and sand-gravel mixtures), and cobbles/boulders with 95, 88, and 91% accuracy, respectively. Application to sites outside the calibration and surveys made at calibration sites at different times were plausible based on observations from underwater video. Analysis of decision trees built with different training data sets suggested that the spectral exponent was consistently the most important variable in the classification. In the absence of theory concerning how spatially variable sediment surfaces scatter high-frequency sound, the primary advantage of this data-driven approach to classify bed sediment over alternatives is that spectral methods have well-understood properties and make no assumptions about the distributional form of the fluctuating component of backscatter over small spatial scales.

1. Introduction

Conventional sampling for grain size of submerged sediment deposits (e.g., grabs, cores, and dredges) is costly and labor intensive. Video and photographic sampling is more cost effective and does not require time-consuming laboratory analyses, which allows sampling at greater frequency and coverage [*Rubin et al.*, 2007; *Van Rein et al.*, 2009; *Buscombe et al.*, 2014a]. However, the use of high-frequency (several hundred kilohertz) acoustic backscatter from swath-mapping systems to characterize sediment and classify by grain size [*Anderson et al.*, 2008; *Brown and Blondel*, 2009; *Brown et al.*, 2011; *Snellen et al.*, 2013] has the potential to provide near-complete coverage, which photographic sampling could not practically achieve, at least within the same time and with the same positional accuracy.

Such density and coverage of sampling are important for characterizing the sedimentary makeup of heterogeneous riverbeds [*Nelson et al.*, 2014], which consist of a patchwork of homogeneous or near-homogeneous sediment over scales ranging from less than one to several tens of square meters [*Dietrich and Smith*, 1984; *Paola and Seal*, 1995; *Nelson et al.*, 2009]. Bed sediment grain size between adjacent patches can vary by an order of magnitude. Capturing such variability with conventional physical or photographic sampling would be prohibitively costly and time consuming.

In shallow water, given the lack of robust classification techniques based on the physics of scattering for high-frequency multibeam systems [*Amiri-Simkooei et al.*, 2009], an alternative phenomenological approach based on statistical analysis of backscatter signals [*Brown and Blondel*, 2009; *Snellen et al.*, 2013] has become popular. Many such methods proposed to date rely on aggregation of data over scales much larger than the typical scales of sediment patchiness on heterogeneous riverbeds. In this contribution, we develop a new statistical method for acoustic sediment classification based on analysis of backscatter, which is both continuous in coverage and of sufficient resolution (order meter) to characterize sediment variability on

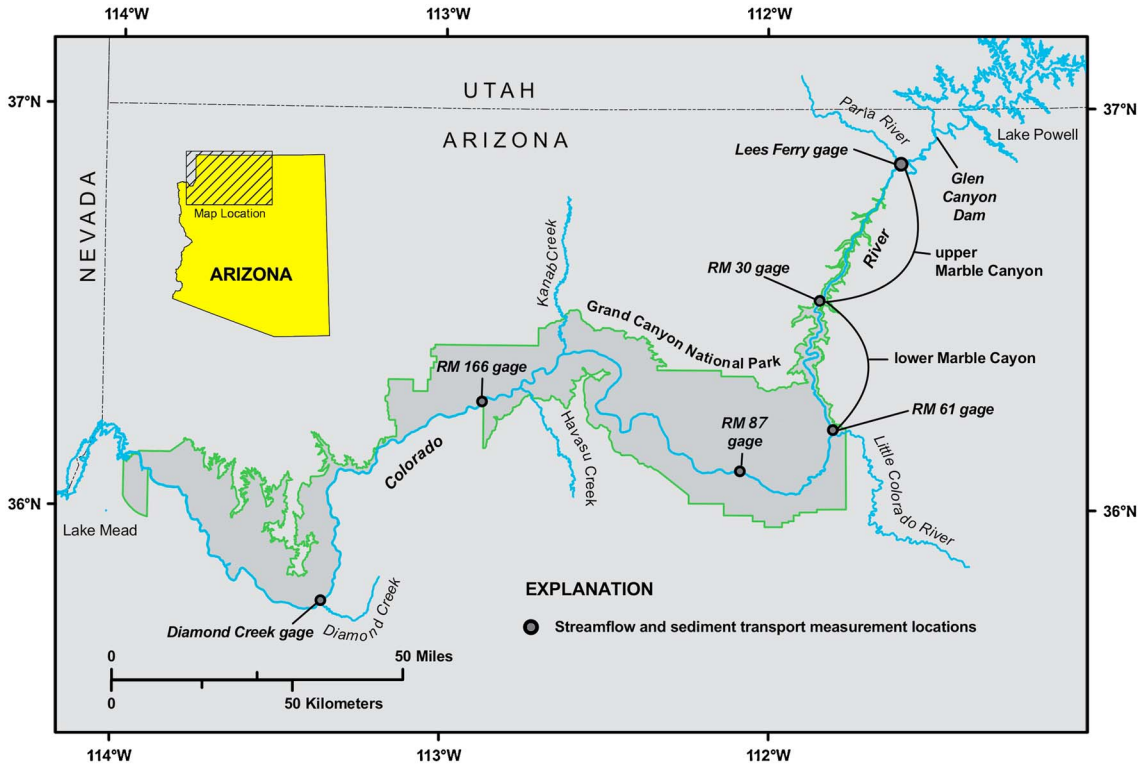

Figure 1. Map of the Colorado River from Glen Canyon Dam to Lake Mead, Arizona, showing streamflow and sediment-monitoring stations (circles). The reaches used for this study are around the RM30, RM61, and RM87 gages.

patchy riverbeds. Being based on backscatter spectra, the classification can be expressed in physical units and is therefore perhaps the most complementary, out of all statistical approaches proposed to date, to physical acoustic scattering models.

In part 1 of this two-part manuscript [*Buscombe et al.*, 2014b, hereinafter Part 1], we detailed general methods to geometrically and radiometrically correct raw high-frequency echoes recorded by the receive beams of a multibeam echo sounder (MBES) in shallow water over a sloping heterogeneous riverbed and described a generalized procedure to construct spatially explicit surfaces of spectral characteristics (length and amplitude scales) from gridded backscatter, using windowing. In this part, we apply these methods to MBES data collected from the bed of the Colorado River in Marble and Grand Canyons.

Sediment dynamics have been studied for several decades on the Colorado River in Grand Canyon National Park [*Howard and Dolan*, 1981; *Rubin et al.*, 2002; *Wright et al.*, 2008], with particular focus on sandbars that principally form in large eddies downstream of tributary debris fans [*Schmidt*, 1990]. Understanding the long-term trend of these sandbars, considered valuable resources by stakeholders and managers of the Colorado River in Grand Canyon, requires detailed monitoring of sand influx, efflux, and changes in sand storage [*Topping et al.*, 2000, 2010; *Grams et al.*, 2013] because of the severe limitations in sand supply imposed by Glen Canyon Dam [*Howard and Dolan*, 1981; *Topping et al.*, 2000; *Hazel et al.*, 2006].

Uncertainties in sand budgets constructed by measurements of sand flux at select gages (Figure 1) and by repeat measurements of bed morphology over multiyear periods [*Topping et al.*, 2010; *Grams et al.*, 2013] need to be constrained in order to determine the effectiveness of sandbar management practices, such as controlled floods [*Rubin et al.*, 2002; *Topping et al.*, 2006; *Wright et al.*, 2008; *Hazel et al.*, 2010] and the fate of sand in Grand Canyon over decadal timescales.

In order to estimate the sand fraction of the sediment budget, it is necessary to estimate the fraction of the morphological changes that are sand and the fraction that are coarser. *Grams et al.* [2013] classified sand and coarse substrates using topographic roughness derived from digital elevation models, but sand bed forms had to be delineated manually; validation was based on grain size observations with positional

Published 2014. American Geophysical Union.

uncertainties up to tens of meters, and the classification skill was estimated to be only 60–70%. One important way in which uncertainties in sand budgets can be constrained is to improve upon this surface sediment classification. Because riverbed morphology on the Colorado River in Grand Canyon is mapped—to a large extent—using MBES [*Kaplinski et al.*, 2009], the primary motivation for the present study is to examine how this might be achieved using completely automated methods based on MBES acoustic backscatter.

Additional motivation for mapping bed sediment at high resolution is the potential insight it may afford into the sediment transport mechanisms, flow, and timescales required for the retention of sand and the building and maintenance of sandbars. *Grams et al.* [2013], observing very high spatial variability of changes in sediment storage derived from repeat bathymetric surveys over timescales of months to years, concluded that local hydraulics exert stronger controls on morphodynamics than changes in upstream sand supply. In order to explore this idea further, it is necessary to examine localized bed sediment composition and the dynamics of grain size changes over time, in order to ascertain the relative flow and bed sediment controls on sediment transport [*Rubin and Topping*, 2001], such as periodic sediment inputs from minor tributaries, the areal coverage of sand [*Grams and Wilcock*, 2014], the dynamics governing persistence of sand patches [*Vericat et al.*, 2008; *Nelson et al.*, 2009], and sand resuspension from the interstices of coarser substrates [*Nelson et al.*, 1995; *Wilcock*, 1997].

This paper is organized as follows. First, we examine the spectral characteristics of backscatter from areas of known sediment type with a dual aim of characterizing the stochastic geometries of backscatter for individual sediment types and developing a rudimentary classification of riverbed sediment. Then, we test and evaluate two simple classification approaches: one based on linear least squares and another based on nonparametric decision tree learning. Finally, we discuss some implications of being able to characterize riverbed sediments at high resolution using swath sonar systems, before drawing some conclusions.

2. Study Area and Data

2.1. Sample Sites

We collected MBES soundings, backscatter, and underwater video camera images in August 2013 from three short (≈500–1000 m) reaches around sediment and flow-gaging stations (Figures 1 and 2). Site 1 (hereafter, RM30), in Upper Marble Canyon at river mile 30, is a relatively straight section of channel within the Redwall Limestone above Shinamu Wash. The mapped section is separated into two halves by a minor constriction, with a scour pool immediately downstream. The rest of the reach has a quasi-rectangular cross section and a relatively uncomplicated flow field. Site 2 (hereafter, RM60), in Lower Marble Canyon at river mile 60, is within the Tapeats Sandstone just above the confluence with the Little Colorado River. The mapped area is composed of two pools separated by a constriction associated with a debris fan. The upper half contains coarse sediment from a debris fan immediately upstream and an eddy-sandbar. The lower half also has a fairly complicated channel geometry including a riffle, bedrock ledges on river right, and a small separation and reattachment bar, in the sense of *Schmidt* [1990], in an eddy on river left with large "hummocky" (strongly three-dimensional) bed forms. Site 3 (hereafter, RM87), in Eastern Grand Canyon at river mile 87, near Phantom Ranch within the Upper Granite Gorge, comprises two sections separated by a riffle and debris fan on river left. The upper section is a relatively straight channel with sand bed forms at a variety of scales, and the lower section has a more complicated geometry with a recirculating eddy and a deep (≈30 m) pool. These sites were chosen to test scattering signatures of bed sediment because they collectively include a majority of the range of sediment, flow, geomorphic, and geological settings found in this canyon river.

2.2. Backscatter Processing

Backscatter data from these three survey reaches were processed using the methods presented in Part 1. A grid size of 0.25 × 0.25 m was chosen for construction of surfaces because this was the smallest scale at which there were, consistently across sites, multiple usable soundings (after quality control procedures, detailed in *Kaplinski et al.* [2014]). A window size of 25 × 25 m (i.e., 10,000 grid samples) was chosen based on examination of individual spectra: smaller windows tended to create nonlinear 1-D spectra, and autocorrelation analyses of elevations that suggested a window size of <30 m^2 was required to maximize the likelihood; the grain size signal is captured independent of the morphological signal. A step size of 0.25 m (1 grid cell) was chosen so the output surfaces were the same spatial resolution as the input surfaces. In this

Published 2014. American Geophysical Union.

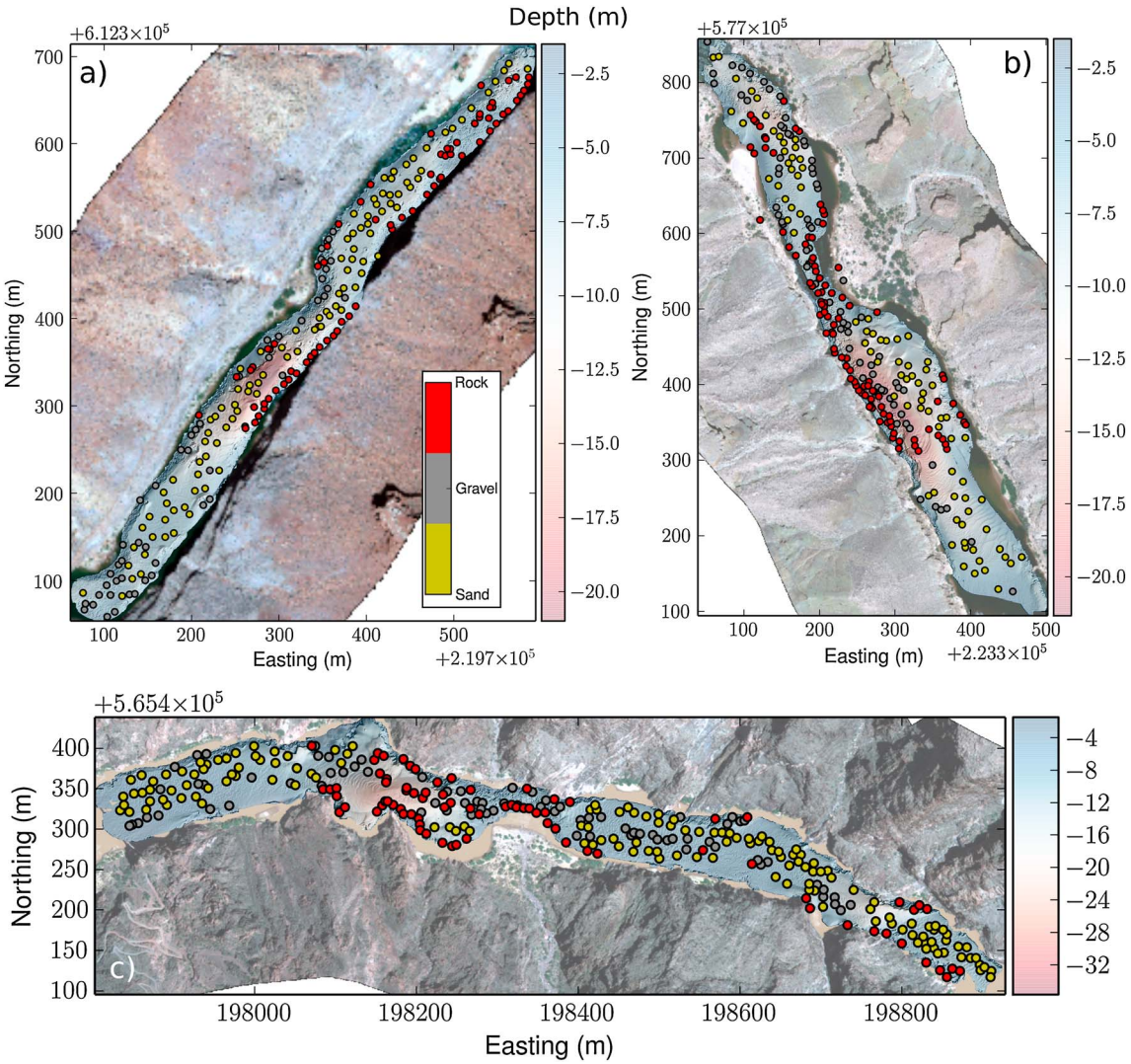

Figure 2. The three study reaches: (a) RM30, (b) RM60, and (c) RM87. Each site shows a (2009) aerial photograph overlain with a grey-shaded relief of the bathymetry, an alpha-map of water depths superimposed, and a set of video-based sediment observations (points) color-coded by sediment type.

configuration, the value in each output grid cell is the ensemble average of 1000 individual spectra for a window passing over that point.

The median backscatter value in each grid was chosen to construct the backscatter surface, because distributions of backscatter were unimodal yet non-normal, so the median was a better measure of central tendency than the mean. While maps of upper percentiles of the cumulative distribution over grids tended to show greater spatial heterogeneity in backscatter, we have no physical basis upon which to choose these higher percentiles and concluded that the greater heterogeneity was merely a reflection of small numbers of observations in the distribution tails (larger standard error on the mean).

For sediment attenuation corrections (Part 1) the median size and concentration of sand and concentration of silt were available every 15 min from gages in the same survey reach [*Topping et al.*, 2010]. It was assumed that the median silt grain size is 30 μm. It was further assumed that sand and silt are homogeneously mixed through the water column. This assumption is physically unreasonable for sand. However, sand suspensions were so dilute so as to make negligible difference to the sediment attenuation calculation, so the assumption of homogeneous mixing was made for numerical convenience. For water attenuation corrections, measurements of temperature and salinity were also available every 15 min from sondes at each gaging station.

Published 2014. American Geophysical Union.

2.3. Sediment Sampling

A cabled video-microscope system collected high-resolution (0.02 mm/pixel) images of the bed and consisted of a video camera fitted with macro lens inside an underwater housing, mounted inside a protective metal housing. The transparent faceplate of the camera housing, fitted with light-emitting diode lights for illumination, made contact with the bed. Data collection involved navigating to each location, turning the boat to face the current, and lowering the video to the bed using an electric winch. The video was relayed instantly to the winch operator. For more details about the system, see *Rubin et al.* [2007]. Coordinates of sample locations in the horizontal plane were measured by tracking the boat using a shore-based robotic laser-tracking system from established survey control points [*Kaplinski et al.*, 2009]. Those coordinates were taken for the camera location ($\pm$1 m in each horizontal direction) given that the horizontal offset between the winch sheave and the target prism was approximately equal to the horizontal offset in the opposite direction calculated from estimated catenary. This method allows bed sediment sampling at moderately high spatial density (up to tens of samples over tens of square meters) [*Rubin et al.*, 2010], and hundreds of images can be obtained in a few hours. Another advantage of this approach over conventional physical sampling is that it samples only those sediments that are exposed on the bed surface, reducing the ambiguity associated with comparisons with MBES backscatter (which, at this high frequency, is also only sensitive to bed surface sediment).

The underwater video camera images were used to characterize bed sediment, at a point, into three categories: (1) "sand": homogeneous sandy surfaces (where no distinction was made between plane bed, ripples, and dunes); (2) "gravel": coarse substrates encompassing homogeneous gravels and sand-gravel mixtures (sand in the interstitial spaces of gravel clasts), in the size range deemed to be movable by dam-regulated flows; and (3) "rock": a category which includes boulder clasts, bedrock ledges, the submerged feet of canyon walls, and cobbles large enough to be considered immovable by dam-regulated flows. These qualitative assessments were made by an experienced field technician operating the winch, based on both the image and the feel of the winch as the video housing hit the bed surface.

Another underwater video camera system—a towed video sled with a wider field of view, equipped with powerful lights and lasers for scale—was also used to observe the bed. These data were collected at a different time (May 2012) in the RM30 and RM60 reaches, along with MBES data, and were used to assess the plausibility of sediment classifications applied to an earlier data set (outside the calibration).

3. Sediment Signatures

For each of the three study sites, the distribution of values from a total of six backscatter metrics were analyzed in areas of known sand, gravel, and rock. Of these six metrics, four were the backscatter stochastic geometries (ω_1, γ_1, σ, and l_0) described in Part 1. We term these quantities stochastic geometries, in the sense of *Fara and Scheidegger* [1961], because they are statistical measures of physical amplitude and length scales, and also because the units of these scales are mixed.

To recap from Part 1, the spectral strength, ω_1, is a measure of power at low frequencies or the magnitude of backscatter fluctuations over relatively large spatial distances. The spectral exponent, γ_1, is a measure of the rate of decay in power as a function of increasing frequency. Complex spatial patterns of backscatter must be described by a range of frequencies, so γ_1, which quantifies the rate of decay in power as a function of frequency, is a useful measure of how complex the data is by quantifying the range of frequencies necessary to describe the data (spectral width). The integral length scale, l_0, is a measure of the persistence (autocorrelation) in the data. The amplitude variance, σ_1, or acoustic roughness, is a measure of the magnitude of backscatter fluctuations over all space (both large and small separation distances) and is therefore calculated over all resolvable (and statistically significant) scales.

The other two metrics were median backscatter B (dB) and the standard error in backscatter SE_B (dB), both also gridded at 0.25 $\times$ 0.25 m. Given the $\pm$1 m uncertainty in the position of sediment observations and under an assumption that sediment type was not varying over smaller scales than this, masks were created for each substrate by applying a circle with 1 m radius to each point. In this way, each sediment observation was related to a 1 m^2 patch of riverbed. The value of each of the six metrics for that patch was calculated as the weighted average of values in grid cells covered by that circle. The weighting was based on distance, D, to the center of the circle (the location of the sediment observation) where each cell's contribution to a weighted average went with e^{-D^2}.

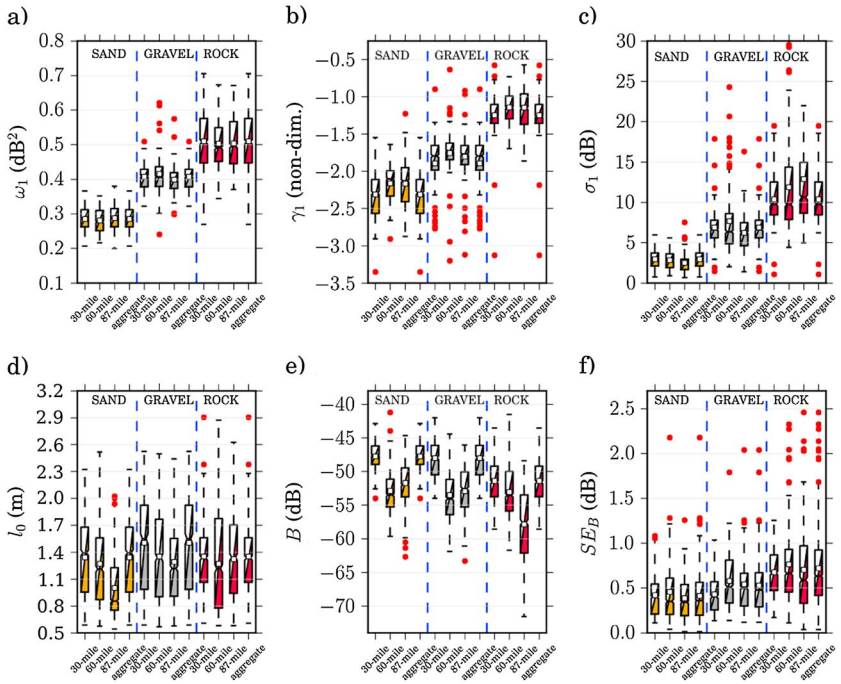

Figure 3. Box-and-whisker plots showing the distributions of (a) spectral strength ω_1, (b) spectral exponent γ_1, (c) acoustic roughness σ_1, (d) integral length scale l_0, (e) median backscatter coefficient B, and (f) standard error of backscatter SE_B, for each site and on aggregate, and for each substrate type. The box extends from the lower to upper quartile values of the data, with a notch at the median and a white square at the mean. The whiskers extend from the box to show the range of the data within 95% of observations, and the outliers (dots) extend beyond the whiskers. The parameters in the top row better distinguish between sediment types than the parameters in the bottom row.

The relationship between distributions of each of the six parameters for each of the three study sites and on aggregate (Figure 3) show a number of interesting trends. Most important, there appears to be some grain size signal in each of these metrics. The strength of this association, however, varies considerably by metric. There are four main attributes that make up a useful parameter for the purposes of sediment classification:

1. The degree of overlap (and central tendency) between the distribution of values per sediment type is small, i.e., the between-substrate variability is high.
2. The degree of overlap (and central tendency) between the distribution of values over different sites is large, i.e., the between-site variability is low.
3. The distribution of values per substrate per site is small, i.e., the within-substrate variability is low.
4. The magnitude of values scales with substrate size on a continuum (small magnitudes equate with small substrates and large magnitudes equate with large substrates).

Of these four, the first three are the most important and the latter is for numerical convenience rather than a limitation on classification per se. Spectral strength (ω_1; Figure 3a), spectral exponent (γ_1; Figure 3b), and spectral variance (σ_1; Figure 3b) have high between-substrate variability, low between-site variability, low within-substrate variability, and the magnitude of values scales with substrate size. Collectively, these metrics have a strong sediment signature. Stronger indeed than the integral length scale (l_0; Figure 3d) that shows a low between-substrate variability; median backscatter (B; Figure 3e) that is typified by high between-site variability and only moderate between-substrate variability; and standard error in backscatter (SE_B; Figure 3f) that shows high within-substrate variability and only moderate between-substrate variability.

On this basis, it was decided to use a combination of ω_1, γ_1, and σ for discriminating between sediment types, and the other three metrics were not considered further. Each of these parameters have a well-defined physical interpretation (Part 1) and, given the nature of their estimation (from the entire power spectrum of median backscatter values aggregated over space) are much less sensitive to outliers than measures based on the intensity of backscatter.

4. Sediment Classification Methods

Toward an aim of classifying each 0.25 × 0.25 m grid cell into three sediment types using the values of ω_1, γ_1, and σ in those cells, two simple classification techniques were developed and tested.

4.1. Least Squares Approach

A linear least squares method was set up that classified each cell based on the relative location of a sample [ω_1, γ_1, and σ] and a calibration matrix consisting of the means of ω_1, γ_1, and σ for each sediment type. The variance u associated with q sediment types is that which minimizes the error:

$$||(uo) + C||^2 \tag{1}$$

where || represents the two-norm (square root of squared vector elements); the calibration $C(1 \times 3)$ is the mean of each of the three stochastic geometries associated with q sediment types, or $C = [\overline{\omega_1(1:q)}, \overline{\gamma_1(1:q)}, \overline{\sigma_1(1:q)}]$; and $o(n \times 3)$ is the n observations of the three stochastic geometries, or $o = [\omega_1(1:n), \gamma_1(1:n), \sigma_1(1:n)]$. Variance u_q is a proportion such that $\sum u_q = 1$. More details are given in Appendix A. This approach is mathematically and conceptually very simple but suffers from the limitation that the sediment types are represented only by the means of the three stochastic geometries; therefore, a considerable loss of information is involved. The predicted variance associated with a particular substrate, u, is converted into a simple confidence metric for that substrate. The confidence for a particular substrate, q, is a function of the proportion predicted for that and all other substrates. For $q = 1{:}n$ substrates this can be expressed as follows:

$$\alpha_q = u_q \sum_{n \subset q \neq n} (1 - u_n) \tag{2}$$

The confidence metric, which takes on values between 0 and 1, would predict 1 (total confidence) only if the proportion of substrate q was 1 and the proportions of all other substrates were zero. Likewise, a value of 0 (no confidence) would only be ascribed to a prediction if the proportion of substrate q was 0 and the proportion of any other substrate was 1. In a (highly unusual) situation where sand, gravel, and rock have equal proportional variance (1/3), then the confidence in each substrate is $\alpha = 0.15$, and variance in any class higher than one third would lead to greater values of α for that class. Equality in two out of three classes would give confidences to each in the range $0.15 < \alpha_q \leq 0.25$, and the sediment type is indeterminate. Therefore, a cell was assigned the sediment class q if $\alpha_q > 0.25$.

4.2. Decision Tree Approach

Given training vectors $C_i \in \mathbb{R}^n$, $i = 1, \ldots, q$ containing n samples of ω_1, γ_1, and σ_1, and a label vector $z \in \mathbb{R}^q$ of associated sediment classes, a decision tree recursively partitions the space into smaller homogeneous subsets using a set of binary condition rules [*Breiman et al.*, 1984] such that the samples with the same labels are grouped together. The basic process is as follows: (1) for each attribute (ω_1, γ_1, or σ_1), find the feature that best divides the training data, C_i, such as information gain from splitting on the attribute; (2) create a decision node that splits on the attribute with the highest information gain; and (3) recurse over the sublists obtained (at descendant nodes) by splitting on the previous decision node until no more splits are possible. The interested reader is referred to reviews by *Murthy* [1998] and *Kotsiantis* [2007].

The method is detailed in Appendix B and computed using the classification and regression tree (CART) algorithm [*Breiman et al.*, 1984] implemented in *Pedregosa et al.* [2011]. Mean-square error (MSE) was used to measure the quality of a split. At least two samples were required to split a decision (intermediate) node, and a minimum of one sample was required to be at a terminal node. No constraints of the maximum depth of the tree (maximum number of decision nodes) were placed.

Decision trees have been used by *Dartnell and Gardner* [2004] for supervised seafloor sediment classification, based on a combination of backscatter and bathymetry information. A variant on the decision tree method for seafloor characterization was developed for use with backscatter by *Stepnowski et al.* [2003]. Here a decision tree approach was used because it is "white box" [*Safavian and Landgrebe*, 1991] (unlike, for example, an artificial neural network), which means the tree of decisions and decision thresholds can be described by simple Boolean logic, easily visualized and conceptually simple compared to some other commonly used supervised machine-learning algorithms such as support vector machines and Naïve Bayes classifiers [*Kotsiantis*, 2007]. Another important factor is that, unlike many of these other methods, data do

Table 1. Percentage Classification of Sand, Gravel, and Rock Based on the 30-60-87 Calibration[a]

Sand, Gravel, and Rock	Classified as ... Sand		...Gravel		...Rock	
Site	Least Squares	Decision Tree	Least Squares	Decision Tree	Least Squares	Decision Tree
30 mile	**98**, 41, 3	97, 1, 1	3, 32, 0	3, **96**, 4	0, 27, **97**	0, 3, 95
60 mile	97, 51, 1	**98**, 8, 0	3, 19, 0	2, **84**, 12	0, 30, **99**	0, 9, 88
87 mile	**97**, 48, 2	92, 6, 1	3, 30, 1	5, **87**, 9	0, 22, **98**	3, 7, 91
Aggregate	**97**, 47, 2	95, 6, 0	3, 27, 0	4, **88**, 9	0, 26, **98**	1, 6, 91

[a]The three values in columns 2–7 refer to sand, gravel, and rock, respectively. The best classifications are in bold.

not need to be preprocessed (scaled, normalized, or otherwise standardized). Therefore, each decision in the tree can be visualized in terms of the units of the input variables, which makes it easier to make a physical as well as a statistical assessment of the model. One advantage the decision tree approach has over the least squares approach is that all observations per substrate are used in the calibration (training data set), not descriptive statistics of each classifying variable per substrate. This means no information is lost in the calibration procedure, and there is no ambiguity associated with deciding what statistical measure is most appropriate to characterize the response of a particular classifying variable for a particular substrate.

A potential disadvantage with the decision tree approach with no constraints, as described above, is that outliers are explicitly modeled in order to force mean square errors to zero. These outliers could be due to errors in the validation data (manual bed classification from video), such as incorrectly labeled substrate owing to human error or inaccurate positioning. Or they could be due to an unusual backscatter signature caused by unknown scattering processes or errors in processing. In order to examine the effects of such outliers, a set of decision trees were constructed with a single constraint that a minimum of 50 samples were required to be at a terminal node.

5. Sediment Classification Results

Classifications were carried out using a calibration based on data aggregated from all three sites (RM30, RM60, and RM87). This calibration is termed 30-60-87 hereafter, and each site was evaluated with this calibration. Then, the performance of the calibration based on two of the three sites was applied to the third site, the evaluation of which is known as leave-one-out validation. Thus, a calibration based on data aggregated from RM30 and RM60 (30-60) was tested on RM87; RM30 and RM87 (30-87) was tested on RM60; and RM60 and RM87 (60-87) was tested on RM30. Both the linear least squares and the decision tree methods were carried out for each set of calibrators.

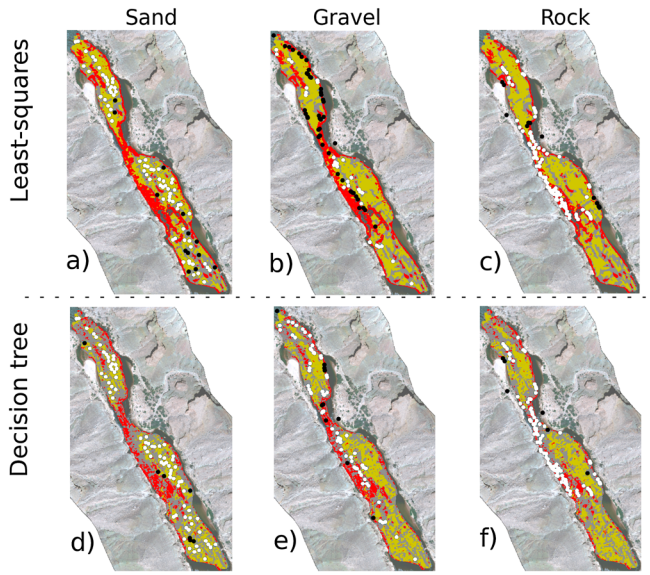

Figure 4. Comparisons of sediment classifications at RM60 based on the (a–c) linear least squares method and the (d–f) decision tree method. The columns show the observations of (from left to right) sand (yellow), gravel (grey), and rock (red). The white and black dots in each pool show correct and incorrect classifications, respectively.

In constructing each of the four classifications (respectively, 30-60-87, 30-60, 30-87, and 60-87), one half of the data were used for training the decision tree and the other half for evaluating. Owing to the unequal numbers of observations in each sediment type, in order to minimize biases in the decision tree, data were combined from pools then one half of observations of each sediment type were selected at random and aggregated to form the training set.

Table 2. Application of the 30-60 Mile Calibration to RM87, 30-87 Mile Calibration to RM60, and 60-87 Mile Calibration to RM30[a]

	Sand		Gravel		Rock	
Observed	Least Squares	Decision Tree	Least Squares	Decision Tree	Least Squares	Decision Tree
Sand	97, 97, **100**	89, 97, 97	3, 3, 0	5, 3, 3	0, 0, 0	6, 1, 0
Gravel	48, 51, 44	6, 8, 1	31, 20, 36	86, 84, **92**	22, 30, 20	8, 9, 7
Rock	2, 1, 3	1, 0, 3	1, 0, 23	11, 13, 13	98, **99**, 74	89, 87, 85

[a]The three values in columns 2–7 refer to the 30-60 Mile Calibration to RM87, the 30-87 Mile Calibration to RM60, and the 60-87 Mile Calibration to RM30, respectively. The best percentage classifications per substrate are in bold.

The results for the application of 30-60-87 on each of the three sites and on aggregate (Table 1) show that the two classification methods are similar in performance for sand and rock; however, for gravel the least squares approach is inferior to the decision tree approach. The least squares approach only classified ≈27% of gravel patches correctly on aggregate, compared to ≈88% with the decision tree. The decision tree correctly classifies ≈95% and ≈91% of sand and rock patches, respectively, on aggregate and is therefore considered a superior approach to least squares overall, based on all three substrate types.

It is possible to create a significantly different spatial pattern in the three substrates using the same data with two different classification approaches. The differences in the classifications from both methods are

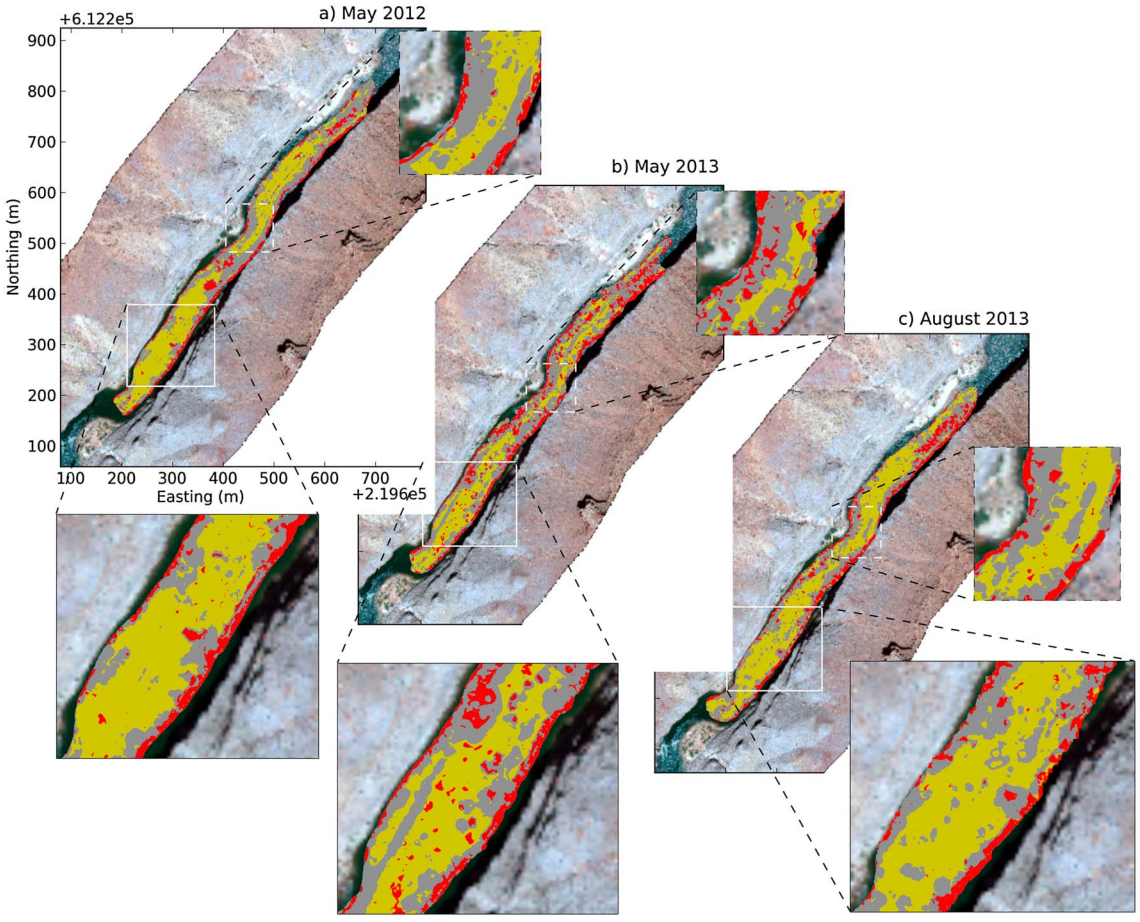

Figure 5. Sediment type at one site (RM30) through time. Pixels (each 0.25 × 0.25 m) classified as sand are yellow; gravels (and sand/gravel mixtures) are gray; and rocks are red. These surveys conducted in (a) May 2012; (b) May 2013; and (c) August 2013 reveal significant changing bed composition and spatial redistribution of sediment over relatively short timescales. Inset panels show close-up details of the sediment type around the flow constriction in the middle of the reach and around the location of the sediment gaging station.

Published 2014. American Geophysical Union.

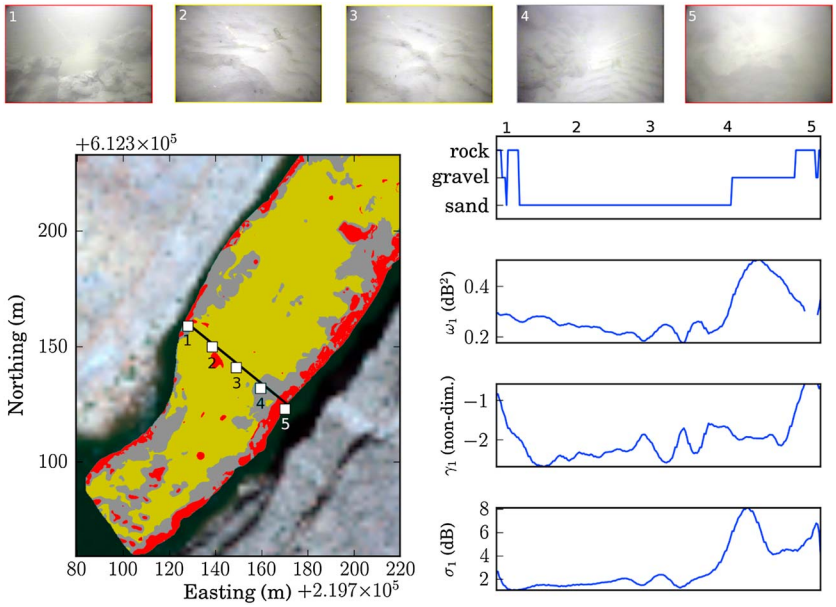

Figure 6. Transect through (left) a sediment classification map at the RM30 site, from a survey conducted in May 2012. The photographs (top) were taken by an underwater video sled at the five stations along the transect line (marked by white squares on the classification map and numbered 1 through 5). The lasers in the photograph are 0.1 m apart. The four panels (right) show continuous (0.25 m) spatial series along the transect of (from top to bottom) sediment classification, ω_1, γ_1, and σ_1. The numbers at the top of the panel of these figures correspond to the approximate location of the photographs.

illustrated for the RM60 site in Figure 4. The classifications are qualitatively similar at the scale of the entire survey reach but significantly different over small scales. The systematic misclassification of gravel as sand and rock by the least squares technique makes the reach appear more sandy and more rocky than it really is.

Scrutiny of the results for the application of the leave-one-out validations (Table 2) show that both methods again classify sands and rocks well but that the least squares method consistently misclassifies gravels. These results indicate that the metrics used in the calibration are not systematically different between different sites and that both calibration approaches are somewhat stable, because removing a third of the training data does not have an appreciable difference on the results. The results also show that there is broadly the same sedimentological and acoustic variety captured by the three spectral metrics in two pools as there is in three. Given the significant geomorphic differences between the three study sites, this is particularly encouraging for application to areas of riverbed outside the calibration.

The 30-60-87 decision tree classification was applied to the RM30 site to two earlier data sets outside the calibration, collected in May 2012 and May 2013. The reach is predominantly sandy and the broad scale distribution of substrates is relatively static in time. The relative spatial distributions of sand and gravelly substrates are variable, however, between surveys (Figure 5). The relative proportions of sand over the three time periods would indicate that the bed in May 2013 was relatively sand starved compared to the other two occasions. The observed bed sediment distribution is highly plausible given our collective experience at this site, in combination with observed general persistence of sand and rock patches. In addition, underwater video sled data is available in May 2012 for the RM30 site (Figure 6): a video transect was taken across channel in the location of the gaging station toward the end of the mapped section above the debris fan. Observations made with the video show a relatively uniform sandy bed in the center of the channel flanked by rocks and large gravels at either end of the transect, and the acoustic classification closely mimics these video observations (Figure 6).

The 30-60-87 decision tree classification was also applied to mapped areas of the channel outside the calibration sites. For example, a small section of the river below the RM30 rapid toward fence fault (Figure 7) was observed with the underwater video sled system during May 2012. A longitudinal transect made on river right close to the shoreline showed a grading from large boulders just below the rapid tending to gravel

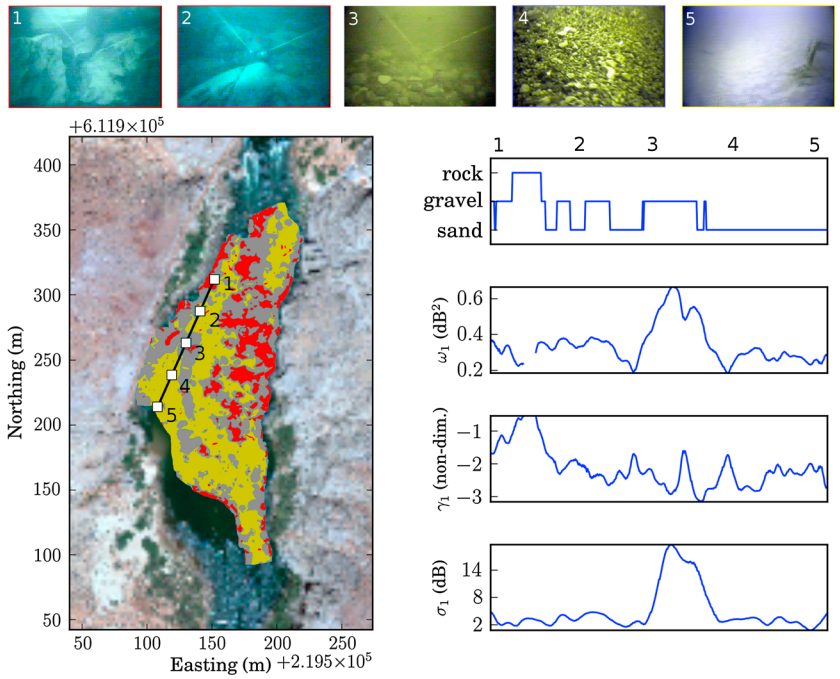

Figure 7. Longitudinal transect through (left) a sediment classification map at a site immediately below RM30 rapid toward fence fault sandbar, from a survey conducted in May 2012. The photographs (top) were taken by an underwater video sled at the five stations along the transect line (marked by white squares on the classification map and numbered 1 through 5). The lasers in the photograph are 0.1 m apart. The four panels (right) show continuous (0.25 m) spatial series along the transect of (from top to bottom) sediment classification, ω_1, γ_1, and σ_1. The numbers at the top of the panel of these figures correspond to the approximate location of the photographs.

then sand toward fence fault sandbar. The acoustic classification picked up this general sediment-fining trend (Figure 7). However, the predicted sediment type did not match the video observations as closely as in the pool upstream (Figure 6). For example, the fine homogeneous gravel at position 4 in Figure 7 is misclassified as sand.

In the 30-60-87 mile calibration, visible as small patches of one class surrounded by a different class in the decision tree surfaces (Figure 8), potential outliers are those samples that occur in small numbers at a given terminal node (Figure 9). The decision tree (Figure 9) is read such that arrows to the left indicate "false" (that the condition stated in the box above is not true) and arrows to the right indicate "true." For exam-

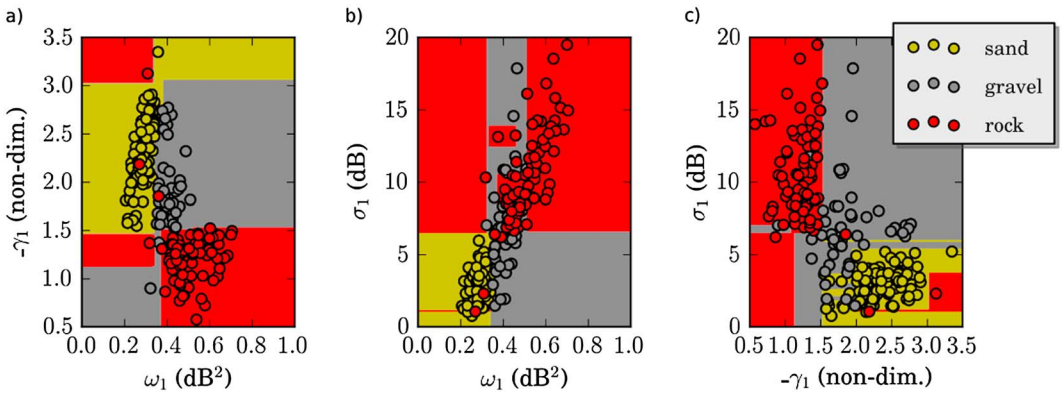

Figure 8. Decision surfaces of paired parameters based on aggregation of observations from RM30, RM60, and RM87 pools: (a) ω_1 and $-\gamma_1$, (b) ω_1 and σ_1, and (c) $-\gamma_1$ and σ_1. The yellow, grey, and red are portions of the parameter space in which the decision tree would classify as sand, gravel, and rock, respectively. This illustrates that a relatively small area of each parameter space is classified as sand.

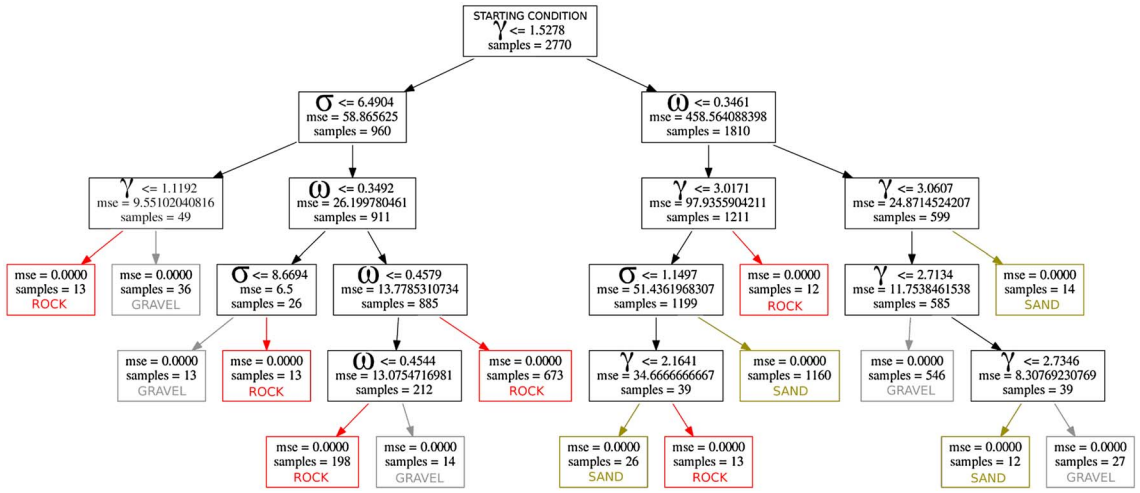

Figure 9. The decision tree classifier based on aggregation of observations from RM30, RM60, and RM87 pools. At each branch the threshold value of the parameter associated with the decision is shown along with the mean square error (MSE) and the number of samples that this applies to. When a decision is made (colored box), the MSE is zero to ensure that the classification fits the training data exactly.

ple, one route through the tree is as follows: (1) Is $\gamma \leq 1.5278$? No (go left); (2) Is $\sigma \leq 6.4904$? No (go left); and (3) Is $\gamma \leq 1.1192$? Yes (go right). The class is "Gravel." Another example route through the tree is)1) Is $\gamma \leq 1.5278$? Yes (go right); (2) Is $\omega \leq 0.3461$? Yes (go right); and (3) is $\gamma \leq 3.0607$? Yes (go right). The class is "Sand."

The tree of decisions made by the resulting 30-60-87 classifier (Figure 10) is greatly simplified compared with the classifier without constraints (Figure 9), with the maximum depth (layers of decisions, each represented by a different row in the tree) reducing from five to four and the number of decision nodes reducing from 15 to 9. The cost of this simplification is greater mean square errors. The resulting decision surfaces (Figure 11) show a more straightforward partitioning, such that each surface has just one area of parameter space for each of the three substrates, and these areas lie in a continuum in greater agreement, compared with Figure 8, with the broader scale patterns in the data.

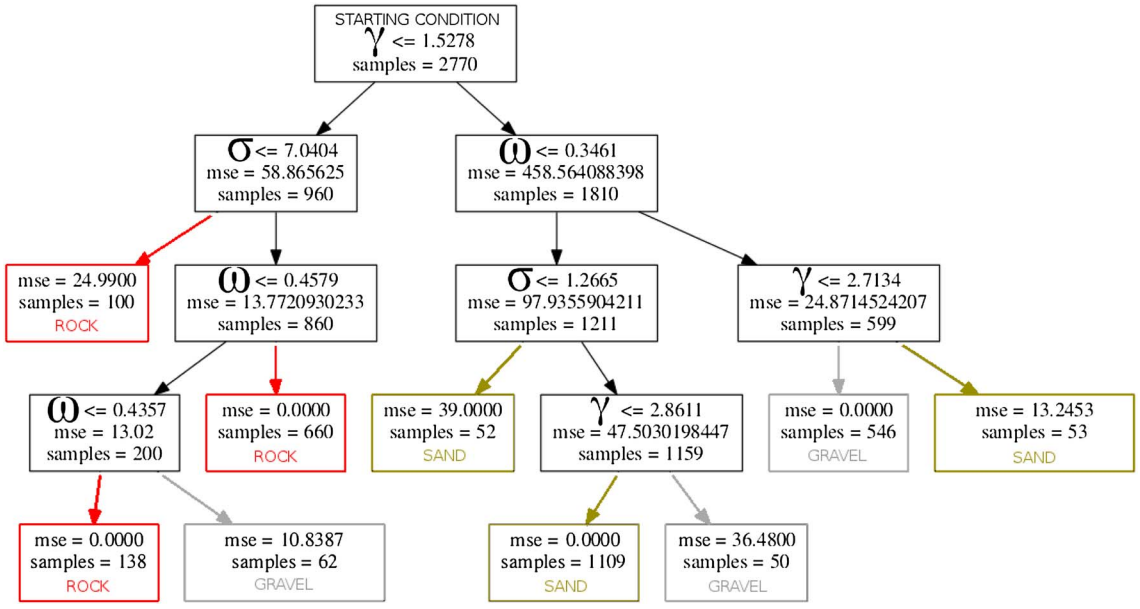

Figure 10. The decision tree classifier based on aggregation of observations from RM30, RM60, and RM87 pools, with a constraint that each decision node must have at least 50 samples. At each branch the threshold value of the parameter associated with the decision is shown along with the mean square error (MSE) and the number of samples that this applies to. When a decision is made (colored box), the MSE is zero to ensure that the classification fits the training data exactly.

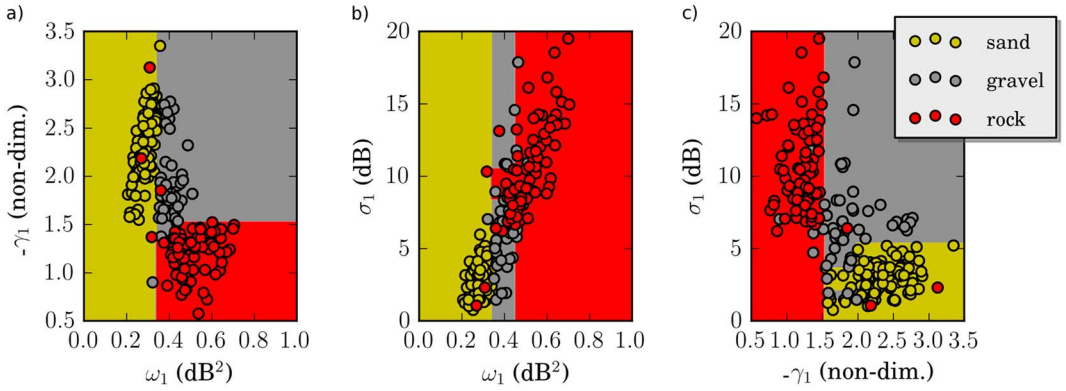

Figure 11. Decision surfaces of paired parameters based on aggregation of observations from RM30, RM60, and RM87 pools, with a constraint that each decision node must have at least 50 samples: (a) ω_1 and $-\gamma_1$, (b) ω_1 and σ_1, and (c) $-\gamma_1$ and σ_1. The yellow, grey, and red are portions of the parameter space in which the decision tree would classify as sand, gravel, and rock, respectively. This illustrates that a relatively small area of each parameter space is classified as sand.

The results for the application of the 30-60-87 decision tree classification with constraints, on each of the three sites and on aggregate (Table 3), reveal that the performance of the simplified classifier is comparable to the original, unconstrained classifier. The classification skill for sand, gravel, and rock is 97, 81, and 95%, respectively for the constrained tree compared to 95, 88, and 91% for the original, unconstrained classifier (Table 1).

6. Discussion

The decision tree approach is deemed better overall than the least squares because of its ability to classify more gravel patches. Scrutiny of the decision surfaces between classifying parameter pairs (Figure 8), which show the relative portions of the parameter spaces in which the tree classifies for each substrate, may reveal why the linear least squares approach is a poor gravel predictor. In general, the decision surfaces for sand and rock are relatively small and well constrained. Only a small portion of each decision surface is covered by the main cloud of observations, because each of the observations are linear. However, the portion of parameter space that overlaps with actual observations for gravel in particular is small. In other words, there are narrow bands in each stochastic geometry that apply exclusively to gravel, compared with large bands of values that apply to sand and rock. The least squares approach is weighted to each of these extremes.

Examination of the tree of decisions made by the (30-60-87) classifier (Figure 9) reveals that the tree could be greatly simplified at the expense of some greater uncertainty. For example, there are six different paths toward a classification of rock, five paths toward gravel, and four paths toward a sand classification. However, 871 out of 922 total rock observations (95%) are reached via only two paths. Similarly, 546 out of a total of 636 gravel observations (86%) are reached by only one path [$\omega_1 \leq 0.3461$ dB^2, $\gamma_1 \leq 2.7134$], and 1160 out of 1212 sand observations (96%) are reached by one path [$\omega_1 \leq 0.3461$ dB^2, $\gamma_1 \leq 3.0171$, $\sigma_1 \leq 1.1497$ dB]. The decision trees associated with two sites (30-60, 30-87, and 60-87) are qualitatively very similar, which explains why the relative proportions of sand, gravel, and rock vary little based on two- and three-site classifications (Table 2). This also indicates that the approach is stable, as does its application to sites and times outside the calibration (Figures 5–7), which are highly plausible.

Table 3. Percentage Classification of Sand, Gravel, and Rock, Based on the 30-60-87 Calibration With a Constrained Decision Tree[a]

Sand, Gravel, and Rock	Classified as ... Sand	...Gravel	...Rock
30 mile	**98**, 1, 1	5, **84**, 11	3, 2, **96**
60 mile	**98**, 2, 0	9, **80**, 11	0, 4, **96**
87 mile	**95**, 3, 2	8, **80**, 12	1, 6, **93**
Aggregate	**97**, 2, 1	8, **81**, 11	1, 4, **95**

[a]The three values in columns 2–7 refer to sand, gravel, and rock, respectively. The best classifications are in bold.

The classification skill of the unconstrained and constrained decision trees were very similar. In the present case, therefore, the simpler, more parsimonious, constrained model is preferable. However, construction of the unconstrained tree is a necessary prerequisite in order to make a subjective decision about the nature of the constraint. In the present case,

Table 4. Gini Scores per Classification[a]

Classification	30-60-87 (v. 2)	30-60 (v. 2)	30-87 (v. 2)	60-87 (v. 2)
Gini score, ω_1	0.124 (0.122)	0.121 (0.122)	0.121 (0.122)	0.846 (0.84r1)
Gini score, γ_1	0.860 (0.861)	0.866 (0.861)	0.866 (0.861)	0.140 (0.149)
Gini score, σ_1	0.016 (0.018)	0.013 (0.018)	0.013 (0.018)	0.014 (0.01)

[a]Values in parentheses (v. 2) refer to the classification with imposed constraint of a minimum of 50 samples at each terminal node.

the decision surface from the constrained tree (Figure 11) was evaluated with regard to the presence or otherwise of obvious outliers and the greater simplicity of the parameter space partitioning that it afforded. It should not be taken for granted, however, that this combination would happen with any data set. In addition, the unconstrained decision tree is the only way to be purely objective if there is any doubt over whether or not certain data are "outliers."

The relative importance of a feature (in this case, the three stochastic geometries) to the classification is computed as the (normalized) total reduction of the criterion brought by that feature. This metric, known as Gini importance, obtains higher values for more important features and is computed as the mean Gini (information) gain (Appendix B) produced by a feature over the entire tree [*Breiman*, 2001]. According to this metric, spectral exponent γ_1 is usually the most important feature in every classification (Table 4). To recap from Part 1, the spectral exponent is a measure of the rate of decay in spectral power as a function of increasing frequency and is sensitive to the range of statistically significant frequencies required to construct the signal (i.e., the complexity of the signal). Out of the three stochastic geometries tested in the decision tree approach used here, it is overall the most powerful individual classifier of substrate type. The one exception to this general rule is for the 60-87 classifications that ranked ω_1 (spectral strength) as the most important. Decision trees have been known to change considerably with small variations in the input data [*Kotsiantis*, 2007], a problem mitigated by using ensembled decision trees or "random forests" [*Breiman*, 2001]. We may adopt such an approach in the future if, given more validation data, we observe greater variability in decision tree classifiers. In this case, however, six out of eight constructed trees were, for practical purposes, almost identical, and we conclude that the sensitivity to input data was moderate and the use of a more complicated classification approach (ensembled decision trees) was not warranted.

Here the sediment classification has been based on very broad substrate categories and is not likely without suitable modification, to be able to discriminate between fine and coarse sands, fine and coarse gravels, and the full spectrum of mixtures of sands and gravels found on the Colorado riverbed in Marble and Grand Canyons. This is illustrated by the mismatch between the transect of video observations and the associated acoustic classification shown in Figure 7. In this field of fast-moving water downstream of the constriction and rapid, errors of the spatial positioning of the video may have been increased due to larger-than-normal boat motions. If this is the case, position 4 could have actually been a few meters to the east, in a patch of bed acoustically classified as gravel. However, it is also likely that the classification of just three substrate types is too rudimentary to encompass the great deal of heterogeneity in sediment types in this surveyed pool. More observations of the full spectrum of substrate types, especially mixtures of sands and gravels, are required. The potential of this or similar techniques to resolve sediment in more detail would require both a larger range of observations for each part of the sedimentary spectrum as well as an approach to maximize the statistical differences between end-members of that spectrum. The present method could be advanced using Fourier filtering to find the frequency band most strongly related to individual substrates. Alternatively, wavelet analysis on fields [*Buscombe*, 2013] of backscatter and bathymetry [*Cataño-Lopera et al.*, 2009] could reveal both the strength of the periodic components in the data as well as the locations associated with these frequency bands. The choice of alternative classification techniques could also be beneficial, as long as classification results can still be expressed and understood in physical units rather than statistical quantities, and uncertainties can be quantified in a straightforward manner. A fuzzy classification [*Legleiter and Goodchild*, 2005; *Wheaton et al.*, 2010; *Nelson et al.*, 2014] would be an appropriate alternative choice to a decision tree approach.

Each approach to acoustic sediment classification has its relative merits. For example, two spatial series of data with identical distributions but different autocorrelations will have different spectra; therefore,

spectral methods have this advantage over techniques to characterize bed sediment based on distributions of backscatter. Neither spectral nor distributional techniques are affected by the spatial arrangement of sediment patches (if the frequency content remains the same, because the power spectrum contain no phase information); therefore, they both hold this advantage over some methods based on image texture [*Reed IV and Hussong*, 1989; *Cutter et al.*, 2003; *Orpin and Kostylev*, 2006; *Blondel and Sichi*, 2009]. The use of both spectral methods and models for backscatter distributions [*Simons and Snellen*, 2009; *Amiri-Simkooei et al.*, 2009] could be highly complementary, with the former revealing the frequency content and spatial scales associated with riverbed scattering and the latter perhaps better reflecting the intrinsic geo-acoustic properties of sediment (such as impedance, porosity, and bulk density).

Neither the spectral analysis of backscatter described in Part 1 nor the sediment classification techniques described here in Part 2 require tunable parameters. We predict that stochastic geometries of noncohesive (sand to rock) sedimentary systems sensed with a 400 kHz MBES should be similar in magnitude and that the same qualitative relationships between sediment size and magnitude of these quantities should apply. However, the calibration required for sediment classification is potentially sensitive to the number of observations upon which it is based, as well as, potentially at least, the specific joint probability distribution of sediment and values of each stochastic geometry. High concentrations of sediment, especially cohesives, could attenuate the acoustic signal so much that the dynamic range of backscatter becomes too low to use for sediment classification, although these thresholds are yet to be determined. Therefore, application to other heterogeneous riverbeds would probably require river-specific calibration using a set of independent observations at least as numerous and spatially uniform as those used here.

The choice of window size is also a potentially important consideration if spectral scattering signatures from quasi-rhythmic bed forms or cobbles are independent of the periodicity in sediment type over the same area. As described in section 2.2, the appropriate window size is that which ensures the morphological (principally bed form) signal is secondary to sedimentology in its effects on the spatial distribution of backscatter. To greatest order, we found that window length scale is best estimated by identifying the plateau (or zero crossing) lag in autocorrelation analyses of elevation profiles.

It is unknown how well the current method would apply to sediments finer than sand. At 400 kHz, these fine sediments are well within the Rayleigh scattering regime, as opposed to coarse sands, gravels, and boulders which are well within the geometric scattering regime. Deposits comprising silts and muds are known to attenuate sound more effectively [*Jackson and Richardson*, 2007]. Both of these factors are potentially limiting, but until independent functional relationships are established between the stochastic geometries used here and both acoustic impedance and surface roughness, the general applicability to all mixtures of riverbed sediments—cohesive and noncohesive—will be uncertain. This will be the subject of further research.

7. Conclusions

The scattering signatures of riverbed sediment were examined using spatially explicit maps of the stochastic geometries of backscatter. It was revealed that length and amplitude scales of backscatter differ over surfaces composed of different sediments, as determined by georeferenced underwater video observations.

Statistics of backscatter magnitudes alone are found to be poor discriminators between sediment types. It is suggested that this is the case because the backscatter magnitude reflects both the interaction of sound with the bed due to physical roughness and acoustic impedance. Backscatter is sensitive to the specific spatial arrangement of scatterers in an insonified area and is therefore a highly stochastic quantity. Metrics from the spectrum of backscatter over small spatial scales, ensemble averaged by windowing, are more discriminatory between sediment because they are less affected by the spatial arrangement of sediment patches. This is because the power spectrum contains no phase information. However, for the above statement to be valid, the frequency content must remain the same, which means that the magnitude of spectral amplitude and length scales may be scale specific.

The variance of the power spectrum and the intercept and slope from a power law spectral form (termed the spectral strength and exponent, respectively) are found to best discriminate between sediment types. This draws parallels with similar findings in seafloor scattering over larger scales, which suggests that these

measures of acoustic roughness inherently mimic physical (grain scale) roughness given appropriate scaling and spatial averaging.

Sediment was classified on a 0.25 × 0.25 m grid cell basis based on these metrics. A classification approach based on a decision tree algorithm was found to outperform one based on linear least squares. A decision tree approach to sediment classification was able to classify patches of homogeneous sands, gravels (and sand-gravel mixtures), and cobbles/boulders with 95, 88, and 91% accuracy, respectively. The difference in the two approaches lay in the inability of the least squares approach to successfully classify more instances of gravel. Examination of the decision surface of the more successful decision tree approach revealed that the delineation of the parameter space, for each pair of classifying variables, was somewhat nonlinear and revealed why the least squares approach misclassified many instances of gravel as sand or rock.

Application to sites outside the calibration and surveys made at calibration sites at different times were found to be physically plausible based on underwater video transects. A decision tree constrained to have a minimum of 50 samples per terminal node (final substrate classification) was found to greatly simplify the decision tree classifier with little effect on the performance for the three substrates. Analysis of the performance of the decision tree approach with different combinations of training data sets revealed the approach to be very stable (relatively invariant of specific data sets used to train the classifier). Gini importance indices suggested that the spectral exponent was consistently the most important variable in the classification. We tentatively conclude that spectral width, or the number of frequencies required to describe the important features of the spatial variation of backscatter, is the most sensitive indicator of different sediment types.

In the absence of theory concerning how spatially variable sediment surfaces scatter high-frequency sound, the primary advantages of this data-driven approach to classify bed sediment over alternatives are that spectral methods have well-understood properties; make no assumptions about the data, such as the distributional form of the fluctuating component of backscatter over small spatial scales, and require limited a priori knowledge of the physical interaction of sound with a heterogeneous riverbed.

Appendix A: Least Squares

The parameter vector $u(1 : q)^t$, where t indicates transpose, is the least squares solution of the following overdetermined linear system:

$$\begin{bmatrix} \overline{\omega_1(1)} & \overline{\gamma_1(1)} & \overline{\sigma_1(1)} & 0 \\ \overline{\omega_1(2)} & \overline{\gamma_1(2)} & \overline{\sigma_1(2)} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \overline{\omega_1(q)} & \overline{\gamma_1(q)} & \overline{\sigma_1(q)} & 0 \end{bmatrix} u^t = \begin{bmatrix} \omega_1(1) & \omega_1(2) & \dots & \omega_1(n) \\ \gamma_1(1) & \gamma_1(2) & \dots & \gamma_1(n) \\ \sigma_1(1) & \sigma_1(2) & \dots & \sigma_1(n) \end{bmatrix} \tag{A1}$$

which is solved using an iterative optimization routine [*Lawson and Hanson*, 1974] that minimizes $\mathrm{norm}(o \times u - C)$ subject to $u \geq 0$.

Appendix B: Decision Tree Learning

The root node contains all the training data (C_i and z), that are split into several intermediate nodes and finally a set of terminal nodes at which a label is ascribed. We used the CART algorithm [*Breiman et al.*, 1984], which constructs a classification tree in three steps: (1) the selection of a node splitting rule, (2) the decision whether to set a node as terminal or to continue splitting (the stopping rule), and (3) the assignment of each terminal node to a given label.

Let the data at node v be represented by Q. For each node, v, split $\zeta = (j, T_v)$, consisting of a feature j (either ω_1, γ_1 or σ_1) and threshold T_v, the data is partitioned into $Q_1(\zeta)$ and $Q_2(\zeta)$ descendant subsets: $Q_1(\zeta) = (C, z)|C_j <= T_v$ and $Q_2(\zeta) = Q \; Q_1(\zeta)$.

The splitting rule reduces the data variability at each $\zeta = (j, T_v)$ into two descendant nodes that are purer [*Safavian and Landgrebe*, 1991]. This is achieved by defining an impurity function at every $\zeta = (j, T_v)$, which is a measure of the heterogeneity of a given node. The impurity at node v, representing a region $\mathbb{R}_v$

with N_v observations, is computed using

$$I(Q,\zeta) = \frac{n_1}{N_v} G\left[Q_1(\zeta)\right] + \frac{n_2}{N_v} G\left[Q_2(\zeta)\right] \tag{B1}$$

where G is the Gini impurity function $G(C_v) = \sum_q p_{mq}(1 - p_{vq})$ [*Breiman et al.*, 1984] and where, if a target is a classification outcome taking on values 0, 1, ..., q, the proportion of class q observations in node v is as follows:

$$p_{vq} = \frac{1}{N_v} \sum_{C_i \in \mathbb{R}_v} \left(z_i \equiv q\right) \tag{B2}$$

Label assignment that occurs at the split that produces the maximum decrease in the node impurity are the parameters that minimizes the impurity:

$$\zeta^* = \text{argmin}_\zeta I(Q,\zeta) \tag{B3}$$

The above is recursed for subsets $Q_1(\zeta^*)$ and $Q_2(\zeta^*)$ until $N_v = 1$, i.e., the stopping rule [*Pedregosa et al.*, 2011], which here is simply when all samples have been assigned labels.

Notation

B	backscatter coefficient (dB).
C	calibration or training data set.
D	distance from arbitrary point to the center of ascribed circle (m).
e	base of the natural logarithm (-).
G	Gini impurity function.
i	index of sample in training set.
I	impurity (-).
j	index of feature (stochastic geometry).
l_0	integral length scale (m).
n	number of sample observations.
N_v	number of observations at node v.
o	sample observations of stochastic geometries.
q	substrate index.
Q	data subset.
$\mathbb{R}$	training set.
SE_B	standard error of backscatter coefficient (dB).
t	matrix transpose.
T_v	threshold at node v.
u	vector of least squares solution (-).
v	node.
z	label vector.
α	confidence (-).
γ_1	spectral strength parameter (-).
ω_1	spectral strength parameter (dB^2).
σ_1	root-mean square variance of $B(\mathbf{X})$ (dB).
ζ	impurity minimum (-).

Acknowledgments

Data and computer code to reproduce the results from this manuscript are available from the authors. The data analyzed and discussed in this manuscript were collected by the dedicated efforts of many field technicians, river guides, and volunteers. Special thanks to David Rubin, Robert Tusso, Peter Wilcock, and Joseph Hazel for data collection and discussions. Thanks to Erich Mueller for stylistic guidance. This work was funded by the Glen Canyon Dam Adaptive Management Program administered by the U.S. Bureau of Reclamation. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. government. Thanks to the Editor, Associate Editor, Chris Sherwood, and two anonymous reviewers for their constructive comments which improved the quality of this paper.

References

Amiri-Simkooei, A., M. Snellen, and D. G. Simons (2009), Riverbed sediment classification using multi-beam echo-sounder backscatter data, *J. Acoust. Soc. Am.*, *126*, 1724–1738, doi:10.1121/1.3205397.

Anderson, J. T., D. V. Holliday, R. Kloser, D. G. Reid, and Y. Simard (2008), Acoustic seabed classification: Current practice and future directions, *ICES J. Mar. Sci.*, *65*, 1004–1011, doi:10.1093/icesjms/fsn061.

Blondel, P., and O. G. Sichi (2009), Textural analyses of multibeam sonar imagery from Stanton Banks, Northern Ireland continental shelf, *Appl. Acoust.*, *70*, 1288–1297, doi:10.1016/j.apacoust.2008.07.015.

Breiman, L. (2001), Random forests, *Mach. Learn.*, *45*, 5–32, doi:10.1023/A:1010933404324.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, Calif.

Brown, C. J., and P. Blondel (2009), Developments in the application of multibeam sonar backscatter for seafloor habitat mapping, *Appl. Acoust.*, *70*, 1242–1247, doi:10.1016/j.apacoust.2008.08.004.

Brown, C. J., S. J. Smith, P. Lawton, and J. T. Anderson (2011), Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques, *Estuarine Coastal Shelf Sci.*, *92*, 502–520, doi:10.1016/j.ecss.2011.02.007.

Buscombe, D. (2013), Transferable wavelet method for grain size-distribution from images of sediment surfaces and thin sections, and other natural granular patterns, *Sedimentology*, *60*, 1709–1732, doi:10.1111/sed.12049.

Buscombe, D., P. E. Grams, and M. A. Kaplinski (2014a), Characterizing riverbed sediments using high-frequency acoustics: 1. Spectral properties of scattering, *J. Geophys. Res. Earth Surface*, doi:10.1002/2014JF003189.

Buscombe, D., D. M. Rubin, J. R. Lacy, C. D. Storlazzi, G. Hatcher, H. Chezar, R. Wyland, and C. R. Sherwood (2014b), Autonomous bed-sediment imaging-systems for revealing temporal variability of grain size, *Limnol. Oceanogr.*, *12*, 390–406, doi:10.4319/lom.2014.12.390.

Cataño-Lopera, Y. A., J. D. Abad, and M. H. García (2009), Characterization of bedform morphology generated under combined flows and currents using wavelet analysis, *Ocean Eng.*, *36*, 617–632, doi:10.1016/j.oceaneng.2009.01.014.

Cutter, G. R., Y. Rzhanov, and L. A. Mayer (2003), Automated segmentation of seafloor bathymetry from multibeam echosounder data using local Fourier histogram texture features, *J. Exp. Mar. Biol. Ecol.*, *285-286*, 355–370, doi:10.1016/S0022-0981(02)00537-3.

Dartnell, P., and J. V. Gardner (2004), Predicting seafloor facies from multibeam bathymetry and backscatter data, *Photogramm. Eng. Remote Sens.*, *70*, 1081–1091.

Dietrich, W. E., and J. D. Smith (1984), Bed load transport in a river meander, *Water Resour. Res.*, *20*, 1355–1380, doi:10.1029/WR020i010p01355.

Fara, H. D., and A. E. Scheidegger (1961), Statistical geometry of porous media, *J. Geophys. Res.*, *66*, 3279–3284, doi:10.1029/JZ066i010p03279.

Grams, P. E., and P. R. Wilcock (2014), Transport of fine sediment over a coarse, immobile riverbed, *J. Geophys. Res. Earth Surf.*, *119*, 188–211, doi:10.1002/2013JF002925.

Grams, P. E., D. J. Topping, J. C. Schmidt, J. E. Hazel, and M. Kaplinski (2013), Linking morphodynamic response with sediment mass balance on the Colorado River in Marble Canyon: Issues of scale, geomorphic setting, and sampling design, *J. Geophys. Res. Earth Surf.*, *118*, 361–381, doi:10.1002/jgrf.20050.

Hazel, J. E., P. E. Grams, J. C. Schmidt, and M. Kaplinski (2010), Sandbar response following the 2008 high-flow experiment on the Colorado River in Marble and Grand Canyons. *U. S. Geol. Surv. Scientific Investigations Rep.,* 2010-5015, 52 pp.

Hazel, J. E., D. J. Topping, J. C. Schmidt, and M. Kaplinski (2006), Influence of a dam on fine-sediment storage in a canyon river, *J. Geophys. Res.*, *111*, F01025, doi:10.1029/2004JF000193.

Howard, A., and R. Dolan (1981), Geomorphology of the Colorado River in the Grand Canyon, *J. Geol.*, *89*, 269–298.

Jackson, D. R., and M. D. Richardson (2007), *High-Frequency Seafloor Acoustics*, Springer, New York.

Kaplinski, M., J. E. Hazel, R. Parnell, M. Breedlove, K. Kohl, and M. Gonzales (2009), Monitoring fine-sediment volume in the Colorado River Ecosystem, Arizona: Bathymetric survey techniques. *U.S. Geol. Surv. Open File Rep.,* 2009-1207, 41 pp.

Kaplinski, M., J. E. Hazel, P. E. Grams, and P. A. Davis (2014), Monitoring fine-sediment volume in the Colorado River ecosystem, Arizona: Construction and analysis of digital elevation models. *U.S. Geol. Surv. Open File Rep.,* 20141052, 29 p.

Kotsiantis, S. B. (2007), Supervised machine learning: A review of classification techniques, *Informatica*, *31*, 249–268.

Lawson, C. L., and R. J. Hanson (1974), *Solving Least Squares Problems*, Prentice-Hall, Upper Saddle River, New Jersey.

Legleiter, C. J., and M. F. Goodchild (2005), Alternative representations of in-stream habitat: Classification using remote sensing, hydraulic modeling, and fuzzy logic, *Int. J. Geog. Inf. Sci.*, *19*, 29–50, doi:10.1080/13658810412331280220.

Murthy, S. K. (1998), Automatic construction of decision trees from data: A multi-disciplinary survey, *Data Min. Knowl. Discovery*, *2*, 345–389, doi:10.1023/A:1009744630224.

Nelson, J. M., M. Shreve, S. R. McLean, and T. G. Drake (1995), Role of near-bed turbulence structure in bed load transport and bed form mechanics, *Water Resour. Res.*, *31*, 2071–2086, doi:10.1029/95WR00976.

Nelson, P. A., J. G. Venditti, W. E. Dietrich, J. W. Kirchner, H. Ikeda, F. Iseya, and L. S. Sklar (2009), Response of bed surface patchiness to reductions in sediment supply, *J. Geophys. Res.*, *114*, F02005, doi:10.1029/2008JF001144.

Nelson, P. A., D. Bellugi, and W. E. Dietrich (2014), Delineation of river bed-surface patches by clustering high-resolution spatial grain size data, *Geomorphology*, *205*, 102–119, doi:10.1016/j.geomorph.2012.06.008.

Orpin, A. R., and V. E. Kostylev (2006), Towards a statistically valid method of textural sea floor characterization of benthic habitats, *Mar. Geol.*, *225*, 209–222, doi:10.1016/j.margeo.2005.09.002.

Paola, C., and R. Seal (1995), Grain size patchiness as a cause of selective deposition and downstream fining, *Water Resour. Res.*, *31*, 1395–1408, doi:10.1029/94WR02975.

Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Lear. Res.*, *12*, 2825–2830.

Reed IV, T. B., and D. Hussong (1989), Digital image processing techniques for enhancement and classification of SeaMARC II side scan sonar imagery, *J. Geophys. Res.*, *94*, 7469–7490, doi:10.1029/JB094iB06p07469.

Rubin, D. M., and D. J. Topping (2001), Quantifying the relative importance of flow regulation and grain size regulation of suspended sediment transport (α) and tracking changes in grain size of bed sediment (β), *Water Resour. Res.*, *37*, 133–146, doi:10.1029/2000WR900250.

Rubin, D. M., H. Chezar, J. N. Harney, D. J. Topping, T. S. Melis, and C. R. Sherwood (2007), Underwater microscope for measuring spatial and temporal changes in bed-sediment grain size, *Sediment. Geol.*, *202*, 402–408, doi:10.1016/j.sedgeo.2007.03.020.

Rubin, D. M., D. J. Topping, H. Chezar, J. E. Hazel, J. C. Schmidt, M. Breedlove, T. S. Melis, and P. E. Grams (2010), 20,000 grain-size observations from the bed of the Colorado River, and implications for sediment transport through Grand Canyon, *9th Federal Interagency Sedimentation Conference,* Las Vegas, Nev.

Rubin, D. M., D. J. Topping, J. C. Schmidt, J. E. Hazel, M. Kaplinski, and T. S. Melis (2002), Recent sediment studies refute Glen Canyon Dam hypothesis, *Eos Trans. AGU*, *83*(273), 277–278, doi:10.1029/2002EO000191.

Safavian, S. R., and D. Landgrebe (1991), A survey of decision tree classifier methodology, *IEEE Trans. Syst. Man Cybern.*, *21*, 660–674, doi:10.1109/21.97458.

Schmidt, J. C. (1990), Recirculating flow and sedimentation in the Colorado River in Grand Canyon, Arizona, *J. Geol.*, *98*, 709–724.

Simons, D. G., and M. Snellen (2009), A Bayesian approach to seafloor classification using multi-beam echo-sounder backscatter data, *Appl. Acoust.*, *70*, 1258–1268, doi:10.1016/j.apacoust.2008.07.013.

Snellen, M., D. Eleftherakis, A. Amiri-Simkooei, R. L. Koomans, and D. G. Simons (2013), An inter-comparison of sediment classification methods based on multi-beam echo-sounder backscatter and sediment natural radioactivity data, *J. Acoust. Soc. Am.*, *134*, 959–970, doi:10.1121/1.4812858.

Stepnowski, A., M. Moszyński, and T. Van Dung (2003), Adaptive neuro-fuzzy and fuzzy decision tree classifiers as applied to seafloor characterization, *Acoust. Phys.*, *49*, 233–244, doi:10.1134/1.1560382.

Topping, D. J., D. M. Rubin, and L. E. J. Vierra (2000), Colorado River sediment transport 1. Natural sediment supply limitation and the influence of Glen Canyon Dam, *Water Resour. Res.*, *36*, 515–542, doi:10.1029/1999WR900285.

Topping, D. J., D. M. Rubin, J. C. Schmidt, J. E. Hazel, T. S. Melis, S. A. Wright, M. Kaplinski, A. E. Draut, and M. J. Breedlove (2006), Comparison of sediment-transport and bar-response results from the 1996 and 2004 controlled-flood experiments on the Colorado River in Grand Canyon, *8th Federal Interagency Sedimentation Conference,* Reno, Nev.

Topping, D. J., D. M. Rubin, P. E. Grams, R. E. Grifths, T. A. Sabol, N. Voichick, R. B. Tusso, K. M. Vanaman, and R. R. McDonald (2010), Sediment transport during three controlled-flood experiments on the Colorado River downstream from Glen Canyon Dam, with implications for eddy-sandbar deposition in Grand Canyon National Park. *U.S. Geol. Surv. Open File Rep.,* 2010-1128, 111 pp.

Van Rein, H. B., C. J. Brown, R. Quinn, and J. Breen (2009), A review of sublittoral monitoring methods in temperate waters: A focus on scale, *Underwater Technol.*, *28*, 99–113, doi:10.3723/ut.28.099.

Vericat, D., R. J. Batalla, and C. N. Gibbins (2008), Sediment entrainment and depletion from patches, *Water Resour. Res.*, *44*, W11415, doi:10.1029/2008WR007028.

Wheaton, J. M., J. Brasington, S. E. Darby, and D. Sear (2010), Accounting for uncertainty in DEMs from repeat topographic surveys: Improved sediment budgets, *Earth Surf. Processes Landforms*, *35*, 136–156, doi:10.1002/esp.1886.

Wilcock, P. R. (1997), The components of fractional transport rate, *Water Resour. Res.*, *33*, 247–258, doi:10.1029/96WR02666.

Wright, S. A., J. C. Schmidt, T. S. Melis, D. J. Topping, and D. M. Rubin (2008), Is there enough sand? Evaluating the fate of Grand Canyon sandbars, *GSA Today*, *18*, 4–10.