ISSN 1330–0016
CODEN FIZBE7

A COMPARATIVE ANALYSIS OF SEA-PARTICLE-SIZE DISTRIBUTION MODELS

DUBRAVKO RISOVIĆ AND MLADEN MARTINIS

Ruđer Bošković Institute, P. O. Box 1016, 10001 Zagreb, Croatia

Received 20 June 1995

UDC 532.582.7

PACS 82.70.Kj, 92.10.Bf

Six most frequently used sea-particle-size distribution models and their merits in describing measured size distributions are evaluated. Comparison is made on the basis of the complexity of the considered models and of quality of the fits to the measured data. The relation between the complexity and the quality of the fits is examined. Two- or three-parameter models seem to be the most adequate for fitting the experimental data. Our analyses show that there is a statistically significant difference in performance of different two-parameter models. The ability to fit simultaneously small- and large-size particle data is examined, and related to measurement errors.

1. Introduction

It is generally known that suspended particles play a significant role in light transfer in the sea water. Light-scattering properties of a given ensemble of particles are to a great extent determined by the particle-size distribution. Thus, the knowledge of the particle-size distribution is important in remote sensing, in studies of photosynthesis, light-field distribution etc. in the sea [1,2]. These applications of light scattering depend on the availability of an adequate numerical model for the particle-size distribution. To describe the measured particle-size distributions, many authors have used different models of variable complexity. To our knowledge, there has been no systematic approach to particle-size distribution modelling nor a comparative evaluation of various models. It is the intent of this paper to estimate the merits of several most frequently used particle-size distribution models, including a recently introduced one, and to compare their performance in the description of measured particle-size distributions.

2. *Particle size-distribution models*

The particle-size distribution, $F(r)$, is defined by the following relation:

$$\mathrm{d}N = F(r)dr, \tag{1}$$

where $\mathrm{d}N$ is the number of particles per unit volume with radii between r and $r+dr$. We assume that r is the radius of a sphere having an equivalent volume to that of the considered particle. This is the usual assumption used for size characterization of irregularly shaped particles justified by the measurement techniques employed, namely the Coulter counter [3].

In the literature, various models were proposed and used to describe the measured particle-size distributions. Of these, the most widely used is the hyperbolic model [4, 5, 6] or the Junge [7] distribution. Other frequently used distributions are the log-normal distribution [8, 9] and the segmented hyperbolic distributions with various number of segments [10, 11, 12] There are also other types of distributions that are not so common, such as the Weibull distribution [13], the exponential distribution [14] and a combination of the hyperbolic and the Gaussian distributions [15, 16]. The method of characteristic vectors [17] presents a purely statistical approach to the problem. In this paper, we consider the most frequently used particle-size distribution models, namely the hyperbolic, log-normal, segmented hyperbolic with two, three and four segments and a recently introduced two component model [18]. The distributions are given by the following relations:

(i) Hyperbolic (H):

$$F(r) = Cr^{-k}, \qquad 3 \le k \le 6. \tag{2}$$

(ii) Log-normal (LN):

$$F(r) = \frac{C}{r} e^{-\alpha[\ln(r/r_0)-(1/2\alpha)]^2}. \tag{3}$$

Here r_0 is the characteristic or mode radius, $\alpha \approx 0.05 - 1$ and $r_0 \approx 10^{-5} - 10^{-2}$.

(iii) Segmented hyperbolic (n -SH):

$$F(r) = \sum_{i=1}^{n} C_i r^{-k_i}, \tag{4}$$

where n is the number of segments ($n = 2, 3, 4$) and $2 \leq k_i \leq 7$.

(iv) Two-component model (TCM):

$$F(r) = C_A r^2 e^{-52 r^{\gamma_A}} + C_B r^2 e^{-17 r^{\gamma_B}}, \tag{5}$$

where γ_A and γ_B are parameters of the distribution, and C_A and C_B are constants related to the concentration of the respective components. r is expressed in dimensionless units (equivalent to one micrometer). The values of the γ-parameters are in the range $0.1 \leq \gamma_A \leq 0.2$ and $0.2 \leq \gamma_B \leq 0.3$.

By parametrization we call a procedure of finding a set of parameters that represent the best fit to the measured discrete data. This procedure is simple if we have a small number of parameters, but can become quite involved if several parameters are to be varied. Therefore, we consider the complexity of the particle size distribution model on the basis of the number of parameters included. We count only those parameters whose variation changes the shape of the distribution [19]. Therefore, multiplicative constants that merely scale the entire distribution without changing its shape are not counted as parameters of the distribution in the sense of complexity. In this notation, the hyperbolic distribution, Eq. (2), will be considered as a one-parameter distribution, with the polydispersity parameter k. Log-normal, Eq. (3), two-component, Eq. (5), and two-segment hyperbolic, Eq. (4), are two-parameter distributions. According to this classification the distributions given by Eqs. (2)-(5) have the complexity range that is between one and four.

3. Comparison of the models: results and discussion

To analyze and to compare the particle-size distribution models, we have used the results of 53 measurements carried out at various locations in the Equatorial and North Pacific [13, 20, 21, 22], the North Atlantic [21, 23], the Gulf of Mexico [10] and the northwestern Mediterranean [21]. The samples' depths ranged from the surface to 3600m. The water types covered are equatorial, subtropical and temperate waters. Measurements were performed using the Coulter counter and, in the case of samples from the Gulf of Mexico, with the aid of electron microscopy. The particle-size range covered by these measurements extends from $r = 0.2$ to $r = 40.3$ μm.

Each of the afore mentioned models, Eqs. (2) – (5), was used to fit the measured data for each sample. Then, the quality of the best fit was compared considering the complexity of the applied model, the mean error per size interval (Coulter counter channel), the absolute error of the total particle number and the mean errors in the fit of small- and large-size fractions of the measured distributions (first three channels and last three channels, respectively).

For a clear and simple comparison between various models, one needs some Q-numbers that will express the performance of a certain model in the description of a given measured particle-size distribution. As a Q-numbers describing the quality of the best fit obtained by a certain model we have chosen the Q-numbers [18], defined as the arithmetic mean of the absolute percentage error of the total number of particles, $|\Delta N_{TOT}|$, and the average percentage error per size interval, $\overline{\Delta N}$. The latter is obtained as the arithmetic mean of the absolute percentage errors for each measured size interval of the considered distribution:

$$Q = \frac{\overline{\Delta N} + |\Delta N_{TOT}|}{2}, \tag{6}$$

$$\overline{\Delta N} = \frac{1}{n} \sum_{i=1}^{n} |\Delta N_i|, \tag{7}$$

where ΔN_i is the percentage error for the i-th size interval and n is the total number of size intervals. Since the particle-size distributions are usually very steep, most of the particles are contained in a few first (smallest) size intervals. It should, therefore, be possible to obtain a very good fit in large-size intervals and a poor fit in a few small size intervals resulting in a relatively small overall average error per size interval, but producing a larger error in the total particle number. However, the correct total number of particles is important in many applications and so is the total number of small particles, since small particles dominate in light scattering (except at small angles in the forward direction). To take into account these effects we have introduced the Q-numbers as defined by Eq. (6).

According to the definition (6), a smaller Q-number means a better approximation. Since the percentage error of the total number of particles is usually much smaller than the average percentage error per size interval, the Q-number, roughly represents the half of the mean percentage error per size interval. Data fits for which Q-numbers are smaller than 4 are considered as an excellent fit, those with $4 \leq Q \leq 7$ as a very good fit and $Q \geq 13$ as a poor fit.

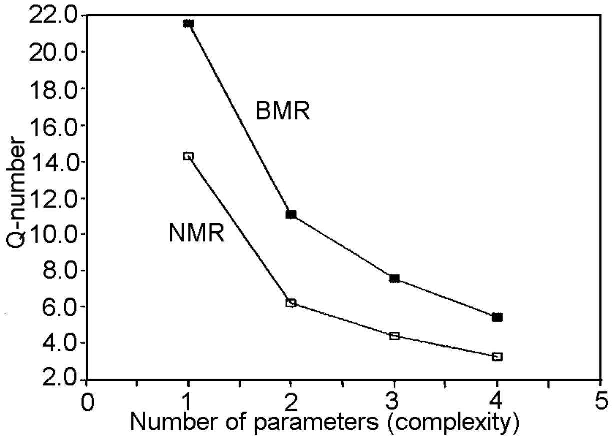

Fig. 1. Influence of the hyperbolic model complexity on the average Q-number for a broad ($r = 0.2 - 40$ μm) and narrow ($r = 1 - 5.5$ μm) measurement range.

The influence of the model's complexity on the quality of the approximation was examined comparing the results obtained using single hyperbolic and segmented hyperbolic distributions with 2, 3 and 4 segments representing complexity levels $1-4$. These models were used to fit 30 samples with identical size ranges of the measurements (Coulter counter's channels) covering a broad particle size range with $r = 0.2$ to 40.3 μm. The average Q-numbers (obtained as the arithmetic mean of the figures of merit for 30 samples) vs. the complexity of the applied model is shown Fig. 1 (curve labeled BMR). The same was done with 16 samples that covered a narrow particle-size range with the corresponding equivalent particle radii ranging from 1.1 to 5.5 μm. The result is the curve labeled NMR in Fig. 1. One notes that in both cases the quality of the approximation increases proportionally to $1/p$, where p denotes the complexity level (i.e. number of parameters) of the applied model.

As expected, the performance is better for measurements that cover a narrower size range than for broad size range distributions, regardless of the complexity of the applied model (Table 1). Therefore, if a measurement covers a broad particle-size range, a model of higher complexity should be used.

TABLE 1.
Average Q-numbers for various models corresponding to a broad (BMR) and narrow (NMR) measured size range. Smaller Q means better fit. s is the standard deviation.

MODEL	H	LN	TCM	2-SH	3-SH	4-SH
Complexity	1	2	2	2	3	4
Q(BMR)	21.6	13.8	9.4	11.2	7.6	5.3
s(BMR)	7.6	7.9	3.1	4.4	2.5	2.0
Q(NMR)	14.3	8.1	6.4	6.3	4.5	3.4
s(NMR)	11.1	1.7	2.2	2.5	1.2	1.0

The performance of the six considered models for a broad measurement range is summarized in Fig. 2. Here the mean figure of merit for each model is given vs. complexity. It can be seen that the best result is obtained with a model of highest complexity, but considering the rather complicated and time-consuming parameterization for such a model, it is not reasonable to use a model whose complexity exceeds level 3. It should also be noted that there is a rather large difference between 2-parameter models. It seems that two-component model has the best performance. The mean Q-value for the two-component model is 9.4 (standard deviation $s = 3.1$) whereas the Q-values for the other 2-parameter models considered, namely the log-normal and the 2-segment hyperbolic models are 13.8 ($s = 7.9$) and 11.2 ($s = 4.4$), respectively.

Statistical testing of these means with the paired t-test [25] showed that the difference between the TCM mean and the other two means is statistically significant at the 95% confidence level, therefore, conclude that the TCM model gives a better overall description of experimental data than other two-parameter models.

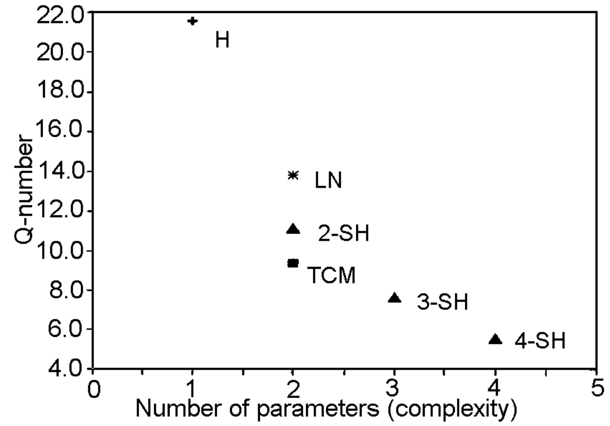

Fig. 2. Average Q-numbers for PSD models considered.

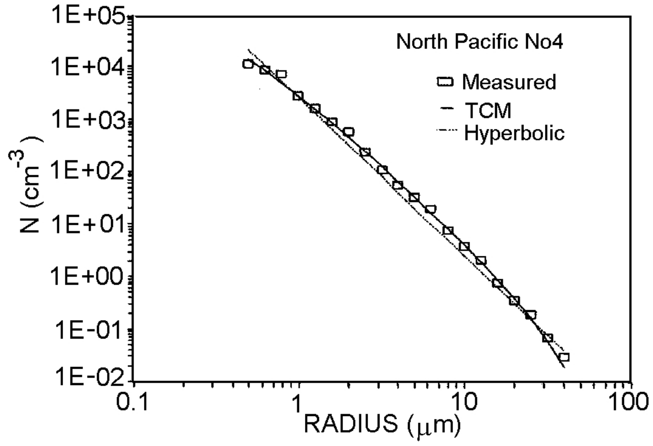

Fig. 3. A typical fit of data of a broad size range with TCM and hyperbolic PSD. Respective Q-numbers are 5.8 and 29.0.

A typical fit of the experimental data [21, 22] with TCM and hyperbolic distributions is shown in Fig. 3. The corresponding distribution of errors along the Coulter-counter channels is shown in Fig. 4. Q-number values are 5.8, 29.0, 9.84 and 4.3 for TCM, hyperbolic, two-segment hyperbolic and three-segment distributions, respectively.

Particle-size distributions are usually measured in a narrower size range than it is required for many applications. One must, therefore, extrapolate to smaller and larger sizes than measured. Typical examples are applications involving light scattering where particles of dimensions smaller than 1 μm play a significant role, but are seldom measured. The same is true of the large-size side of the particle-size distribution (PSD). To investigate the behavior of PSD models in small- and large-size ranges which are important for numerical extrapolation, we have calculated the mean percentage error in the fit of three smallest and three largest measure size ranges (channels). For this comparison, we have used only those distributions that cover a broad size range. The results are shown in Table 2.

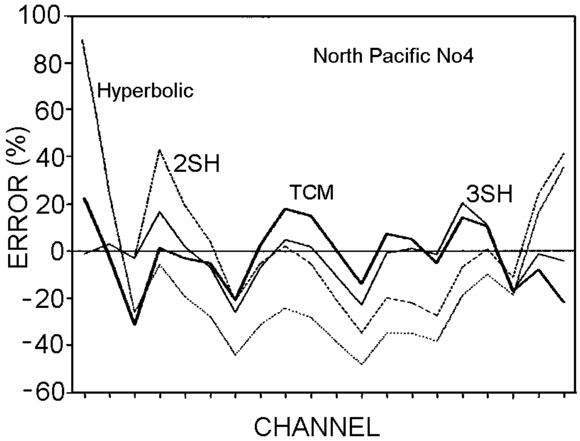

Fig. 4. Typical error distributions for TCM and hyperbolic and segmented hyperbolic models (data from Fig. 3). Channels cover size range 0.5 to 40.3 μm.

TABLE 2.
Average percentage errors per size interval for various models corresponding to measured small- (SSR) and large-size (LSR). The average was taken over three smallest (largest) channels. SSR: 0.2 μm $\leq r \leq$ 2.8 μm; *LSR:* 8 μm $\leq r \leq$ 40 μm.

MODEL	H	LN	TCM	2-SH	3-SH	4-SH
Complexity	1	2	2	2	3	4
$\lvert\Delta N\rvert$(SSR)	24.3	24.9	19.5	18.7	13.2	9.4
s(SSR)	11.8	19.8	10.8	9.3	8.8	6.6
$\lvert\Delta N\rvert$(LSR)	95.9	36.2	30.8	45.4	27.6	19.6
s(LSR)	93.2	37.2	42.4	45.1	22.7	16.8

We can see that almost all models considered produce significantly smaller errors in small-size ranges than in large-size ranges, and that standard deviations are much larger for large-size ranges. This seems to be in agreement with evidence, showing that the size measurement error is larger for large particles than for small particles, and increases with particle size. This is attributed to a much smaller total number of large particles and to the influence of their shape and the way they cross the orifice of the Coulter counter on the measurement result [26, 27]. The paired t-test seems to partially support this conclusion. It turns out that these differences for n-segmented hyperbolic distributions ($n = 1 - 3$) are statistically significant at the 95 % confidence level. However, for the TCM and the log-normal model, these differences are statistically significant at only the 70 % confidence level. This ambiguity, therefore, hinders any decisive conclusions regarding the cause of this difference between small- and large-size errors (at least with the investigated sample size).

Comparison between performances of various model gives a similar result as before: the 4-segment hyperbolic model giving the best result and the TCM model being better than the other 2-parameter models. However, the tests show that there is no statistically significant difference at the 95 % confidence level between the small-size mean error values of the TCM and 2-segment hyperbolic distributions. It should also be noted that the TCM model gives a significantly lower mean error in large-size ranges than the 2-segment hiperbolic distribution. Namely, the mean error for the last three channels in the large-size range for the TCM is 30.84 % ($s = 42.4$) and for the 2-segment hyperbolic 45.4 % ($s = 45.2$). The paired t-test shows that this difference is statistically significant only at the 75 % confidence level. There is no statistically significant difference between the TCM and 3-segment hyperbolic distribution.

4. *Conclusions*

The evaluation of six particle size distribution models of various complexity applied to fit the 53 measured distributions from various sites has shown the following:

a) Models of greater complexity (expressed by the number of parameters involved) do indeed produce a better fit than simpler models. The Q-numbers decreases almost inversely with the complexity of the model; (smaller Q means a better fit). However, since the amount of the numerical work needed to fit the data increases sharply with the complexity of the model involved, one should prefer less complicated models. It is our opinion that optimum balance between the work involved and the results obtained is reached when the 2-parameter model or at most a 3-parameter model is used. Of course, if we consider a narrow range of measured sizes, even a one-parameter model (such as simple hyperbolic distribution) could produce satisfactory results. The influence of joints in segmented hyperbolic models, which is an inherent drawback of these models, has not be investigated, though one may expect that it would contribute to the error in pertinent channels.

b) Considering the 2-parameter models, we have found that, on the average, the two-component model (TCM) gives a significantly better fit (at the 95 % confidence level) than the 2-segment hyperbolic distribution and a much better fit than the log-normal distribution.

c) All models considered give on the average a better fit for small-size than for large size ranges. This could be attributed to larger size measurement errors connected with large particles, but we did not find sufficient statistical evidence to entirely confirm this hypothesis.

d) For small sizes, we have not found a statistically significant difference between the 2-segment hyperbolic and the TCM models whereas the log-normal model has a significantly worse performance.

e) For large-size ranges, we have found that the TCM performs significantly better than other two-parameter models. Therefore, one may expect that owing to a potentially better large-size range extrapolation (which is necessary for light-scattering calculations), the TCM should (at least theoretically) give better results in computing light scattering at small angles, which is determined by very large particles. Comparison with the experimental results has confirmed this expectation [24].

References

1) N. G. Jerlov, "Optical Oceanography", Elsevier Publishing Co., New York, 1968, p. 194;

2) Ocean Optics VIII, SPIE vol. 637 (1986);

3) R. W. Sheldon and T. R. Parsons, *A Practical Manual on the Use of Coulter Counter in Marine Science*, Coulter Electronics, Toronto 1967, p. 66;

4) H. Bader, J. Geophys. Research **75** (1974) 2822;

5) R. Reuter, Oceanologica Acta **3** (1980) 317;

6) D. E. Campbell and R. W. Spinrad; Estuarine, Coastal and Shelf Science **23** (1987) 53;

7) C. E. Junge, *Air Chemistry and Radioactivity*, Academic Press, New York, 1963, p. 382;

8) H. J. Pak, R. V. Zaneveld and G. F. Beardsley, Jr., J. Geophys. Research **76** (1971) 5065;

9) S. L. Oshchepkov and E. A. Sorokina (in Russian), Izv. Ac. Sci. USSR, Atmospheric and Oceanic Physics **25** (1989) 870;

10) J. E. Harris, Deep-Sea Research **24** (1977) 1055;

11) O. B. Brown and H. R. Gordon, Applied Optics **13** (1974) 2874;

12) J. C. Kitchen, and J. R. V. Zaneveld; J. Geophys. Research **95** (1990) 20237;

13) K. L. Carder, G. F. Beardsley, Jr. and H. Pak, J. Geophys. Research **76** (1971) 5070;

14) J. R. Zaneveld, and H. Pak; J. Optical Soc. of America **63** (1973) 321;

15) M. Jonasz, Tellus **35B** (1983) 345;

16) M. Jonasz and H. Prandke, Tellus **38B** (1986) 144;

17) J. C. Kitchen, D. Menzies, H. Pak and J. R. V. Zaneveld, Limnology and Oceanography **20** (1975) 775;

18) D. Risović, Deep-Sea Research **40**, No.7 (1993) 1459;

19) M. G. Kendall and A. Stuart, *The advanced theory of statistics, Vol. 1: Distribution theory*, Charles Griffin and Co. Ltd., London, 1963, p. 432;

20) K. S. Shifrin et al. (in Russian). Izv. Ac. Sci. USSR; Atmospheric and Oceanic Physics **10** (1974) 25;

21) R. W. Sheldon, private communication (1990);

22) R. W. Sheldon, A. Praksh and W. H. Sutcliffe, Jr., Limnology and Oceanography **17** (1972) 327;

23) V. A. Del Grosso, SPIE Vol. 160, Ocean Optics V (1978) 34;

24) D. Risović, PhD Thesis, University of Zagreb, Zagreb 1995;

25) M. Kilpatrick, *Business Statistics using Lotus 1-2-3*, J. Wyley & Sons, Inc. New York, 1987;

26) D. C. Golibersuch, Biophys. J. **13** (1973) 265;

27) V. H. Kachel et al. Z. f. Ges. Experim. Medizin, Exp. Chir. **153** (1970) 331.

USPOREDBA MODELA RASPODJELE VELIČINE ČESTICA U MORU

DUBRAVKO RISOVIĆ AND MLADEN MARTINIS

Institut Ruđer Bošković, P. P. 1016, 10001 Zagreb

UDK 532.582.7

PACS 82.10.Kj, 92.10.Bf

Analizirano je šest najčešće primijenjenih modela veličinske raspodjele čestica u morskoj vodi i uspoređene su njihove značajke u opisu izmjerenih raspodjela veličine čestica. Modeli su uspoređeni s obzirom na njihovu složenost i slaganje s eksperimentalnim podacima. Istražena je ovisnost složenosti modela i uspješnosti u opisivanju izmjerenih podataka. Pokazuje se da su dvo- i tro-parametarski modeli najadekvatniji za opis eksperimentalnih podataka. Analiza je pokazala da postoje statistički značajne razlike u učinbi dvoparametarskih modela. Istražena je mogućnost istovremene uskladbe s mjernim podacima za male i velike veličine čestica i ukazano je na njihovu povezanost s mjernim pogreškama.