Size distributions of *Scenedesmus obliquus* cells: experimental results from optical microscopy and their approximations using the φ-normal distribution

OCEANOLOGIA, 40 (4), 1998.
pp. 345–353.

KEYWORDS
φ-normal distribution
Algal cells

Barbara Pawlak,
Józef Kopeć
Marine Physics Department,
University of Szczecin,
Wielkopolska 15, 70–451 Szczecin, Poland; e-mail: pawlak@sus.univ.szczecin.pl

Manuscript received August 12, 1998, in final form December 14, 1998.

Abstract

3400 measurements of algal cell size using the microscopic technique are discussed. The algal population is observed to evolve. The size distribution is well approximated by the φ-normal distribution. There is variability in the form of the size distribution.

1. Introduction

Let $F(x)$ be the distribution function of the random variable X describing the size parameter in the investigated population of suspensions, and $N(x) = N_0\,[1 - F(x)]$ the dispersion distribution. N_0 is the number of cells in unit volume of suspension. The empirical distribution function $F_n(x)$ and the number N_0 can be determined independently (Bricaud and Morel, 1986). The observed distribution of suspensions is an instantaneous distribution. Strong turbulent motion (mostly in the case of emulsions and mineral suspensions) or growth processes (in cells) are responsible for the rapid variability in dispersion distribution. The variability of distribution type and of its parameters with respect to time, and the variability of the physical properties of the suspension (depending on the distribution) are not well understood. This is probably due to experimental difficulties and the lack of simple methods of constructing functions that adequately approximate empirical distributions, especially for densities with larger numbers of extreme values. Even though a number of papers on this topic have appeared – an extensive list of references is given in Jonasz and Fournier (1996) – the known theoretical functions are, in our opinion, unsuitable for analysing temporal changes in the dispersion distribution. We have recently attempted such an analysis of laboratory emulsions and suspensions in natural waters of low salinity by making use of the variations in optical properties of suspensions (Kopeć and Pawlak, in press). In the present work we took direct microscopic measurements of the size parameter of algal cells. Our observations have shown that four out of five distribution functions examined possess more than one inflection point, which precludes the application of many of the theoretical distribution functions commonly used for approximations. We therefore utilised φ-normal distributions, which differ from the normal distribution in that they contain a correction term. The coefficients of this are determined using an optimisation method, the concept of which is taken from the paper by Jonasz and Fournier (1996).

2. φ-normal distributions

Let φ be a differentiable function increasing strictly within the interval (a, ∞), such that $\lim\limits_{x\to\infty} \varphi(x) = \infty$, $\lim\limits_{x\to a+} \varphi(x) = -\infty$. The distribution of random variable X is φ-normal, since the distribution of the variable $Y = \varphi(X)$ is normal, *i.e.* its distribution function takes the form $F(y) = P(Y < y) = \Phi\left(\frac{y-m}{\sigma}\right)$. As a result of strict monotonicity, we have $P(X < x) = P[\varphi(X) < \varphi(x)]$, hence for the distribution function and the density of the X variable for $u = \frac{\varphi(x)-m}{\sigma}$, we obtain

$$\begin{aligned} F(x) &= P\left(Y < \varphi(x)\right) = \Phi\left[\frac{\varphi(x)-m}{\sigma}\right] = \Phi(u), \\ G(x) &= \Phi'(u)\frac{\varphi'(x)}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\,u^2\right)\varphi'(x). \end{aligned} \tag{1}$$

Obviously, when $\varphi(x) = x$, distribution (1) becomes a normal one, and when $\varphi(x) = \ln x$ it converts into a log-normal one. Since φ does not depend on either m or σ, the parameters can be estimated in the usual way using the maximum likelihood method. In particular, if x_1, ..., x_r are the centres of class intervals in a separating series and k_1, ..., k_r are the corresponding numbers of cells in each class, the estimators of parameters m and σ are

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{r} k_i\,\varphi(x_i), \qquad \hat{\sigma}^2 = \sum_{i=1}^{r} k_i\,[\varphi(x_i) - \hat{m}]^2, \qquad N = \sum_{i=1}^{r} k_i. \tag{2}$$

By differentiating formulas (1) we can readily infer that the inflection points of the distribution function F satisfy the equation

$$\sigma^2\varphi''(x) - [\varphi(x) - m]\varphi'^2(x) = 0. \tag{3}$$

Hence, every inflection point of function φ is an inflection point of the distribution function F if $\varphi'(x) = 0$ or $\varphi(x) = m$.

Pawlak (1998) demonstrated that if the empirical distribution functions deviate significantly from the normal distribution functions only within some small interval of the argument, a good approximation of empirical distributions is then given by the φ-normal distribution, where $\varphi(x) = X - R(x)$. As $R(x)$ is a correction term to the normal distribution, it should be limited, disappear at $\pm\infty$, ensure the monotonicity of the φ function ($R'(x) \leq 1$) and the existence of inflection points. Furthermore, it should deviate from x (generating a normal distribution) to the highest degree within the measurement interval. From the mathematical point of view the rational function

$$R(x) = A\,\frac{x-B}{(x-B)^2+C}\,, \qquad C > 0, \tag{4}$$

can be regarded as the simplest function of this kind.

In order to satisfy the above postulates, parameters A, B and C should fulfil additional requirements. This is readily accomplished for the function

$$\varphi(x) = x - \frac{\sqrt{C}}{E}\,\frac{z+p}{z^2+1}, \qquad z = \frac{x-B}{\sqrt{C}}, \tag{5}$$

since then

$$\varphi'(x) = 1 - \frac{E(z)}{E}, \qquad E(z) = \frac{1-2pz-z^2}{(1+z^2)^2}, \tag{6}$$

and the extreme value of $E(z)$ is easily determined from the equation

$$z^3 + 3pz^2 - 3z - p = 0. \tag{7}$$

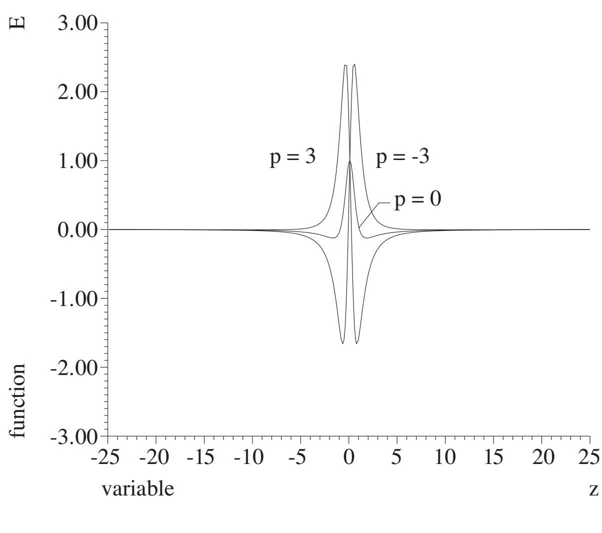

Fig. 1. Diagrams of the function $E(z)$ for different values of the parameter p

This allows the value of E to be established, for which the conditions for monotonicity and the existence of inflection points are fulfilled. Plots of the function $E(z)$ for several values of p are presented in Fig. 1.

3. Method and results of measurements

A population of the alga *Scenedesmus obliquus*, prepared in distilled water and stored for five days, was subjected to investigation. 3400 measurements of cell sizes were made using the microscopic technique described in a number of papers, *e.g.* Gurgul, 1991; Dera, 1992; Gurgul *et al.*, 1992. 5 series of measurements corresponding to the consecutive days of storage were obtained. The measurements of cell diameters were accurate to within 1 μm. The results were grouped in separating series, the width of each class being equal to 2 μm. Thus, 8 size classes were obtained with centres at points 5, 7, 9, ..., 19. The number of cells was appended to each class. The number of cells in unit volume was obtained by measuring their number in the standard volume $v_0 = 0.45$ mm^3 on a Bürcker table. A separating series, an empirical distribution function and the number of cells in volume v_0 were found each day.

Table 1. Separating series of cells sizes in *S. obliquus* algae

Day	No.	1	2	3	4	5	6	7	8	
	Class	$\langle 4,6\rangle$	$\langle 6,8\rangle$	$\langle 8,10\rangle$	$\langle 10,12\rangle$	$\langle 12,14\rangle$	$\langle 14,16\rangle$	$\langle 16,18\rangle$	$\langle 18,20\rangle$	Totals
	Centre of class interval	5	7	9	11	13	15	17	19	
1		48	86	250	237	40	45	18	14	738
2	number	35	103	170	176	60	42	23	1	610
3	of cells	32	107	190	211	63	31	8	0	642
4		43	132	266	219	84	40	1	0	785
5		41	136	171	188	58	28	3	0	625

The results are set out in Tab. 1. The agreement between the empirical distributions and the normal distribution was examined using the χ^2 test. The values of these statistics recorded in the samples are given in Tab. 2. The line below these gives the critical significance levels – the values for which the critical values of the χ^2 statistics are equal to the recorded ones. Since these are markedly lower than the usual significance levels (0.05–0.02), the hypothesis about the agreement with the normal distribution must be rejected. This makes it impossible to use tests based on this distribution

Table 2. The parameters of theoretical distributions. The last line describes the maximum deviation of function generating distributions from a normal distribution

Distributions		1	2	3	4	5
Parameters	P	1	1.5	–1	3	–0.5
	B	10.76029	14.80349	13.20129	13.30937	12.79585
	C	4.289131	2.690551	1.05509	0.631764	0.363530
χ^2		3.02562	1.255	1.77	1.805	7.363
critical significance level		0.80562	0.974	0.94	0.937	0.289
normal		177 1.4×10^{-35}	40 4.31×10^{-7}	20.8 0.002	16.6 0.01	17.0 0.009
$\max \mid \varphi(x) - x \mid$		2.941	1.77	0.973	1.01	0.452

(test for differences, variance analysis) to examine the variability of mean values. For the same reasons the hypothesis about the agreement with the log-normal distribution must also be rejected for all distributions, since the critical significance levels for this distribution are very much lower than for the normal one ($< 4 \times 10^{-6}$). In order to examine agreements with φ-normal distributions, we determined coefficients B and C for selected values of p by minimising the χ^2 statistics. This was done by calculating χ^2 for parameters $B = B_0 + r \cos\left(\frac{2\pi i}{k}\right)$, $C = C_0 + r \sin\left(\frac{2\pi i}{k}\right)$, $i = 1, 2, ..., k$ with k chosen arbitrarily. That point in the circle in which χ^2 reaches a minimum becomes the centre of a new circle. The value of r decreases automatically during every iteration up to some fixed minimum value, the finding of which is the condition for halting the procedure. At each iteration, m and σ were calculated using the maximum likelihood method. All the distributions obtained display agreement with 6 degrees of freedom, with the empirical distributions at a significance level of at least 0.25. Tab. 2 gives the values of constants, χ^2 statistics and the significance levels for optimum distributions. The values of χ^2 and the critical significance levels for the normal distribution are appended for comparison. Apart from the first-day distribution, optimum values of constants were obtained for $E = \max E(z)$. For that first distribution E was made equal to $1.8 \min E(z)$, because it was the only distribution for which there was no inflection point with a vanishing derivative. This is illustrated in Fig. 2. For a better picture of the accuracy of the approximation, the empirical and theoretical counts in the relevant class intervals for the distributions from the second, third and fourth day of observations are given in Tab. 3. For the other distributions, the approximations are not quite so good.

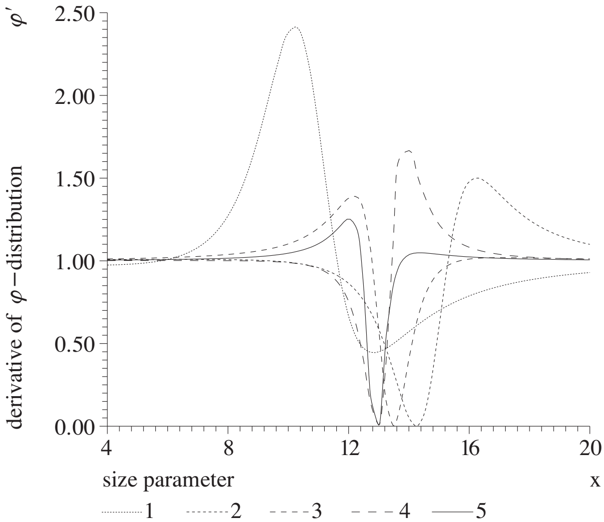

Fig. 2. Derivatives of the function $\varphi(x)$. The numbers correspond to the consecutive days of observation

Table 3. Empirical (LE) and theoretical (LT) numbers of cells in a class for distributions 2, 3, and 4

Distributions		Class intervals								Totals
		$\langle 4,6\rangle$	$\langle 6,8\rangle$	$\langle 8,10\rangle$	$\langle 10,12\rangle$	$\langle 12,14\rangle$	$\langle 14,16\rangle$	$\langle 16,18\rangle$	$\langle 18,20\rangle$	
2	LE	35	103	170	176	60	42	23	1	610
	LT	35	100	178	167	63	43	23	1	
3	LE	32	107	190	211	63	31	8	0	642
	LT	35	101	192	208	64	33	8	1	
4	LE	43	132	266	219	84	40	1	0	785
	LT	42	142	259	215	85	40	2	0	

4. Evolution of the population and distributions

Tab. 4 shows that the population evolved during the period of observation. It also shows that on different days there were substantial differences in the proportions of large and small cells in the population. The proportion of large cells falls from 6 to 0.5%, while that of small ones rises from 22 to 29%. The mean cell size decreases on consecutive days from 10.2 to 9.58 μm. The number of cells in unit volume does not vary in such a uniform manner; the maximum was reached on the fourth day of observation. The standard deviation reached a maximum on the second day and a minimum on the fourth. The usual test for mean differences shows that they are significant, *e.g.* between the first and second, and the fifth mean values at the 0.0014 level. The differences between the first and second, and the fourth mean values, and between the third and fifth mean values differ from zero at a level < 0.01. The differences between the others cannot be regarded as significant. For significance levels < 0.01 the differences between mean values can be regarded as significant, despite deviations of the relevant distributions from the normal. The distributions therefore do differ significantly. Clearly, agreement of distributions does not necessarily follow from the non-significance of differences in mean values.

Table 4. Mean values and proportion of cells of different sizes to the overall number of observations

Day	Number of cells in unit volume	Mean value m	Standard deviation σ	Small < 9 [μm]	Medium 10–15 [μm]	Large > 15 [μm]
				% numbers of observations		
1	350	10.11653	2.839896	22	72	6
2	390	10.13443	2.844315	26	68	6
3	346	9.906542	2.462299	25	73	2
4	448	9.746497	2.39494	25.5	74	0.5
5	342	9.5824	2.487732	29	70	1

For the ultimate verification of the hypothesis about the significance of differences between all distributions, every empirical distribution was compared with the optimum φ-normal distribution for the other four distributions. In all cases, the χ^2 statistics were > 20, which, with 6 degrees of freedom, yields a critical significance level of < 0.0028. It therefore follows that the optimum φ-normal distribution is a good approximation of the relevant empirical distribution, but does not display agreement with any other empirical distribution. We thus conclude that the hypothesis about the agreement between empirical distributions on different days must be rejected for all distributions.

More about the relation between the empirical distributions and the normal distribution can be said on the basis of Fig. 3 and from the data in the last line of Tab. 2. That contains the maximum differences between the φ functions generating the φ-normal distributions used here and the function $y = x$ generating the normal distribution. The table shows that the absolute value of this deviation decreases from 2.9 to 0.45. However, it is evident from Fig. 3 that generating functions oscillate around the

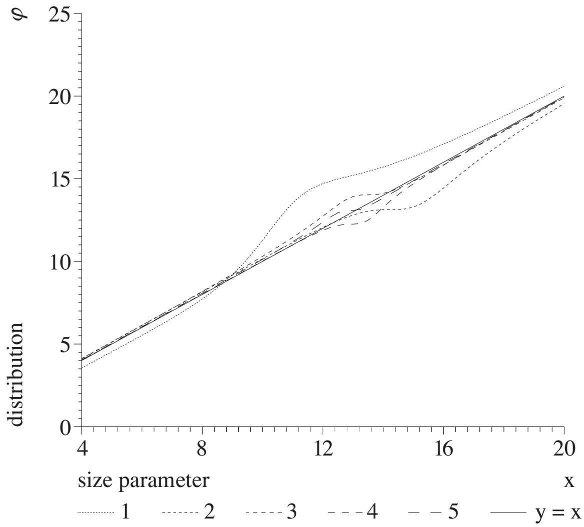

Fig. 3. Functions $\varphi(x)$ generating theoretical distributions ($y = x$ generates a normal distribution). The numbers correspond to the consecutive days of observation

function generating the normal distribution with decreasing amplitude. The hypothesis that a normal distribution is a boundary distribution in this case becomes reliable. However, the distributions recorded at the onset of the observation period differ from it significantly.

5. Conclusions

- A population of *S. obliquus* algae stored for 5 days in distilled water at room temperature was subject to significant changes. The mean value of the size parameter decreased from 10.2 to 9.58 μm. This difference was significant at the 0.0014 level. The proportion of large cells fell from 6 to 0.5%, while that of the small ones rose from 22 to 29%.
- Except in one case, the size parameter distributions were very well approximated by φ-normal distributions with generating function coefficients obtained by minimising the χ^2 statistics. From the statistical point of view no distribution could be regarded as a normal one.
- The distributions from the consecutive days differed substantially from one another with respect to the generating function parameters and were statistically significant.

- The value of χ^2 for approximating the distributions on consecutive days of observations by a normal distribution decreased from 177 to 17, and the critical significance level rose to 0.01. The maximum deviation of the functions generating distributions on the consecutive days of observations from the function generating a normal distribution decreased from 2.9 to 0.45. In our opinion this justifies the hypothesis that a normal distribution is a boundary distribution in this and in other, similar processes; however, the distributions observed in the initial developmental phases of a population can differ significantly from it.

References

Bricaud A., Morel A., 1986, *Light attenuation and scattering by phytoplanktonic cells: a theoretical modelling*, Appl. Opt., 25 (4), 571–580.

Dera J., 1992, *Marine optics*, PWN, Warszawa, 432 pp, (in Polish).

Gurgul H., 1991, *Dispersion systems in the sea*, Wyd. Nauk. Uniw. Szczecin, Szczecin, 248 pp., (in Polish).

Gurgul H., Kopeć J., Lizak J., Wierzbicka-Kharraz I., 1992, *Dispersion distributions of sediments in waters of South Baltic*, Acta Phys., 2, 71–88, (in Polish).

Jonasz M., Fournier G., 1996, *Approximation of the size distributions of marine particles by a sum of log-normal functions*, Limnol. Oceanogr., 41, 744–754.

Kopeć J., Pawlak B., *Application of approximate formulae for analysis of spectral distribution of light attenuation coefficient*, Oceanol. Acta, (in press).

Pawlak B., 1998, *φ-normal distributions of sizes in an algal population*, Conference: *Physical problems of ecology of natural waters*, Uniw. Szczecin, Szczecin, 8 pp., (in Polish).