SMS-618, Particle Dynamics, Fall 2003 (E. Boss)

Calculus of Particle Size distributions

For sediments, grain size is typically distributed according to a power (or Φ) scale (Table 1). $\Phi=-\log_2(D)$ or $D=2^{-\Phi}$, where D is the particle diameter in mm. It is based on the idea that since particles span several orders of magnitude, a power scale will provide a better description of all scales. It is also consistent with the observation, that, in general, the volume (or mass) per logarithmic bin is approximately equal across all bins (this is often called the Junge distribution). This is, off course, a gross simplification to a complex reality. It has, however, been the starting point for much good science.
Using sequential weighing of material going through sieves of different sizes, a histogram of % mass as function of size is generated (e.g. Fig. 1). For large material dry sieving is performed while for clays (which when dried up will clump to large aggregates) are sieved when wet in suspension with a deflocculating agent (e.g. Calgon).

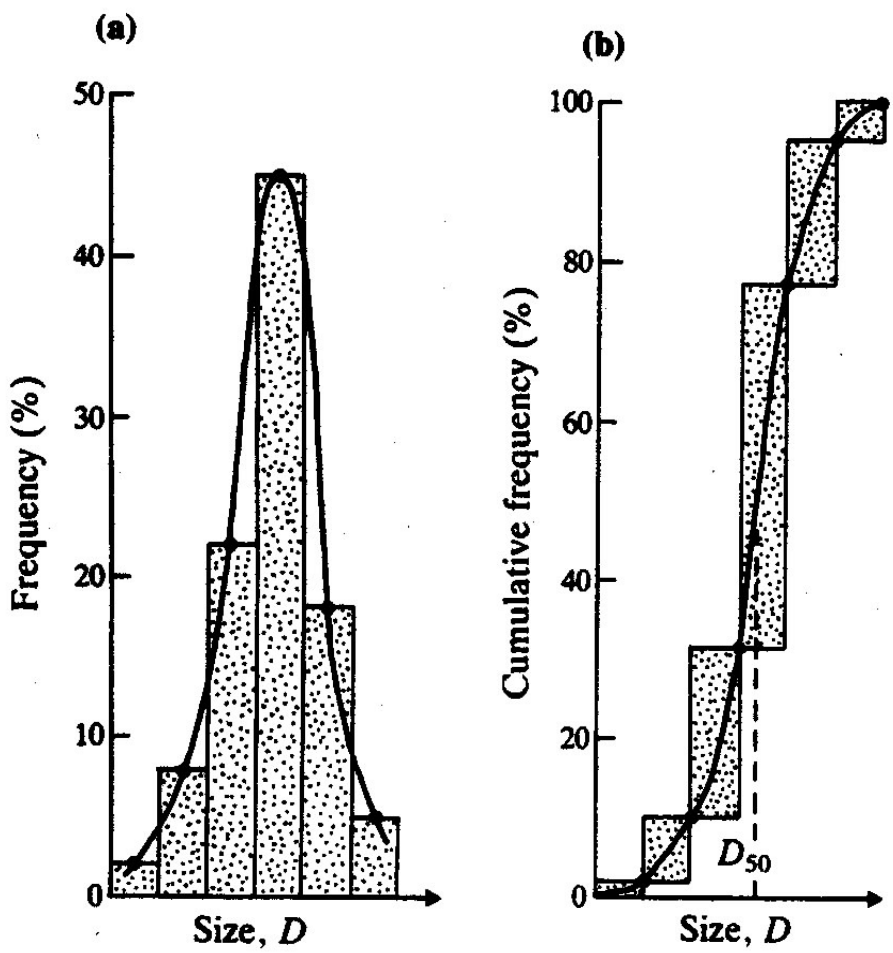

Figure 1. A frequency histogram (left) and the associated cumulative frequency historgram (based on Figure 2.2. in Allen, 2001) .

The boundaries of each bar in the histogram are the nominal size of the filter (sieve). By summing the histograms from the smallest size up, the cumulative distribution is generated (Fig. 1 right panel). It describes the percent mass below a given size.

Mathematically, the % mass distribution can be written, for the sieving example above, as follows:

$$f(D) = \begin{cases} 8\% & D_{\min} < D < D_1 \\ 15\% & D_1 < D < D_2 \\ \vdots & \vdots \\ 3\% & D_{\max} < D < \infty \end{cases},$$

while the cumulative distribution is:

$$F(D) = \begin{cases} 8\% & D_{\min} < D < D_1 & = \mathrm{f}(D_{\min} < D < D_1) \\ 23\% & D_1 < D < D_2 & = \mathrm{f}(D_{\min} < D < D_1) + \mathrm{f}(D_1 < D < D_2) \\ \vdots & \vdots & \vdots \\ 100\% & D_{\max} < D < \infty & \sum_N f(D_{j-1} < D < D_j) \end{cases}.$$

In the example above *f*(D) is the particulate (mass) size distribution (PDF) while *F*(D) is the cumulative size distribution. For the sake of comparison the PDF is described by its statistical (parametric or nonparametric) properties (e.g. mean size, median size, mode size, standard deviation, percentiles, kurtosis, skewness, etc'). Note that the errors of these statistics can be quite large and depends on how fine the scale is and how many particles we have in each bin.

The frequency distribution provides us with:
Median particle size, d_{50}, which is the size at which 50% by weight is finer.
Mean particle size, $d_m=\sum(p_iD_i)/100$, where p_i=percentage by weight of grain of size D_i.
Standard deviation, $\sigma=0.5(D_{84}+D_{16})$, where subscript denotes the position of the percentile. In terms of the Φ values (Finite bins):
Mean: $\Phi_m=\sum(p_i \Phi_i)/100$, where p_i=percentage by weight of grain of size Φ_i
Standard deviation: $\sigma_\Phi=0.5(\Phi_{84}+\Phi_{16})$
Skewness: $\alpha=(\Phi_m+\Phi_{50})/\sigma_\Phi$
Kurtosis: $\beta=0.5(\Phi_{95}+\Phi_5-\sigma_\Phi)/\sigma_\Phi$

An analytic function with a few parameters is often fit to the size distribution. The most often used PSDs are the normal, log-normal and hyperbolic distributions (see appendix). How to fit a function to the data is not a trivial matter and is often done without the necessary care. We never know *f(D)* at a specific *D*. *F(D),* however, is known, within the (aggregated) measurement uncertainties, at the boundaries between each size bin. Thus, it is *F(D)* that should be fit to an analytical function using its values at D_j. Lets assume we want to fit the analytical function *G(D)* to our N observations $F(D_j)$ each of which has an uncertainty of $\pm\delta F(D_j)$. We derive a cost function (in the least-squares sense):

$$\cos t = \sum_{j=1}^{N} \left(\frac{\left(G(D_j) - F(D_j)\right)}{\delta F(D_j)} \right)^2$$

We then find the parameters of the analytical function *G(D)* that minimize this cost function.

Once *G(D)* has been derived, we can obtain the fitting function of the PSD *f(D)=dG(D)/dD*. We often have to translate from a mass (or volume distribution, assuming all particles have the same density, most often ρ=2.65gr/cm^3) to an area or number concentration. Assuming the particles are spherical, the number concentration is obtained from the volume concentration distribution by dividing with the volume:

$$N(D)=g(D)/\{4\pi D^3/3\}.$$

Once a fitting function has been derived it is important to *quantify* how well it fits the data. This could be done using the average, worst, median difference between the data and fit {i.e. the statistics of the residual *G(D)-F(D),* taking into account $\delta F(D_j)$}. The shape of the residual as function of *D* will indicate the shortcoming of the fitting function if it does not look random. Since F(D) is (or is proportional to) a cumulative distribution function the Kolmogorov-Smirnov test can be used to evaluate the likelihood that G(D) is indeed the underlying distribution (e.g. Press et al.'s Numerial Recipes). This test is based on max{$|G(D)-F(D)|$}.
An analytical function with more fit parameters is very likely to provide a better fit to the data than one that has less. It is important to remember, however, that an important reason to do analytical fits is to provide the maximal informational content of the data with a minimum set of parameters.

An important point we need to remember throughout this analysis is that each size filter has its own biases. Those biases are easy to deal with when the particles are all perfect spheres but are hard to deal with when we deal with elongated particles. For example, when we deal with long and skinny particles, depending on which 2-D projection is presented to the sieve, the particles will or will not make it through.

Another important point is that each particle sizing technique measures a different property of the particle. Some measure a proxy of the volume of each particle (e.g. Coulter Counter), some the cross-sectional area of all particles of approximately the same size (LISST), some the volume of all particles of approximately the same size (acoustic and optics in the Rayleigh limit, when wavelength >> D). Comparing distribution generated by different techniques require making assumption regarding shape (e.g. to convert from cross-sectional area to volume). In this respect sphere are not an ideal shape but rather an extreme shape; a sphere has the *smallest* surface are to volume ratio of all 3-D shapes.

PHI SIZE CONVERTER

GRAIN SIZE mm	PHI f		
256	-8.00	BOULDER	
215	-7.75		G
181	-7.50		
152	-7.25		R
128	-7.00	Cobble	
108	-6.75		A
90.5	-6.50		
76.1	-6.25		V
64.0	-6.00		
53.8	-5.75		E
45.3	-5.50		
38.1	-5.25		L
32.0	-5.00		
26.9	-4.75		
22.6	-4.50		
19.0	-4.25	Pebble	
16.0	-4.00		
13.5	-3.75		
11.3	-3.50		
9.51	-3.25		
8.00	-3.00		
6.73	-2.75		
5.66	-2.50		
4.76	-2.25		
4.00	-2.00		
3.36	-1.75		
2.83	-1.50	Granule	
2.38	-1.25		
2.00	-1.00		
1.68	-0.75		
1.41	-0.50	Very Coarse	
1.19	-0.25		
1.00	0.00		
0.841	0.25		
0.707	0.50	Coarse	
0.595	0.75		S
0.500	1.00		
0.420	1.25		A
0.354	1.50	Medium	
0.297	1.75		N
0.250	2.00		
0.210	2.25		D
0.177	2.50	Fine	
0.149	2.75		
0.125	3.00		
0.105	3.25		
0.0884	3.50	Very Fine	
0.0743	3.75		
0.0625	4.00		

GRAIN SIZE microns	PHI f			
52.6	4.25			
44.2	4.50	Coarse		
37.2	4.75			
31.3	5.00		S	
26.3	5.25			
22.1	5.50	Medium	I	
18.6	5.75			
15.6	6.00		L	
13.1	6.25			
11.0	6.50	Fine	T	
9.29	6.75			
7.81	7.00			
6.57	7.25			
5.52	7.50	Very Fine		M
4.65	7.75			
3.91	8.00			U
3.28	8.25			
2.76	8.50			D
2.32	8.75			
1.95	9.00	CLAY		
1.64	9.25			
1.38	9.50			
1.16	9.75			
0.977	10.00			
0.821	10.25			
0.691	10.50			
0.581	10.75			
0.488	11.00			
0.411	11.25			
0.345	11.50			
0.290	11.75			
0.244	12.00			
0.205	12.25			
0.173	12.50			
0.145	12.75			
0.122	13.00			
0.103	13.25			
0.0863	13.50	COLLOIDS		
0.0726	13.75			
0.0610	14.00			
0.0513	14.25			
0.0432	14.50			
0.0363	14.75			
0.0305	15.00			
0.0257	15.25			
0.0216	15.50			
0.0181	15.75			
0.0153	16.00			

Table 1. The Φ chart. Converts between size (mm or micron) and Φ value. Conversion is based on $D[mm]=2^{-\Phi}$.

Appendix

Analytical functions that are often applied to cumulative particles size distributions.

I. Hyperbolic size distribution (two fit parameters, A and α, and two boundaries D_{min} & D_{max},often determined by sampling method):

$$g(D)=\begin{cases}0 & D<D_{min}\\ AD^{-\alpha} & D_{min}<D<D_{maz}\\ 0 & D_{max}<D\end{cases}\rightarrow G(D)=\int_0^D g(D')dD'=\begin{cases}0 & D<D_{min}\\ \dfrac{AD^{-\alpha+1}}{-\alpha+1} & D_{min}<D\end{cases}$$

for $\alpha\neq1$.

II. Normal size distribution:

$$g(D)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\left(D-\overline{D}\right)^2}{2\sigma^2}\right]$$

where $\overline{D}$ and σ are the two parameters of this distribution, the mean and standard deviation respectively:

$$\overline{D}=\int_{-\infty}^{\infty}Dg(D)dD,\ \sigma=\int_{-\infty}^{\infty}\left(D-\overline{D}\right)^2g(D)dD.$$

III Log-Normal size distribution:

$$g(D)=\frac{1}{D\sigma_g\sqrt{2\pi}}\exp\left[-\frac{\left(\log D-\log\overline{D}\right)^2}{2\sigma_g{}^2}\right]$$

It is logD not D that is normally distributed so:

$$\log\overline{D}=\int_{-\infty}^{\infty}\log Dg(D)dD,\ \sigma^2{}_g=\int_{-\infty}^{\infty}\left(\log D-\log\overline{D}\right)^2g(D)dD.$$

In this case, the median equals the geometric mean: $D_g=\overline{D}=(D_1D_2D_3\ldots D_n)^{1/n}$.

σ_g is the standard deviation of logD, the geometric mean standard deviation.

The mode (where the peak of the distribution is), median and mean are related to the geometric mean by:

$\log(D_{Mode})=\log(D_g)-\sigma_g{}^2$

$\log(D_{mean})=\log(D_g)$

$\log(\langle D\rangle)=\log(D_g)+0.5\sigma_g{}^2$

For such a distribution the cumulative distribution is a straight line on a log(G(D))-logD plot.

IV zeroth order logarithmic distribution (ZOLD)

$$g(D) = \exp\left[-\frac{\left(\log D - \log D_{Mode}\right)^2}{2\sigma_0{}^2}\right] \Big/ \sqrt{2\pi}\,\sigma_0 D_M \exp\left[\sigma_0^2/2\right]$$

Which is a two parameter (D_{Mode}, σ_0) distribution.
The relationship between mean and mode is given by:

$\log D_{Mean} = \log D_{Mode} + 1.5\sigma_0{}^2$,

and the standard deviation is given by:

$\sigma = D_{Mode}[\exp(4\,\sigma_0{}^2) - \exp(3\sigma_0{}^2)]^{1/2}$.

More can be found in:

Allen, J. R. L., 2001. Principles of physical sedimentology. Blackburn press.
Dyer, K. E., 1986. Coastal and Estuarine Sediment Dynamics, Wiley.
Kerker M., 1969. The Scattering of light and other EM radiation, Academic Press.
Shifrin, K., 1983.Physical Optics of Ocean Water, AIP.