Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

Estimating confidence intervals for gravel bed surface grain size distributions

Brett C. Eaton[1], R. Dan Moore[1], and Lucy G. MacKenzie[1]

[1]Geography, The University of British Columbia, 1984 West Mall, Vancouver, BC, Canada

Correspondence: Brett Eaton (brett.eaton@ubc.ca)

Abstract. Most studies of gravel bed rivers present at least one bed surface grain size distribution, but there is almost never any information provided about the uncertainty of the percentile estimates. We present a simple method for estimating the confidence intervals about the grain size percentiles derived from standard Wolman or pebble count samples of bed surface texture. Our approach uses binomial probability theory to generate confidence intervals for all grain sizes in the distribution. We find that the standard sample size of 100 observations is associated with errors ranging from about ±15% to ±30%, which may be unacceptably large for many applications. In comparison, a sample of 500 stones produces an uncertainty ranging from about ±9% to ±18%. In order to help workers develop appropriate sampling approaches that produce the desired level of precision, we present simple equations that approximate the proportional uncertainty associated with the median size and the 84th percentile of the distribution as a function of the sample size and the standard deviation of the distribution, assuming that the underlying distribution is log-normal. However, the true uncertainty of any sample can only be accurately estimated once the sample has been collected, so these simple equations complement – but do not replace – the basic uncertainty analysis using binomial probability theory.

1 Introduction

A common task in geomorphology is to estimate one or more percentiles of a particle size distribution, denoted D_p, where D represents the particle diameter (mm) and the subscript p indicates the percentile of interest. Such estimates are typically used in calculations of flow resistance, sediment transport, and channel stability; they are also used to track changes in bed condition over time, and to compare one site to another. In fluvial geomorphology, commonly used percentiles include D_{50} (which is the median) and D_{84}.

Various methods for measuring bed surface sediment texture have been reviewed by previous researchers (Church et al., 1987; Bunte and Abt, 2001b; Kondolf et al., 2003). While some approaches have focused on using qualitative approaches such as facies mapping (e.g. Buffington and Montgomery, 1999), or visual estimation procedures (e.g. Latulippe et al., 2001), the most common means of characterizing the texture of a gravel bed surface is still the cumulative frequency analysis of some version of the pebble count (Wolman, 1954; Leopold, 1970; Kondolf and Li, 1992; Bunte and Abt, 2001a). Pebble counts are sometimes completed by using a random walk approach, wherein the

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

operator walks along the bed of the river, sampling those stones that are under the toe of each boot and recording the b-axis diameter. In other cases, a regular grid is superimposed upon the sedimentological unit to be sampled, and the b-axis diameter of all the particles under each vertex is measured. In still other cases, computer-based photographic analysis identifies the b-axis of all particles in an image of the bed surface. Data are typically reported as cumulative grain size distributions for 0.5ϕ size intervals (e.g., 8 - 11.3 mm, 11.3 to 16 mm, 16 - 22.7 mm, 22.7 - 32 mm, and so on), from which the grain sizes corresponding to various percentiles are extracted. Attempts to characterize the uncertainty of this approach have focused on estimating the uncertainty of D_{50}, and have typically assumed that the underlying distribution is log normal (Hey and Thorne, 1983; Church et al., 1987; Bunte and Abt, 2001b). Attempts to characterize the uncertainty associated with other percentiles besides the median have relied on statistical analysis of extensive field data sets (Marcus et al., 1995; Rice and Church, 1996; Green, 2003; Olsen et al., 2005), and do not provide an easy means of calculating the sample size required to achieve a given confidence level.

Operator error and the technique used to randomly select bed particles have frequently been identified as important sources of uncertainty (Hey and Thorne, 1983; Marcus et al., 1995; Olsen et al., 2005; Bunte et al., 2009), but the largest source of uncertainty in many cases is likely to be sampling variability, which is a function of sample size. Unfortunately, the magnitude of the confidence interval is seldom calculated and/or reported, and the implications of this uncertainty are – we believe – generally under-appreciated. To address this issue, we believe that it should become standard practice to calculate and graphically present the confidence intervals about surface grain size distributions.

The objective of this note is introduce a robust, distribution-free approach to computing confidence intervals for percentile estimates. We then use this approach to demonstrate that the higher percentiles, such as D_{84}, are subject to substantial uncertainty for typically used sample sizes, and that this uncertainty translates into significant uncertainty in estimates of sediment entrainment thresholds. We then provide recommendations regarding sample sizes for estimating particle size percentiles.

2 Statistical basis

The key to our approach is that the estimation of any grain size quantile D_p can be treated as a binomial experiment during which the b-axis diameter of n particles is measured, some of which will be smaller than the true value of D_p for the population of grains on the bed, and some of which will be larger. For repeated samples from the population, the number of measured stones that will be smaller than the true value of D_p will vary about a mean value $n \cdot p$, just as the number of heads observed during n tosses of a fair coin will vary about a mean value of $0.5n$. The binomial distribution can be used to derive confidence intervals for any estimate of D_p for a sample that can be expected to contain the true value of D_p for the entire population.

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

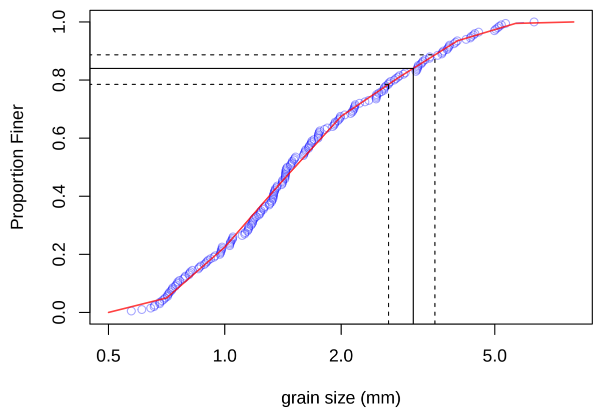

Figure 1. A grain size distribution from a stream table experiment based on a sample size of 200 observations. Blue circles indicate individual grain size measurements, and the red line is the cumulative frequency distribution for binned data using the standard 0.5 ϕ bins. Dashed lines indicate the interpolation procedure for translating the estimated confidence intervals for the binned data as percentiles (i.e., the horizontal lines) into the corresponding grain size quantiles (i.e., the vertical lines) that bound the estimate of the D_{84} (represented as black solid lines).

In order to illustrate our approach for estimating confidence intervals, we will use grain size data from a recent laboratory experiment, comprising 200 measurements of b-axis diameters; since we preserve each measurement rather than grouping them into size classes, the data can be treated as a binomial experiment, analogous to flipping a coin, wherein each measurement represents the outcome of a single coin flip. These data are sorted in rank order and then used to compute the quantiles of the (Fig. 1). The difference in granularity between the raw data and the standard binned data is illustrated on the figure by adding a cumulative frequency curve based on binned data using the standard 0.5ϕ size classes.

A variety of approaches has been proposed in the statistical literature for estimating quantiles from a sample (Hyndman and Fan, 1996). The differences among methods are greatest for smaller sample sizes, and decrease as n increases. The first step in all approaches is to sort the measured values from lowest to highest and use these to define order statistics $d_{(i)}$ such that $d_{(1)} \leq d_{(2)} \leq ... \leq d_{(n)}$, where, for example, $d_{(1)}$ is the minimum value of d_i.

2.1 Exact solution for a confidence interval

Suppose we wish to compute a specific quantile, say D_p, from our sample of sediment particles. The probabilities of drawing a specific number of particles, k, that are smaller than D_p (i.e., $d_{(k)} < D_p$ and $d_{(k+1)} > D_p$) can be computed from the binomial distribution:

$$Pr(k,n,p) = p^k (1-p)^{n-k} \frac{n!}{k!(n-k)!} \quad (1)$$

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

To define a confidence interval, we first specify the confidence level, usually expressed as $100 \cdot (1-\alpha)\%$. For 95% confidence, $\alpha = 0.05$. Following Meeker et al. (2017), we then find lower and upper values of the order statistics ($d_{(l)}$ and $d_{(u)}$, respectively) such that the coverage probability (P_c) is as close as possible to $1-\alpha$, but no smaller. Coverage probability is defined as:

$$P_c = B(u-1,n,p) - B(l-1,n,p) \tag{2}$$

where $B(j,n,p)$ is the cumulative distribution function for j "successes" in n trials for probability p. The goal, then, is to find integer values l and u that satisfy the condition that $P_c \geq 1-\alpha$, with the additional condition that l and u be approximately symmetric about the expected value of k, $n \cdot p$. The lower and upper confidence limits are then given by $d_{(l)}$ and $d_{(u)}$.

We have created an R function (`QuantBD`) that determines the upper and lower confidence limits, and returns the coverage probability, which is included in the supplementary material for this paper. Our function is based on a script published online by W. Huber [1], which follows the approach described in Meeker et al. (2017). For $n = 200$, $p = 0.84$ and $\alpha = 0.05$ (i.e., 95% confidence level), $l = 159$ and $u = 180$, with a coverage probability (0.953) that is only slightly greater than the desired value of 0.95. This implies that the number of particles in a sample of 200 measurements that would be smaller than the true D_{84} should range from 159 particles to 180 particles, 19 times out of 20. This in turn implies that the true D_{84} could correspond to sample estimates ranging from the 80^{th} percentile (i.e., 159/200) to the 90^{th} percentile (i.e., 180/200). We can translate the bounds into corresponding grain size values using our ranked grain size measurements: the lower bound of 159 corresponds to a measurement of 2.7 mm, and the upper bound corresponds to a measurement of 3.7 mm.

2.2 Approximate solution for equal-area tails

One disadvantage of the exact solution described above is that the areas under the tails differ, as evident from Fig. 2. Meeker et al. (2017) described an alternative approach based on interpolation for finding lower or upper limits for one-sided intervals. This approach can be applied to find two-sided intervals by finding one-sided intervals, each with a confidence level of $1-\alpha/2$. By interpolating between the integer values of k, we can find real numbers for which the binomial distribution has values of $\alpha/2$ and $1-\alpha/2$, which we refer to as l_e and u_e. The corresponding grain sizes can be found by interpolating between measured diameters whose ranked order brackets the real numbers l_e and u_e.

The values of l_e and u_e are indicated on Fig. 2 by dashed vertical lines. As can be seen, the values of l and u generated using the equal tail approximation are shifted to the left of those found by the exact approach. Consequently, the approximate confidence limits are also shifted to the left of the exact approach. The corresponding grain sizes representing the confidence interval are 2.7 mm and 3.6 mm, which are similar to the exact solution presented above.

[1] https://stats.stackexchange.com/q/284970!, last accessed on 19 September, 2019

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

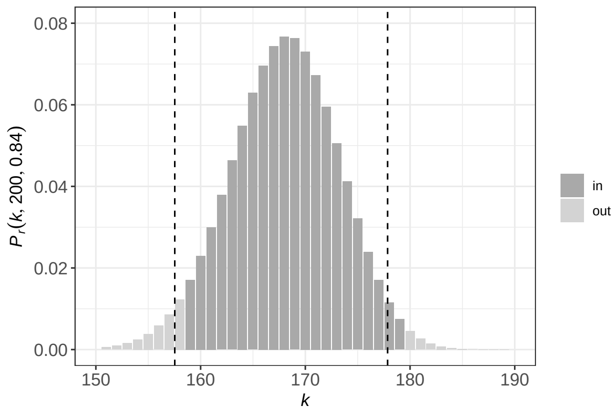

Figure 2. Binomial distribution values for $n = 200$ and $p = 0.84$, displaying the range of k values included in the coverage probability. The dark grey bars indicate which order statistics are included in the confidence interval, and light grey indicates order statistics that lie outside the interval. The vertical dashed lines indicate confidence limits computed by an approximate approach that places equal area under the two tails.

2.3 Approximate solution for binned data

We have adapted the approximate solution described above to allow estimation of confidence limits for binned data, which is accomplished by our R function called `WolmanCI`. We use the equal area approximation of the binomial distribution to compute upper and lower limits of k, and then transform these ordinal values into percentiles by normalizing by the number of observations. Using our sample data, the ordinal confidence bounds $l_e = 157.03$ and $u_e = 177.36$ thus become the percentiles 79% and 89%, respectively.

To estimate the confidence limit in terms of grain sizes, we simply interpolate from the empirical cumulative frequency distribution based on the classed sediment diameters to find the corresponding quantiles. Note that the linear interpolation is applied to $log_2(d)$, and that the interpolated values are then transformed to diameters in mm. This interpolation procedure is represented graphically on Fig. 1. The dashed horizontal lines represent percentile values of l_e/n and u_e/n, while the solid horizontal line represents the percentile of interest (i.e., $p = 0.84$). Our binned sample data yield a confidence interval for the D_{84} that ranges from 2.7 mm to 3.5 mm.

Clearly, the binomial probability approach requires that the sample distribution be known in order to calculate the confidence intervals in units of length. While this is problematic when attempting to predict the statistical power associated with a given sample size, n, before actually collecting the sample, it is possible to use any previously collected distribution to calculate and plot confidence intervals of the bed surface grain sizes, provided the number of observations used to generate the distribution is known. The approach can also be used to estimate the confidence intervals about any previously published grain size distribution, and to assess whether or not a given set of distributions is statistically different or not.

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

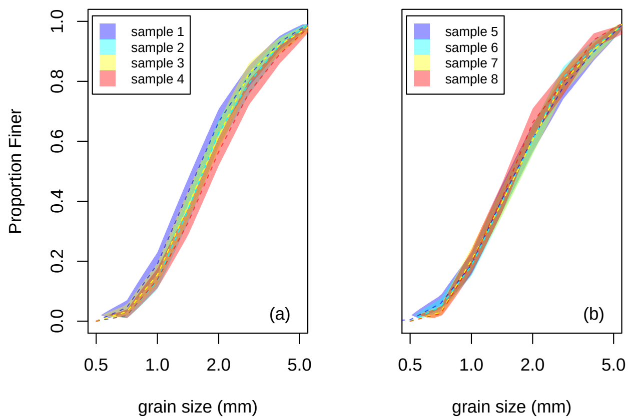

Figure 3. All grain size distributions from a stream table experiment based on a sample size of about 400 observations. The estimated grain sizes are shown, along with the 95% confidence intervals.

3 Confidence interval testing

The approximate method presented in the preceding section can easily be tested numerically by sub-sampling a large population of observations, determining the distribution of resulting percentile size estimates produced by the sub-samples, and comparing it to the confidence interval based on binomial theory. We have eight samples of about 400 observations each from a stream table experiment. Based on the overlap in confidence intervals for the eight samples, the distributions do not appear to be statistically different (see Fig. 3). Therefore, the data have been pooled to form a single data set of 3411 observations. For the purposes of our uncertainty analysis, we let these 3411 observations define the population of interest and then take repeated, random sub-samples (with replacement) of 100 observations from the larger data set. For each sub-sample, we generate the cumulative frequency distribution and then estimate the bed surface D_{16}, D_{50}, and D_{84}.

As seen in Fig. 4, the spread of the estimates from the repeated sub-sampling of the data set is generally similar to the confidence intervals based on binomial theory; the predicted confidence interval containing 50% of the observations (shown in blue) corresponds approximately to the upper and lower quartiles of the box plots, and the 95% confidence interval corresponds approximately to the overall spread of the numerical estimates. A more direct comparison shows that the calculated 50% confidence intervals contain 54% of the grain size estimates from the sub-samples, while the 95% confidence intervals contain 97% of the estimates.

The close match between the confidence intervals calculated from binomial theory and the distribution of percentiles based on sub-sampling supports the validity of the proposed approach for computing confidence limits about the cumulative grain size distribution. Since these confidence limits are straightforward to calculate, we argue that it should be standard practice to plot them on all grain size distribution graphs, particularly those that purport to show a difference between two distributions.

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

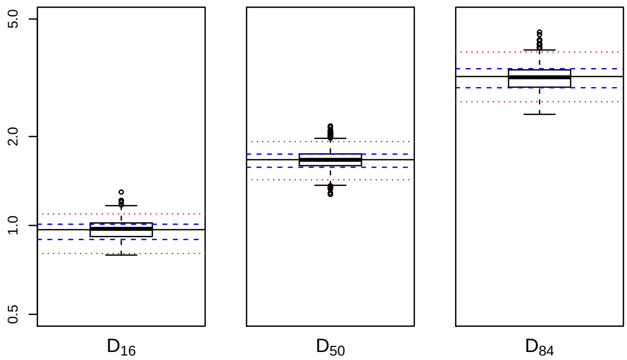

Figure 4. The box-plots represent the distribution of estimates for the D_{16}, D_{50}, and D_{84} of the same bed surface, based on repeatedly selecting 100 measurements from the larger population of observations. The 99% confidence interval estimated using binomial theory is shown in red, the 50% confidence interval is shown in blue, and the 'true' percentile for the population is shown in black, for comparison.

4 Reassessing previous analyses

In order to demonstrate the importance of understanding the uncertainty, we have reanalyzed the results of several previous papers that have compared bed surface texture distributions, but which have not considered uncertainty associated with sampling variability. In most cases, these re-analyses confirm the authors' interpretations, and strengthen them by highlighting which parts of the distributions are different and which are similar, thus allowing for a more nuanced understanding. In some cases, however, the re-analyses demonstrate that the observed differences do not appear to be statistically significant, and suggest that the interpretations and explanations of those differences are not supported but the authors' data. In either case, we believe that adding information about the confidence intervals is a valuable step that should be included in every surface grain size distribution analysis.

Figure 5 plots data published by Kondolf (1997), which were used to compare the bed surface grain size distribution estimated using a pebble count method, and from a truncated bulk sample of the bed surface. Re-plotting the analysis by Kondolf (1997) demonstrates that the coarse tail (i.e., $D_i > 22.6$ mm) of their bulk sample of the bed surface is statistically similar to the coarse end of the distribution for a pebble count, once the sediment finer than 4 mm is excluded from the analysis of the bulk sediment. Interestingly, the finer half of the two distributions appear to be statistically different. While Kondolf (1997) reached essentially the same conclusion, the use of confidence bands about the distributions highlights the statistical similarity of the coarse tail, and can be used to suggest that the transition occurs at a grain size of about 22.6 mm.

The data published by Bunte et al. (2009) include pebble counts of about 400 stones for different channel units in two mountain streams (see Fig. 6). Adding the confidence bands to the distributions emphasizes the advantages of taking larger sample sizes, since the confidence bands are narrower than those for a sample of only 100 stones

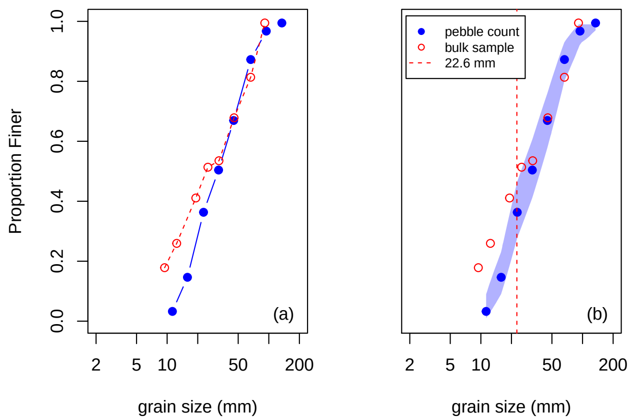

Figure 5. Comparing the bulk surface sample and pebble count distributions, published by Kondolf (1997, their Fig. 3). Panel A shows the traditional grain size distribution representation. Panel B uses the confidence band calculated for the pebble count to highlight where the distributions are statistically similar and where they are different.

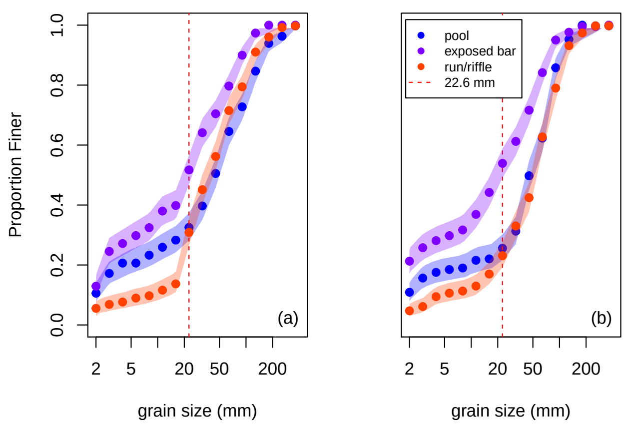

Figure 6. Comparing pebble counts from different channel units. Panel A presents data reported by Bunte et al. (2009) for Willow Creek. Panel B presents data for North St. Vrain Creek.

(e.g., Fig. 5). It also emphasizes that the key difference for the bed texture in pools and in runs or riffles is the fraction of sediment less than about 22.6 mm; the distributions of sediment coarser than this are not statistically different for either stream. This observation suggests that the differences in bed surface texture are likely due to the deposition of finer bed-load sediment in pools on the waning limb of the previous flood hydrograph, and that the bed surface texture of both kinds of mainstem units during flood events could be quite similar. The analysis also clearly demonstrates that size distributions of the exposed channel bars in these two streams are statistically different from both the pools and the runs/riffles. From these plots we can conclude that the bed roughness (which is typically indexed by the bed surface D_{50} or by sediment coarser than that) is similar for the mainstem units (i.e., pools,

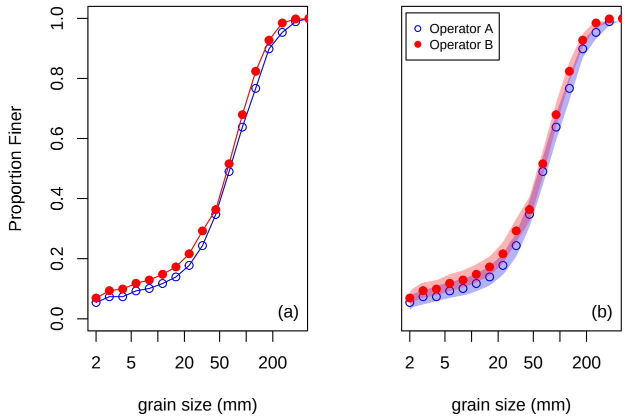

Figure 7. Comparing pebble counts of the same bed surface by different operators. The data plotted were published by Bunte and Abt (2001a). Panel A shows the traditional grain size distribution representation. Panel B uses the confidence band calculated for the pebble count to demonstrate that the two distributions are not statistically different.

and runs/riffles), but that exposed bar surfaces in these two streams are systematically less rough. These kinds of inferences could have important implications for decisions about the spatial resolution of roughness estimates required to build 2D or 3D flow models; it is also possible to reach the same conclusions based on the original data plots in Bunte et al. (2009), but the addition of confidence bands supports the robustness of the inference.

A more fundamental motivation for plotting the binomial confidence bands is illustrated in Fig. 7, which compares the bed surface texture estimated by two different operators using the standard heel-to-toe technique to sample more than 400 stones from the same sedimentological unit. These data were published by Bunte and Abt (2001a) (see their Fig. 7). Based on their original representation of the two distributions (Fig. 7, Panel A), Bunte and Abt (2001a) concluded that

> "operators produced quite different sampling results ... operator B sampled more fine particles and fewer cobbles ... than operator A and produced thus a generally finer distribution."

However, once the confidence bands are plotted (Fig.7, Panel B), it is clear that the differences do not appear statistically significant. A similar analysis of the heel-to-toe sampling method and the sampling frame method advocated by Bunte and Abt (2001a) shows that the distributions produced by the two methods are not statistically different, either. In both cases, the uncertainty associated with sampling variability appears to be greater than any difference between operators or between sampling methods, and thus one cannot claim these differences as evidence for statistically significant effects. It may be the case that there are significant differences among operators or between sampling methods, but larger sample sizes would be required to reduce the magnitude of sampling variability in order to identify those differences.

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

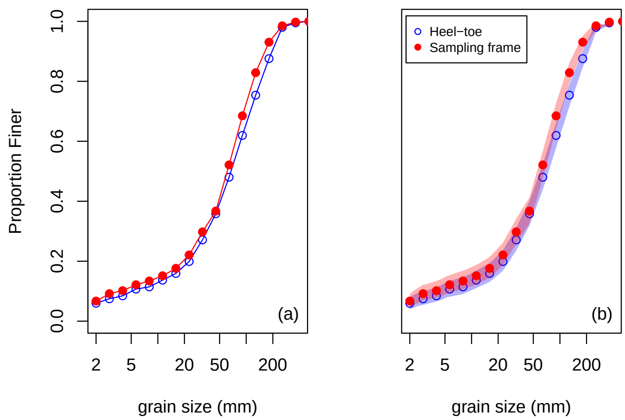

Figure 8. Comparing sampling methods for the same bed surface and operator. The data plotted were published by Bunte and Abt (2001a), and were collected by operator B. Panel A shows the traditional grain size distribution representation. Panel B uses the confidence band calculated for the pebble count to demonstrate that the two distributions do not appear to be statistically different.

Indeed, Hey and Thorne (1983) found that operator errors were difficult to detect for small sample sizes (wherein the sampling uncertainties were comparatively large), but became evident as sample size increased, so the issue at hand is not whether there are important differences between operators, but whether the differences in Fig. 7 are statistically significant. Interestingly, Hey and Thorne (1983) were able to detect operator differences at sample sizes of about 300 stones, whereas Bunte and Abt (2001a) did not detect statistical differences for samples of about 400 stones, indicating either that Hey and Thorne (1983) had larger operator differences than did Bunte and Abt (2001a), or smaller sample uncertainties due to the nature of the sediment size distribution.

5 Determining sample size

Our method for estimating uncertainty requires only the cumulative distribution and the number of measurements used to construct the distribution. Therefore, confidence intervals can be constructed and plotted for virtually all existing surface grain size distributions (provided that the number of stones that were measured is known, which is almost always the case), and future sampling efforts need not be modified in any way in order to take advantage of our method.

The actual uncertainty of an estimated grain size percentile cannot be predicted using our method before the cumulative distribution has been generated. This problem is well recognized, and has been approached in the past by making various assumptions about the distribution shape (Hey and Thorne, 1983; Church et al., 1987; Bunte and Abt, 2001a, b), or using computational approaches (Marcus et al., 1995; Rice and Church, 1996; Green, 2003; Olsen et al., 2005), but in all cases it is still necessary to know something about the spread of the distribution – regardless

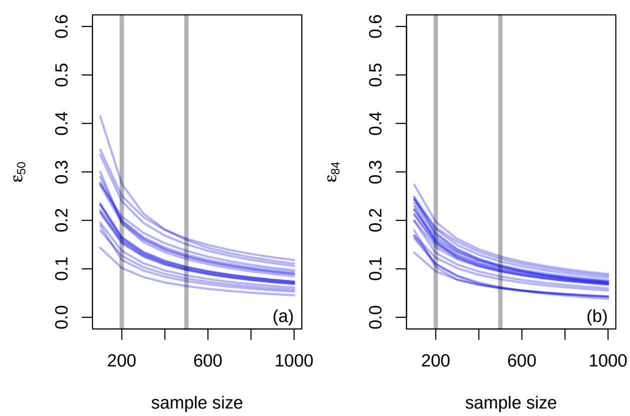

Figure 9. Estimated uncertainty for estimates of D_{50} (Panel A) and D_{84} (Panel B) are plotted against sample size. Curves were generated for all the bed surface samples analysed in this paper (Kondolf et al., 2003; Bunte and Abt, 2001a; Bunte et al., 2009), and for bed surface samples collected by BGC Engineering and students from The University of British Columbia (unpublished data). Vertical lines highlight the range of uncertainties for sample sizes of 200 and 500 stones.

of its assumed shape – in order to predict the level of uncertainty associated with a given sample size. It is perhaps the difficulty of predicting sample uncertainty that has led to the persistent use of the standard 100-stone sample.

5.1 Uncertainty based on field data

Here, we demonstrate the effect of sample size on uncertainty. We begin by calculating the uncertainty of estimates for D_{50} and D_{84} for all the surface samples used in this paper, for eight samples collected by BGC Engineering from gravel bed channels in the Canadian Rocky Mountains, and for samples from two locations on Cheakamus River, British Columbia, collected by undergraduate students from the Department of Geography at The University of British Columbia. Uncertainty (ϵ) is expressed as a proportion of the estimate, calculated as follows:

$$\epsilon = 0.5\left(\frac{D_{upper} - D_{lower}}{D_{est}}\right) \quad (3)$$

where D_{upper} is the upper 95% confidence bound calculated for a given sample size, D_{lower} is the lower confidence bound, and D_{est} is the estimated size for the percentile of interest. For the sake of simplicity, we have assumed that uncertainty is symmetrically distributed about D_{est}, but this is not true for all distribution shapes. Therefore, we can be approximately 95% confident that the interval $D_{est}[1 \pm \epsilon]$ includes the true value of the percentile.

Fig. 9 presents the range of uncertainties for various gravel bed surface samples, including those shown in Figs. (3), (5), (6), and (7). For a sample size of 100 stones, the uncertainties are relatively large, with a mean uncertainty

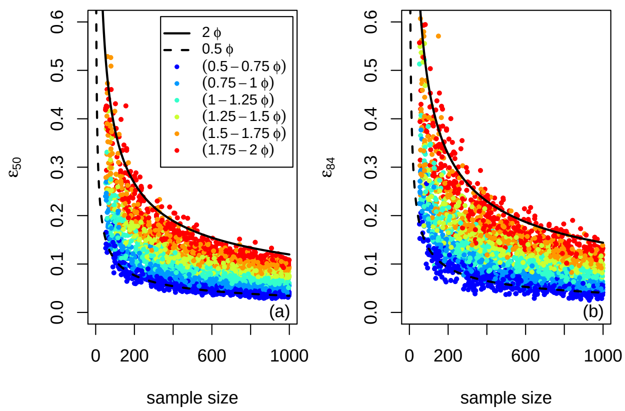

Figure 10. Estimated uncertainty for estimates of D_{50} (Panel A) and D_{84} (Panel B) are plotted against sample size for a simulated set of log normal surface distributions with a range of standard deviations. The markers are color-coded by standard deviation. The bounding curves for $SD_{log} = 0.5\ \phi$ and $SD_{log} = 2.0\ \phi$ are shown for reference, calculated using Eq. (5) and Eq. (7).

across all of the distributions of ±25% for D_{50} and of ±21% for D_{84}. The mean uncertainty drops to ±18% for D_{50} and ±15% for D_{84} for a sample of 200 stones, and to ±11% (D_{50}) and ±9% (D_{84}) for 500 stones.

5.2 Uncertainty for Log-normal distributions

We can also approach this problem by assuming that bed surface texture distributions are approximately log-normal, but have varying degrees of gradation, indicated by a standard deviation expressed in ϕ units (SD_{log}). As a point of comparison, if we estimate the SD_{log} for the samples analyzed in the previous section by assuming that $SD_{log} = \log_2 D_{84} - \log_2 D_{50}$, then SD_{log} ranges from 0.8 to 1.8, with a median value of 1. For those samples, the largest values of SD_{log} were associated with samples from channels on steep gravel bed fans and on bar top surfaces, while samples characterizing the bed of typical gravel bed streams had values close to the median value.

We generated a relation between uncertainty and sample size by first simulating 3000 log-normal grain size distributions with D_{50} ranging from 22.6 mm to 90.5 mm, n ranging from 51 to 999 stones, and SD_{log} ranging from 0.5 ϕ to 2 ϕ. We then used least-squares regression to fit models of the form

$$\ln(\epsilon) = a \cdot n + b \cdot SD_{log} + c \tag{4}$$

where a, b, and c are the estimated coefficients. The empirical model describing the uncertainty of D_{50} has an adjusted R^2 value of 0.95, with the variable n explaining about 47% of the total variance, and SD_{log} explaining 47% of the variance. The model for D_{84} has an adjusted R^2 value of 0.9 with the variables n and SD_{log} explaining the similar amounts of the total variance (46% and 45%, respectively).

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

Table 1. Coefficient values for estimating uncertainty in D_{50} and D_{84} as a function of SD_{log} and sample size (n) using Eqs. (5) and (7)

Coef.	0.75ϕ	1.00ϕ	1.25ϕ	1.50ϕ	1.75ϕ	2.00ϕ
A	0.278	0.486	0.694	0.901	1.109	1.317
B	0.531	0.742	0.952	1.163	1.374	1.584

After back-transforming from logarithms, the equation describing the uncertainty in D_{50} can be expressed as:

$$\epsilon_{50} = A \cdot n^{-0.498} \tag{5}$$

where the coefficient A is given by:

$$A = \exp(-0.346 + 0.832 SD_{log}) \tag{6}$$

The equation for estimating uncertainty in D_{84} are:

$$\epsilon_{84} = B \cdot n^{-0.51} \tag{7}$$

where B is given by:

$$B = \exp(-0.1 + 0.842 SD_{log}) \tag{8}$$

Table 1 provides values of A and B for a range of standard deviations.

6 Practical implications of uncertainty

The implications of uncertainty can be important in a range of practical applications. Here, we translate uncertainty in grain size percentiles into uncertainty in the critical discharge for significant morphologic change using data for Fishtrap Creek, a gravel bed stream in British Columbia that has been studied by the authors (Phillips and Eaton, 2009; Eaton et al., 2010a, b). The estimated bed surface D_{50} for Fishtrap Creek is about 55 mm, which we estimate becomes entrained at a shear stress of 40 Pa, corresponding to a discharge of about 2.5 m^3s^{-1}(Eaton et al., 2010b). If we assume that significant channel change can be expected when D_{50} becomes fully mobile (which occurs at about twice the entrainment threshold), then we would expect channel change to occur at a shear stress of 80 Pa, which corresponds to a critical discharge of 8.3 m^3s^{-1}, based on the stage-discharge relations published by Phillips and Eaton (2009).

Since we used the standard technique of sampling 100 stones to estimate D_{50} and since the standard deviation of the bed surface distribution is about 1.0ϕ, we can assume that the uncertainty will be about ±16%, based on Eqs. (5 and 6), which in turn suggests that we can expect the actual surface D_{50} to be as small as 46 mm or as large as

Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4
Manuscript under review for journal Earth Surf. Dynam.
Discussion started: 13 February 2019

64 mm. This range of D_{50} values translates to shear stresses that produce full mobility that range from 67 Pa to 93 Pa. This in turn translates to critical discharge values for morphologic change ranging from 5.9 $m^3s^{-1}$to 11.1 m^3s^{-1}, which correspond to return periods of about 1.5 years and 7.2 years, based on the flood frequency analysis presented in Eaton et al. (2010b). Specifying a critical discharge for morphologic change that lies somewhere between a flood that occurs virtually every year and one that occurs about once a decade, on average, is of little practical use, and highlights the cost of relatively imprecise sampling techniques.

If we had taken a sample of 500 stones, we could assert that the true value of D_{50} would likely fall between 51 mm and 59 mm, assuming an uncertainty of ±7%. The estimates of the critical discharge would range from 7.2 m^3s^{-1} to 9.5 m^3s^{-1}, which in turn correspond to return periods of 2 years and 4.1 years, respectively. This constrains the problem more tightly, and is of much more practical use for managing the potential geohazards associated with channel change.

Operationally, it takes about 20 minutes to sample 100 stones from a typical gravel bed river, and a bit over an hour to sample 500 stones, so the effort required to sample the larger number of stones is far from prohibitive. Furthermore, computer-based analyses using photographs of the channel bed may be able to identify virtually all of the particles on the bed surface, and generate even larger samples. The statistical advantage of the potential increase in sample size are obvious, and justify further concerted development of these computer-based methods, in our opinion.

7 Conclusions

Based on the statistical approach presented in this paper, we developed a suite of functions in the R language that can be used to estimate the uncertainty of any percentile in a cumulative grain size distribution (see the supplemental material for the source code). The approach uses binomial theory to generate uncertainty estimates for any cumulative grain size distribution based on pebble count data, and requires only that the total number of stones used to generate the distribution is known. Approaches were developed for cases in which individual grain sizes are known and in which data are binned (e.g., into ϕ classes).

By estimating the uncertainty for each percentile in the distribution, the uncertainty can be displayed graphically as a polygon surrounding the distribution estimates. When comparing two different distributions, this means of displaying grain size distribution data highlights which distributions appear statistically different, and which do not.

Our analysis of various samples collected in the field demonstrates that the uncertainty depends on the shape of the distribution, with more widely graded sediments having higher uncertainty than narrowly graded ones. Our analysis also suggests that typical gravel bed river channels have a similar gradation, and that the typical uncertainty of the D_{50} varies from ±25% for a sample size of 100 observations to about ±11% for 500 observations.

When designing a bed sampling program, it is useful to estimate the precision of the sampling strategy and to select the sample size accordingly; to do so, we must first assume something about the spread of the data (assuming a log-normal distribution), and then verify the uncertainty after collecting the samples. Simple equations for predicting uncertainty (as a percent of the estimate) are presented here to help workers select the appropriate sample size for the intended purpose of the data.

Code and data availability. Both the analysis code and the data used to create all of the figures in this paper are available online (http://doi.org/10.5281/zenodo.2551824). The R package used to estimate the confidence intervals for grain size distributions is also available online (http://doi.org/10.5281/zenodo.2551826)

Author contributions. B.C. Eaton drafted the manuscript, created the figures and tables, and wrote the code for the associated modelling and analysis in the manuscript; R.D. Moore developed the statistical basis for the approach, wrote the code to execute the error calculations, reviewed and edited the manuscript, and helped conceptualize the paper; and L.G. MacKenzie collected the laboratory data used in the paper, tested the analysis methods presented in this paper, and reviewed and edited the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

References

Buffington, J. M. and Montgomery, D. R.: A procedure for classifying textural facies in gravel-bed rivers, Water Resources Research, 35, 1903–1914, 1999.

Bunte, K. and Abt, S. R.: Sampling frame for Improving pebble count accuracy in coarse gravel-bed streams, Journal of the American Water Resources Association, 37, 1001–1014, 2001a.

Bunte, K. and Abt, S. R.: Sampling surface and subsurface particle-size distributions in wadable gravel-and cobble-bed streams for analyses in sediment transport, hydraulics, and streambed monitoring, Gen. Tech. Rep. RMRS-GTR-74, US Department of Agriculture, Rocky Mountain Research Station, Fort Collins, CO, 2001b.

Bunte, K., Abt, S. R., Potyondy, J. P., and Swingle, K. W.: Comparison of three pebble count protocols (EMAP, PIBO and SFT) in two mountain gravel-bed streams, Journal of the American Water Resources Association, 45, 1209–1227, 2009.

Church, M., McLean, D. G., and Wolcott, J. F.: River bed gravels: sampling and analysis, in: Sediment Transport in Gravel-bed Rivers, edited by Thorne, C., Bathurst, J., and Hey, R., pp. 43–88, John Wiley & Sons Ltd., 1987.

Eaton, B. C., Andrews, C., Giles, T. R., and Phillips, J. C.: Wildfire, morphologic change and bed material transport at Fishtrap Creek, British Columbia, Geomorphology, 118, 409–424, 2010a.

Eaton, B. C., Moore, R. D., and Giles, T. R.: Forest fire, bank strength and channel instability: the 'unusual' response of Fishtrap Creek, British Columbia, Earth Surface Processes and Landforms, 35, 1167–1183, 2010b.

Green, J. C.: The precision of sampling grain-size percentile using the Wolman method, Earth Surface Processes and Landforms, 28, 979–991, 2003.

Hey, R. D. and Thorne, C. R.: Accuracy of Surface Samples from Gravel Bed Material, Journal of Hydraulic Engineering, 109, 842–851, https://doi.org/10.1061/(ASCE)0733-9429(1983)109:6(842), 1983.

Hyndman, R. J. and Fan, Y.: Sample quantiles in statistical packages, The American Statistician, 50, 361–365, 1996.

Kondolf, G. M.: Application of the pebble count: notes on method, purpose, and variants, Journal of the American Water Resources Association, 33, 79–87, 1997.

Kondolf, G. M. and Li, S.: The pebble count technique for quantifying surface bed material size in instream flow studies, Rivers, 3, 80–87, 1992.

Kondolf, G. M., Lisle, T. E., and Wolman, G. M.: Bed Sediment Measurement, chap. 13, pp. 347–395, John Wiley, 2003.

Latulippe, C., Lapointe, M. F., and Talbot, T.: Visual characterization technique for gravel-cobble river bed surface sediments; validation and environmental applications Contribution to the programme of CIRSA(Centre Interuniversitaire de Recherche sur le Saumon Atlantique), Earth Surface Processes and Landforms, 26, 307–318, 2001.

Leopold, L. B.: An improved method for size distribution of stream bed gravel, Water Resources Research, 6, 1357–1366, 1970.

Marcus, W. A., Ladd, S. C., Stoughton, J. A., and Stock, J. D.: Pebble counts and the role of user-dependent bias in documenting sediment size distributions, Water Resources Research, 31, 2625–2631, 1995.

Meeker, W. Q., Hahn, G. J., and Escobar, L. A.: Statistical Intervals, John Wiley & Sons, Hoboken, New Jersey, USA, 2nd edn., 2017.

Olsen, D. S., Roper, B. B., Kershner, J. L., Henderson, R., and Archer, E.: Sources of variability in conducting pebble counts: their potential influence on the results of stream monitoring programs, JAWRA Journal of the American Water Resources Association, 41, 1225–1236, 2005.

Phillips, J. C. and Eaton, B. C.: Detecting the timing of morphologic change using stage-discharge regressions: a case study at Fishtrap Creek, British Columbia, Canada, Canadian Water Resources Journal, 34, 285–300, 2009.

Rice, S. and Church, M.: Sampling surficial fluvial gravels: The precision of size distribution percentile estimates, Journal of Sedimentary Research, 66, 654–665, 1996.

Wolman, M. G.: A method of sampling coarse river-bed material, EOS, Transactions American Geophysical Union, 35, 951–956, 1954.