Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2019-4-RC1, 2019

Earth Surface Dynamics
Discussions
Open Access

Interactive comment on "Estimating confidence intervals for gravel bed surface grain size distributions" *by* Brett C. Eaton et al.

Kristin Bunte (Referee)

kbunte@engr.colostate.edu

Received and published: 16 March 2019

General comments Eaton et al. (2019) present what seems to be a new way to compute confidence intervals around grain-size distributions that is based on the binomial approach. Encouraging the routine computation of confidence intervals around sampled grain-size distributions is a worthwhile undertaking and often a monitoring requirement for detecting change in rivers beds over time or space. The study by Eaton et al. sets out to provide such a tool. However, the authors do not succeed in making their tool easily accessibleâĂŤin fact as presented, their approach remains a black box to most potential users. The manuscript does not provide more than general statistical background information and no step-by-step explanations are given on how a potential user could apply the authors' approach to his/her field data. The reader is not much the wiser even after downloading the supplemental material which contains computer code but still no instructions on how to apply the computations. For a user whose basic work tool is spreadsheet computation, the study by Eaton et al. (2019) provides no help for computing confidence intervals.

Computation of confidence bands around grain-size distributions without assuming an underlying distribution type is not a new idea. Fripp and Diplas (1993) presented a binominal approach to compute the relation between sample size and error around individual percentiles. The study by Church and Rice (1996) applied a bootstrap approach to a large pebble count of 3500 particles and computed error bands around various percentiles of the grain size distribution. The grain-size distributions did not fit a particular distribution type, but the bootstrap confidence limits were reasonably close to those computed assuming an underlying skewed log-normal distribution. Petrie and Diplas (2002) cautioned that "... the binomial distribution considers only two possibilities for each particle sampled: (1) the particle is within a specific size class (e.g., smaller than a certain size) or (2) the particle is not within the specified size class. The binomial distribution is then inadequate to use for representing entire size distributions." To overcome this limitation and to compute confidence bands around the cumulative frequency distribution from a pebble count with data binned into size classes while considering distribution characteristics of the distribution, Petrie and Diplas (2000) developed a multinomial approach.

While the study presented by Eaton et al. (2019) is successful in raising awareness that the n=100 sample size is too low to attain reasonable accuracy for pebble counts in most gravel beds and that sample sizes of 400 or 500 particles are required to enable statistical evaluations about sameness or difference, the study does not succeed in presenting its computational approach in an easy to understand way. Providing computer code in R-language is not helpful for most users, hence the authors' computations cannot be repeated or applied by users who are not expert statisticians but are seeking to determine confidence limits around their sampled grain-size distributions.

The authors display the confidence bands that they drew with their binomial approach around grain-size distributions sampled in other studies (Kondolf, 1992; Bunte et al. 2009, Bunte and Abt, 2011) and go on to discuss whether the now-drawn confidence bands warrant the interpretations made in the original studies. In the final sections of the study, the authors show general relations between sampling error, as computed with their binomial approach, and sample size as well as distribution sorting.

Suggestions to the authors for improving their manuscript 1) Reference prior work and build on it Eaton et al. (2019) should discuss prior studies that likewise compute errors around percentiles without assuming an underlying distribution type and explain the improvements and advantages offered in the study presented. What reason is there for a user to select the authors' approach if the authors do not explain WHY their approach constitutes an improvement?

2) Provide explanations and instructions In order for readers to apply the binomial approach to their own data, the authors need to provide a step-by step explanation on how to use their approach rather than referring to a book on statistics, pointing to a website, and offering computer code in R-language. Offering a reader access to computer code is a courtesy, but not a substitute for a step-by step explanation, especially not for a very hands-on and applied topic of monitoring bed-material changes.

3) Comparison of results to those from prior work: How do percentile errors computed from the authors' binomial approach compare to percentile errors computed from other approaches? Apart from a similarity of sampling errors around the D50 and D84 that the authors computed from their binomial as well as a bootstrap approach for a symmetrical grain-size distribution (the authors' flume experiment), the authors do not show how their binomial approach to computing confidence bands relates to confidence bands computed from other approaches. The authors should apply their binomial approach together with the approaches suggested by Fripp and Diplas (1993), Petrie and Diplas (2000), and Rice and Church (1996) as well as simply to sample-size equations for an error around the mean to a few pebble-count distributions that differ in

their sorting and skewness (esp. the extent of a fine tail) and then assess differences and similarities between results.

4) Explain whether or how confidence intervals computed from the binomial approach are affected by sorting and skewness of a sampled grain-size distribution While the authors show that confidence bands increase in width with a distribution's sorting coefficient, the authors do not explain how exactly sorting (and skewness) of a sampled grain-size distribution (e.g., a tail of fines) flow into the computation of confidence intervals based on the binomial approach. The binomial approach introduced by Fripp and Diplas (1993) does not seem to involve sorting or skewness of the sampled distribution, suggesting that confidence intervals from a binomial approach are similar for all percentiles within a sampled grain-size distribution with a known sample size and number of size classes.

5) Have a user in mind and offer a procedure that is reasonably easy to be applied by the user The authors provide a study that is of interest to users who are involved in relations of sample size to error. However, the study is geared towards a statistically expert audience rather than the needs of non-expert potential users. If the authors' work is to be applied for monitoring purposes by staff from environmental agencies or consulting and by those whose main interest is not statistical but who need to apply such relations, then the authors need to provide detailed explanation and instruction. A spreadsheet implementation of their computations of a percentile error would be considerably more helpful than code in R-language.

6) Editing suggestions Figures provided by the authors are generally fine, but considering that the study discusses plotted details of whether or not confidence bands overlap, a larger figure size would be helpful. It would also be helpful to place the figures below their first mention in the text, not simply at the top of the page with a mention somewhere below on the page. With respect to writing style and typos (etc.), the manuscript is well written and clean.

Specific Comments p.2, l. 15: "...but the largest source of uncertainty in many cases is likely to be sampling variability, which is a function of sample size." How do the authors know that sampling variability (do they mean statistical uncertainty due to a poorly sorted channel bed?) rather than methodological differences (e.g., measurements of particle sizes, spatial heterogeneity, differences in the sampled channel width or leaving poorly accessible stream locations unsampled) is the most likely factor causing uncertainty? The comparative study by Bunte et al. (2009) showed that differences in sampling outcomes due to methodological variability can be huge.

p. 2, line 21: "... We then use this approach to demonstrate that the higher percentiles, such as D84, are subject to substantial uncertainty for typically used sample sizes, and that..." 1) Given this statement, it is odd then that the confidence bands drawn by the authors around the size-distributions from two streams sampled by Bunte et al. (2009) and another stream sampled by Bunte and Abt (2001) are all narrower for the D84 than for the D75 and the D95. 2) That statement is not backed by results from other studies: Rice and Church (1996) have shown for a very large pebble count that uncertainty was lowest for the D75 and D84 sizes, followed by the D50 and D95 sizes, and highest for the smallest percentiles. Green (1993) corroborated this finding; on average, the D73 could be determined with the least uncertainty. Similarly, Bunte and Abt (2001a) found in their field study that uncertainty was lowest for the D50 and the D75, slightly higher for the D84, D95 and D25, and percentiles lower than D25 were subject to the highest uncertainties.

p. 3, l. 3: "...since we preserve each measurement rather than grouping them into size classes, the data can be treated as a binomial experiment, ..." Does that mean that the binominal computations is not applicable to field data binned in 0.5 phi units which results from measuring particle size using a 0.5-phi template?

In Eq. 1, Pr and p are not defined

p. 4, line 10-19: The description of the methodology is too vague. To allow a reader to replicate the computations, authors need to provide step-by-step guidance. Reference to websites and other studies is not sufficient for a paper that would like to introduce a new approach to computing confidence bands.

p. 4, line 21: That statement comes out of the blue . . .what areas? What tails?? Fig. 2 does not provide much help either.

p. 5, line 1-5: Again, step-by-step instructions are needed to allow a reader to replicate the authors' approach.

p. 6, line 5-6 ". . .Based on the overlap in 5 confidence intervals for the eight samples, the distributions do not appear to be statistically different (see Fig. 3). . .. 1) Confidence bands plotted by the authors for their stream table sediment overlap for samples 2 and 3, but not for samples 1 and 4 (Fig. 3, panel A). 2) With respect to their multinomial approach, Petrie and Diplas (2000) stated that error bands are identical for all particle-size distributions as long as the value for alpha (e.g., 0.05) and the number of sampled size classes remain the same. For the authors' 8 samples from the stream table sediment surface, I assume that the same number of size classes were collected in each of the 8 samples and that the same alpha value was applied to all computed confidence bands. If the statement by Petrie and Diplas (2000) was true for the error bands conducted by the authors, then why do the error bands plotted in Fig. 3 differ between samples? 3) The authors use as basis for their analyses a sand-rich sand-gravel mixture with a D50 near 1.5 mm. The lengths of b-axes appear to have been determined to a precision of two decimals (e.g., 0.53 mm). It is difficult to imagine how a pebble count was performed and particle sizes were measured on sediment this small.

p. 7, Fig. 4: 1) While the box of box and whisker plots typically shows the quartiles, there is less standardization of what the whiskers represent. Please indicate what the whiskers in this plot represent. It can't be the overall spread because "outliers" are plotted as dots. Please define. 2) What parameter is plotted on the y-axis? Please clarify. 3) It would have been useful to show the 95% confidence limits, too.

p. 6, line 9-19. The authors state that they found a close match between the confidence bands computed from the binomial approach and a bootstrap approach (Fig. 4) for an unskewed grain-size distribution (i.e., their stream table sediment). The comparison plot by Petrie and Diplas (2000) for a pebble count from the Mamquam River shows that the confidence bands computed with the approach by Fripp and Diplas (1993) are between ïĆś0.02 and 0.06 phi-units higher than those from the bootstrap approach computed by Rice and Church (1996). Is the binomial approach by Fripp and Diplas (1993) similar or different to the authors' binomial approach? Does a binomial approach yield wider confidence bands than a bootstrap approach?

p. 7, lines 11-19: In the authors' reassessment of particle-size distributions from Kondolf (1997) and from Bunte et al. (2009), the authors need to clearly state to what percentage confidence the plotted confidence bands refer? I assume they are 95% confidence bands. Please clarify.

p. 8, Fig. 6: The study by Eaton et al. (2019) has drawn confidence limits around grain-size distributions from three Rocky Mountain gravel-bed streams sampled by Bunte et al. (2009) and Bunte and Abt (2001). 1) Based on visual examination of the error bands plotted in Fig. 6, I'd say that for Willow Creek, the error bands for riffles and pools are different except for the narrow range between 20 and 50 mm within which they cross. 2) The plotted confidence intervals for Willow Creek and the St. Vrain are jagged around the sampled distribution and seem to widen notably for the flatter sections of the cumulative size distribution but neck down for the steeper sections. The authors offer no explanation for this phenomenon.

p. 10, line 9-14: The authors write: "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),. . ."

If computation of the width of the confidence interval for any percentile of interest requires only knowledge of the sampled distribution and sample size n, and if the computation is conducted for each percentile individually, then how does the spread or sorting of the sampled distribution influence the computed confidence interval? Please CLARIFY!

p. 11, Fig. 9 and p. 12, Fig. 10: 1) The units in which the error is computed needs to be clearly stated. Somewhere down in the text the reader gets a hint that the error pertains to a percentage error in mm units. 2) The findings that percentile errors decrease with sample size and with the distribution sorting is in and of itself nothing new. What is new here is that the error is computed from the authors' binomial approach (assuming an underlying log-normal distribution for Fig. 10). To allow a reader to see whether there is a difference between errors computed from the authors' binomial approach and other approaches (e.g., Fripp and Diplas (1993) or simply errors around a mean), the computed relations between errors and n should be compared to errors computed with other approaches. 3) For comparison with other studies that compute percentile errors in terms of absolute +- error in phi-units it would be helpful if the error-n relations in Fig. 9 had a second y-axis with error in terms of the absolute +- error in phi-units. 4) It would be useful if the relation of error to n was also provided for the error around the D16.

p. 12, line 8: The authors state that for a given n and sorting, errors are largest for steep gravel fans and bar top surfaces and smaller for typical gravel beds with a sorting near 1. That is a useful comment. It would be even more useful to elaborate a little bit here on what kind of sorting values to expect for different morphological or sedimentary channel units and hence what a user needs to expect in terms of the error - sample size relation.

p. 14, line 12-13: I am afraid that the authors' time estimates refer to dry deposits of mainly mid-sized gravels. The time requirements for a 500-particle pebble count increases to about 5 hours when sampling in poorly wadeable conditions, in the presence of abundant algae and large woody debris, under overhanging bushes, and with particles being next to irretrievable from the bed because they are tightly wedged within neighboring particles or small particles placed in tiny pockets between large clasts. The necessity for a large sample size remains, but usersâĂŤand their funding agencies!âĂŤneed to commit to realistic time requirement.

p. 14, Line 19: “a suit of functions in R language that can be used to estimate the uncertainty of any percentile in a cumulative grain-size distribution (see the supplemental material. . .)“ I personally find “a suite of functions in R language” not useful, and I am not sure how many other potentially interested users have access to something in R-language. If the authors would like to provide helpful support to a user, then please provide a spreadsheet.

Typos etc. p. 2, l.5: The value should be 22.6 (=2ˆ0.5*16), not 22.7. p. 3 L. 5. . .compute the quantiles of the (Fig. 1). Something is off in that sentence. p. 4, Footnote: The access date is in the future.

Interactive comment on Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2019-4, 2019.