Emmanuel Boss and Nils Haentjens contributed to this page.

Particle size distributions (PSDs), the descriptions of how particle number, area, or volume depend on particle size, are useful tools in oceanography used in applications such as characterizing the ecosystem (Cermeno and Figueiras, 2008) or computing the carbon flux to depth (Guidi et al., 2009). While PSDs have been extensively used in the literature, surprisingly little has been written about how they are actually derived from measured data, so we therefore felt the need to write this short note.

Empirical PSDs

How one constructs a PSD depends on the tool used to size the assembly of particles. Single-particle analyzers, such as the Coulter Counter or cytometers, provide information on the size of each individual particle passing through the instrument (typically based on volume or cross-sectional area and assuming an equivalent sphere). Other methods, such as the Laser In Situ Scattering and Transmission meter (LISST, Agrawal and Pottsmith, 2000), provide a PSD of the bulk assembly of particles by inverting a bulk measurement (near-forward angular scattering in the case of the LISST) to obtain the most likely underlying PSD.

The process of building a PSD from the size information of individual particles is as follows. Choose a number of bins (M) and denote their boundaries $b_1, b_2, , b_{M+1}$. Place a particle in the particular bin for which its diameter (D) obeys $b_j \leq D < b_{j+1}$ (by "placing it" is meant that the number of particles in that bin is incremented by one). Ideally, the size characterizing each bin is based on the mean size of the particles in that bin. Typically, however, that size is based on the boundaries of the bin (e.g. the arithmetic or geometric mean of the bin boundaries). Thus the "discrete" PSD, denoted by $N(D_j)$, gives the number of particles with mean diameter D_j (units of number per volume of water). To obtain a continuous PSD, $n(D_j)$, (for example for the purpose of comparison between different instruments each having different bin sizes), one divides the discrete PSD by the bin width: $n(D_j) = N(D_j)/(b_{j+1} - b_j)$ (units of number per volume per size). To obtain a volume or area size distribution, the number of particles in each bin is multiplied by the average volume or area of a particle in that bin, respectively.

Continuous particle number size distributions in the surface ocean are often approximated by a power-law distribution (e.g. Jackson et al., 1997):

$$n(D) = AD^{-\xi} \quad [\text{number cm}^{-3} \, \mu\text{m}^{-1}]. \tag{1}$$

Power-law differential distributions are observed to have an exponent (ξ above) varying between 3 and 4 in the surface ocean (Jackson et al., 1997; Sheldon et al., 1972). An exponent of 4 implies that volume is constant within bins increasing in size by a power-law rule (e.g. Sheldon et al., 1972). A problem in the above equation is that, in principle, we should never exponentiate with a fraction a quantity that has physical units. Often, (likesection1) will be written instead as function of a nondimensional ratio, e.g.

$$n(D) = A(D/D_0)^{-\xi},$$

with D_0 being a reference diameter. In what follows we assume, without loss of generality, that $D_0 = 1 \, \mu\text{m}$, and D is reported in μm and hence this normalization is implicitly assumed.

Before the PSD can be built certain decisions need to be made. The upper and lower bounds for particle size range, the number of bins, and the rule according to which bins are allocated. Traditionally, due to the rapid decrease in particle concentration with size, bin sizes have been chosen to follow a power-law scaling (Sheldon et al., 1972; Jackson et al., 1997; Agrawal and Pottsmith, 2000). That is to say, a subsequent bin is q times larger than the previous bin (for other possible choices see the Appendix below). This choice has the advantage that bins are of equal size on a logarithmic D (abscissa) axis and that oceanic volume distributions are nearly flat (Sheldon et al., 1972), which provides a quick check on the data. The downside is that over a decade in size the number of particles per bin still decreases rapidly. For example, for a choice of ten bins over a decade, the number of particles between the first and last bin fall by a factor of ~ 1000. This means that to reduce counting errors to 10% at the largest bin (counting errors scale like $\sqrt{N}$) for such a choice, more than 100,000 particles are required per sample, which is typically unrealistically large for cytometry.

Parametric Description of a PSD

Assume we want to produce a size distribution for oceanic plankton (e.g. Fig. 2 in Lombard et al., 2019). Denote the boundaries of the bins by $b_1, b_2, , b_{M+1}$. Assuming that the bins they bound grow following a power-law:

$$q = \frac{b_3 - b_2}{b_2 - b_1} = \frac{b_4 - b_3}{b_3 - b_2} = \cdots = \frac{b_{M+1} - b_M}{b_M - b_{M-1}},$$

which is satisfied if for any j,

$$b_j = b_1 q^{j-1} \implies b_{M+1} = b_1 q^M \implies q = \sqrt[M]{\frac{b_{M+1}}{b_1}}.$$

Thus, if we have the lowest and largest boundaries of the PSD (b_1 and b_{M+1}) and the number of bins in the PSD (M), we can compute the boundaries of all other bins. The volume of material associated with spherical particles distributed as a power-law with a differential power law exponent of 4 (the "canonical" value of Sheldon et al., 1972) is

$$V(b_j < D < b_{j+1}) = \frac{\pi A}{6} \int_{b_j}^{b_{j+1}} D^{-1} dD = \frac{\pi A}{6} \ln q,$$

which, as discussed above, is the same for all bins.

The average size of a particle associated with each bin for such a PSD is

$$\overline{D}(b_j < D < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(D) D \, dD}{\int_{b_j}^{b_{j+1}} n(D) \, dD} = \frac{\int_{b_j}^{b_{j+1}} D^{-3} dD}{\int_{b_j}^{b_{j+1}} D^{-4} dD} = \frac{3(b_{j+1}^{-2} - b_j^{-2})}{2(b_{j+1}^{-3} - b_j^{-3})} = \frac{3b_{j+1}(1+q)}{2(1+q+q^2)} = \frac{3b_1 q^j (1+q)}{2(1+q+q^2)}.$$

(2)

This, however, is different from the typical size chosen to represent bins in the literature. Typically the size associated with PSD bins is computed as the geometric mean of the bin boundaries:

$$\overline{D}(b_j < D < b_{j+1}) = \sqrt{b_{j+1} b_j} = b_j \sqrt{q} = b_1 q^{j-1} \sqrt{q},$$

which is larger than the size computed in (likesection2). While the two may be close (their ratio is constant and depends on q), they are not identical. Choosing the arithmetic mean to represent the bin is even more biased.

This issue is of importance because the mean size of the bin is used to convert between number, area, and volume size distributions. For example the LISST particle size output is a volume distribution (in parts per million, ppm). Converting it to a size distribution requires a division of the volume in each bin by the volume of the average particle in that bin. The choice of the average diameter, because it is cubed, can result in a significant bias.

In coastal areas the differential PSD power-law slope of particles has an exponent closer to $\xi = 3$ (e.g. Jackson et al., 1997). In such cases, where we expect the power-law exponent not to be 4, the mean size representing a bin changes from that of (likesection2) to (Boss et al., 2001):

$$\overline{D}(b_j < D < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(D)D dD}{\int_{b_j}^{b_{j+1}} n(D)dD} = \frac{\int_{b_j}^{b_{j+1}} D^{1-\xi}dD}{\int_{b_j}^{b_{j+1}} D^{-\xi}dD} = \frac{(1-\xi)(b_{j+1}^{2-\xi} - b_j^{2-\xi})}{(2-\xi)(b_{j+1}^{1-\xi} - b_j^{1-\xi})}.$$

Normalizing a PSD by the total number of particles between the lower and upper boundaries provides a probability distribution for a particle to be within a specific bin. For a power-law PSD with exponent ξ:

$$p(b_j < D < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(D)dD}{\int_{b_1}^{b_{M+1}} n(D)dD} = \frac{b_{j+1}^{1-\xi} - b_j^{1-\xi}}{b_{M+1}^{1-\xi} - b_1^{1-\xi}}.$$

The cumulative distribution, $P(D < D_c) = \int_{b_1}^{D_c} p(D)dD$ is therefore

$$P(b_1 < D < D_c) = \frac{\int_{b_1}^{D_c} n(D)dD}{\int_{b_1}^{b_{M+1}} n(D)dD} = \frac{D_c^{1-\xi} - b_1^{1-\xi}}{b_{M+1}^{1-\xi} - b_1^{1-\xi}},$$

which is useful to answer questions such as how numerically abundant are certain planktonic species compared to all other groups. A similar calculus is used to derive the probability distribution for particle area or volume.

Jackson et al., 1997, further discuss the case of computing the PSD for the solid fraction of particles if the particles are fractals (as is the case for oceanic aggregates). This requires assumptions regarding the change of fractal dimension with size and hence is not discussed further here (see, for example, Maggi, 2007; Kahlifa and Hill, 2006).

Ecological Size Spectra

In ecology, size spectra are typically represented by an abundance or biomass size spectrum with the size axis often represented as mass or volume (e.g., the review by Blanchard et al., 2017). A power-law function is often fit to the spectrum whose value is interpreted based on ecological theory. In such a case the number distribution is represented as

$$n(V) = AV^{-\xi_2} \quad [\text{particles cm}^{-3}\,\mu\text{m}^{-1}],$$

and we expect that $\xi_2 = \xi/3$, which is less steep than is a function of size as seen in (likesection1). Because it is less steep, using the *arithmetic* mean for bin size is reasonable for the canonical distribution, more so than the *geometric* mean. In this case,

$$\overline{D}(b_j < V < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(V)DdV}{\int_{b_j}^{b_{j+1}} n(V)dV} = \frac{\int_{b_j}^{b_{j+1}} DV^{-\xi_2}dD}{\int_{b_j}^{b_{j+1}} V^{-\xi_2}dD} = \frac{(1-\xi_2)(b_{j+1}^{4/3-\xi_2} - b_j^{4/3-\xi_2})}{(4/3-\xi_2)(b_{j+1}^{1-\xi_2} - b_j^{1-\xi_2})}.$$

Correctly Fitting a Power Law to a PSD

Suppose one is interested in fitting a power-law model to a PSD, for example as a simple descriptor of a relative contribution of large vs. small particles in a given sample. What is the correct way to fit the PSD such that the exponent will be appropriately computed whether it is computed from a number, area, or volume distribution?

The answer is as follows. The number distribution $n(D_i)$, where D_i is the size representing the ith bin, has an uncertainty $\delta(D_i)$. For example, if the uncertainty is due to counting alone, $\delta(D_i) = \sqrt{n(D_i)}$ (another source of uncertainty may be instrument sensitivity, particularly at the small end of the PSD). If the metric for fitting is to minimize the root-mean-square error, we want to find the fit parameters A and ξ that minimize the cost function

$$\chi = \sum_{i=1}^{M} \left[\frac{n(D_i) - AD_i^{\xi}}{\delta(D_i)} \right]^2 .$$

A more robust fit that reduces the weight of outliers may be found by minimizing

$$\chi = \sum_{i=1}^{M} \frac{|n(D_i) - AD_i^{\xi}|}{\delta(D_i)} .$$

If the appropriate uncertainties are used for each type of size distribution (be it number, area, or volume), the exponent found should be the same as we would expect, e.g. the exponent for the differential volume distribution equals that of the differential number distribution plus three.

On the other hand, if one simply fits a type-I regression line to $\log\{n(D_i)\}$ vs. $\log\{D_i\}$, the implicit assumption is that the relative uncertainty in $n(D_i)$ is constant and the exponent obtained will not be consistent between the different size distributions.

References

Agrawal, Y. C. and H. C. Pottsmith. 2000. Instruments for particle size and settling velocity observations in sediment transport. *Mar. Geol.* 168(14), 89114.

Boss, E., M. S. Twardowski, and S. Herring. 2001. Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. *Appl. Opt.* 40(27), 48854893.

Cermeno, P. and F. G. Figueiras. 2008. Species richness and cell-size distribution: size structure of phytoplankton communities. *Mar. Ecol. Prog. Ser.*, 357, 7985.

Guidi, L., L. Stemmann, G. A. Jackson, F. Ibanez, H. Claustre, L. Legendre, M. Picheral and G. Gorski. 2009. Effects of phytoplankton community on production, size, and export of large aggregates: a world-ocean analysis. *Limnol. Oceanogr.* 54, 19511963.

Jackson G. A., R. Maffione, D. K. Costello, A. L. Alldredge, B. E. Logan, and H. G. Dam. 1997. Particle size spectra between 1 m and 1 cm at Monterey Bay determined using multiple instruments. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 44(11), 17391767.

Khelifa A., and P. S. Hill. 2006. Models for effective density and settling velocity of flocs. *J. Hydraul. Res.* 44, 390401.

Lombard, F., E. Boss, A. M. Waite, M. Vogt, J. Uitz, L. Stemmann. H. M. Sosik, J. Schulz, J-B. Romagnan, M. Picheral, J. Pearlman, M. D. Ohman, B, Niehoff, K. O. Mller, P. Miloslavich, A. Lara-Lpez, R. Kudela, R. M. Lopes, R. Kiko, L. Karp-Boss, J. S. Jaffe, M. H. Iversen, J-O. Irisson, K. Fennel, H. Hauss, L. Guidi, G. Gorsky, S. L. C. Giering, P. Gaube, S. Gallager, G. Dubelaar, R. K. Cowen, F. Carlotti, C. Briseo-Avena, L. Berline, K. Benoit-Bird, N. Bax, S. Batten, S. D. Ayata, L. F. Artigas and W. Appeltan, 2019. Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Front. Mar. Sci.* 6:196. doi: 10.3389/fmars.2019.00196F.

Maggi, F., 2007. Variable fractal dimension: A major control for floc structure and flocculation kinematics of suspended cohesive sediment. *J. Geophys. Res.* 112(C7), C07012.

Sheldon, R. W., A. Prakash, and W. H. Sutcliff. 1972. Size distribution of particles in ocean. *Limnol. Oceanogr.* 17: 327340.

Appendix: Building PSDs with a Different Rule for the Bin Size

In principle, one could use a different bin size convention from the canonical use of bins increasing as a power law. For example, if one desires a PSD where, for any ξ, the volume in each bin is the same, the process is as follows. The total volume (assuming spheres) is

$$\int_{b_1}^{b_{M+1}} A\frac{\pi D^3}{6} D^{-\xi} dD = \frac{A\pi(b_1^{4-\xi} - b_{M+1}^{4-\xi})}{6(4-\xi)} = V_0 \,.$$

For every bin to have the same volume we have

$$\int_{b_j}^{b_{j+1}} A\frac{\pi D^3}{6} D^{-\xi} dD = \frac{A\pi(b_j^{4-\xi} - b_{j+1}^{4-\xi})}{6(4-\xi)} = \frac{V_0}{M} \,,$$

which implies that

$$b_1^{4-\xi} - b_2^{4-\xi} = b_2^{4-\xi} - b_3^{4-\xi} = \cdots = b_M^{4-\xi} - b_{M+1}^{4-\xi} = \frac{b_1^{4-\xi} - b_{M+1}^{4-\xi}}{M} \,.$$

With a choice of b_1, b_{M+1}, and M, we can compute all the other bin boundaries.

On the other hand, if one wanted to have bins of constant numbers of particles (for example to have similar counting errors in all bins), we would require the same number of

particles in each bin. Assuming we have N_0 particles in M bins spanning from b_1 to b_{M+1} gives

$$\int_{b_1}^{b_{M+1}} AD^{-\xi}dD = \frac{A(b_1^{1-\xi} - b_{M+1}^{1-\xi})}{1-\xi} = N_0 \, .$$

For every bin to have the same number of particles requires that

$$\int_{b_j}^{b_{j+1}} AD^{-\xi}dD = \frac{A(b_j^{1-\xi} - b_{j+1}^{1-\xi})}{1-\xi} = \frac{N_0}{M} \, ,$$

which implies that

$$b_1^{1-\xi} - b_2^{1-\xi} = b_2^{1-\xi} - b_3^{1-\xi} = \cdots = b_M^{1-\xi} - b_{M+1}^{1-\xi} = \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M}$$

or

$$b_2^{1-\xi} = b_1^{1-\xi} - \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M}, \cdots, b_{j+1}^{1-\xi} = b_j^{1-\xi} - \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M} \, .$$

Again, with a choice of b_1, b_{M+1}, and M, we can compute all of the other bin boundaries.