Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2019-4-RC2, 2019

Earth Surface Dynamics Discussions

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

Anonymous Referee #2

Received and published: 21 March 2019

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.

Main comments Literature review: 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.

Calculations: I do not completely understand how some of the calculations are implemented and more details are needed in the main text. I have broken these comments into the main sections of the paper:

In section 2.1, how is equation (1) used? Please provide a stepwise explanation on how 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.

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

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?

Specific comment denoted by page (P) and line numbers (L)

P 1 L 21-22 For facies mapping, my understanding of the Buffington approach is not that it is meant to be purely qualitative as implied here. They have visual classification of patches that are then verified by numerous pebble counts on the patches. So their approach likely provides a more accurate representation of the grain size distribution because they use many pebble counts in a single reach.

P3 L5 Missing word(s) here.

P4 L 12-16 Please state if this text is for a specific sample (e.g. the data shown in Figure 1), right now it seems to be written as if it applies to all grain size measurements but I don't think that is actually the case?

P4 L 15-16 Please explain what you mean by 19 times out of 20. I'm not clear why these exact numbers are chosen instead of a percent of trials. It is also not clear how this percent of trails was calculated or how the range of 159-180 was determined.

P4 L 21-23 Stating that the area under the tails differs is pretty vague. Do you mean tails of the distribution? How are the tails of the distribution defined? Please state why these different areas are problematic. Similarly, upper and lower limits of what exactly? What do you mean by a one-sided interval and how does this relate to your calculations? I can guess what you mean but the lack of language specificity here makes your text somewhat difficult to follow.

Figure 3 More details are needed as to how the grain size data were collected, through

a random sample or grid count? Were the samples in different locations on streamtable and using the same or different operators? It is a little difficult to see the confidence bounds in this figure to assess overlap of various distributions, not sure though how you can easily address this problem.

P 7 L 8 typo here

Figure 5. I appreciate this reanalysis but I don't think that you can say that the distributions are statistically similar or different without a similar confidence bound on the bulk sample data. Previous studies have demonstrated that bulk samples also have considerable uncertainty depending on the size of the actual bulk sample and the portion of the sample that is occupied by the largest grain sizes. So the bulk sample is also not free from uncertainties and this needs to be acknowledged.

P 8 L 3-5 The statement that fine sediment would be deposited preferentially in the pool rather than in the run/riffle during the waning limb of the preceding hydrograph needs some references to support it.

P 12 L 6-7 Please explain why you are assuming the standard deviation of the distribution is related to logD84-logD50.

P 12 L 10-12 I do not entirely why you are simulating log-normal samples with this given range of D50 values and SDlog values? How were these distributions simulated by defining D50 and SDlog beforehand? Figure 10 does not seem to be referenced or explained anywhere in the text.

P 13 L. 14-22 More details are needed as to how you estimated that this grain size is entrained at a certain shear stress and discharge. Did you use Shields equation? What critical Shields stress did you assume? How did you then translate this shear stress into a discharge beyond using a stage-discharge relation; did you have a measured channel bed slope and are you assuming stage is equivalent to the average flow depth in a reach? What is the basis of the assumption that D50 becomes fully mobile at twice

the shear stress needed to initiate D50 movement? Some rational and supporting references are needed to support this argument. I am also a little confused about this uncertainty in grain size because all of these sizes (46, 55, 64 mm) are essentially in the same half-phi bin. I may be mistaken but if you have binned your data into half phi intervals for this analysis, wouldn't you expect a similar, although likely smaller, level of uncertainty in the D50 anyway? This uncertainty would occur because you are determining the measured streambed D50 value (55 mm) by interpolation between the two percentiles straddling the 50th percentile value, and these two bounding percentiles correspond to grain size bins 45 and 64 mm. But you do not actually have any grain size resolution finer than half phi bin size. So when you calculate a median grain size of 55 mm, you are interpolating this grain size to a finer resolution than you actually have data. Doesn't this already seem to imply that your uncertainty in D50 might be somewhere within a half phi bin size when you only have binned data, depending of course on how the actual grain sizes are distributed within that half phi bin?

P 15 L 12-13 Although I certainly agree that having more than 100 sampled particles would be better for uncertainties in most studies, these time estimates assume a team of people performing pebble counts. Having conducted a very large number of pebble counts on my own, these can take much longer than 20 minutes. The time also really depends as to whether you are binning grain sizes or measuring individual b axes. Finally, setting up and finding grains on a grid also adds to the pebble count time, so I would argue that this 20 minute estimate is a minimum.

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