

Species-specific phytoplankton growth rates via diel DNA synthesis cycles.

IV. Evaluation of the magnitude of error with computer-simulated cell populations

Jeng Chang, Edward J. Carpenter

Marine Sciences Research Center, State University of New York, Stony Brook, New York 11794, USA

ABSTRACT: A computer model was constructed to evaluate the magnitude of error embedded in the estimation of phytoplankton growth rate via cell cycle analysis. Attempts have been made to assess errors from different parts of computational procedures, errors caused by variations in curve fitting, and errors arising due to situations that do not match the assumptions used in the original derivations. Results indicated that errors involved in measuring the fraction of a population in the combined S, G2, and M phase have minimal influence on the accuracy of estimated growth rate. When errors in the estimation of the duration of cell cycle phase are included, an averaged deviation of 11 % was observed. If the additional errors in the measurement of single-cell DNA content are also considered, the averaged deviation in estimated growth rates increases to 22 %. Changing the sampling interval from 1 h to 2 h created only a minor increase in relative error. Among different methods for defining maximum values of phase fractions, the periodic curve fitting technique performed the best. For special situations that do not match original assumptions, a right-skewed S phase fraction curve caused the largest deviation, but the deviation was less than 41 %.

INTRODUCTION

Mitotic division is the primary means for a phytoplankton population to increase its size. Monitoring cells in a population traveling through various cell cycle phases thus can be used as a powerful tool to estimate species-specific growth rates of phytoplankton in natural environments (Weiler & Chisholm 1976). To estimate a growth rate via cell cycle analysis, 2 pieces of information are usually required. One is the fraction of proliferating cells (f_x) which are involved in a 'terminal event' of the cell cycle (Fig. 1). The other is the traverse time for a cell to complete the terminal event (T_x). Subsequently, the daily mean growth rate (μ) is calculated by (McDuff & Chisholm 1982):

$$\mu = \frac{1}{n T_x} \sum_{i=1}^n \ln [1+f_x(t_i)] \quad (1)$$

where t_i = time when sample i is taken; and n = total number of samples taken during a 24 h cycle. Recently, a well-defined procedure to estimate growth rate has been developed by Carpenter & Chang (1988) based on a functional model of the cell cycle (Fig. 1). A test of

this technique on 2 species of dinoflagellates, *Heterocapsa triquetra* and *Prorocentrum minimum*, yielded growth rates similar to that from daily increases in cell counts (Chang & Carpenter 1988, Antia et al. 1990). However, in these studies, the growth rates estimated from cell counts may deviate from the true growth rate due to possible cell loss in the cultures.

To determine the actual magnitude of deviation in growth rate estimation associated with this technique, a straightforward approach is to build a computer simulation program based on the functional model of Carpenter & Chang (1988), and then test the method on the data generated by the program. An advantage of such a program is that all the parameters about a cell cycle are known precisely. As a result, performing growth rate estimations on computer-generated data allows direct comparisons between the estimates and the true values defined in the program. Determination of the magnitude of errors embedded in the method thus becomes possible. Secondly, the program can be altered easily to simulate cell populations which proliferate according to rules that violate the model's assumptions. Conducting estimations on such

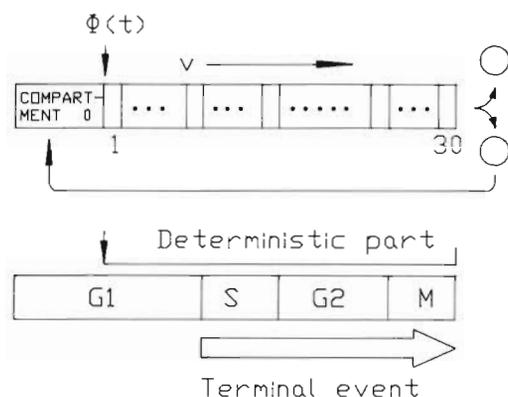


Fig. 1. Functional model of the cell cycle used in the derivation of the growth rate estimation procedure (Carpenter & Chang 1988). Top panel illustrates the basic structure of the cycle used in the computer simulation. Bottom panel shows the relationship between the mathematical model and the cell cycle phases. The part of the cell cycle after the arrow in the G1 phase is usually considered deterministic (Smith & Martin 1973). This is analogous to a constant moving speed, v , for cells in-between compartments 1 and 30. A release factor Φ is used to adjust the flow of cells from compartment 0 to 1 so a partially synchronized population can be simulated by the program. A terminal event is defined as the period between any point on the cell cycle and the end of mitosis (Mitchison 1971). The thick arrow in the bottom panel indicates the terminal event used in this study, that is, the combination of S, G2, and M phases

sets of data can reveal the degree of tolerance of the method under various conditions.

Errors in growth rate estimation may come from the fact that the method of cell cycle analysis is derived from a continuous model whereas limited samples at discrete time points are the best that we can obtain in reality. This difference will cause error in the 2 independent estimates in Eq. (1). One estimate is for T_X which can be called the 'duration term', and the other is for the rest of the right hand side of the equation which can be named the 'mean fraction term'. Hence a general form for growth rate estimation becomes

$$\text{Growth rate} = \frac{\text{Mean fraction term}}{\text{Duration term}} \quad (2)$$

The influence of sampling intervals on the estimation of mean fraction terms and growth rates has been studied by McDuff & Chisholm (1982). How sporadic samples affect the estimation of the duration term will be examined in the present report.

Other errors may emerge from some of the assumptions in the functional model. For example, the model assumes that the shape of the proliferating cell cohort is symmetrical (Carpenter & Chang 1988), but this may not be true for all phytoplankton populations in nature. The emulation program was altered to imitate the effects of an asymmetrical cell cohort, ultradian growth mode (Chisholm 1981), or cells arrested in G2 phase

(Vaulot et al. 1987). The magnitude of error under each condition was determined.

In addition, because DNA content must be measured fluorometrically, the measurement CV (coefficient of variation) will introduce additional errors to both duration and mean fraction terms. Phase fractions extracted from a DNA histogram usually carry an error of 10 to 20 % (Baisch et al. 1982). Therefore, the magnitude of errors in growth rate estimation due to this factor needs to be assessed. Computer-simulated populations were used to generate DNA histograms with measurement CV's attached. The size of relative errors estimated this way will be helpful in understanding growth rate data from natural phytoplankton populations.

METHODS

Computer simulation. Simulated partially-synchronized cell populations were constructed according to Carpenter & Chang's (1988) functional model with 31 compartments (Fig. 1). The traveling speed, v , was set to 2 compartments per hour. Compartments 1 to 30 were used as the deterministic segment. To simulate many possible shapes of a cell cohort, distributions of the beta family were selected to represent the release factor, $\Phi(t)$. In this simulation, the beta distribution was not treated as a probability density function. Rather, it was used as a function of time to generate a variable traveling speed, $\Phi(t)$, between compartment 0 and 1 (Fig. 2). By altering the parameters ν and w in a beta distribution, symmetrical, skewed, or double-peaked cell cohorts can be conveniently generated (Hastings & Peacock 1975). With these restrictions, Carpenter & Chang's (1988) functional model becomes

$$\frac{dn_i}{dt} = v(n_{i-1} - n_i) \quad i = 2, 3, \dots, 30$$

$$\frac{dn_0}{dt} = 2vn_{30} - \Phi(t)n_0 \quad (3)$$

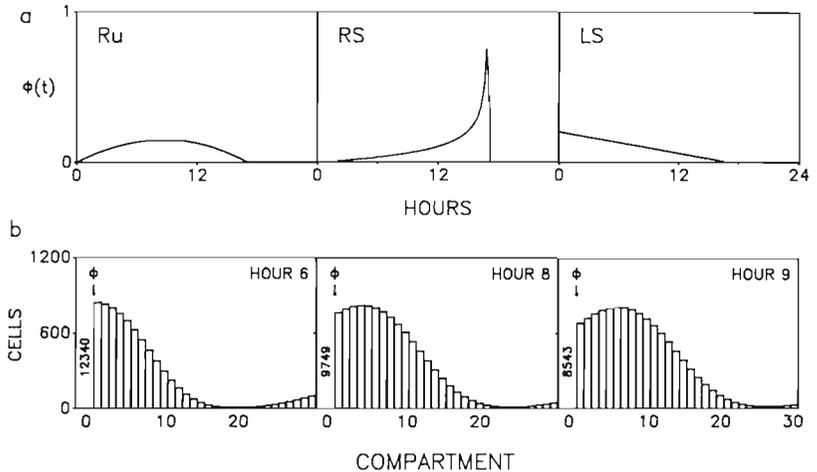
$$\frac{dn_1}{dt} = \Phi(t)n_0 - vn_1$$

where

$$\Phi(t) = \begin{cases} A \cdot \left[\frac{t}{t_{\text{end}}} \right]^{\nu-1} \left[1 - \frac{t}{t_{\text{end}}} \right]^{w-1} & \text{if } 0 \leq t \leq t_{\text{end}} \\ \int_0^1 u^{\nu-1} (1-u)^{w-1} du & \\ 0 & \text{if } t_{\text{end}} < t \leq 24 \end{cases}$$

Here, 'hours' was chosen as the unit for t in the program. The term t_{end} , which is a time point between Hour 0 and

Fig. 2. Shapes of $\Phi(t)$ and the process of generating a cell cohort. (a) Variation of $\Phi(t)$ with time in the simulation of the rudimentary case (Ru), and the populations with right-skewed (RS) and left-skewed (LS) f_S . In all 3 cases, t_{end} was set to Hour 17. (b) The formation of a cell cohort in the rudimentary case. Number of cells in compartment 0 is written near the left border of each panel. For the relationship between compartment numbers and the cell cycle, see Fig. 1



24, controls the spread of a cohort. The term A is the time averaged $\Phi(t)$ between time 0 and t_{end} . As a result, the number of cells in a cohort as well as the growth rate of a such simulated population are determined jointly by t_{end} , A , and the length of the deterministic segment (Figs. 1 and 2). To prevent from becoming infinity, $\Phi(0)$ was set to $(5A)$ when $\nu < 1$. Similarly, $\Phi(t_{\text{end}})$ was subjected to the same restriction when $\omega < 1$ (Fig. 2a). Definitions of other terms are listed in Table 1.

The fourth order Runge-Kutta method was used to solve Eq. (3) numerically in order to generate a time series of cell numbers in each compartment (Conte & de Boor 1972). A step size of 0.2 h was selected, and accuracy was controlled by the technique of step doubling described by Press et al. (1986). After an arbitrary initial condition, which is to put 310 cells in compartment 0 at $t=0$, has been given, numerical integration was carried out on an IBM PC/AT computer until the age distributions were identical for 2 consecutive days and the population size exceeded 3×10^4 cells. The population size, $N(t)$, was computed by

$$N(t) = \sum_{i=0}^{30} n_i(t) \quad (4)$$

The true growth rate of this simulated population thus was computed from the daily increase of population size along the course of the numerical integration. The final output from the simulation program was S and $G2+M$ phase fractions at 0.2 h intervals in the last 24 h of the numerical integration period. The term 'phase fraction' is defined as the fraction of a population which is undergoing a certain cell cycle phase at time t .

Determining the magnitude of error. Estimated growth rates were calculated according to Eq. (1) with the sum of the S and $G2+M$ phases as the terminal event (Carpenter & Chang 1988). In each case tested, an estimated μ was compared to its corresponding true μ to determine the magnitude of deviation. The relative

Table 1. Symbols used in the text. The unit of each variable in the simulation is included in parentheses

A	Averaged Φ between time 0 and t_{end} (compartment h^{-1})
a_0, a_1, a_2	Coefficients of the periodic function
b_1, b_2	Coefficients of the periodic function
f_{G1}, f_S, f_{G2M}	$G1, S$, and $G2M$ phase fraction
f_X	Fraction of a population undergoing the terminal event X
N	Total number of cells in the population (cells)
n	Number of samples taken in a 24 h cycle
n_i	Number of cells in compartment i (cells compartment $^{-1}$)
n_S, n_{G2M}	Number of cells in S , and $G2 + M$ phase
T_S, T_{G2M}, T_X	Duration of $S, G2 + M$ phase, and the terminal event X (h)
t	Time (h)
t_{end}	Time in a 24 h cycle when the last cell in a cell cohort is released by Φ (h)
t_1, t_2	Time when n_S , and n_{G2M} reaches maximum (h)
ν	Traveling speed of a cell in the cell cycle (compartment h^{-1})
ν	Parameter in a beta distribution
ω	Parameter in a beta distribution
X	Terminal event of the cell cycle
μ	Daily mean growth rate (d^{-1})
Φ	Release factor of compartment 0 (compartment h^{-1})

error of the estimated μ is expressed as a percentage of the true μ , that is,

$$\text{Relative error (\%)} = \frac{(\text{Estimated } \mu) - (\text{True } \mu)}{\text{True } \mu} \cdot 100 \% \quad (5)$$

Several ways were used to estimate the duration and the mean fraction terms in Eq. (2). To estimate error due to the mean fraction term, the originally defined phase durations were used as duration terms to com-

Table 2. Settings used to assess errors from various sources in the estimation of growth rate

Source of error	Sampling interval	Method of obtaining phase fractions	Method of defining peaks on fraction curves
Mean fraction term	2 h	Directly from the output of the program	Use predetermined phase durations
Mean fraction and duration term – w/o measurement CV	2 h	Directly from the output of the program	1. Direct method 2. Polynomial regression 3. Polynomial regression with iteration
Mean fraction and duration term – with measurement CV	Both 1 and 2 h	From the simulated DNA histograms	1. Polynomial regression 2. Polynomial regression with iteration 3. Periodic curve fitting

pute growth rates (Table 2). Phase fractions from the simulated populations were sampled every 2 h.

In a separate set of calculations, the estimated T_X 's were used to evaluate the error caused by the duration term. The estimation of T_X was accomplished by

$$T_X = T_S + T_{G2M} = 2(t_2 - t_1) \quad (6)$$

where t_1 and t_2 = times when S and G2+M cells [$n_S(t)$ and $n_{G2M}(t)$] reach their maximum values during a 24 h cycle. Three variations in determining t_1 , t_2 , and T_X were compared; they are: (1) the direct method which selects maximum values among sporadic phase fractions directly as t_1 and t_2 ; (2) the polynomial regression method which fits second degree to fourth degree polynomial functions to sporadic phase fractions followed by calculating peak positions, t_1 and t_2 , from the

Table 3. Special settings, which are different from those described in the rudimentary case or Eq. (3), in the simulation of various cell division cycles

Description	Special settings
Ultradian growth mode	Reducing the number of compartments to 16. $A = 0.2$ to 1.5
Good synchrony	$t_{end} = 4$ $A = 0.02$ to 1.5
Asymmetrical cohort: Left-skewed f_S	$\nu = 1, \omega = 2, A = 0.1$
Right-skewed f_S	$\nu = 2, \omega = 0.5, A = 0.11$
Double-peaked f_S	$\nu = \omega = 0.1, A = 0.2$
G2 phase arresting ^a	$\frac{dn_{24}}{dt} = \nu n_{23} - \Phi(t) n_{24}$ $\frac{dn_{25}}{dt} = \Phi(t) n_{24} - \nu n_{25}$ $A = 0.21$
^a When a complete arrest is required, set $\Phi(t)$ to 0	

fitted curves; and (3) the polynomial regression with iteration method which applies the iterative method of Carpenter & Chang (1988) in addition to the polynomial curve fitting (Table 2). Detailed procedures about these methods are described in Chang & Carpenter (1988).

Different cases tested. For all situations tested, one setting served as a basis of comparison by emulating a phased population of *Heterocapsa triquetra* in Chang & Carpenter (1988). In this rudimentary case, the ν and ω in $\Phi(t)$ were both set to 2 in order to create a symmetrical cell cohort. A moderate synchrony was chosen by setting t_{end} to Hour 17. The S and the G2+M duration were set to 3 and 11.5 h, respectively. Next, 0.1 compartments per hour was used for A in Eq. (3), and the growth rate of the rudimentary case became 0.596 d^{-1} . Programs emulating other situations were constructed either by changing parameters in the rudimentary case or by modifying some of the equations in Eq. (3). Special settings for each situation are listed in Table 3.

The computational error from simulated populations matching all assumptions in the functional model was examined first. Settings were the same as in the rudimentary case. Different growth rates were created by varying A. Growth rates higher than 1.0 d^{-1} were achieved by reducing the total number of compartments to 16, and the resultant deterministic period was 7.5 h. The simulated populations with better synchrony were created by changing t_{end} in the rudimentary case to Hour 4. In another set of tests, the relationship between size of error and phase durations was examined with the rudimentary case. The ratio of $T_S : T_{G2M}$ was varied from 0.07 to 8.67 while the sum of T_S and T_{G2M} was fixed at 14.5 h.

Situations which do not match model assumptions exactly include asymmetrical cohorts, double-peaked cohorts, and cohorts arrested in G2 phase. Asymmetrical and double-peaked cohorts were created by varying ν and ω (Table 3). Permanent G2 arrest was accom-

plished by blocking the cell flux from compartment 24 to 25 when the population had reached a certain size. To simulate the occurrence of temporarily blocked cells in G2 phase, the out-flux from compartment 24 was subjected to the restriction of a 2nd release factor which behaved exactly the same as the 1st one (Table 3).

Errors introduced by measurement CV. As for examining the effect on estimated μ due to measurement CV, simulated DNA histograms were generated from 5 populations: the rudimentary case, a population with good synchrony, a population under ultradian growth mode, a population with an asymmetrical cohort, and a population with a slow growth rate. We used the same procedures of computer simulation. Every hour, a sample of 300 cells was selected randomly from each population for the last 24 h cycle of the numerical integration period. The DNA content of each of the 300 cells was assigned according to that cell's position on the cell cycle as well as the predetermined phase durations. At the same time, normally distributed random noise was attached to each DNA value in order to imitate the real experimental procedures. The random noise was manipulated to generate a measurement CV of 10 % in each histogram, and to cause the mean DNA content of G2+M cells to vary between 1.9 and 2.1 times that of G1 cells.

The parametric method proposed by Fried (1976) was applied to deconvolve histograms into phase fractions (Dean 1985). The duration term was estimated by 3 methods (Table 2): (1) the polynomial regression method; (2) the polynomial regression with iteration method; and (3) the periodic curve fitting method (see below). The resulting estimates of growth rate were compared to true μ 's. Subsamples of phase fraction data, formed by omitting a data point every other hour, were also employed in the same procedures to test the effects of the sampling interval on growth rate estimation.

Periodic curve fitting method. The sporadic phase fractions obtained from DNA histograms were fitted by a periodic function of the following form:

$$f(t) = a_0 + a_1 \cos\left(\frac{2\pi}{24}t\right) + b_1 \sin\left(\frac{2\pi}{24}t\right) + a_2 \cos\left(\frac{2\pi}{12}t\right) + b_2 \sin\left(\frac{2\pi}{12}t\right) \quad (7)$$

where a_0 , a_1 , a_2 , b_1 , and b_2 are coefficients of each term. The series of sampling times were converted to corresponding trigonometrical values, then a multiple linear regression program in the SPSS statistical package (SPSS Inc., Chicago) was used to calculate the coefficients. This procedure has been checked against a Fourier analysis program (IBM application program, System/360 scientific subroutine package, 360A-CM-03X). Both programs generated identical coefficients.

RESULTS

Behavior of a simulated population – the rudimentary case

The settings in the rudimentary case simulated characteristics of a phased population successfully. Population size increased exponentially for 5 d before the program was instructed to stop. Within each 24 h cycle, cell number increased in a staircase-like manner (Fig. 3a). The final output of the simulation program was not affected by the initial conditions selected. With all other settings the same, distributing 310 cells evenly over the cell cycle (i.e. 10 cells per compartment) or placing all 310 cells in compartment 0 generated identical phase fractions. The number of cells in each cell cycle phase changed in a cyclic manner which indicated that one cohort of cells was generated on each day and traveled through the cell cycle (Fig. 3). If the

Fig. 3. Cell cycle characteristics of the rudimentary case. (a) Growth curve of cell number during a 24 h period. The curve is normalized to the population size at Hour 0. (b) Cell number curve (—) and fraction curve (---) of G2+M phase generated by the simulation program. Each curve is normalized to its value at Hour 0. The 2 arrows mark peak positions on individual curves. (c) G1 and S phase fraction curves. Filled circles on the S curve mark data taken at a 2 h intervals. Dashed curve is fitted by polynomial regression based on the sporadic data points. (d) G2+M phase fraction curve (—) from the simulation. Data taken at a 2 h intervals (●) and the corresponding regression curve (---) are also shown

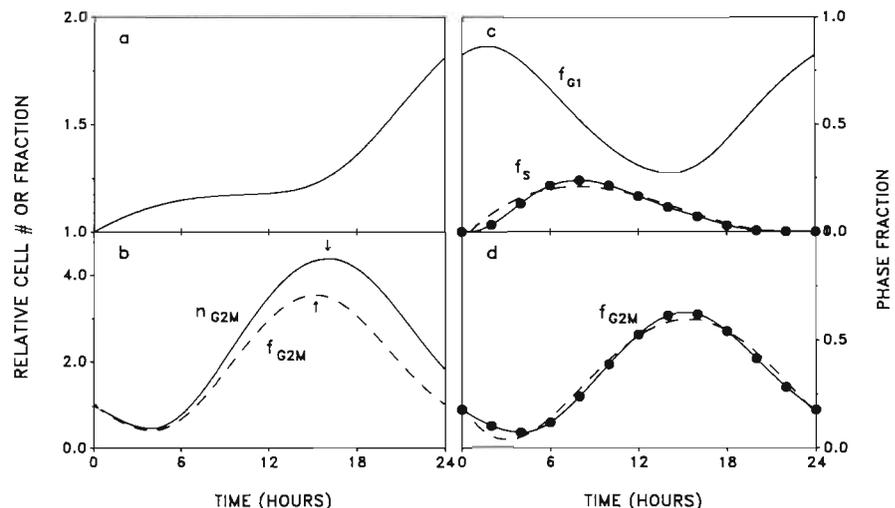


Table 4. The rudimentary case. Predetermined and estimated values of t_1 , t_2 , and $T_S + T_{G2M}$. Effect of measurement CV is not included

	t_1 (hour of day)	t_2 (hour of day)	$T_S + T_{G2M}$ (h)
Predetermined values	–	–	14.5
From cell number curves generated by the program	8.0	16.0	16.0
From fraction curves generated by the program	7.8	15.2	14.8
From samples taken at 2 h intervals from fraction curves	8.0	16.0	16.0
From polynomial curve fitting on sporadic samples	7.8	15.3	15.0
From iterative method on the fitted curves	8.0	16.3	16.6

G2+M phase fraction curve was used to monitor the cell cycle, its peak position was different from that on the cell number curve (Fig. 3b). When the population size was increasing drastically between Hour 15 and 24, the f_{G2M} curve began to deviate more and more from the n_{G2M} curve. As a result, the peak on f_{G2M} was shifted back by 0.8 h (Fig. 3b, Table 4). The peak on the f_S curve, however, was only shifted back by 0.2 h since few cell divisions occurred when the cohort was traveling through S phase.

Magnitude of error without effect of measurement CV

In the rudimentary case, error in the estimated growth rates mainly came from the duration term. If the predetermined $T_S + T_{G2M}$, 14.5 h, was used in the calculation, the estimated μ was 0.59 d^{-1} . Compared to the true μ , 0.596 d^{-1} , such a small deviation (-1%) indicated that error due to the mean fraction term was negligible. When Eq. (6) was used to estimate $T_S + T_{G2M}$, however, a more noticeable deviation emerged (Table 4). The $T_S + T_{G2M}$ calculated directly from the cell number curve was 1.5 h longer than the predetermined $T_S + T_{G2M}$. The accordingly estimated μ was 0.53 d^{-1} , representing an 11% underestimate.

Polynomial curve fitting was an effective way of defining the peaks on phase fraction curves. If samples were taken on a 2 h interval, times associated with the highest values were not good estimates for positions of peaks on f_S and f_{G2M} (Table 4), but fitted curves possessed peaks very close to those on the original f_S and f_{G2M} curves (Fig. 3c, d). The iterative method also worked well in the rudimentary case. The method estimated t_1 on the cell number curve of the S phase without error. As for the estimate of t_2 on n_{G2+M} , an error of 0.3 h was observed (Table 4).

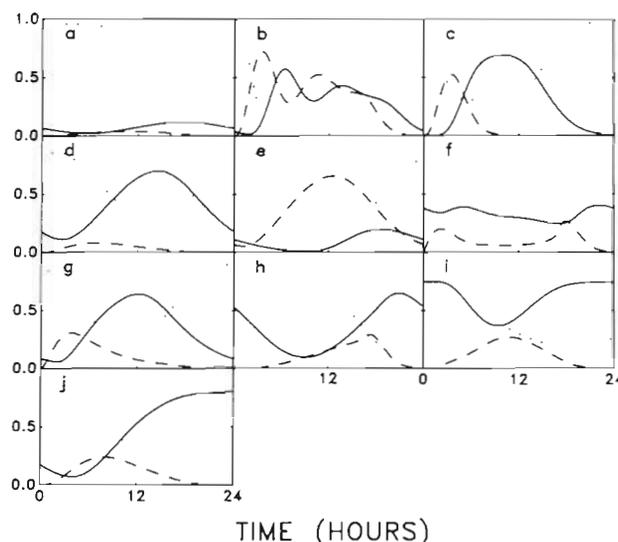


Fig. 4. G1 (···), S (---), and G2+M (—) phase fraction curves of representative cases tested. (a) A simulated slow growing population with moderate synchrony. $\mu = 0.143 \text{ d}^{-1}$ (b) A rapid growing population with a short deterministic segment. $\mu = 1.706 \text{ d}^{-1}$ (c) A population with good synchrony. $\mu = 0.529 \text{ d}^{-1}$ (d) A population with very short S duration. $\mu = 0.596 \text{ d}^{-1}$ (e) A population with very long S duration. $\mu = 0.596 \text{ d}^{-1}$. (f) A population with double-peaked cell cohort. $\mu = 0.595 \text{ d}^{-1}$ (g) A population with left-skewed S phase fraction curve. $\mu = 0.596 \text{ d}^{-1}$ (h) A population with right skewed S phase fraction curve. $\mu = 0.602 \text{ d}^{-1}$. (i) A population with its cell cohort temporarily arrested in G2 phase. $\mu = 0.575 \text{ d}^{-1}$ (j) A population with its cell cohort permanently arrested in G2 phase. $\mu = 0.167 \text{ d}^{-1}$

Altering A in the rudimentary case either decreased or increased growth rate up to 0.96 d^{-1} . When A was reduced, fraction curves had smaller amplitude, and peaks were less obvious since less cells participated in the daily cohort (Fig. 4a). On the other hand, reducing the length of the deterministic segment created very high growth rates. Fraction curves showed multiple

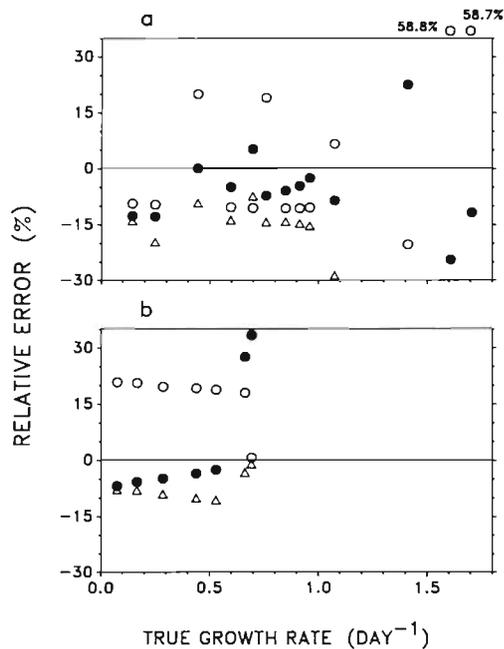


Fig. 5. Relative error of estimated μ 's under various true growth rates. Sampling intervals were 2 h. The duration term was obtained by: locating t_1 and t_2 directly from sporadic data points (\circ), polynomial curve fitting (\bullet), or curve fitting plus the iteration procedure (Δ). (a) Simulated populations with moderate synchrony. Growth rates higher than 1.0 d^{-1} were achieved by reducing the length of the deterministic segment of the cell cycle. (b) Populations with good synchrony

peaks (Fig. 4b), indicating that more than one cohort was formed in each 24 h cycle. In the simulated populations with good synchrony, narrow peaks were observed on phase fraction curves (Fig. 4c). A somewhat flattened tip, however, was noticed on G2+M phase fraction curves.

In various cases with different growth rates, the magnitude of the error varied according to the method used in estimating the duration and the mean fraction terms. Relative error due to the mean fraction term alone ranged from -2.8% to -0.1% . With estimated $T_S + T_{G2M}$'s, errors were larger but showed no correlation with true μ 's except under quite high growth rates (Fig. 5). Among the 3 ways of defining t_1 and t_2 , locating them directly from sporadic sample points with a 2 h interval (the direct method) gave the largest error in general. The other feature of this approach was its instability. The size of relative error might change abruptly between 2 cases with similar true μ 's. In most cases, polynomial curve fitting generated better estimates of μ . Relative error increased suddenly when the true μ exceeded a certain value (Fig. 5). In all cases tested, the iterative method generated more underestimated μ 's than polynomial regression alone did. In several cases, which involved fraction curves with multiple peaks, the iterative method failed in generating estimates.

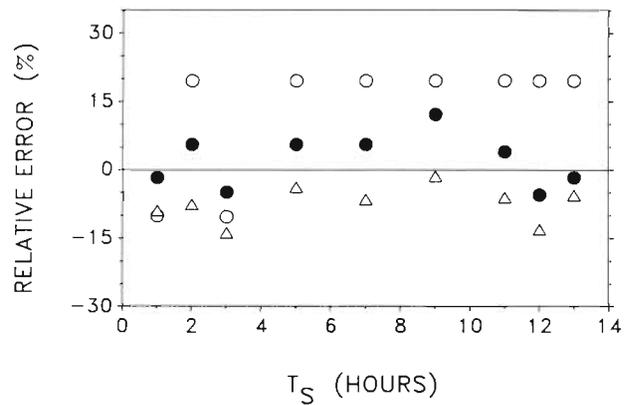


Fig. 6. Relative error of estimated μ 's under various S phase durations. Growth rates of all cases were fixed at 0.596 d^{-1} . Sampling intervals were 2 h. Symbols as in Fig. 5

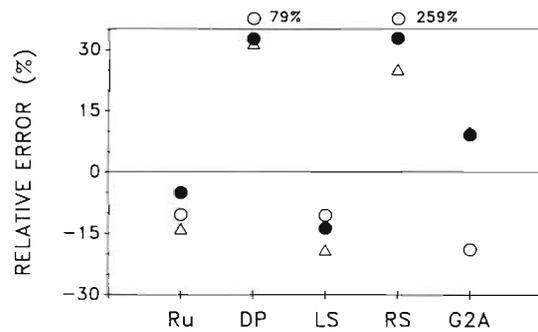


Fig. 7. Relative error of estimated μ 's from populations with either an asymmetrical cell cohort or a cell cohort temporarily arrested in G2 phase. Sampling intervals were 2 h. Symbols as in Fig. 5. Ru: the rudimentary case; DP: a population with double-peaked cell cohort. RS: a population with right-skewed f_S . LS: a population with left-skewed f_S . G2A: a population with its cell cohort temporarily arrested in G2 phase

When the growth rate was fixed, a longer S phase created an S fraction curve with higher amplitude (Fig. 4e). In contrast, the area under the S phase fraction curve was greatly reduced when its duration was short (Fig. 4d). Relative error varied little with the duration of the S phase. Similarly, error due to the mean fraction term was negligible. Among the 3 ways of estimating $T_S + T_{G2M}$, polynomial regression alone generated better estimated μ 's (Fig. 6).

For cell cohorts with nonstandard shape, that is, asymmetrical (Fig. 4g, h) or double-peaked (Fig. 4f), 2 situations (double-peaked and right-skewed) generated sizable errors. When polynomial regression alone was used to determine $T_S + T_{G2M}$, estimated μ 's for both populations with double peaked and right-skewed f_S were about 33 % too high. In the case of a left-skewed f_S , the duration term underestimated the true μ , but to a much lower extent (Figs. 7 and 8).

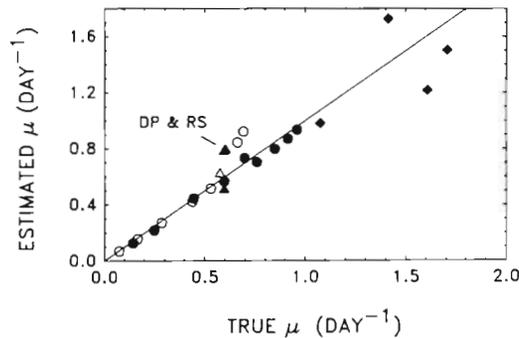


Fig. 8. Relationship between estimated μ 's and true μ 's for emulated populations with moderate synchrony (\bullet), good synchrony (\circ), short deterministic segment (\blacklozenge), asymmetrical cohort (\blacktriangle), and temporary G2 phase arresting (\triangle). Continuous line marks points with 1:1 relationship between estimated μ 's and true μ 's. Abbreviations as in Fig. 7

If a simulated population is permanently arrested in G2 phase, a plateau on the G2+M phase fraction curve was noticed (Fig. 4j). Estimating $T_S + T_{G2M}$ was impossible since no peak existed on f_S and f_{G2M} . If the cohort was just temporarily arrested in G2 for less than 1 d (Fig. 4i), reasonable estimates of the true μ could still be obtained (Fig. 7). The absolute deviation of all estimated μ 's via the polynomial regression method is shown in Fig. 8.

Effect of measurement CV on growth rate estimation

The computer-simulated DNA histograms possessed characteristics similar to histograms from real phytoplankton cultures. Deconvolving these histograms generated quantitative estimates for G1, S, and G2+M phase fractions. Plotting phase fractions against time summarized the 5 simulated cell division cycles (Figs. 9 to 13). Phase fractions extracted from DNA histograms

showed a certain degree of scatter. When these phase fractions were used to restore phase fraction curves via curve fitting, G1 and G2+M curves did not differ significantly from the original curves except in the population under ultradian growth mode. The restored S fraction curves, on the other hand, were not as successful. Apparently some S cells were mistakenly assigned to G1 and G2+M phases by the deconvolution procedure so that S phase fractions were systematically underestimated.

In the rudimentary case, obtaining S phase fractions from DNA histograms obscured the true shape of the $f_S(t)$. As a result, the mean fraction term was underestimated and the position of the peak (t_1) was shifted. When samples were taken on an hourly basis, and polynomial regression was used to define t_1 , the peak on the restored f_S differed from that on the true f_S by 0.2 h. With the 2 h sampling interval, however, the difference increased to 1.8 and 4.6 h, respectively. When the periodic function described by Eq. (7) was used to restore $f_S(t)$, with various sampling intervals, the S peak deviated from the true position by 0.2 to 2 h (Fig. 9).

The restored S phase fraction curve was closer to the true f_S when the degree of synchrony improved (Fig. 10). To obtain a meaningful polynomial regression, only the data points between Hours 0 and 9 were used for the fitting procedure. No such subjective screening was necessary for using the periodic function in curve fitting. Although a false hump was generated in the flat region of the S curve around Hour 16, it did not interfere with locating t_1 .

When ultradian growth occurred, the high scatter of data points made it very hard to generate fitted curves which had similar shapes to the true curves (Fig. 11). In the G2+M phase fractions, polynomial regression failed to extract the double peaked shape out of the

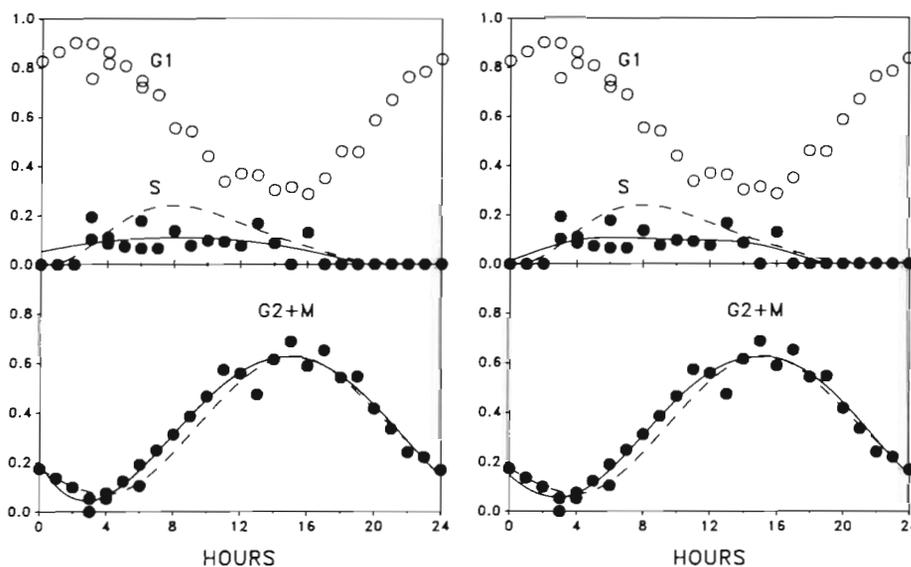


Fig. 9. Phase fractions for the rudimentary case. Left column, upper panel: (\circ) G1 phase fractions extracted from DNA histograms at different times of the 24 h cycle; (\bullet) extracted S phase fractions. (—) A fitted polynomial; (---) true S fraction curve. Left column, lower panel: (\bullet) extracted G2+M phase fractions. (—) fitted polynomial; (---) true G2+M fraction curve. Right column: same as left column but continuous curves (—) are fitted with the periodic function

Fig. 10. Phase fractions for the population with good synchrony. Symbols as in Fig. 9

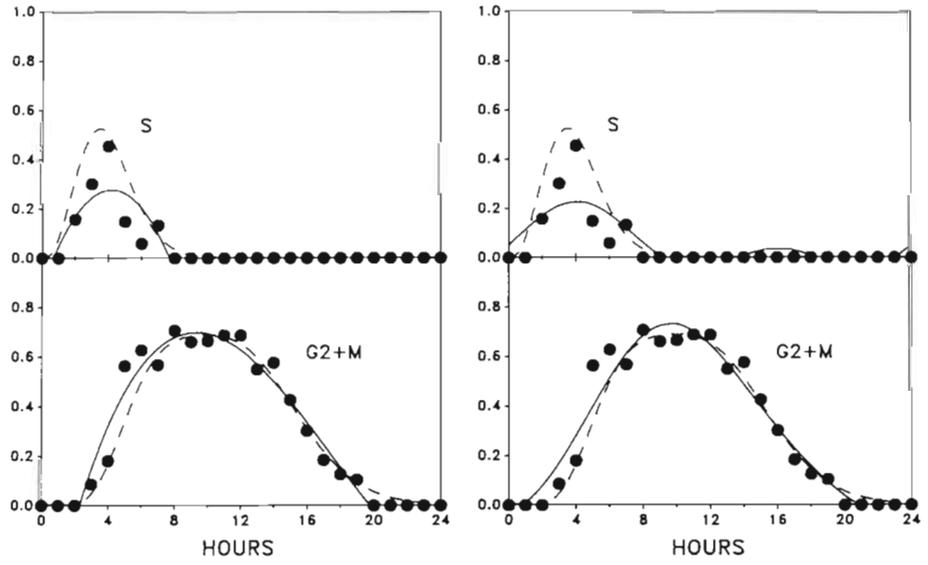


Fig. 11. Phase fractions for the population under ultradian growth mode. Symbols as in Fig. 9

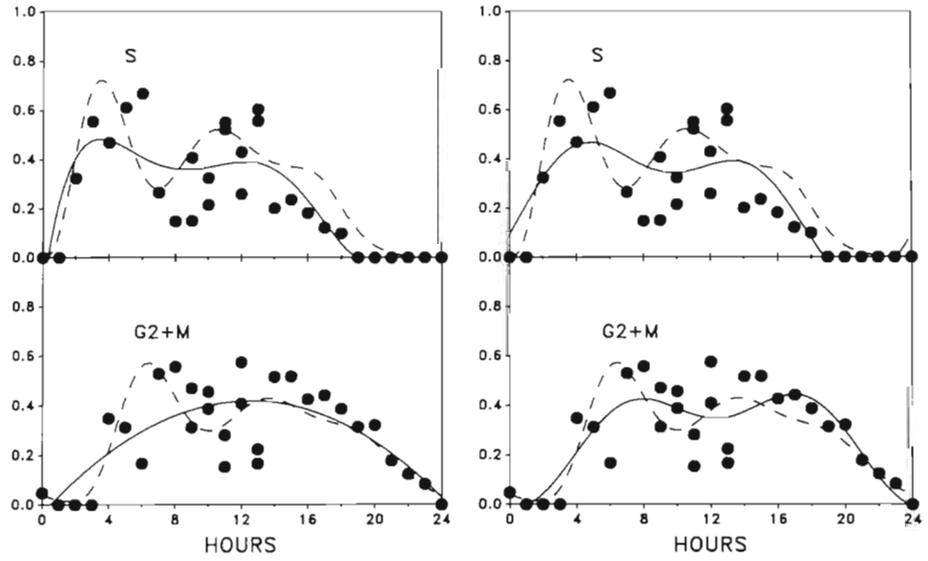
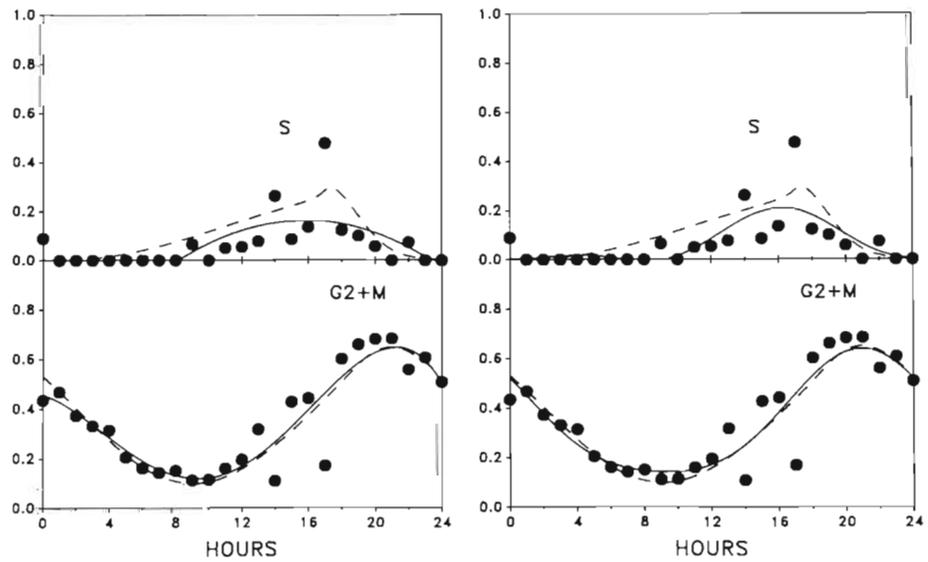


Fig. 12. Phase fractions for the population with an asymmetrical cohort. Symbols as in Fig. 9



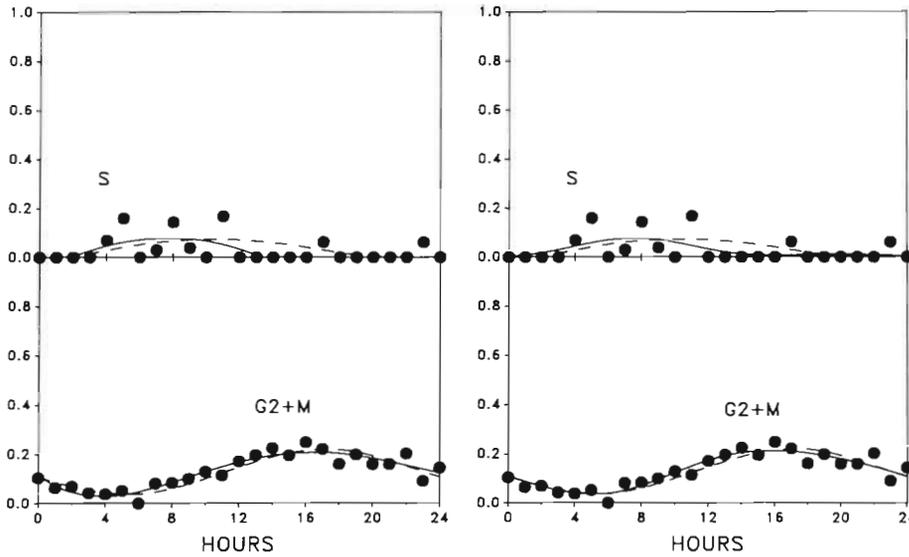


Fig. 13. Phase fractions for the population with a slow growth rate. Symbols as in Fig. 9

data points, so the restored f_{G2M} had a peak that mimicked none of the original peaks.

In the simulated population with an asymmetrical cohort, neither polynomial nor periodic curve fitting restored the right skewed f_S correctly (Fig. 12). The underestimation of S phase fractions still existed. When samples were taken at odd hours, a polynomial regression failed to generate a f_{G2M} curve with a local maximum. Hence the iterative method could not be applied.

In the slow-growing population, the underestimation of S phase fractions concentrated on the later half of the $f_S(t)$ (Fig. 13, Hours 12 to 16). Both curve fitting techniques thus generated a peak shifting to an earlier time.

Estimated μ 's via cell cycle analysis were reasonably close to the true μ 's (Figs. 14 and 15). Substantial underestimation ranging from -50 to -73% was observed in the ultradian growing population when polynomial regression was involved in defining t_1 and t_2 . Such a deviation did not occur when the periodic function was used in the curve fitting. The sampling interval seemed to have little effect on the magnitude of errors. Most estimated μ 's were within 40% of the corresponding true μ 's.

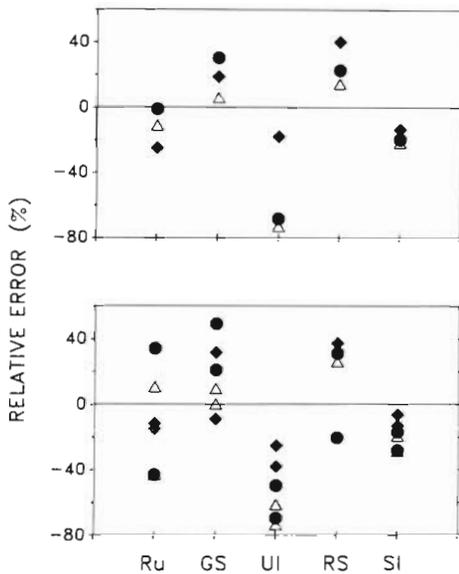


Fig. 14. Relative error of estimated μ 's after the effect of measurement CV is included. Upper panel: estimated with a 1 h sampling interval. Duration term was obtained by: polynomial regression method (\bullet), polynomial regression with iteration method (Δ), or periodic curve fitting (\blacklozenge). Lower panel: estimated with a 2 h sampling interval. Ru: the rudimentary case; GS: a population with good synchrony; UI: a population under ultradian growth mode; RS: a population with right-skewed f_S . SI: a population with a slow growth rate

DISCUSSION

The fact that $T_S + T_{G2M}$ estimated by Eq. (6) failed to generate completely accurate estimates from cell number curves (Table 4) deserves our attention

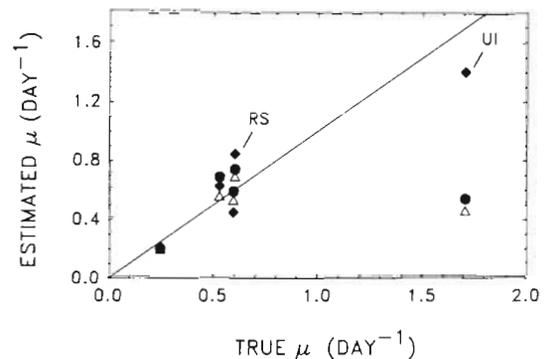


Fig. 15. Relationship between estimated μ 's and true μ 's after the effect of measurement CV is included. Growth rates shown are estimated with a sampling interval of 1 h. Symbols and abbreviations as in Fig. 14. Continuous line marks points with 1:1 relationship between estimated μ 's and true μ 's

because the derivation of Eq. (6) was based on n_S and n_{G2M} (Carpenter & Chang 1988). Possibly, it was caused by the lack of perfect symmetry of the cell cohort generated by the program. As indicated by the shape of the cell cohort (Fig. 2b), the symmetrical $\Phi(t)$ used in the rudimentary case (Fig. 2a, Ru) did not generate a perfectly symmetrical cohort. This is due to the fact that the cell number in the reservoir, or compartment 0, cannot be maintained at a constant level (Fig. 2b). After $\Phi(t)$ has passed its peak at $t = \frac{1}{2}t_{end}$, relative few cells are left in the reservoir. Compared to $\Phi(t)$ on the opposite side of the peak, $\Phi(t)$ with the same magnitude now releases less cells. Note that although this observation is true for the computer simulation, there is no evidence that cell cohorts in all natural phytoplankton populations are essentially asymmetrical.

When phase fractions were extracted from DNA histograms, a substantial underestimation of f_S has been found from the simulated populations tested (Figs. 9 to 13). Similar observations have been documented by Baisch et al. (1982) who indicated that, regardless of the analytical procedure used to extract phase fractions, the underestimation of f_S is inevitable. A direct consequence of this fact is that the mean fraction term in growth rate calculations will always be underestimated. A second, and more severe, effect is that the S phase fraction curves obtained this way possess peaks at incorrect time points.

In growth rate estimation, errors caused by the mean fraction term are quite small. This is true even if the S phase fractions were underestimated by deconvolution procedures. These observations support the statement that errors involved in the estimated mean fraction term are small as long as the sampling interval is shorter than the duration of the terminal event (McDuff & Chisholm 1982). In our computer simulation, the terminal event has a duration of either 7.5 or 14.5 h, but the longest sampling interval used was only 2 h.

The estimation of the duration term contributes more to the total deviation in estimated μ 's. Part of the reason is that any error in locating t_1 and t_2 is increased 2-fold by the constant in Eq. (6). For example, if the estimated t_2 is accurate, but the estimated t_1 bears an error of 1 h, the resultant $T_S + T_{G2M}$ will have an error of 2 h. For this reason, estimating t_1 and t_2 directly from the raw data points with a 2 h sampling interval should be avoided. According to this method, the possible values of t_1 and t_2 are restricted to integral multiples of 2 h. This kind of resolution causes large relative error in estimates of μ , especially in populations with asymmetrical cell cohorts (Fig. 7).

On the other hand, the constant 2 in Eq. (6) provides the advantage of estimating the duration of terminal events longer than 24 h with samples taken from a single 24 h cycle. If temporary G2 arrest occurs (Fig. 4i), this method successfully estimates the growth rate (Fig.

7). Turning to the growth rate estimation in populations arrested permanently in the G2 phase, a large relative error is expected because the actual $T_S + T_{G2M}$ is infinite. One way to identify if this situation is occurring is from the shapes of phase fraction curves. If cells start to be arrested in G2 phase, G2 phase fractions will continuously increase with time, and the f_{G2M} at Hour 24 will be much higher than f_{G2M} at Hour 0 (Fig. 4j).

Curve fitting is an effective way to smooth out the phase fraction curves and to locate peaks on them. In general, the restored f_S and f_{G2M} curves possess peaks quite close to those on corresponding true fraction curves (Figs. 3 and 9 to 13). In most of the cases tested, growth rates estimated via the curve fitting procedures have smaller relative error than those from the direct method (Figs. 5 to 8).

Whether a polynomial or a periodical function was used to fit the sporadic phase fractions, similar growth rate estimates were generated (Figs. 14 and 15). The only exception is the ultradian growing population. The periodic function apparently has a better ability to sense the multi-peaked shape embedded in the data points; therefore, it generated more accurate estimates (Figs. 11 and 15). Another advantage of periodic curve fitting is that all data points can join the curve fitting without considering the shape of a fraction curve (Fig. 10). Further, periodic curve fitting may give better results when a peak appears near the beginning or ending point of a 24 h sampling period (Antia et al. 1990). The only disadvantage of the periodic curve fitting method was that in rare cases, it generated 2 peaks on a plateau region.

Although the iterative method restored the t_1 and t_2 on cell number curves in the rudimentary case (Table 4), little or no improvement has been observed on growth rate estimation under various other situations (Figs. 5, 6, 7 and 14). A major reason is that fraction curves obtained via polynomial regressions do not match original phase fraction curves exactly (Fig. 3c, d). Sometimes peaks on regression curves appear at time points later than the true peaks on cell number curves. When iteration is applied in such cases, instead of restoring the true peaks, it will locate time points farther to the right. In other cases, especially when high growth rates are generated due to multiple cohorts in a single 24 h cycle (Figs. 4b and 11), curve fitting techniques are unable to imitate the sharp slopes on the original fraction curves. Consequently, the slope which the iterative method intends to locate does not exist on a fraction curve generated by polynomial regression, and the iteration fails. Under such circumstances, we believe that the iterative method will improve growth rate estimation only when a better way of restoring original phase fraction curves appears.

Changing sampling intervals from 1 to 2 h does not

increase the relative error substantially (Figs. 14 and 15). This observation suggests that the 2 hour interval is acceptable if the experimental condition does not allow a more frequent schedule.

As long as some periodicities exist in phase fraction curves, it is still possible to estimate $T_S + T_{G2M}$ for populations growing in the ultradian mode (Fig. 15). The magnitude of errors is much larger when growth is fast and multiple cell cohorts are created each 24 h cycle (Figs. 5a and 8). A denser sampling schedule may improve the shape of a fitted curve, and reduce the relative error (Fig. 14).

Accuracy of estimation decreased when cell cohorts were asymmetrical (Figs. 7 and 14). A skewed cohort means that when a peak appears on a phase fraction curve, its mode is not located at the midpoint of that phase. Hence the difference between t_1 and t_2 is not equal to $\frac{1}{2}(T_S + T_{G2M})$. Using the time lag between 2 medians on phase fraction curves, as in some related studies (Braunwarth & Sommer 1985), may improve the accuracy in the estimated duration term. Nevertheless, if a fraction curve never touches the zero line, it becomes harder to define the position of the median. The accuracy of estimated μ 's suffers more from the right-skewed f_S population (Figs. 7 and 8). The relatively large errors associated with the population with double-peaked f_S seems to come from the fact that the right half of its f_S resembles a right-skewed f_S (Fig. 4f). In both cases, the method overestimated true μ 's by about 0.2 d^{-1} . As estimates, they are still within an acceptable range (Figs. 8 and 15).

In conclusion, computer simulations indicate that the method derived by Carpenter & Chang (1988) is a reliable technique to estimate species-specific phytoplankton growth rates (Figs. 8 and 15). This technique is also quite tolerant of situations that do not match the original assumptions in the derivation. To obtain a general picture of errors embedded in this technique, an indicator can be computed by averaging the magnitude (ignoring negative signs) of the relative errors. If phase fractions are measured with complete accuracy, and the polynomial regression method is used to define t_1 and t_2 , the averaged magnitude of error is 11%. If the effect of measurement CV is included, the averaged error in estimated μ becomes 22% with the use of periodic curve fitting. Among the 5 cases tested, the relative error ranged from -38% to +40% (Fig. 14). In comparison, when the ^{14}C method is used to estimate phytoplankton growth rate in chemostat cultures, the range of error is between 0 and 100% (Li & Goldman 1981).

Acknowledgements. This research was supported by NSF grant OCE8816584 to E.J.C. We thank E. M. Cosper, A.

Okubo, R. M. Cerrato, L. N. Edmunds, S. W. Chisholm, and R. E. Wilson for helpful discussions. Contribution No. 756 from the Marine Sciences Research Center.

LITERATURE CITED

- Antia, A. N., Carpenter, E. J., Chang, J. (1990). Species-specific phytoplankton growth rates via diel DNA synthesis cycles. III. Accuracy of growth rate measurement in the dinoflagellate *Prorocentrum minimum*. Mar. Ecol. Prog. Ser. 63: 273–279
- Baisch, H., Beck, H.-P., Christensen, I. J., Hartmann, N. R., Fried, J., Dean, P. N., Gray, J. W., Jett, J. H. (1982). A comparison of mathematical methods for the analysis of DNA histograms obtained by flow cytometry. Cell Tissue Kinetics 15: 235–249
- Braunwarth, C., Sommer, U. (1985). Analysis of the in situ growth rates of Cryptophyceae by use of the mitotic index technique. Limnol. Oceanogr. 30: 893–897
- Carpenter, E. J., Chang, J. (1988). Species-specific phytoplankton growth rates via diel DNA synthesis cycles. I. Concept of the method. Mar. Ecol. Prog. Ser. 43: 105–111
- Chang, J., Carpenter, E. J. (1988). Species-specific phytoplankton growth rates via diel DNA synthesis cycles. II. DNA quantification and model verification in the dinoflagellate *Heterocapsa triquetra*. Mar. Ecol. Prog. Ser. 44: 287–296
- Chisholm, S. W. (1981). Temporal patterns of cell division in unicellular algae. In: Platt, T. (ed.) Physiological bases of phytoplankton ecology. Can. Bull. Fish. Aquat. Sci. 210: 150–181
- Conte, S. D., de Boor, C. (1972). Elementary numerical analysis: an algorithmic approach. McGraw-Hill Kogakusha, Tokyo
- Dean, P. N. (1985). Methods of data analysis in flow cytometry. In: van Dilla, M. A., Dean, P. N., Laerum, O. D., Melamed, M. R. (eds.) Flow cytometry: instrumentation and data analysis. Academic Press, London, p. 195–221
- Fried, J. (1976). Method for the quantitative evaluation of data from flow microfluorometry. Comput. Biomed. Res. 9: 263–276
- Hastings, N. A. J., Peacock, J. B. (1975). Statistical distributions. Butterworth, London
- Li, W. K. W., Goldman, J. C. (1981). Problems in estimating growth rates of marine phytoplankton from short-term ^{14}C assays. Microb. Ecol. 7: 113–121
- McDuff, R. E., Chisholm, S. W. (1982). The calculation of in situ growth rates of phytoplankton populations from fractions of cells undergoing mitosis: a clarification. Limnol. Oceanogr. 27: 783–788
- Mitchison, J. M. (1971). The biology of cell cycle. Cambridge University Press, London
- Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. (1986). Numerical recipes: the art of scientific computing. Cambridge University Press, New York
- Smith, J. A., Martin, L. (1973). Do cells cycle?. Proc. natn. Acad. Sci. U.S.A. 70: 1263–1267
- Vaulot, D., Olson, R. J., Merkel, S., Chisholm, S. W. (1987). Cell-cycle response to nutrient starvation in two phytoplankton species, *Thalassiosira weissflogii* and *Hymenomonas carterae*. Mar. Biol. 95: 625–630
- Weiler, C. S., Chisholm, S. W. (1976). Phased cell division in natural populations of marine dinoflagellates from ship-board cultures. J. exp. mar. Biol. Ecol. 25: 239–247

This article was presented by Dr M. Levandowsky, New York, USA

*Manuscript first received: December 22, 1989
Revised version accepted: May 17, 1990*