Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2019-4-AC2, 2019

Earth Surface Dynamics Discussions

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 comments provided by Reviewer 2 are presented below, along with our responses. Several of key points raised by Reviewer 1 (and addressed in detail in our responses to R1's comments, which are already posted) come up again. This highlights the importance of those points, and we are very grateful for the feedback.

1.1 Comment 1

> The submitted paper focuses on estimating uncertainties in measured grain size distributions using statistical analysis of grain size data from experiments, field measurements and synthetic data. I think that the authors make an important main point, which is that uncertainties in grain size distributions should be reported especially when used to assess grain size changes over time or in space. Although I am supportive of the overall goals, topics, and messages of this manuscript, I think that there are many details missing from the methods. This makes it difficult to evaluate how this calculation is actually applied, the assumptions involved, and finally how it compares to previously published studies on uncertainties in grain sizes. I suggest adding these details such that your paper can be understood by a broader audience.

1.2 reply by authors

We are quite gratified to see that Reviewer 2 is supportive of including error analysis into the treatment of grain size data. That is the primary motivation for this paper, and for the development of the `bicalc` package. The general criticisms raised are very similar to the comments we had from Reviewer 1, and we are making major changes to help address them. This includes a comprehensive re-write of the statistical basis section to better describe how our approach actually works, expansion of the results section to clarify the links to previous work, the development of reference appendices providing supporting information, the creation of a spreadsheet that implements the normal approximation to our technique (as described by Fripp and Diplas, 1993) to estimating confidence limits, and the development of a 'how-to' guide for using the R package. We appreciate all of the suggestions that are made in this review, and we are confident that the revised version will reach a broader audience.

1.3 Comment 2

> I would really like to see a more detailed review of what previous studies have done to quantify uncertainties in the D50 and other percentiles of the grain size distributions. Do approaches without an assumed grain size distribution exist? If so, what is wrong with these approaches that motivates this current study? I'm a bit confused because in the introduction you state that there is no easy way to estimate the required sample size. In the abstract you also write that you propose a simple approach to estimate sample size, but this also relies on assuming a log-normal distribution as in previous studies highlighted on p 2 lines 8-9. What is the difference between your approach that assumes a log normal distribution to estimate sample size and other log normal approaches? It is not entirely clear to me in reading the introduction what is new in this study compared to previous approaches. A more in depth review of previous approaches and a statement of how this new approach is different would really help.

1.4 reply by authors

As described in our reply to Reviewer 1, we are extending our discussion of previous approaches by re-writing the paper to leverage the previous work by Diplas and colleagues as the starting point, and we describe in more detail how we replicated the bootstrap approach of Rice and Church to estimate the uncertainty of samples with various sizes drawn from our population of 3411 b-axis measurements. Basically, we believe that our approach is entirely consistent with that proposed by Fripp and Diplas (1993), and replicates the empirical results presented by Rice and Church (1996), but with a simple set of calculations rather than the laborious collection of 1000's of b-axis measurements. The main issue that we try to address in this paper is not that previous methods are flawed, but rather that we as a community have failed to use those

approaches to quantify sampling uncertainty (despite the precedents in the literature). As a result, there are published results that are clearly not statistically defensible, and it is our impression that many people continue to collect relatively small samples with limited appreciation of what that means in terms of uncertainty.

In our revisions, we will also emphasize that we think are the main contributions of this paper, which are:

- to describe clearly how surface sampling can be described as a binomial experiment, analogous to a traditional coin toss experiment (a point that the lead author of the paper failed to fully appreciate until we started working on this paper);
- to present a simple set of tools based on binomial theory with which anybody can easily calculate the confidence interval about any sample estimate of a grain size percentile that will contain the population percentile size for a given confidence level, α (this now includes a better description of how the method actually works, reference tables, a guide to using the R package, and a spreadsheet that implements a version of Fripp and Diplas's (1993) normal approximation to the binomial solution);
- to demonstrate the importance of considering uncertainty when comparing samples of the bed surface, or when making calculations based on those samples; and finally
- to make some assumptions about distribution shape so that we can provide some general guidance on the sample size required to reach a desired level of sampling precision.

This last point involves making assumptions about the underlying distribution (i.e. we assume a log normal grain size curve), but that is simply to generate synthetic data

with which to model the effect of sample size and the spread of the distribution on the precision of a percentile estimate. We will make it clear that any distribution form could have been used, but that we chose a log-normal distribution because (1) is the simplest to describe (i.e. it can be described by a mean and standard deviation), (2) it has been used previously by others, and (3) many gravel beds are approximately log-normal. We are attempting to more clearly emphasize our central message in our revisions (i.e. that we need to calculate confidence intervals, and that we can do so easily with binomial theory), and de-emphasize the point about sample size (since the actual precision of a sample in terms of grain size can only be determined once the sample has been collected).

1.5 Comment 4

The reviewer made several comments about our calculations that we would like to address:

> In section 2.1, how is equation (1) used? Please provide a step wise explanation nohow someone would perform these calculations and what information is needed. Right now it is somewhat difficult to understand how equation (2) is actually solved. Although I appreciate the inclusion of the R code that is part of this paper, a simple explanation of your detailed methodology is really needed in the main text to properly evaluate your methods. What are 'successes', please define. I am also somewhat confused about the definition of p, earlier you state it is the percentile of a distribution but on P 4 L6 is it called a probability.

We have completely re-written the statistical basis, including an overview section that walks the user through the idea of a binomial experiment, the probabilities of a particular outcome (and the relation of those probabilities to the grain size percentiles for the

population being sampled), and the relation between confidence interval of the sample percentiles and the confidence interval bounding an estimate of a grain size percentile.

> In section 2.2, please also provide more details on this approach, one brief sentence on interpolation really does not make this calculation clear.

We are re-writing the entire section to improve clarity.

> Section 3 and Figure 4 How many times did you create a sample with 100 grains to make these distributions in Figure 4? It seems like the results could really vary with the number of 100 grain samples? Also, some explanation of the boxplots is needed to evaluate the results. What are the horizontal lines at the top and bottom ends of the distributions? This information is needed to validate that the two predictions actually provide similar results. Can you provide the actual numeric values of the 99%confidence interval bounds for the two methods in the figures to enable quantitative comparisons?

We are expanding and re-writing the entire section to improve it. This is the section where we repeat the kind of boot strap error estimates presented by Rice and Church (1996). We ended up taking 5000 samples from the population to ensure that the distributions of estimates stabilized.

2 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.