Particle-Size Distribution Data Presentation: *Why NOT to use Log-Density Distribution!*

Comments *by Prof. M. Kostic*, Northern Illinois University < www.kostic.niu.edu >
2003 NASA Faculty Fellow at Glenn Research Center, Cleveland, OH < www.grc.nasa.gov >

General Comments:

It is very important to properly evaluate and present an aerosol *particle property distribution*, as a function of a characteristic particle property itself.

The most common way of particle distribution data evaluation and presentation is to represent the so called *'cumulative frequency distribution,'* as a corresponding cumulative particle frequency (like cumulative number of particles, N, but also any other particle property, like cumulative particle volume, cumulative particle area, etc.) of all particles which have smaller than or equal to values of a characteristic property (like characteristic size or equivalent diameter, D_p , for example), as a function of that characteristic particle property. The cumulative frequency distribution, $N_{cum}(D_p)$, is normalized with reference to the total number of all particles, N_{Total}, so that its value, as a ratio of a total, varies from 0 to 1 or from 0 to 100%, see (Eq. 1) and Fig. 1.

$$N_{cum}(D_p) = \frac{N_{0<Size<D_p}}{N_{Total}} \qquad \text{(Eq. 1)}$$

It is often of interest to evaluate and represent the so called *range* or *spectral particle distribution* as a *histogram*, or range frequency over a given range or a 'bin'. Such frequency is naturally dependent on the given range size and any generalization, normalization or curve fitting is meaningless for comparison and thus inappropriate [*Sommer (2001)*]. However, representing range frequency per unit of the corresponding range, in limit, becomes a continuous function of that characteristic particle property (e.g., D_p) which is independent on the "bin" range size, and thus it could be generalized, normalized, and discrete experimental values could be curve fitted. Such a function represents an important aerosol property, the particle *density distribution*, see (Eq. 2) and Fig. 2.

$$n_d(D_p) = \frac{dN_{cum}(D_p)}{dD_p} = \lim_{\Delta D_p \to 0} \frac{\Delta N}{\Delta D_p} = \lim_{D_{p,i+1} - D_{p,i} \to 0} \frac{N_{i+1} - N_i}{D_{p,i+1} - D_{p,i}} \qquad \text{(Eq. 2)}$$

The (Eq. 2) has an important physical meaning, since its integral over any range ($\Delta D_p = D_{p2} - D_{p1}$) will represent the particle frequency (ΔN_{12}) over that range, see Fig. 2, i.e.:

$$\Delta N_{12} = \int_{D_{p1}}^{D_{p2}} n_d(D_p) \cdot dD_p \qquad \text{(Eq. 3)}$$

Also, the higher value of density distribution at a given particle size, the higher value of particle frequency (i.e., ΔN) over the same narrow range interval around a particle size. The maximums of the density distribution function represent the fact that particle distribution tends to concentrate around those maximum values, also known as *modal* values, see Fig. 2. Some aerosols have maximum of particles per unit of range (thus maximum density distribution) around a certain single value (*monomodal* aerosols), while the other aerosols may have more than one density distribution maximums (*polymodal* aerosols).

If a particle distribution spans over several orders of magnitudes, it is advantageous to present the particle distribution data in a semi-logarithmic plot, with the particle size on a logarithmic abscissa coordinate. Some researchers present their density distribution data, not only in the semi-logarithmic coordinate system, but also define the so called *logarithmic density distribution*, as:

$$n_{d\_\log}(D_p) = \frac{dN}{d\log D_p} = \lim_{\Delta D_p \to 0} \frac{\Delta N}{\Delta \log D_p} \qquad \text{(Eq. 4)}$$

Note that (combining Eqs 2 &4), we obtain:

$$n_{d\_\log}(D_p) = \frac{dN}{d\log_B D_p} = \ln B \cdot D_p \left.\frac{dN}{dD_p}\right|_{B=10} = 2.303 \cdot D_p \cdot n_d(D_p) \qquad \text{(Eq. 5)}$$

Using the *logarithmic density distribution* (Eq. 4 & 5) introduces certain experimental, scaling, and dimensional complications, for example:

$$\log(D_{p2}) + [-\log(D_{p1})] = \log\frac{D_{p2}}{D_{p1}} \qquad \text{(Eq. 6)}$$

$$\log(D_{p2}) + \log(D_{p1}) = \log(D_{p2} \cdot D_{p1}) \qquad \text{(Eq. 7)}$$

The dimension of the Left-Hand-Side of (Eq. 6) is [log(Length)], i.e. [log(m)], while the Right-Hand-Side is dimensionless, i.e. [1]. However, dimension of the LHS of (Eq. 7) is the same as (Eq. 6), [log(Length)], i.e. [log(m)], while the dimension of the RHS is [log(Length^2)], i.e. [log(m^2)]. These anomalous outcomes are due to the fact that the logarithmic function (like trigonometric ones) fundamentally operates on dimensionless numbers only (unless the formulas are strictly empirical and used without any generalization!). Scaling is also an issue since the logarithmic function is bigger or smaller than unity (1), or even negative, depending whether its argument is smaller or bigger than the log-base, B, or smaller than 1, respectively. Thus, significant amplification or reduction (or both) of calculated values may occur only because values of the argument may vary over several orders of magnitude depending on an arbitrary choice of units used in measurements or conversions during data reduction.

There is no physical nor practical reason to evaluate and use the *logarithmic density distribution* (Eq. 4), since it no longer represents the true particle *density distribution* (need to be called 'true' now), like (Eq. 2), but instead it is a density distribution of the logarithm(!) of the particle characteristic parameter, namely the logarithm of a particle size D_p. For example, for a uniform density distribution, $n_d(D_p)=C=const$, the *logarithmic density distribution* (Eq. 5), will become, $n_{d\_log}(D_p)=2.303\cdot C\cdot D_p$, i.e will become a linearly increasing function of particle size, D_p (which then appears curved in semi-log coordinate system), with a meaningless maximum at the end of the particle size range, see Fig. 3. This 'artificial scaling' with value of independent coordinate (particle size, see Eq. 5) of actual density distribution, will not only skew the distribution, but also shift the modal vales and thus further misrepresent the actual particle distribution, since the maximums of the two density distributions are not at the same modal values ($D_{p,mod} \neq D_{p,mod\_Log}$), see (Eq.8) and Figs. 2-4 and 6:

$$\frac{d}{dD_p}\left(n_{d\_\log}(D_p)\right)=\frac{d}{dD_p}\left(\ln B\cdot D_p\cdot n_d(D_p)\right)\Big|_{B=10}=2.303\cdot\underbrace{\left\{n_d(D_p)+D_p\cdot\underbrace{\frac{d}{dD_p}\left[n_d(D_p)\right]}_{=0\,for\,D_p=D_{p,\mathrm{mod}}}\right\}}_{=0\,for\,D_p=D_{p,\mathrm{mod\_}Log}\neq D_{p,\mathrm{mod}}}=0$$

(Eq.8)

Therefore, it is often necessary and useful to graphically represent the particle true *density distribution* (Eq. 2) in semi-logarithmic coordinate system, but it is meaningless and unnecessary (does not serve any purpose) to evaluate the *logarithmic density distribution* (Eqs. 4 & 5), since it may introduce dimensional and scaling problems, and misrepresentation of actual or true *density distribution* (Eq. 2), including shifting of the *modal* values, see again (Eq.8) and Figs. 2-4 and 6.

GRC-PAGEMS Data Presentation:

After reviewing GRC-PAGEMS sample data and its presentation in form of $dN/d\log D_p$ versus D_p [(*Penko, 2003)*], the followings are relevant observations and comments:

The 'raw' data are obtained by counting 'monodisperse' particles ΔN_i separated in the DMA separator (for a given setting of the DMA rod high-voltage) using the CPC counter. This particle count, ΔN_i, corresponds to a certain particle size, $D_{p,i}$, or more precisely to a corresponding narrow range, $\Delta D_{p,i}$, around $D_{p,i}$, which in turn corresponds to the set high-voltage applied to the DMA rod, see Fig. 5. It would be more appropriate to know the corresponding, equivalent 'monodisperse' size range, $\Delta D_{p.i}$, and calculate the corresponding density distribution as:

$$n_{d,i}=\frac{\Delta N_i}{\Delta D_{p,i}}$$

(Eq. 9)

However, the actual PAGEMS data are reduced using the following correlation to calculate the *logarithmic density distribution*:

$$n_{d\_\log,i}(D_{p,i}) = \frac{(\Delta N_i + \Delta N_{i+1})/2}{(\log D_{p,i+1} - \log D_{p,i})} = \frac{(\Delta N_i + \Delta N_{i+1})}{2 \cdot \log\left(\frac{D_{p,i+1}}{D_{p,i}}\right)} \qquad \text{(Eq. 10)}$$

The corresponding 'true' *density distribution* will be:

$$n_{d,i}(D_{p,i}) = \frac{(\Delta N_i + \Delta N_{i+1})/2}{(D_{p,i+1} - D_{p,i})} \qquad \text{(Eq. 11)}$$

There are two issues related to accuracy of the PAGEMS data representation using (Eq. 10). First, if the two successive data sets are used to average number of particles between the two 'monodisperse' particle sizes, why is the average of the particle size range, $D_{p\_avg,i} = (D_{p,i+1}+D_{p,i})/2$, not used as the corresponding independent (abscissa) variable, i.e?:

$$n_{d\_\log,i}([D_{p,i+1} + D_{p,i}]/2) = \frac{(\Delta N_i + \Delta N_{i+1})/2}{(\log D_{p,i+1} - \log D_{p,i})} = \frac{(\Delta N_i + \Delta N_{i+1})}{2 \cdot \log\left(\frac{D_{p,i+1}}{D_{p,i}}\right)} \qquad \text{(Eq. 12)}$$

or,

$$n_{d,i}([D_{p,i+1} + D_{p,i}]/2) = n_{d,i}(D_{p\_avg,i}) = \frac{(\Delta N_i + \Delta N_{i+1})/2}{(D_{p,i+1} - D_{p,i})} \qquad \text{(Eq. 13)}$$

This representation (Eqs. 12 & 13) will only shift the density distribution along the abscissa from $D_{p,i}$ for $(D_{p,i+1}-D_{p,i})/2$, to $D_{p\_avg,i}$ value. More important and critical issue is whether the two successive data sets properly account for the number of particles in the $(D_{p,i+1}-D_{p,i})$ range. If the two successive data set overlap or skip the particle size range, the averaged particle count, $(\Delta N_i+\Delta N_{i+1})/2$, will be erroneous, as well as the calculated density distribution, see Fig. 5. This critical sequencing issue could be completely eliminated if the equivalent 'monodisperse' particle sizes range, $\Delta D_{p.i}$, is known and used in (Eq. 9) to determine the true particle *density distribution*.

Finally, the same PAGEMS sample data are presented as original *logarithmic density distribution* using (Eq. 10), then properly shifted (centered) using (Eq. 12), and finally as the true *density distribution* using (Eq. 11) and (Eq. 13) for proper shifting along abscissa (last two are multiplied with 50 to better scale on the plot with the first two); all four cases are presented in the same semi-logarithmic coordinate system for convenient comparison, see Fig. 6.

Based on the above analysis, it is recommended to calculate the true particle density distribution using (Eq. 9), and in the absence of the corresponding ‘monodisperse’ particle sizes range, $\Delta D_{p,i}$, to use (Eq. 13) if the two successive data sets’ average particle counts, $(\Delta N_i+\Delta N_{i+1})/2$, properly accounts for the number of particles in the $(D_{p,i+1}-D_{p,1})$ range.

Acknowledgements:

Writing of this article was motivated by personal discussions with Steve Culler and Terry Sanders, and especially by Dr. Paul Penko at NASA Glenn Research Center.

Reference:

Karl Sommer (2001), "40 Years of Presentation Particle Size Distribution - Yet Still Incorrect?" *Part. Part. Syst. Charact.* **18***, p 22-25.*

Paul Penko (2003), Personal Communications, *NASA Glenn Research Center, Cleveland, OH.*

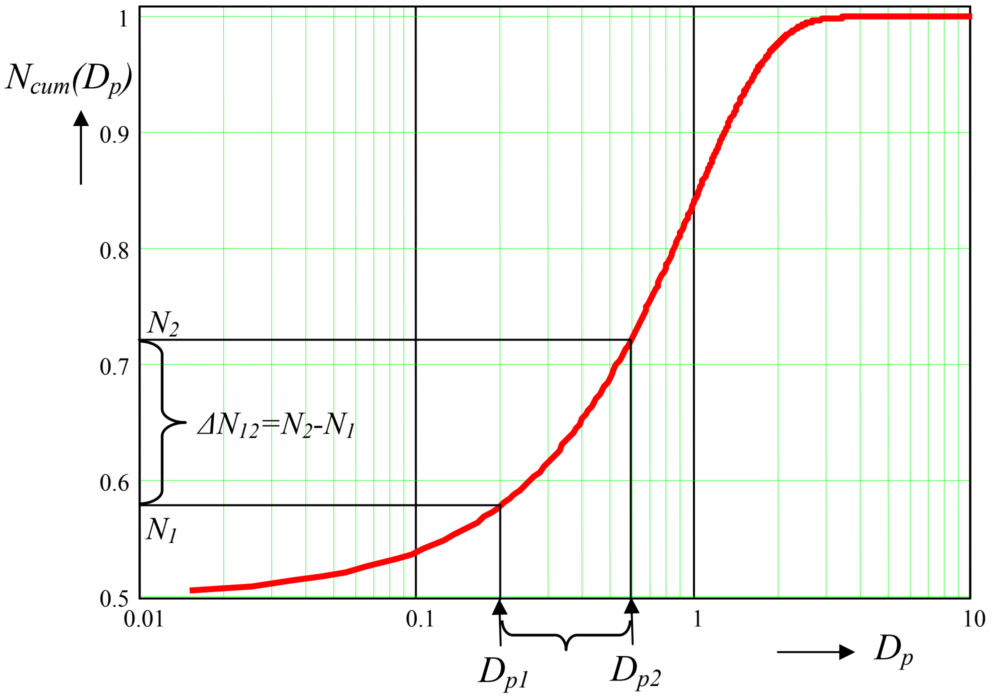

Fig. 1: Cumulative frequency distribution (for normal Gaussian distribution)

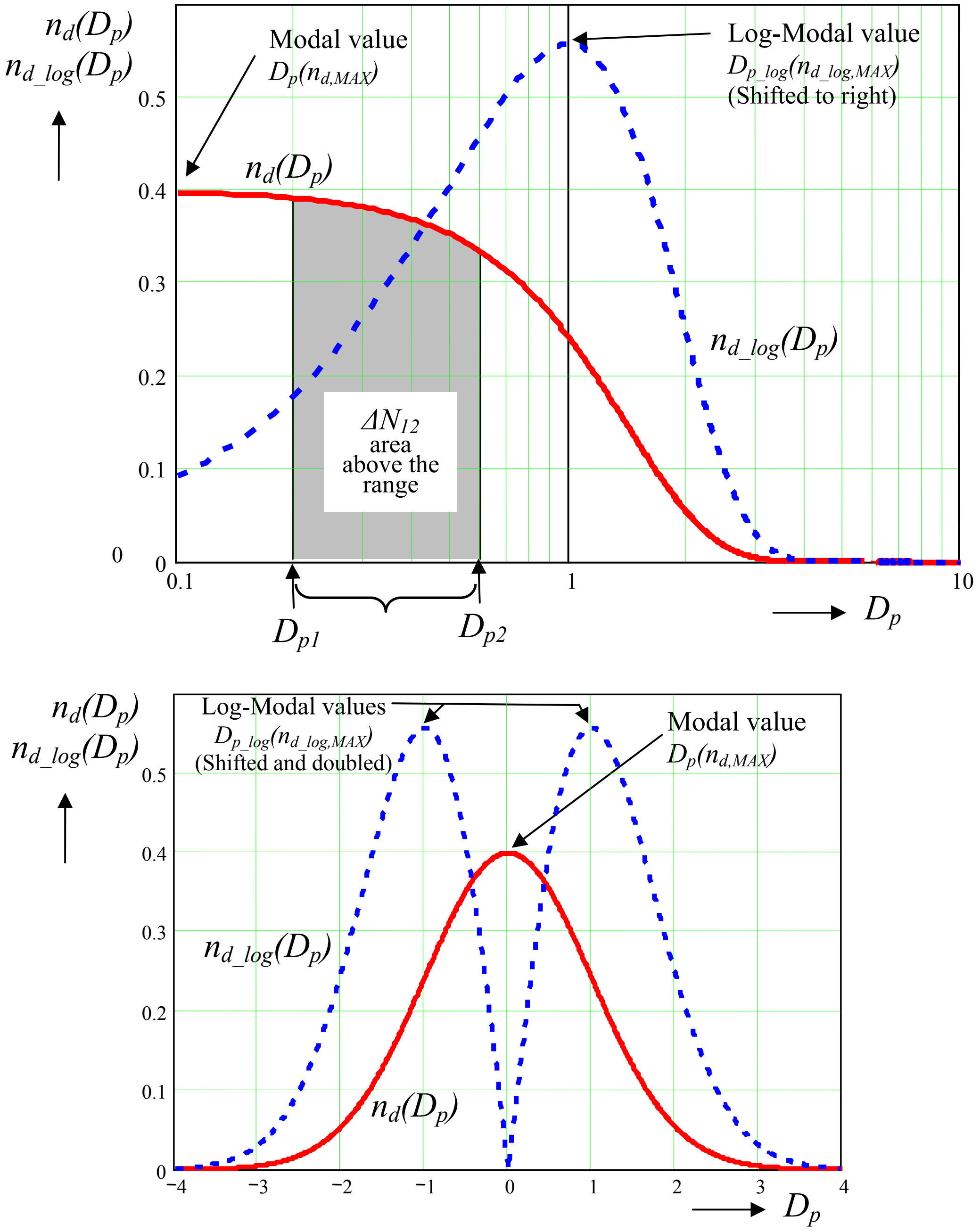

Fig. 2: Density distribution (Eq. 2) and logarithmic density distribution (Eq. 4), both for the same normal Gaussian distribution (note that log-distribution is skewed, up- and down-scaled, and zero modal-value is doubled (mirrored) and shifted to positive and negative regions, in semi-log at coordinates top and regular coordinates at bottom).

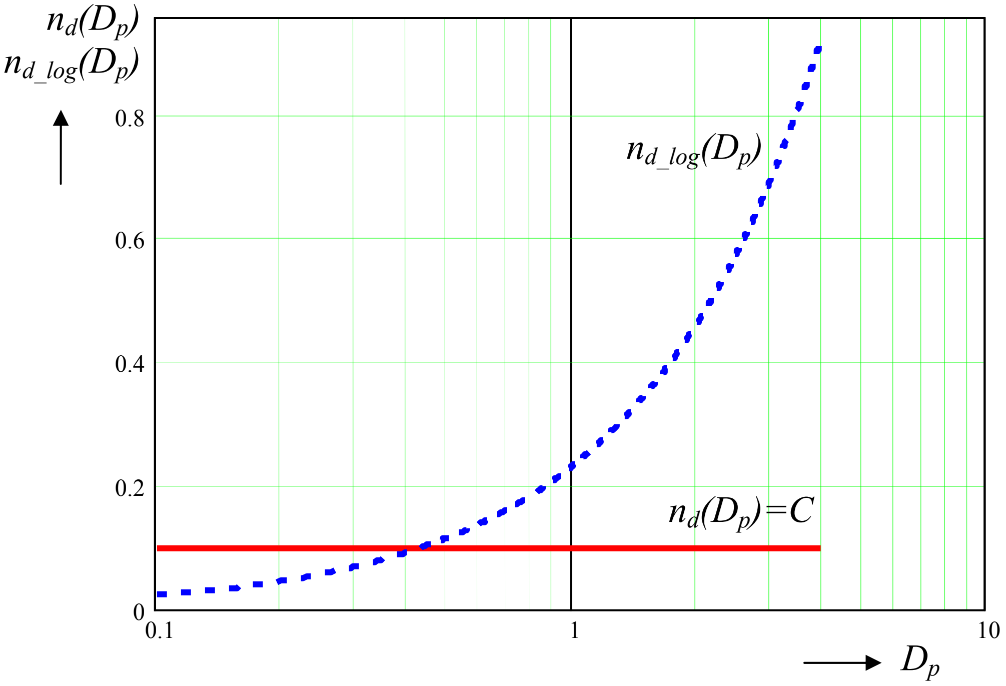

Fig. 3: Uniform density distribution is misrepresented by logarithmic density distribution as increasing function with maximum (modal point) at the end of range.

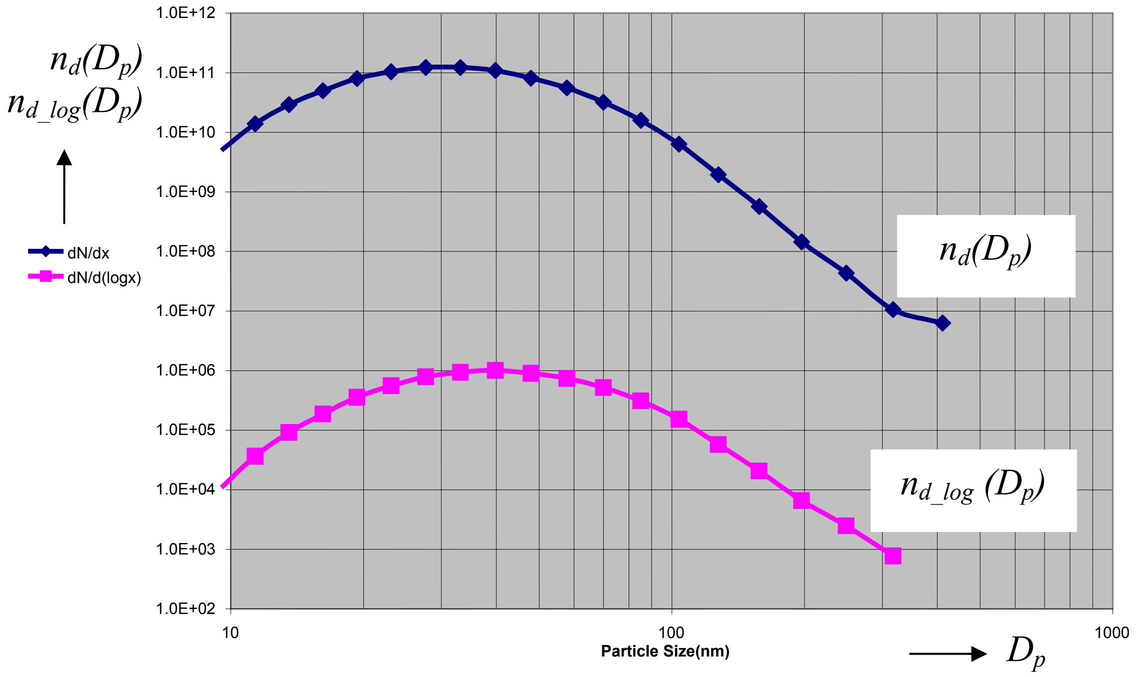

Fig. 4: True and logarithmic density distributions for a PAGEMS sample data [*Penko (2003)*] (notice skewing, scaling and shifting of modal value).

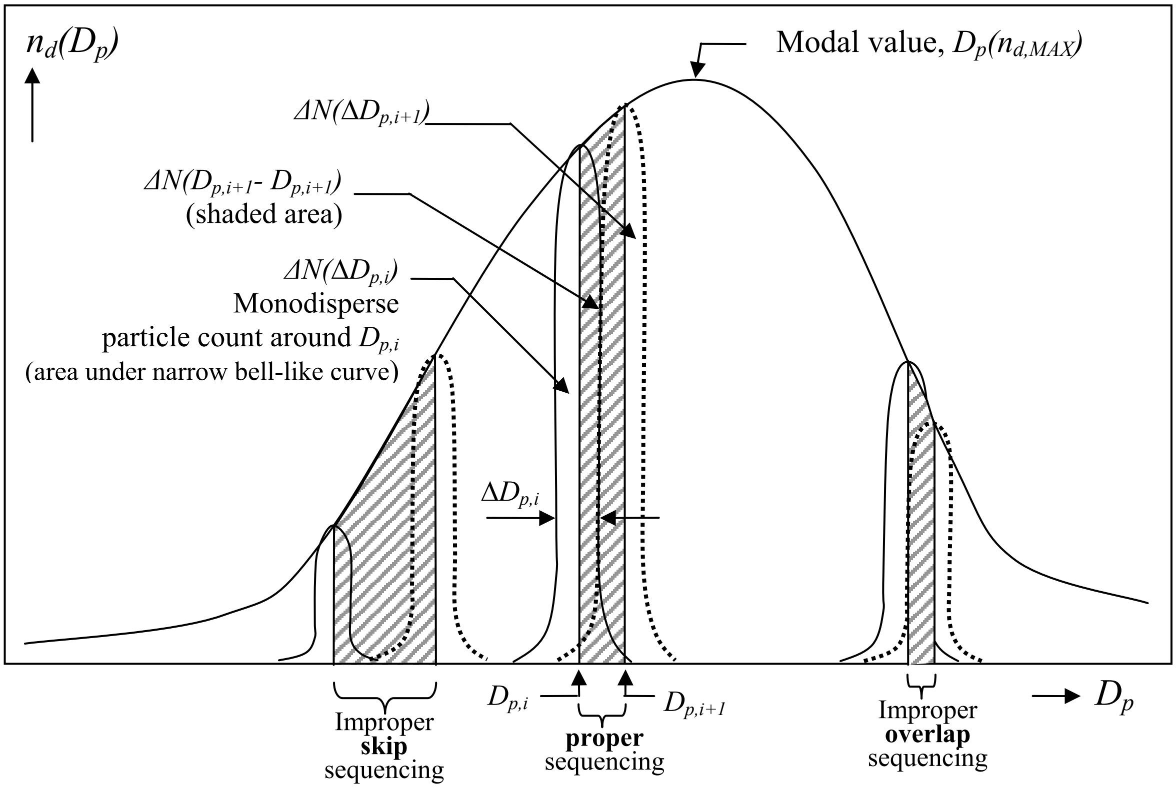

Fig. 5: Influence of proper "monodisperse" sequencing on particle density distribution accuracy

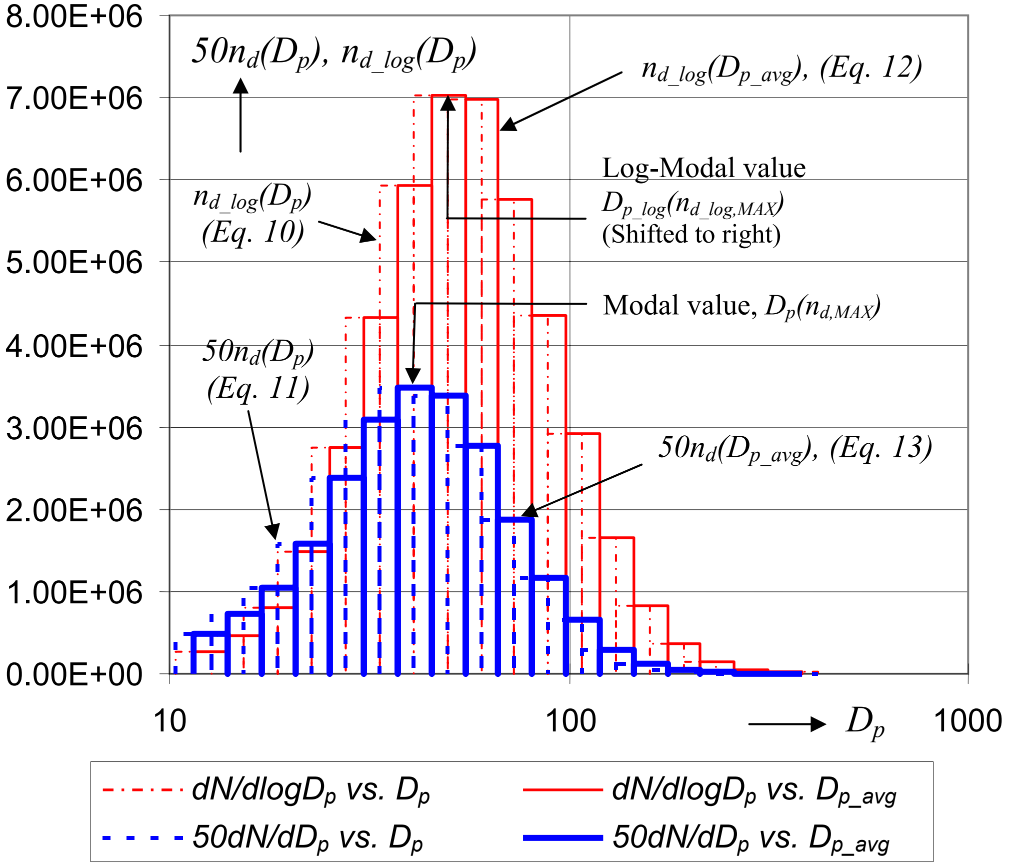

Fig. 6: True and logarithmic density distributions with and without shifting particle range (centering at D_{p_avg}) for PAGEMS sample data [*Penko* (2003)].