Maximising the accuracy of image-based surface sediment sampling techniques

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: GRAHAM, D.J. ... et al, 2010. Maximising the accuracy of image-based surface sediment sampling techniques. Water Resources Research, 46, W02508.

Additional Information:

- This article was accepted for publication in the journal, Water Resources Research [© 2010 American Geophysical Union]. Further reproduction or electronic distribution is not permitted. The definitive version is available from: http://www.agu.org/journals/wr/

Metadata Record: https://dspace.lboro.ac.uk/2134/5891

Version: Accepted for publication

Publisher: © American Geophysical Union (AGU)

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Maximizing the accuracy of image-based surface sediment sampling techniques

David J. Graham¹, Anne-Julia Rollet², Hervé Piégay³, Stephen P. Rice¹

¹Department of Geography, Loughborough University, Loughborough, LE11 3TU UK.

²Aix-Marseille université, CEREGE – CNRS UMR 6635 France.

³University of Lyon, CNRS, UMR 5600, Site ENS-lsh, Lyon, France

Corresponding author: David Graham (D.J.Graham@lboro.ac.uk)
Recent years have seen increased interest in automated methods, utilizing photographs collected with a hand-held digital camera, for determining the grain-size distribution of coarse river sediments. Such methods are as precise as traditional field methods, and have considerable time and cost advantages. Nevertheless, several unresolved issues pertaining to their deployment remain to be addressed. Using datasets collected from seven gravel-bed rivers, this paper examines four key issues: (i) the minimum area required to obtain a representative sample; (ii) the effect of lower-end truncation on grain-size percentiles; (iii) the effect of river-bed structure such as imbrication and hiding; and (iv) the potential benefits of using individual particle measurements rather than the number (or mass) of particles per size class to calculate percentiles. It is demonstrated that sampling areas of between 50 and 200-times that of the largest grain are adequate to achieve percentile errors of <10% (in mm). The appropriateness of lower-end truncation depends on the study aims and sediment properties. It has a limited effect on higher percentiles, except where sand is a major constituent. Understanding the influence of bed structure on image-derived size information is complicated by the absence of error-free benchmarks against which accuracy may be evaluated, but it is likely that other errors are more important. The use of individual particle measurements to calculate percentiles in preference to classified data is shown to have a small, but appreciable, effect on precision. These results will assist practitioners in making appropriate operational decisions to maximize data quality using image-based grain-size data capture.
1. Introduction

The grain-size distributions of coarse sediments exposed on river beds, and elsewhere, are important in a variety of geomorphic, ecological and engineering contexts. Characterization of coarse-grained size distributions using traditional field methods like pebble counts, paint-and-pick and adhesive sampling [Wolman, 1954; Lane and Carlson, 1953; Diplas and Fripp, 1992] is time-consuming and costly, and destroys the surface being sampled. It is therefore often impracticable, using these techniques, to resolve important spatial variability and temporal changes in surface grain size. However, recent advances in image-analysis have seen the rapid development and adoption of alternative techniques that use the analysis of surface images to overcome the limitations of traditional methods [Graham et al., 2005b; Marcus and Fonstad, 2008]. These methods, which build on earlier photographic sampling procedures [e.g. Adams, 1979; Ibbeken and Schleyer, 1986], are non-invasive and can provide grain-size information at a quality comparable to that from traditional techniques. Their key advantage is the rapidity with which data may be collected and analyzed. This rapidity facilitates data collection at higher spatial and temporal resolutions, and across larger areas than are possible using traditional methods, all of which are hugely beneficial in understanding geomorphic processes.

There are two major approaches to image-based granulometry at present. The first provides ensemble grain-size parameters (e.g. the D_{50} for an area) based on evaluating the spectral characteristics or semivariance structure of imagery, principally from airborne platforms [Carbonneau et al., 2004; Chandler et al., 2004; Verdu et al., 2005]. This is proving to be a valuable method for mapping grain-size variations at a reach scale. [Carbonneau, 2005; Hedger et al., 2006]. Rubin [2004] and Buscombe and Masselink [2009] have developed similar systems for examining sand-bed grain-size parameters at scales of a few centimeters.
This technology has been extended for application on gravel-sized particles [Adams et al., 2007; Ruggiero et al., 2007].

The second approach applies semi- and fully-automated object-detection algorithms to images collected by ground-based photography to obtain grain-size measurements of the individual grains in an image and thereby produce a complete grain-size distribution. Butler et al. [2001], Reid et al. [2001] and Sime and Ferguson [2003] demonstrated that this approach could provide detailed grain-size information at bar and reach scales. Subsequently, Rollet [2007] applied a semi-automated procedure to assess the impact of a sediment deficit associated with dam construction on downstream grain-size characteristics along a 40 km river reach (incorporating 109 bars). Graham et al. [2005a, 2005b] developed a fully-automated and transferable algorithm which optimizes results for a variety of sediment types.

This paper focuses on these discrete-grain techniques and uses field data to examine several important issues related to their deployment that have not, to date, been fully assessed. These issues relate to procedures for minimizing errors associated with each of the three stages of photographic data capture: collection of images; extraction of grain sizes; and analysis of the results (Fig. 1). Whilst some work has examined these issues in the context of the manual digitization of photographs [e.g. Kellerhals and Bray, 1971; Ibbeken and Schleyer, 1986; Bunte and Abt, 2001a], several issues are unresolved in the context of automated methods. Although these issues are discussed here in the context of developing better use of automated image-based methods, several of them are also of relevance for the accurate estimation of grain-size distributions using traditional methods. In particular this paper examines the effect of several procedures on the grain-size distributions: the choice of sampling area; the effect of applying a truncation to small grains; and the use of discrete particle measurements versus measurements that bin particles into size classes. We also examine how hiding and
imbrication of surface particles affect grain-size distributions. These are important considerations because of the anticipated growth in the adoption of image-based techniques by scientists and practitioners. We utilize data collected on seven gravel-bed rivers of contrasting bed-material lithology and character: Ettrick Water (Scotland), Afon Ystwyth (Wales), River Lune (England), the Ain River (France), and the Fraser, Chilliwack and Peace Rivers (Canada).

2. Sampling area and sample size

2.1 Rationale

All grain-size measurement approaches are limited, to a greater or lesser extent, in the range of grain sizes that can be measured. Such limitations may result in truncation of the upper or lower part of the grain-size distribution, or introduce a size-dependent bias in the measured distribution as a result of inadequate characterization of fine or coarse grains. The size of particles that can be measured by image-based techniques is limited by the area covered in the image (at the coarse end) and the resolution of the image (at the fine end). Truncation of the coarse end of the distribution may be avoided by photographing a larger area (either by increasing the area of individual images, or montaging multiple images). Issues surrounding truncation at the fine end of the distribution are discussed in the next section.

In order to avoid inadvertent truncation of the coarse end of the grain-size distribution, all sampling methods must include sufficient grains to adequately represent the population being studied. One approach to achieving this is to photograph the entire area of interest using multiple images. This would allow the continuous variation in grain size across the surface to be established, and permit investigation of the ways in which surface texture varies at scales.
from the individual grain to an entire bar. The ability to undertake such analyses is a
significant benefit of image-based sampling over traditional methods. More often, however, a
more focused sampling strategy will be employed to characterize the properties of a particular
facies or assess spatial trends in grain-size attributes. In such cases it is desirable to use the
smallest sample sizes that are adequate to represent the populations of interest. This is to
minimize sampling effort and to reduce the probability of inadvertently sampling across areas
within which there is a textural gradient.

There is a considerable body of literature on the sample sizes required for representative grid
surface sampling [e.g. Fripp and Diplas, 1993; Rice and Church, 1996; Petrie and Diplas,
2000] and volumetric sampling of the subsurface [De Vries, 1970; Church et al., 1987; Fripp
and Diplas, 1993; Rice and Haschenburger, 2004; Haschenburger et al., 2007]. Much less
attention has been paid to criteria for representative areal sampling.

The number of grains required in an areal sample is considerably larger than in a grid sample
of similar precision [Petrie and Diplas, 2000]. However, because areal samples are based on
predefined areas (cf. grid samples, which consist of a predefined number of grains), guidance
on appropriate sample sizes is most usefully given in terms of the area to be sampled rather
than the number of grains to be included. This sampling area has commonly been defined as a
function of the area of the largest particle of interest. Diplas and Fripp [1992] proposed that a
sampling area at least 100 times the area of the largest particle should be used to obtain a
precision equivalent to a 100-grain grid sample. Fripp and Diplas [1993] argued that a
sampling area 400 times the that of the largest particle is required to meet the ‘low precision’
criterion of De Vries [1970], giving a relative error of 10%. A more sophisticated approach
was introduced by Petrie and Diplas [2000], based on the multinomial distribution and the
application of a two-stage sampling procedure. This is rather complex to apply in practice,
and requires the collection of an initial grid sample, which negates many of the benefits of photographic areal sampling. Nevertheless, the results of their analysis indicated that even the 100-times criterion of Diplas and Fripp [1992] is overly conservative, and substantially smaller sampling areas are acceptable.

The relative ease with which photographic samples can be collected facilitates the collection of many samples in a short time. Where rapid fully-automated measurement procedures are used, taking only a minute or two per image, it is feasible to combine size information from multiple images to achieve acceptable truncation at the fine end whilst sampling an area large enough to be representative of the population. This may be appropriate in many cases, but where characterizing small-scale spatial variability in grain-size is the objective of the study it is essential that the area sampled has experienced similar local hydraulic conditions, and must therefore be as small as possible (whilst still being large enough to representatively sample the full grain-size distribution) [Rollet et al., 2002; Rollet, 2007]. Furthermore, where manual or semi-automatic measurement of the images is used, increasing the number of images adds substantially to the processing time and may negate many of the benefits of photographic data collection (although expensive field time will still be reduced). Other types of areal sampling procedure (e.g. paint-and-pick; clay; wax) may sometimes be desirable, but these are particularly time and labor intensive in both the field and laboratory.

For their fully-automated image-based analysis method, Graham et al. [2005a] determined that the smallest grain of interest (i.e. the lower-end truncation) should have a diameter larger than 23 pixels in the image, although this limit can probably be relaxed slightly for processes where there is operator intervention. If the criterion is applied in a study with a truncation at 8 mm and using an 8 MP camera, the largest area that could be included in a single image is 0.97 m². Assuming that the area of a grain is equal to the square of its $b$-axis and applying the
100-times criterion of Diplas and Fripp [1992], the $D_{\text{max}}$ particle size must be smaller than 99
mm if a representative sample is to be obtained. If the 400-times criterion of Fripp and Diplas
[1993] is applied, the $D_{\text{max}}$ size must be less than 49 mm. Seven photographs are required
when sampling a surface containing large cobbles (up to 256 mm) to meet the 400-times
criterion.

These issues mean that the development of clear guidelines on the sampling area required to
meet the precision requirements of a particular study is highly desirable, preferably in a form
readily applied in the field.

2.2 Empirical assessment of minimum area for representative sampling

The ease of photographic sampling facilitates the empirical evaluation of the various size
criteria for areal samples. The effect of sample area on the precision of key percentiles was
assessed using eight patches, each with an area of 12.75 $m^2$, on two bars at the Afon Ystwyth,
Wales. Each patch (representing the population grain size against which samples with
different areas are compared) was selected to have no visually discernable systematic
variation in grain-size distribution across it. Subsequent analysis revealed that four of the
patches did have statistically significant variations in either the $D_{50}$ or $D_{90}$ across them, but in
all cases the magnitude of these variations was extremely small. The $D_{50}$ of the patches
ranged from 4.4 Psi (21 mm) to 5.1 Psi (35 mm) and they were moderately sorted (inclusive
graphic standard deviation between 0.7 and 1.0 Psi). Every patch was divided into 50 sample
areas, each with an area of 0.255 $m^2$. The sample areas were photographed vertically with a
hand-held digital camera and the images analyzed using the procedure of Graham et al.
[2005a, 2005b], the grain-size distributions being truncated at 8 mm.
For each patch, a random sample was chosen and the difference in its D50 and D90 relative to
the whole patch (population) was calculated. This was repeated for increasingly larger sample
areas, aggregating percentiles for two randomly selected samples, then three, then four etc.
Each repeat produces an estimate of the error in percentile values (relative to the population)
associated with a particular sample area. Because the results of this procedure vary slightly
each time it is undertaken, it was repeated 10 times for each patch to overcome random
effects. Areal samples are often collected for the purpose of comparison with data derived
from grid sampling, so percentile errors were calculated in both area-by-number and grid-by-
number form (the method for calculating grid-by-number percentiles where the area and b-
axis of each grain are known is described by Graham et al. [2005b]). For each permutation,
the ratio between the sampled area and the area of the largest grain in the population was also
calculated. Figures 2 (area-by-number percentiles) and 3 (grid-by-number percentiles) present
covariant plots of the magnitude of the percentile errors (expressed as percentages in mm)
against these ratios and provide an indication of the sample area (relative to the area of the
largest grain) required for the errors to stabilize to low values.

Figure 2.

Figure 3.

As expected, the results indicate a strong relation between sampled area and the precision of
the D50 and D90 (summarized in Table 1). Where data are required in area-by-number form
(Fig. 2) they suggest that the sample area should be at least 100-times greater than that of the
largest particle to achieve D50 errors of less than 5% in mm (0.07 Psi). To achieve equivalent
precision in the D90, the sample area should be greater than 200-times that of the largest
particle. To achieve errors of less than 10% in mm (0.14 Psi) sample areas of around 100-
times that of the largest grain are acceptable. Larger sampling areas are required to achieve
similar errors in grid-by-number percentiles because of the greater weight placed on larger
grains in such distributions (Fig. 3).

Caution should be exercised before extrapolating these results to other locations with different
textural and sorting characteristics. Nevertheless, the results do suggest that the small sample
areas proposed by Petrie and Diplas [2000] may be insufficient. For higher precision, larger
percentiles and grid-by-number data, areas as large as 400-times that of the largest grain
(which for practical purposes may be approximated by the square of the $b$-axis length) may be
required. Practitioners will need to balance the competing demands of precision and field time
within the context of their project requirements.

3. The importance of truncation

3.1 Rationale

In practice, it is relatively easy to avoid problems associated with inadequate characterization
of the coarse fraction because the presence of large grains is easily recognized and steps can
be taken to adjust the sampling strategy to include them. Bias or truncation at the fine end of
the distribution is likely to be a more significant problem because its effects may not be so
easily recognized. In the context of image-based techniques, the minimum resolvable grain
size is controlled by the resolution of the image. Although continual increases in digital
camera resolution will result in improvements in the ability to recognize small grains, there
will always be a lower limit to the size of grains that can be identified reliably and measured
accurately using image-based methods. For the fully-automated technique of Graham et al.
[2005b], measurement error increased substantially for grains with an apparent $b$-axis smaller
than 23 pixels in the image. Such issues are not limited to new sampling techniques. It is well established, for example, that grid-by-number samples [Wolman, 1954] commonly result in under representation of fine grains as a result of operator bias [Diplas and Fripp, 1992; Fripp and Diplas, 1993; Marcus et al., 1995]. As a result, it is common practice to truncate grid-by-number samples at a size above which the operator is confident that bias is no longer present (often 8 mm [Rice, 1995]).

Although size-dependent bias may be acceptable in some circumstances (e.g. where having biased data is considered better than no data), in most cases a truncation is required to ensure that only representatively sampled grains are included in the derived size distribution [Church et al., 1987]. Where samples are being compared, it is essential that a common range of representatively sampled grains is used, which will usually require a truncation to be applied to one or both data sets. However, the application of truncations can be problematic because truncation modifies, to a greater or lesser extent, the entire cumulative grain-size distribution curve used to determine percentiles [Bunte and Abt, 2001b]. Indeed, Fripp and Diplas [1993, p. 478] argued that “the hazards of using truncated samples cannot be overstated”. Some applications, such as calculation of the bedload transport fluxes [Reid et al., 1996], are highly sensitive to small variations in percentiles. What then is the magnitude of the error introduced by truncation (either intentional or inadvertent) of the distribution?

3.2 Empirical assessment of the impact of truncation

The effect of lower-end truncation was tested using a dataset consisting of 74 grid-by-number samples, each of around 300 grains, collected from three Canadian rivers (Fraser, Chilliwack, Peace). Grids with intervals greater than twice the largest grain diameter were laid out with tapes and particles to be sampled were identified as those lying below relevant tape
graduations. For each sample, the number of grains smaller than 1 Psi (2 mm) was recorded and larger particles were assigned to half Psi classes using templates. The $D_{50}$ of the samples ranges from 4.2 Psi (18.4 mm) to 7.1 Psi (137 mm) and sorting from moderately well sorted to very poorly sorted (inclusive graphic standard deviation ranges from 0.6 Psi to 2.1 Psi).

Grain-size percentiles were calculated for the raw data (Fig. 4), and for the data truncated at 1 Psi (2 mm), 3 Psi (8 mm) and 5 Psi (32 mm). Percentiles were calculated in Psi units and spline interpolation was used to smoothly interpolate between class boundaries. The differences between the percentiles from the raw and truncated data were then calculated, giving the error at each percentile for each sample (upper panels in Fig. 5). The mean errors at every 5th percentile (lower panels in Fig. 5) provide a clear indication of the magnitude of the likely errors associated with any particular truncation and percentile. Several of the samples had significant proportions of sand within them (commonly inducing a substantial bimodality), resulting in substantial errors in lower percentiles when truncated. To examine this effect, those samples with > 5% sand ($n = 8$) were separated from the remainder of the samples (indicated by dashed lines in Fig. 5), and the mean errors in the percentiles examined separately for each group.

**Figure 4.**

**Figure 5.**

For samples without significant quantities of sand, for an 8 mm (3 Psi) truncation, the mean error in the $D_{50}$ is -0.06 Psi (-4.2% in mm), and for a 32 mm (5 Psi) truncation the error is -0.2 Psi (-14.9% in mm). For those with >5% sand, the mean error in the $D_{50}$ with a 32 mm (5 Psi) truncation is -0.3 Psi (-23.1% in mm). As would be expected, the errors for lower percentiles are substantial for sandy samples (Fig. 5). Whether this is acceptable depends on the
application, but for many purposes (e.g. as a proxy for roughness or to determine median
grain size) truncated data are likely to provide acceptable estimates of the true grain-size
percentiles. Indeed, where minor quantities of sand have been draped over the surface during
waning flows, it may be advantageous to ignore this component of the grain-size distribution
for some applications.

4. The effect and significance of river-bed structure

4.1 Rationale

In most surface sampling techniques, grains are removed from the substrate and their true
dimensions are therefore revealed prior to measurement. In contrast, photographic techniques
operate on grains in situ, and only the exposed parts of the grain can be measured. Given that
it is those parts of grains that are visible in photographs that actually interact with the flow,
this may not be considered a significant problem for some applications. Nevertheless, this
property of photographic methods has the potential to underestimate the true size of
individual grains as a result of three factors related to the structure of the river bed: (i) partial
burying of grains (the ‘iceberg’ effect); (ii) overlapping of grains as a result of imbrication;
and (iii) foreshortening, where the size of a grain appears smaller than it really is when
viewed from an angle (Fig. 6). These potential sources of bias have been recognized since
photographic sampling was first used [e.g. Kellerhals and Bray, 1971; Ibbeken and Schleyer,
1986; McEwan et al., 2000], but the reported magnitude of the resulting errors have varied
substantially (Table 2). Furthermore, assessing of the effects of sediment structure is
challenging because of errors associated with the methods used to collect the control data.
4.2 Empirical assessment of the impact of river-bed structure

We quantified the magnitude of these structural effects at the Ain River, France, using 10 patches, each of 0.6 m², at which we observed moderate imbrication. Each patch was photographed vertically with a digital camera, spray painted, and all painted grains larger than 8 mm were returned to the laboratory and graded using square-hole sieves, the number of grains in each size class being recorded. The photographs were digitized manually on screen by a single operator using Adobe® Illustrator® (Fig. 7) and the apparent grains then measured automatically using Scion Image. The photographs were also analyzed using the semi-automatic process of Rollet et al. [2002]: (i) grayscale images were converted into binary form by the application of a threshold, the value of which was set at a level that the operator judged to be optimal; (ii) the binary image was skeletonized by the application of a watershed segmentation algorithm [Digabel and Lantuéjoul, 1977; Soille, 2002]; and (iii) the resulting objects were labeled and measured. Because the size measured by square-hole sieves is influenced by grain flatness (the ratio of the c-axis to the b-axis), a correction must be applied to image-based measurements to permit direct comparisons with the data derived by sieving [Church et al., 1987]. The appropriate correction is shape dependent, but Graham et al. [2005b] demonstrated that the correction factor is rather insensitive to the range of shapes commonly associated with fluvial sediments and that a factor of 0.8 (equivalent to a c-axis to b-axis ratio of 0.71) is generally appropriate for conversion from image-based to sieve-based measurements. After the application of the sieve correction, both the digitized and sieved data were truncated at 8 mm. It was judged that this was the effective limit of both reliable digitization of the image and collection of painted grains in the field. The errors associated with the image-based measurements were assessed with reference to seven commonly-used
Mean error, or procedure bias, $b$, is defined as $b = \frac{1}{n} \sum (\Psi_{xi} - \Psi_{xp})$ where $n$ is the sample size (number of patches multiplied by number of percentiles used), $\Psi_{xi}$ is a percentile derived from the image, and $\Psi_{xp}$ is the equivalent percentile derived from the paint-and-pick data. The mean square error, $E_{ms}$, is defined as $E_{ms} = \frac{1}{n} \sum (\Psi_{xi} - \Psi_{xp})^2$ and the irreducible random error, $e$, an indication of scatter, as $e = \sqrt{E_{ms} - b^2}$.

**Figure 7.**

Consistent with earlier work [e.g. Butler et al., 2001; Graham et al., 2005b], the number of grains identified in the images is substantially fewer in all size classes than obtained by paint-and-pick sampling (Fig. 8). These studies found that, despite undercounting, the precision of the grain-size percentiles was excellent because the depletion was approximately uniformly distributed.

**Figure 8.**

Contrary to the expectation that the sediment structure will lead to an underestimation of grain sizes by image-based approaches, percentiles of the manually-digitized grains tend to overestimate those from paint-and-pick sampling when presented in their original area-by-number form. Typical grain-size distributions obtained by paint-and-pick sampling and manual digitizing are illustrated in Figure 9 for a single sample. The relation between the percentiles derived by the two methods for all samples is illustrated in Figure 10a. The mean error in the seven reference percentiles is 0.12 Psi (Table 3), but errors are not uniformly distributed. Higher percentiles tend to have larger and positive errors (a mean of 0.28 psi for the $\Psi_{90}$), but the lowest percentiles have small negative errors (a mean of -0.011 psi for the
There is a moderate amount of scatter, especially for higher percentiles, with an irreducible random error of 0.21 psi. The semi-automatic sizing method produces better results, with a mean error of 0.006 psi, and an irreducible random error of 0.099 psi (Fig. 10b; Table 3).

Three key questions arise from these results. First, why does manual digitizing tend to result in an overestimation of larger percentiles when sediment structure (hiding, overlapping, foreshortening) might be expected to lead to the size of all grains being underestimated? This effect is likely to result from three factors that work in combination to modify the grain-size distributions from both the manually digitized and paint-and-pick data (the effects of which are illustrated schematically in Fig. 11):

(a) Paint-and-pick sampling is likely to overestimate the number of small particles. This is because it is difficult to unambiguously differentiate between small grains that are part of the surface layer and those that are really part of the subsurface. Whilst the use of spray paint to identify surface grains helps, drifting and penetration into interstices may result in some subsurface grains receiving paint [Church et al., 1987]. The over-collection of small grains skews the paint-and-pick grain-size distribution towards the fine end, leading to an underestimation of the true proportion of coarse grains and reducing the size of the larger percentiles (Fig. 11a).
(b) Manual digitizing is likely to underestimate the number of small particles. This is because they are harder to see than large grains because they are often located in interstices, and they are more difficult to digitize. The effect is to skew the digitized grain-size distribution towards the coarse end because of the relative overabundance of easily identified larger grains (Fig 11b).

(c) Structural effects may also influence the digitized grain-size distributions. These effects may not operate uniformly on different size fractions. Selective reduction in the size of coarse grains would result in a skew to the fine end and a reduction of larger percentiles (Fig. 11c). Selective reduction in the size of the small grains would increase the larger percentiles (Fig 11d). Careful inspection of the images suggests that small grains, many of which lie in the interstices of larger particles, are more likely to exhibit foreshortening by presenting their c-axis to the camera than larger grains, resulting in a reduction in their apparent size.

Figure 11.

A second question is, why does the semi-automated approach generate better results than manual digitizing, which might be expected to be the ‘gold-standard’ in image-based analysis? It has not been possible to provide a definitive solution to this problem. However, the quality of the results from the semi-automatic method is consistent with the findings from previous studies that have used fully- and partially-automatic methods [McEwan et al., 2000; Butler et al., 2001; Sime and Ferguson, 2003; Graham et al., 2005b]. These have identified substantial under-counting of the number of particles and some attribution of individual grains to the wrong size class (largely as a result of over- or under-segmentation leading to the splitting or merging of grains). Nevertheless, the level of undercounting was reported to be...
largely consistent between size classes and the levels of over- and under-estimation of the sizes of individual grains approximately compensated for one another, producing size distributions with small or moderate errors. The results of this study suggest that there is a bias in percentiles produced by the semi-automatic measurements relative to those from manual measurements, but this bias results in an improvement in the automated measurements relative to the paint-and-pick control sample. Whilst this result is fortuitous, it is entirely consistent with earlier work. The errors produced by the semi-automatic process used here are very similar to the area-by-number errors associated with the fully automated procedure of *Graham et al.* [2005b] (Table 3).

The third question is, why are the results from this experiment different to those from previous work (which invariably report negative biases)? The answer is probably a function of the way in which the data are presented and reported. Most studies have reported their errors in grid-by-number form (Table 2). This requires the data to be transformed from the area-by-number form in which it is collected, which may be accomplished using the procedure of *Kellerhals and Bray* [1971]. Once transformed in this way, the data from this study do exhibit the negative bias observed in previous studies (Fig. 12; Table 3).

**Figure 12.**

Given that the data are positively biased in area-by-number form, this negative bias in grid-by-number form cannot be accounted for simply as an artefact of sediment structure. The apparent change in bias is likely to result from the interaction of the nature of the grain-size distributions and the transformation. *Graham et al.* [2005c] have demonstrated that the conversion from area-by-number to grid-by-number data is extremely sensitive to variations in the number of large grains. In an area-by-number distribution grains of all sizes have an
equal weighting, whereas in a grid-by-number distribution the weighing is proportional to the
area of the grain, so a single 128 mm grain has the same influence on the distribution as
around 250 8-mm grains. Adding a single large grain to the distribution can therefore
substantially change the shape of the cumulative distribution curve, and thus the percentiles
derived from it. So, the grid-by-number distributions are strongly controlled by the (relatively
small) number of large grains, whilst the area-by-number distributions are dominated by the
large numbers of smaller grains. It is possible that the size of the larger manually digitized
grains is underestimated as a result of structural effects, leading to the slight overall negative
bias. The even greater negative bias of the semi-automatic method is likely to result from
over-segmentation of a few of the larger grains.

The purpose of undertaking these experiments was to understand the importance of hiding,
imbrication and foreshortening on derived grain-size distributions. In practice, it has been
found that, whilst structural effects may be an important source of error for photographic
sizing methods, it is difficult or impossible to fully disentangle them from the range of other
factors operating, at least at the Ain River. This difficulty leads us to believe that structural
effects are probably small relative to the other factors operating. Although restricted to a
single river with a relatively limited range of structural properties, this conclusion is
consistent with the results of Graham et al. [2005b] which showed little evidence of structural
effects at three rivers with contrasting sediment characteristics. Nevertheless, the significance
of structural effects would benefit from further work, perhaps using artificially-created bed
structures composed of sediment with a known size distribution and a variety of camera
orientations to assess the importance of foreshortening.

5. Interpolated versus grain-by-grain derived percentiles


5.1 Rationale

Grain measurements associated with traditional sampling methods are most commonly made with sieves or templates [e.g. Bunte and Abt, 2001b], although rulers, calipers, or pebble boxes are sometimes used [e.g. Ibbeken and Denzer, 1988; Marcus et al., 1995; Green, 2005]. Sieves and templates have the advantage that the operator does not have to visually identify the particle b-axis, removing a potential source of operator error. However, because these methods result in each grain being classified as belonging to a particular size class, the calculation of percentiles from sieve- or template-derived data requires interpolation between class boundaries. This is not the case using image-based methods (or measurements by ruler, caliper or pebble box) which provide information about the size of individual grains, and percentiles may be determined directly from the grain-by-grain data (with interpolation only required between the two grains lying on either side of the percentile of interest). Grain-size percentiles obtained using interpolation between size-class boundaries are therefore not exactly equivalent to those that use direct measurement of each grain.

Because of its computational simplicity, the most common method of calculating percentiles for classified data is to use linear interpolation between class boundaries. Where the cumulative curve is convex, linear interpolation results in an underestimation of the percentiles; where it is concave, linear interpolation results in an overestimation of percentiles. This effect is illustrated in Fig. 13a for a single sample from the River Lune, England. Size data for individual grains were obtained using the automated image-processing procedure of Graham et al. [2005a, 2005b] to produce a grain-by-grain cumulative curve. The same data were placed into 1 psi classes as the basis for the interpolated curve. Different results are obtained when percentiles are calculated in millimeter and psi units. In Figure 13, percentiles are calculated by linear interpolation in both millimeter and psi units and then
plotted in psi units. Cumulative curves calculated in millimeter units have the greatest
deviation from the true distribution (Fig. 13b).

Although the interpolated cumulative grain-size curves consist of linear segments, their
overall ‘smoothness’ depends on the number of grain-size classes used to construct them. The
larger the number of classes used, the closer the derived cumulative distribution curve will be
to the true (grain-by-grain) curve. This means that the values of the calculated percentiles are
dependent on the number of size classes used. In Fig. 13, large 1 psi classes are used (rather
than the more usual half or quarter psi classes) to highlight the differences between the grain-
by-grain and interpolated curves more clearly.

These problems may be overcome to some extent by fitting a smooth curve through the
classified data. This may be achieved graphically using a draftsman’s curve, or
computationally using a spine-fitting algorithm. By producing a smooth cumulative curve,
spline interpolation results in percentiles that are closer to the true percentiles (Figure 14).
The results are improved for calculations made in both millimeter and psi units, although psi
units still produce better results. An additional advantage of spline interpolation is that it is
less sensitive to variations in the number of size classes used to record the data.

Figure 13.

Figure 14.

5.2 Empirical assessment of the impact of interpolating percentiles

To examine the likely errors associated with interpolation, a dataset consisting of 37 samples
for which $b$-axis measurements of every single grain are available was used. Data were
collected photographically, and grain measurements made using the automated procedure of Graham et al. [2005b]. The photographic samples were collected at three rivers: River Lune, England (n = 7); Afon Ystwyth, Wales (n = 15); Ettrick Water, Scotland (n = 15). Each sample had an area of 1.2 m$^2$ and the number of grains larger than 3 Psi (8 mm) ranged from 335 to 1253 (mean = 734). Percentiles were calculated in area-by-number form (truncated at 3 Psi) using spline and linear interpolation with 1, 0.5, and 0.25 Psi size classes and compared with percentiles calculated on a grain-by-grain basis. The results of these analyses are presented in Fig. 15. As expected, errors are largest for percentiles calculated using linear interpolation and with larger size classes. Substantial improvements are associated with moving from 1 psi to 0.5 psi classes for both linear and spline interpolation. Improvements associated with the use of 0.25 psi classes are less impressive. Similarly, the benefits of spline interpolation are greatest for larger size classes. For 0.5 psi classes, spline interpolation only marginally outperforms linear interpolation.

Figure 15.

Although the errors associated with interpolation are small where 0.5 or 0.25 psi classes are used, it is generally more appropriate to use unclassified data to avoid the need for interpolation when determining grain-size percentiles. This is an advantage of photographic measurement techniques. Where data are obtained unavoidably in classified form, interpolation should always be undertaken in psi/phi units rather than millimeters. In cases where only classified data are available, it is recommended that spline interpolation be used in preference to linear interpolation because it is less affected by the choice of size class.

6. Summary
The use of image analysis to extract surface grain-size information from images collected with a hand-held digital camera is now well established [Butler et al., 2001; Sime and Ferguson, 2003; Rollet et al., 2002; Graham et al., 2005a, 2005b; Rollet, 2007], and can provide data with a precision equivalent to conventional field-based methods. Such methods have significant advantages in terms of time and cost savings, and they do not disturb the substrate, making them particularly appropriate for ecological applications and monitoring studies [Graham et al., 2005b]. Nevertheless, there are still unresolved issues associated with image collection, extraction of grain-size information, and analysis of the resulting data (Fig. 1). This paper represents the first attempt to address several issues related to the deployment of image-based methods (although many of the results are also of relevance to traditional grain-sizing methods). It is anticipated that this information will assist practitioners in making appropriate operational decisions to maximize the quality of the information obtained by image-based methods.

It has been found that sampling areas of less between 200- and 400-times the area of the largest grain are required for characterizing the full grain-size distribution of the Afon Ystwyth, producing errors in percentiles of less than 5% (in mm). Where information on the coarse part of the distribution is not required, areas of between 100- and 200-times that of the largest grain are adequate for characterizing the median grain size. The areas can be halved if errors of less than 10% (in mm) are acceptable. These results do not support the theoretical work of Petrie and Diplas [2000], who argued that sample areas smaller than 100-times that of the largest grain are acceptable. Although the results are for a single site with a relatively restricted range of sediment characteristics, it seems likely that larger sample areas (as recommended by Diplas and Fripp [1992] and Fripp and Diplas [1993]) are appropriate.
The investigations of the impact of truncation at the fine end of the distribution indicate that this has a relatively small impact on higher percentiles where the sample contains less than 5% sand. For these samples, the mean error in the \(D_{50}\) associated with truncation at 8 mm was -4.2% (in mm) (based on 66 samples). For samples with larger quantities of fines, errors were larger, but may still be acceptable depending on the application. These results indicate that truncation should be applied with care, with due regard to the purpose of the study. Where the fine part of the grain-size distribution is of particular interest (e.g. in assessment of habitat suitability for salmon spawning), truncation may be inappropriate. However, comparisons between data sets must always be made on a like-with-like basis, and truncation may be essential to facilitate the comparison of datasets collected by different methods. For image-based sampling methods, these results highlight the need to collect images at an appropriate resolution bearing in mind the minimum resolvable grain size required for the application [Church et al., 1987]. The relation between camera resolution, minimum resolvable grain size and area photographed are discussed by Graham et al. [2005a, 2005b].

Attempts to assess the impact of sediment structure on grain-size distributions obtained by photographic sampling have been complicated by a range of other factors in operation. The analysis has highlighted the fact that no technique provides a definitive assessment of the surface grain-size distribution. All techniques are subject to biases of one kind or another, and these make assessments of procedure performance challenging because there is no universally appropriate benchmark against which errors may be evaluated. In this regard, fully-automated image-based methods have a significant advantage over alternatives because they do not require any user intervention and are thus free of operator-introduced bias. Such biases are likely to be greatest where inexperienced operators are used, whereas image-based data collection may be undertaken successfully with limited training. The analysis suggests that
the magnitude of errors, associated with structural effects, are small relative to other factors, but further work is desirable to explore this issue further.

Although the errors introduced by interpolating percentiles between the boundaries of size classes are likely to be small, measurements of individual grains enable percentiles to be calculated directly and more accurately (provided that the measurements of the individual grains are accurate). A particular disadvantage of interpolated percentiles is that the results are partially dependent on how big the size classes are. Because they measure grains individually, image-based methods therefore offer advantages over techniques that allocate grains to size classes, without the time overhead associated with measurements using a ruler, calipers or pebble box. Where size data are unavoidably obtained in classified form, errors are minimized by calculating percentiles in psi/phi units (rather than millimeters) and by using spline rather than linear interpolation.

Acknowledgements

Thanks to Gordon Grant and four anonymous referees whose comments have resulted in improvements to the manuscript. Thanks also to the numerous assistants over many years who have helped us collect the data presented. The work was facilitated by a travel grant to DJG by Loughborough University.

References


Verdu, J. M., R. J. Batalla, and J. A. Martínez-Casasnovas (2005), High-resolution grain-size characterization of gravel bars using image analysis and geo-statistics. Geomorphology, 72, 73-93.

Figure captions

Figure 1. The three steps in deriving grain-size information from ground-based digital photographs, illustrating some key procedural questions associated with each. Questions highlighted in bold are addressed in this paper.

Figure 2. Covariant plot of percentile errors (area-by-number) versus the ratio between sampled area and the area of the largest grain in the population ($D_{\text{max}}$). The shaded region indicates an error of less than 5% in mm (equivalent to an error of less than 0.07 Psi). (a) Errors in the median grain size ($D_{50}$). (b) Errors in the $D_{90}$ grain size.

Figure 3. Covariant plot of percentile errors versus the ratio between sampled area and the area of the largest grain ($D_{\text{max}}$). The shaded region indicates an error of less than 5% in mm (equivalent to an error of less than 0.07 Psi). (a) Errors in the median grain size ($D_{50}$). (b) Errors in the $D_{90}$ grain size.

Figure 4. The grain-size distribution of 74 grid-by-number samples. Solid lines represent samples with $<5\%$ sand ($n = 66$), and dashed lines represent samples with $>5\%$ sand ($n = 8$). Data are for Fraser ($n = 19$), Chilliwack ($n = 11$) and Peace ($n = 44$) Rivers. Mean sample size $= 304$ grains (standard deviation $= 25$ grains). All samples recorded at 0.5 Psi intervals down to 1 Psi (2 mm), and the presence of smaller grains tallied.

Figure 5. The effect of truncation at (a) 1 Psi (2 mm), (b) 3 Psi (8 mm), and (c) 5 Psi (16 mm) for the 74 grid-by-number samples. Upper panels show the difference between true and truncated percentiles for each sample. Lower panels show the mean error in every 5th percentile, with error bars representing the standard error. In both panels, the solid line
represents samples with <5% sand (n = 66), the dashed line represents samples with >5% sand (n = 8). The shaded area indicates an error of less than 0.5 Psi.

Figure 6. Illustration of the three potential sources of grain size underestimation resulting from sediment structure and the use of photographic sampling methods. (a) Partial burying of grains. (b) Overlapping of grains as a result of imbrication. (c) Foreshortening of grains as a result of projection onto a horizontal plane.

Figure 7. Example 0.6 m² sediment patch and manually digitized grain boundaries.

Figure 8. The number of grains identified in 0.5 psi size classes using image-based methods compared to paint-and-pick sampling. (a) Manual digitizing with Adobe® Illustrator®. (b) Semi-automated grain identification using the procedure of Rollet et al. [2002].

Figure 9. Area-by-number grain-size distribution for one of the sample patches derived by paint-and-pick sampling and manual digitization. (a) Histogram illustrating the number of grains in 0.5 Psi sieve classes. Note that manual digitizing finds a larger proportion of the large grains relative to the smaller grains. (b) The resulting cumulative grain-size distribution curves.

Figure 10. Area-by-number percentiles defined by image analysis and paint-and-pick sampling. (a) Manual digitization vs paint-and-pick sampling. (b) Semi-automated analysis vs paint-and-pick sampling.

Figure 11. Schematic representation of the effect on cumulative grain-size curves of biases operating on paint-and-pick and manually digitized data. (a) Overestimation of the number of small grains by paint-and-pick sampling. (b) Underestimation of the number of small grains
by manual digitization. (c) Underestimation of the size of large grains by manual digitization, as a result of sediment structure. (d) Underestimation of the size of small grains by manual digitization as a result of sediment structure.

Figure 12. Grid-by-number percentiles defined by image analysis and paint-and-pick sampling. (a) Manual digitization vs paint-and-pick sampling. (b) Semi-automated analysis vs paint-and-pick sampling.

Figure 13. Example of the errors associated with using linear interpolation on classified data to calculate grain-size percentiles. Solid lines indicated the true grain-by-grain distribution derived using the photographic method of Graham et al. [2005b] and with a lower truncation of 3 Psi (8 mm). Dotted lines indicate the distribution based on placing the same data in 1 Psi classes and interpolating percentiles in mm units. Dashed lines indicate the distribution based on 1 Psi classes and interpolation percentiles in Psi units. (a) Cumulative grain-size distributions. (b) Deviations from the true grain-size distribution associated with linear interpolation.

Figure 14. Example of the errors associated with using spline interpolation on classified data to calculate grain-size percentiles. Solid lines indicated the true grain-by-grain distribution derived using the photographic method of Graham et al. [2005b] and with a lower truncation of 3 Psi (8 mm). Dotted lines indicate the distribution based on placing the same data in 1 Psi classes and interpolating percentiles in mm units. Dashed lines indicate the distribution based on 1 Psi classes and interpolation percentiles in Psi units. (a) Cumulative grain-size distributions. (b) Deviations from the true grain-size distribution associated with spline interpolation.
Figure 15. The errors associated with the use of linear and spline interpolation in psi units on classified data to calculate grain-size percentiles for 37 samples. The errors presented are absolute errors; actual errors may be positive or negative. Data are derived from photographs using the method of Graham et al. [2005b] with a lower truncation of 3 Psi (8 mm) and are presented in area-by-number form. Note that the vertical scale is the same in all plots.
Table 1. Sampling areas required for areal sampling relative to the population $D_{\text{max}}$ grain size.

<table>
<thead>
<tr>
<th>Precision</th>
<th>Area-by-number data</th>
<th>Grid-by-number data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D_{50}$</td>
<td>$D_{90}$</td>
</tr>
<tr>
<td>5% in mm (0.07 Psi)</td>
<td>100 times</td>
<td>200 times</td>
</tr>
<tr>
<td>10% in mm (0.14 in Psi)</td>
<td>50 times</td>
<td>100 times</td>
</tr>
</tbody>
</table>
Table 2. Reported bias in image-based grain-size measurements.

<table>
<thead>
<tr>
<th>Source</th>
<th>Method of grain measurement</th>
<th>Absolute bias (Psi)&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Percentage bias (mm)&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Basis of comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kellerhals and Bray [1971]</td>
<td>Unreported</td>
<td>0.26 Psi</td>
<td>20%</td>
<td>Grid-by-number</td>
</tr>
<tr>
<td>Adams [1979]</td>
<td>Manual measurement</td>
<td>0.1 Psi</td>
<td>7 %</td>
<td>Grid-by-number/area-by-number&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ibbeken and Schleyer [1986]</td>
<td>Photo-sieving: manual</td>
<td>0.24 Psi</td>
<td>18%</td>
<td>Area-by-weight</td>
</tr>
<tr>
<td>Bray [1972], cited in Church et al. [1987]</td>
<td>Unreported</td>
<td>0.29 Psi (for D50)</td>
<td>22% (for D50)</td>
<td>Grid-by-number</td>
</tr>
<tr>
<td>Butler et al. [2001]</td>
<td>Semi-automated digital analysis</td>
<td>0.13 – 0.33 Psi</td>
<td>9.4 – 26.7%</td>
<td>Selected grains</td>
</tr>
<tr>
<td>Sime and Ferguson [2003]</td>
<td>Semi-automated digital analysis</td>
<td>1.11 – 1.34 Psi&lt;sup&gt;3&lt;/sup&gt;</td>
<td>115 – 153%&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Grid-by-number</td>
</tr>
<tr>
<td>Graham et al. [2005b]</td>
<td>Fully automated digital analysis</td>
<td>0.007 – 0.03 Psi</td>
<td>0.5 – 2.1%</td>
<td>Area-by-number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.10 – 0.17 Psi</td>
<td>7.2 – 12.5%</td>
<td>Grid-by-number</td>
</tr>
</tbody>
</table>
Notes:

1 Bias is invariably negative (i.e. photographic methods underestimate the true size).


3 *Graham et al.* [2005b] found a coding error in the algorithm used by *Sime and Ferguson* [2003] for converting from area-by-number to grid-by-number. It is likely that stated biases are significantly overestimated, and likely to be closer to those of *Graham et al.* [2005b].
Table 3. Errors associated with manual digitizing and semi-automated image analysis.

<table>
<thead>
<tr>
<th></th>
<th>Area-by-number</th>
<th>Grid-by-number</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_{10} mean error</td>
<td>-0.011</td>
<td>-0.028</td>
</tr>
<tr>
<td>D_{16} mean error</td>
<td>-0.004</td>
<td>-0.030</td>
</tr>
<tr>
<td>D_{25} mean error</td>
<td>0.018</td>
<td>-0.029</td>
</tr>
<tr>
<td>D_{50} mean error</td>
<td>0.091</td>
<td>0.049</td>
</tr>
<tr>
<td>D_{75} mean error</td>
<td>0.217</td>
<td>0.078</td>
</tr>
<tr>
<td>D_{84} mean error</td>
<td>0.252</td>
<td>0.048</td>
</tr>
<tr>
<td>D_{90} mean error</td>
<td>0.279</td>
<td>-0.047</td>
</tr>
<tr>
<td>Mean error or Bias, b</td>
<td>0.120</td>
<td>0.006</td>
</tr>
<tr>
<td>Mean-square error, $E_{ms}$</td>
<td>0.058</td>
<td>0.010</td>
</tr>
<tr>
<td>Irreducible random error, $e$</td>
<td>0.209</td>
<td>0.099</td>
</tr>
</tbody>
</table>

Note: Grid-by-number data are obtained by converting the area-by-number data using the procedure of Kellerhals and Bray [1971].
Step 1
Collection of photos

- What sampling area is required?
- What internal camera settings (focal length, resolution etc.) are required? [Graham et al., 2005a]
- What external camera configuration (elevation, flash etc.) is required? [Graham et al., 2005b]
- What type of illumination is required? [Graham et al., 2005b]
- What image resolution is required? [Graham et al., 2005a]

Step 2
Extraction of grains

- What is the size of the minimum grain identifiable? [Graham et al., 2005a]

Consequently:
- What effect does truncation have on the resulting size distribution?
- Does bed structure affect the grain-size distribution?
- How should particles touching the edges of the photo be treated? [Graham et al., 2005b]
- What errors are associated with the grain extraction & measurement algorithms? [Graham et al., 2005a,b]

Step 3
Analysis of results

- Does the method of calculating percentiles affect the results?
Error in $D_{50}$ (% in mm)

Error in $D_{90}$ (% in mm)

Ratio of sample area / population $D_{max}$
a

Error in
percentile (Psi)

Mean error in
percentile (Psi)

b

Error in
percentile (Psi)

Mean error in
percentile (Psi)

c

Error in
percentile (Psi)

Mean error in
percentile (Psi)
a) Burial
- true
- apparent

b) Overlapping
- true
- apparent

c) Foreshortening
- true
- apparent
Paint-and-pick sampling

True distribution
Apparent distribution

Manual digitization

True distribution
Apparent distribution

Manual digitization

True distribution
Apparent distribution

Manual digitization
Linear interpolation

Absolute error

Percentile

Spline interpolation

Mean error + SE

Maximum error

0 0.1 0.2 0.3

0 0.1

0 50 100

Percentile

Linear interpolation

Absolute error

(σd)