Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2019-4-AC1, 2019

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

Brett C. Eaton et al.

brett.eaton@ubc.ca

Received and published: 24 April 2019

1 General Comments

The general comments from reviewer 1 are presented below, and are discussed at some length. We attempt to address all the key issues raised there, and to highlight how we are responding to those comments in our revisions. We thank the reviewer very much for such a careful and helpful review of our paper.

1.1 Comment 1

> 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 accessible: in 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.

1.2 reply by authors

This is very useful feedback for us. Our intention is indeed to provide a user-friendly tool that implements binomial statistical theory (which is the basis for the approximation presented by Fripp and Diplas, 1993, mentioned later by this reviewer) to calculate confidence bands about grain size distributions to prevent type 1 statistical errors (i.e. false positives, where differences between distributions is asserted where it is not statistically justified). The reviewer is incorrect in contending that we do "not provide more than general statistical background information", and that we do not provide "step-by-step explanations" on how to implement the analysis. The manuscript lays out the precise statistical basis for the calculations we make using the different methodologies appropriate for different kinds of data (i.e. raw observations and binned data). However, it is clear that we have failed to communicate this appropriately for our target audience.

We are in the process of completely re-writing the introductory section of the paper to better explain

1. how the binomial distribution can be applied to both raw data comprising n measurements of b-axis diameters and also to the typical binned data collected in the field; and

2. how the binomial theory can be used to generate confidence intervals about an estimate of a given grain size percentile.

The process is summarized using an example shown in a new figure we are integrating into the paper (Fig 1, below). In that figure, panel A presents the cumulative distribution of the population defined by 3411 individual measurements of b-axis diameters. Panel B shows a sample of 100 stones taken from that population and indicates the difference between the population median grain size and the sample median. Using the binomial equation, we calculate the uncertainty in terms of grain size percentile of the sample (a step that requires the implementation of the computer code that we provided, but which can be approximated using the normal distribution, as described by Fripp and Diplas, 1993). The percentile uncertainty is indicated in panel C using grey shading and arrows. Panel D indicates how this uncertainty in percentiles can be translated into a confidence interval for the sample that will contain the true population median grain size 95% of the time.

In addition, we are preparing an appendix to the paper that describes how to use the simpler normal approximation to the binomial distribution to calculate the confidence interval, and we are developing a spreadsheet implementation of that approach.

We have also created an appendix containing reference tables of the uncertainty (expressed as sample percentiles (corresponding to panel C of Fig 1) for a range of percentiles of interest (i.e. D_{10}, D_{15}, $D_{20} \ldots D_{90}$), sample size (n), and acceptable confidence limit (α).

Finally, we are developing a basic introduction to the R Package that demonstrates how to enter standard grain size data, set the confidence interval parameters (i.e., what D_i values are of interest, and what α is acceptable), run the analysis to generate the confidence intervals, and then export the results as a text file that could be imported into any standard spreadsheet application. We appreciate that many people still use spreadsheets as their go-to analysis tool, and we are trying to accommodate those users. However, it is our experience that there is strong demand for the kind of R-based analysis tools that can be incorporated into scripting languages like R, so we have focused on the `bicalc` package as the primary means of distributing our tools.

1.3 Comment 2

> 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 binomial 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.

1.4 reply by authors

This is also very useful information for us, and we have read the papers with interest. The work by Diplas and colleagues is particularly relevant and strengthens our paper. There are some important differences, but the analysis by Fripp and Diplas (1993) can be used as a jumping off point for our analysis. They use a normal approximation to the binomial equation and apply it to binned data like that typically collected in the field. We are re-writing our manuscript to use that paper as the basis from which we start, we describe that approach in the appendix we are writing, and we will implement a version of it in a spreadsheet that we are developing to accompany this paper.

The paper by Rice and Church (1996) was the inspiration for the bootstrap re-sampling that we presented in our original paper (i.e. it is the boxplot in our original figure 4). However, we have clearly not done justice to the analysis presented therein, so we are expanding that section. Since our intention initially was to present a tool using relatively standard statistical approaches for generating confidence bands around distributions, we focused on validating the approach rather than comparing it to previous attempts. To be clear, Rice and Church (1996) did not present a *method* for estimating uncertainties about different grain size percentiles, they presented *estimates* of those uncertainties based on resampling measurements from a population of over 3000 b-axis diameter values (as described clearly by Petrie and Diplas, 2000). Those estimates can only strictly be applied to the population of stones that they collected,

and only approximately to populations with similar characteristics. Our analysis in section 5.1 and Figure 9 show that many, but not all, gravel beds are similar to their data from Mamquam River. To apply their method to another river would require that a sample large enough to nearly perfectly define the population be taken (i.e. about 3500 stones), and then resampled with different sample sizes to estimate the uncertainty associated with the given sample size. Since the binomial theory approach replicates their sampling estimates (see Petrie and Diplas, 2000, to confirm this), then it is far more efficient to implement the binomial theory than to undertake a laborious empirical estimation. In our revised paper, we will make these issues clear.

The work by Petrie and Diplas with multinomial theory is primarily focused on determining the sample size required for a given level of accuracy for estimating the shape and relative position of the cumulative grain size distribution, using binned data. Our approach and intent is different: we develop our statistical theory using individual measurements of b-axis diameters, and we develop confidence bounds to be plotted when comparing distributions to avoid type 1 and 2 statistical errors. In this context, the binomial approach is most appropriate – a point made clearly by our successful validation of our approach against the statistics of repeated resampling of a known population (i.e. our Figure 9). Given the data that we use, it is not true to say "...the binomial distribution considers only two possibilities for each particle sampled: (1) the particle is within a specific size class (or (2) the particle is not within the specified size class." Our implementation of binomial theory is based on the interpretation that a measured stone is either (a) greater than a percentile of interest for the population, or (b) less than or equal to the percentile of interest, with no reference to or limitation imposed by having binned data. In this context, the estimation of j percentiles involves the execution of j independent binomial experiments with assumed probabilities corresponding to the percentile of interest. To test the difference between our approach and the traditional binned data, we will use the interpolation scheme described in our paper to directly compare the distributions based on all measurements, and the binned data (this was done implicitly in Fig 1 of our original paper, but we are much more explicit about this in our revisions).

While current practice in the field is still to collect binned data, the automated techniques for grain size analysis that are standard practice in most experimental laboratories, and which are being increasingly deployed in the field promise to deliver much more data than can be collected manually, and will obviate the need for binned data. Our methodology is best leveraged in that context, using the automated data analysis approach possible using languages like R and Python. Therefore, our differentiation between binned data and the underlying b-axis diameter measurements is not simply a technical one, it is based on our perceptions of the future data types that will be commonly used.

1.5 Comment 3

> 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.

1.6 reply by authors

We are very grateful for the feedback about the relative difficulty in understanding our approach, and about the need for addition means of implementing our tools for estimating the confidence bands. We are trying to respond to the first point by re-writing the section of the paper presenting the method, and to the second by developing reference materials in two appendices, a "how-to" guide for using the R code, and a spreadsheet implementing the normal approximation to our solution, as described by Fripp and Diplas (1993).

2 Recommendations for improving the paper

The reviewer made several helpful suggestions for improving the paper, listed below:

> 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 therefor a user to select the authors' approach if the authors do not explain WHY their approach constitutes an improvement?

As described (in part) above, we are improving the links between our paper and the previous work. We are also re-iterating in the paper that our main purpose is to produce a user-friendly introduction to the basic method for estimating confidence bounds using binomial theory (as approximated by Fripp and Diplas, 1993; and validated empirically by Rice and Church's empirical analysis, as well as our own parallel empirical analysis of our own relatively large population of b-axis measurements). We will point out that our approach is statistcally conventional, has precedents in the literature, and is consistent with empriical analyses. We will also more strongly articulate that the key message of the paper is that all grain size curves ought to be plotted with confidence intervals, particularly when two distributions are being compared. The method by which the confidence interval is calculated is less important than the fact that it is calculated at all. Current practice seems to be to ignore all statistical uncertainty (despite the precedents in the literature); we hope this paper will make it easy for researchers to actually conduct the statistical analysis, and include it in their analysese.

> 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.

With this particular comment in mind, we have re-written the manuscript and generated various reference materials. We are currently "test-driving" the new material with graduate student who do not have extensive statistical backgrounds.

> 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 D84that the authors computed from their binomial as well as a bootstrap approach for asymmetrical 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 difference sand similarities between results.

In our revised paper, we make the links to the cited literature clear, and we make it more clear that we did in fact replicate the approach described by Rice and Church, and then compare it to the binomial methodology we describe.

> 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 co-efficient, 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.

The revised description of our methodology and the new figures we are developing (e.g. Fig. 1, attached), will address this point.

> 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.

We are developing resources that address this point, and we are particularly thankfull for this feedback, since our main purpose is to make it easy for people to use our approach. Sometimes we forget that distributing an R package is a great solution for only a sub-set of the community we hope to reach.

> 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 some-where below on the page. With respect to writing style and typos (etc.), the manuscript is well written and clean

We are re-working our figures, and will leave it to the editorial staff to properly place the figures in the final version of the manuscript.

3 Specific comments

The reviewer also provides a list of specific comments that will improve the paper. We are currently working to integrate those specific comments.

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

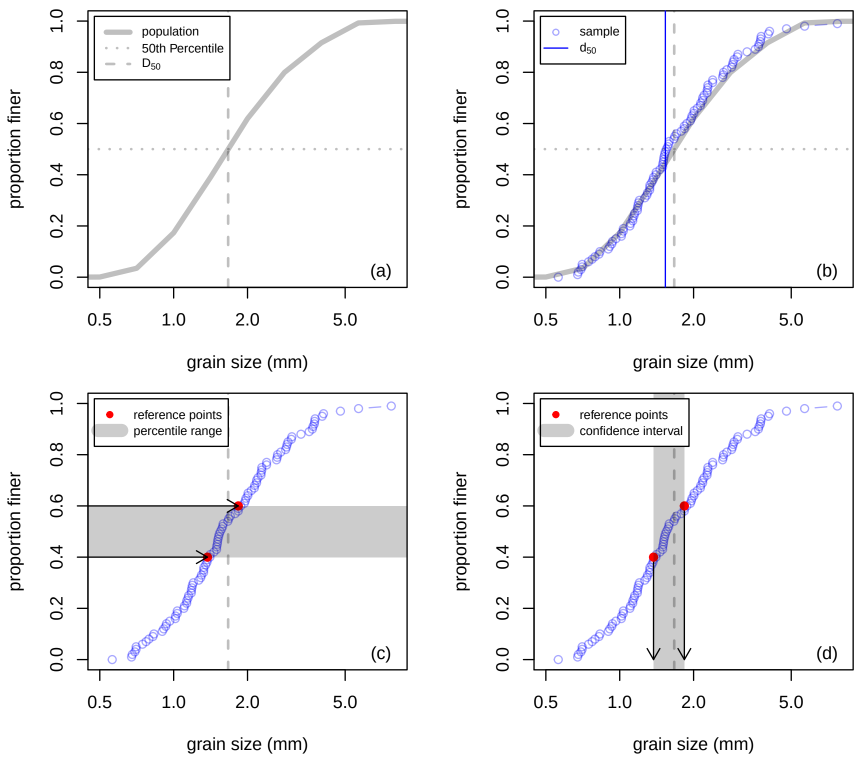

Fig. 1. Example of binomial equation used to generate confidence bounds.