J. Math. Biol. 45, 313–336 (2002)
Digital Object Identifier (DOI):
10.1007/s002850200160

Mathematical Biology

Julien Arino · Jean-Luc Gouzé · Antoine Sciandra

A discrete, size-structured model of phytoplankton growth in the chemostat

Introduction of inhomogeneous cell division size

Received: 9 February 2000 / Revised version: 10 October 2001 /
Published online: 23 August 2002 –

Abstract. We introduce inhomogeneous, substrate dependent cell division in a time discrete, nonlinear matrix model of size-structured population growth in the chemostat, first introduced by Gage *et al.* [8] and later analysed by Smith [13]. We show that mass conservation is verified, and conclude that our system admits one non zero globally stable equilibrium, which we express explicitly. Then we run numerical simulations of the system, and compare the predictions of the model to data related to phytoplankton growth, whose obtention we discuss. We end with the identification of several parameters of the system.

Introduction

A chemostat is a continuous culture device in which organisms (bacteria, phytoplankton) grow, submitted to a flow of nutrient. Chemostats have been extensively studied mathematically (see [14]), mainly using ordinary differential equations. The global behaviour of a chemostat is well known, using models such as the Monod model or the Droop model. But these models describe the behaviour of the total biomass of the system.

One of the possible ways to gain further insight on the behaviour of chemostats is to use structured models, describing the evolution of the population in more detail. Moreover, it is now possible to obtain long time series of structured data from chemostats (*e.g.*, with the automated device described in [5], which will be described more thoroughly in Section 4). Therefore comparison of model behaviour to the data is possible, and should lead to refinements of the models [4].

However, structured models of the chemostat are less known. Most works on the latter use partial differential equations. In [10] a size structured model is presented,

J. Arino*, J-L. Gouzé: INRIA Sophia Antipolis, BP 93, 06902 Sophia Antipolis Cedex, France. e-mail: Jean-Luc.Gouze@inria.fr

⋆*Present address:* Department of Mathematics and Statistics, University of Victoria, B.C., Canada. e-mail: jarino@math.uvic.ca

A. Sciandra: Station Zoologique, BP 28, 06234 Villefranche sur Mer, France.
e-mail: sciandra@obs-vlfr.fr

Key words or phrases: Chemostat – Structured population models – Discrete model – Inhomogeneous division size

which is extended to the competition case in [6], while [11] introduces cell cycle structuration. However, the mathematical analysis of these models can be hard (one often has to reduce the model to an ODE system by considering moments of the density), so as their numerical integration.

Another kind of structured modelling is the use of discrete time, discrete structure systems (see [7, chapters 1 and 3] for a review of structured discrete time models). While the mathematical analysis of such systems can be as tedious as in the continuous case, they have the advantage that they are easy to simulate.

Such a model, time discrete and structured in biomass (size) classes, was introduced for the chemostat in [8] by Gage *et al.*. They showed that a stable distribution of the biomass is reached, in which the biomass is constant and equal for all size classes. They also studied numerically the influence of various factors (number of size classes, flow rate) on the convergence speed. Later, Smith [13] corrected a mistake in the formulation of the model, showed mathematically that the equilibrium is globally stable, and also introduced competition between two species, showing that the competitive exclusion principle holds. Recently, he and Zhao [15] extended this result to the n-species case.

But this model is based on a very strong assumption: all cells are born with the same biomass b, and cells surviving dilution divide when they reach biomass $2b$. Gage *et al.* obtain biomass spectra which resemble the experimentally observed spectra, but by using the assumption that the biomass in each class is log-normally distributed [17]. Hence variability is added *a posteriori* to the model.

The aim of the present work is to extend the model of Gage *et al.* to the case where cell division (and consequently, cell birth) can happen for cells in several biomass classes, the effective size at division being distributed following some probability density. Hence cells need not necessarily exactly double their biomass in order to divide. This is closely related to the continuous models (in time and in structure) of [10] and [11], where the division rate is defined using a probability density. Using this approach, the division process is included in the model, instead of being artificially added onto the results.

This paper is organised as follows: in Section 1, we derive our model from the one of [8], introducing division for cells of different sizes. The reader interested only in the mathematical formulation of the system should turn directly to Section 1.4 where it is written down. We then analyse the model in Section 2; using the fact that it conserves the mass, we are able to closely follow the analysis of [13] and to conclude existence of a globally stable non trivial equilibrium, which we characterise. Sample equilibrium distributions are then shown in Section 3, as well as examples of the transient behaviour of the model. Biological data is then discussed in Section 4, and compared to model predictions. In particular, an identification procedure is applied to obtain values for some of the system parameters.

1. Model formulation

We suppose that the system under consideration is a well-stirred chemostat. In Gage *et al.*, the following biological assumptions are made.

(H1) In a constant environment (*i.e.*, if the concentration of limiting nutrient is constant and high), the growth of a cell is exponential.
(H2) Cells are born with a biomass b, they grow, then divide when they reach a biomass $2b$.
(H3) When a cell divides, it divides into two daughter cells, whose individual biomass is exactly one half of the biomass of the original cell.

In the present model, we replace hypothesis (H2) with a new one, to relax the division size hypothesis.

(H2′) The division biomass $2b$ (accordingly, the birth biomass b) is not a constant. There exists a distribution of division biomass, describing the individual cellular division biomass.

Before we proceed with the proper formulation of the model, let us give here some precisions concerning mathematical terms that will be used throughout the following sections. First of all, since the model will be formulated in discrete time, we define T, the iteration period (or time step). Let then E ($0 \leq E < 1$) be the dilution rate per iteration period. We also define the state variable S_t, substrate concentration in the chemostat at time t.

The formulation of the model is carried out in several steps. First of all, an appropriate description of cellular size must be made. This is the object of Section 1.1. Then, in Section 1.2, cell growth is taken care of. Finally, preceding the definitive formulation of the system in Section 1.4, modelling of cellular division is explained in Section 1.3.

1.1. Description of cellular size

We suppose that the minimal individual cellular biomass is b_{min}. Then the model will describe cells whose individual biomass belongs to an interval $[b_{min}, b_{max}]$, where b_{max} is a parameter dependent constant that will be discussed later. We split this interval into r biomass (size) classes. Hence a given cell will be a member of a certain class, depending on its biomass.

We then define the state variables $x_t = (x_1(t), \ldots, x_r(t))^T$ to be the total cellular biomass in each of the r size classes, at time t. Consequently, denoting $\mathbb{1} = (1, \ldots, 1)^T$ and $U_t = \mathbb{1}^T x_t$, one has the total cellular biomass in the system at time t.

In [8,13], all size classes are alike. However, since we will need to be able to account for cellular division, we need introduce some inhomogeneities. Therefore, we suppose that the r biomass classes are divided as follows: there are r_g (*growth*) size classes, during which the dynamics is the same as in [8,13]: the cells in these classes can either proceed to the next biomass class (if they grow of a sufficient amount) or stay in the class (if their growth is not sufficient). There are no births nor divisions in these classes. Following this stable stage, there are r_d (*division*) classes, during which cell division occurs. Correspondingly, we assume that there are r_b (with $r_b = r_d$) size classes prior to the stable stage, in which the dividing cells "fall" (see Figure 1.1). Hereafter, we will call the latter classes *birth* classes.

Note that in order to be able to track cells in a satisfactory manner, we impose a one to one correspondence between division and birth classes.

Therefore, $r = r_b + r_g + r_d$ is the total number of classes, and the structure of the model can be decomposed as follows:

- $i = 1, \ldots, r_d$ birth classes.
- $i = r_d + 1, \ldots, r_d + r_g$ growth classes.
- $i = r_d + r_g + 1, \ldots, r$ division classes.

In what follows, such indices will be referred to as *absolute* indices. When dealing with growth or division classes, we will also use *relative* indices, *i.e.*, the index of growth (respectively division) classes among growth (respectively division) classes. For example, the division class with relative index d_r is the size class of absolute index $d_a = r_b + r_g + d_r$. The terms absolute and relative will be dropped if no confusion is possible.

There can possibly be no growth classes, but the number of division (and hence birth) classes has to be at least one. As will become clear later, if $r_b = r_d = 1$, then the model restricts to the one of Gage *et al*, whatever the number r_g of growth classes.

To describe the way to split the biomass interval $[b_{min}, b_{max}]$ between the r sub-intervals, we now need to define the size classes in a more precise manner. To do so, we use a constant, M, representing the size increment for a cell moving from class i to class $i + 1$. Since hypothesis (H1) must be fulfilled, M has to account for an exponential growth of the cells (if the substrate concentration is high). Suppose that a cell has an exponential growth between biomass b and $2b$. If we want to "track" this biomass as it progresses along, say n classes, then using $M = 2^{1/n}$ does the trick. Indeed, define $M^{i-1}b$ to be the biomass of the cell in class i. Then if a cell progresses from one class to the next at each time step, we have an exponential growth of the biomass from b to $2b$.

We want to allow more flexibility in the size doubling hypothesis, *i.e.*, cells need not exactly be twice their birth biomass at division. In particular, we assume that a cell can more than double its birth mass. Thus we need to choose $M > 2^{1/r}$ so that $M^r b_{min} > 2b_{min}$. We also impose a one to one correspondence between division classes and birth classes. Then M is determined as follows.

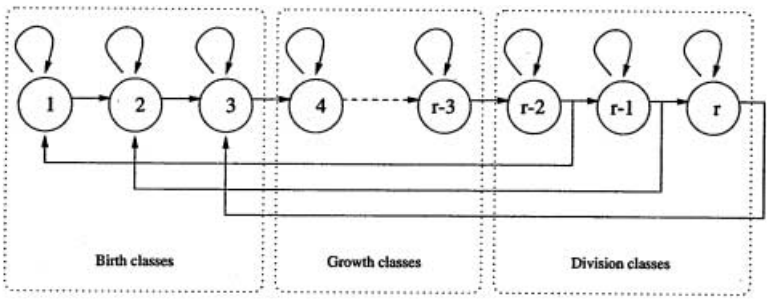

Fig. 1.1. Structure of the model. Example with $r_d = 3$ $(= r_b)$

Let d_r be the (division class relative) index of a division class. For the one to one correspondence to hold, the result of the division of a cell in class d_r must be two cells in birth class k, with $k = d_r$. Thus we must have

$$M M^{r_b+r_g+d_r-1} b_{min} = 2M^{d_r-1} b_{min}$$

The factor M in the left-hand side of this equation results from the modelling hypothesis that division follows growth, which will be discussed in more detail in Section 1.3. Now in particular, this equation has to be true for $d_r = k = 1$, so we must have

$$M^{r_b+r_g+1} = 2$$

so finally

$$M = 2^{1/(r_b+r_g+1)}$$

and the mean biomass of cells in class i is given by

$$M^{i-1} b_{min}$$

1.2. Description of cellular growth

What we have so far is a static description of cellular size, linking the size of a cell with its position among the size classes. But this was done using the assumption that $S = +\infty$ (*i.e.*, in unlimiting growth conditions). Now in practice, the substrate concentration in the chemostat is limiting. So we now need to determine how cells effectively grow, *i.e.*, since the model is discrete, to determine the proportion of cells in one size class which will move to the next size class, given a certain substrate concentration. It is our feeling that the obtention of the proportion of growing cells should be a little more detailed than it is in [8, 13], so we proceed here, using a slightly different justification as well.

We first need to define a growth function, to model nutrient uptake and consequent cell growth. This function, noted f, is taken to be an increasing, bounded function of S, *i.e.*, that it verifies $f(0) = 0$, $f'(S) > 0$ and $f''(S) < 0$. Since this will come of use later, we define $m = \lim_{S\to\infty} f(S)$, the maximal growth rate per iteration period.

Now since the biomass classes have a certain width, all the cells that belong to one biomass class do not have the exact same biomass. Therefore, we have to determine what fraction of the biomass in a class corresponds to cells that grow enough to become members of the next size class, and what fraction does not. In order to do so, we proceed as follows. Suppose that there is no washout (*i.e.*, $E = 0$), and consider class i at time t. The number of cells in this class is given by $n_i(t) = x_i(t)/(M^{i-1} b_{min})$. This number stays the same after growth, but a certain number $n_i^s(t)$ of cells will stay in the class, while a number $n_i^p(t)$ will progress to the next class. This can be written

$$n_i(t) = n_i^s(t) + n_i^p(t)$$

Now we can also write that following growth, the biomass $x_i(t)$ has become

$$\bar{x}_i(t+1) = n_i^s(t)M^{i-1}b_{min} + n_i^p(t)M^i b_{min}$$

Remark that $\bar{x}_i(t+1) \neq x_i(t+1)$, since it neglects the biomass moving in class i from class $i-1$. This equation states that the new biomass in class i at time $t+1$ is the sum of the biomass of the cells which have stayed in class i and of the cells which have progressed to class $i+1$. Using this equation, we can then easily compute the number $n_i^s(t)$:

$$n_i^s(t) = \frac{\bar{x}_i(t+1) - n_i^p(t)M^i b_{min}}{M^{i-1}b_{min}}$$

Now adding n_i^p to both sides of this equation and using $n_i(t) = x_i(t)/(M^{i-1}b_{min})$, we find that

$$\frac{x_i(t)}{M^{i-1}b_{min}} = \frac{\bar{x}_i(t+1) - n_i^p(t)M^i b_{min}}{M^{i-1}b_{min}} + n_i^p(t)$$

Therefore,

$$n_i^p(t) = \frac{\bar{x}_i(t+1) - x_i(t)}{M^{i-1}(M-1)b_{min}}$$

In this equation, $\bar{x}_i(t+1) - x_i(t)$ is the increase of biomass in class i due to cell growth. This is given by $x_i(t)f(S_t)$. Hence we have that the number of cells of class i that will move to class $i+1$ after growth is given by

$$n_i^p(t) = \frac{x_i(t)f(S_t)}{M^{i-1}(M-1)b_{min}}$$

Finally, the proportion of cells in class i moving to class $i+1$ is the ratio of the number of passing cells to the total number of cells in class i:

$$P_t = \frac{n_i^p(t)}{n_i(t)} = \frac{f(S_t)}{M-1}$$

1.3. Description of cellular division

We now turn our attention to the truly novel part of this model: the description of cellular division. Indeed, the model of Gage *et al.* supposes that, in our notations, $r_b = r_d = 1$, so that division occurs for cells in the last size class, *i.e.*, for cells with a given biomass.

On the contrary, we want (H2′) to be accounted for. So we suppose that cells that are in a division class and grow sufficiently, can either divide, with a certain proportion, or proceed to the next division class. Furthermore, we suppose that this proportion is a function of the substrate concentration, and that it is size dependent (*e.g.*, one could assume that bigger cells divide even in low substrate concentrations, while smaller cells do not). The method used to model this is depicted in Figure 1.2: one first determines the proportion of cells in a given division class

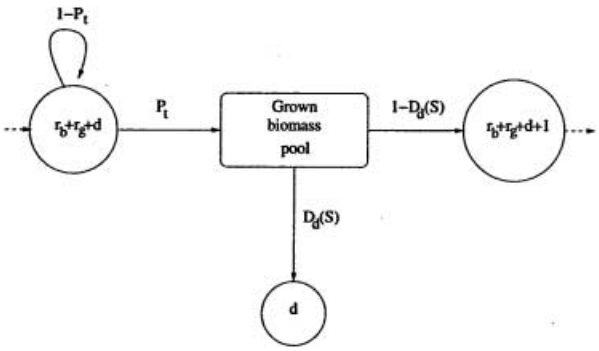

Fig. 1.2. A schematic representation of division in division class d (indices shown are absolute indices). The bottom class labelled d is birth class d, while the upper two classes are division classes

(division class relative index d in the figure) that would proceed to the next size class, using the function P_t obtained in Section 1.2. Among this pool of grown cells, a certain proportion of cells will effectively divide, each yielding two cells in (birth) size class d, while the rest will proceed to the next (division) size class.

Hence let us denote by $D_i(S)$ the proportion of cells in division class i that divide. Then $1 - D_i(S)$ is the proportion of cells of class i that grow instead of dividing. The way these functions depend on the substrate concentration will not be specified at this point, but one can think of Holling type II functional response, or of sigmoidal functions. All that we require is that these functions be continuous and bounded, that is for $i = 1, \ldots, r_d - 1$ and for all $S \in \mathbb{R}_+$, $0 \leq D_i(S) \leq 1$, where both inequalities have to be strict for some S. This can be formulated the following way: we suppose that there exists a nonempty subset $S_{int} \subset \mathbb{R}_+$, $S_{int} \neq \{0\}$, defined by

$$S_{int} = \{\bar{s} \in \mathbb{R}_+ : \forall i < r_d, D_i(\bar{s}) \in (0, 1)\} \tag{1.1}$$

Note that we do not require that S_{int} be a connected set. Figure 1.3 shows the possible nature of the S^i_{int} set for a given division class i. The set S_{int} then consists of the intersection of all such sets (for all division classes except the last). This subset will be useful in Theorem 2.5.

In order to constrain the cell sizes, we suppose that in the last division class all cells divide, *i.e.*, that $D_{r_d}(S) = 1$ for all S.

From a biological point of view, these hypotheses mean that there exist substrate concentrations such that: the only size class in which all cells divide is the last one; there are no division classes in which no cells divide.

There is indirect biological evidence underlying the hypothesis that cell division size is substrate dependent. In a recent study, Sciandra *et al.* have carried out experiments on *Cryptomonas* sp. [12], in which they show that, depending on the substrate concentration, the mean cell size (diameter) differs. A counter intuitive result in this paper is that the mean cell size decreases as the input substrate concentration increases. This confirms the results of Turkia and Lepistö [16], who carried

measurements of the diatom *Aulacoseira* Thwaites in Finnish lakes, and found similar correlations between the mean cell size and nutrient concentrations. Although this phenomenon has yet to be explained, it does indeed prove that cell size varies with substrate concentration. Supposing that the size at division is a function of the substrate density is then a natural consequence of this observation.

The type of division functions that we hypothesise will be shown in more detail in Section 3 (numerical simulations), and identification of the parameters of these functions, in the steady state case, will be carried out in Section 4.

1.4. The model

When constructing the parts of the model dealing with cell size, cell growth and cell division, we could suppose that there was no washout, since these processes are independent of it. Now to state the model in its definitive form, we need take the latter into account. As in [8], we suppose that dilution influences substrate and cells alike, so the model is written as follows, for $t \geq 0$:

$$x_{t+1} = (1 - E)A(S_t)x_t \tag{1.2a}$$

$$S_{t+1} = (1 - E)[S_t - f(S_t)U_t] + ES^0 \tag{1.2b}$$

with initial conditions $x_0 \in \mathbb{R}^n_+$ and $S_0 \in \mathbb{R}_+$. In (1.2), $U_t = \mathbb{1}^T x_t$ is the total biomass at time t ($\mathbb{1} = (1, \ldots, 1)^T$), and $A(S_t)$ is a $r \times r$ transition matrix given by Table 1.1.

This matrix has $1 - P_t$ on its diagonal. Its sub-diagonal is MP_t in the birth and growth blocks, and $MP_t(1 - D_i)$ in the division block. Finally, there is an upper diagonal part consisting of MP_tD_i, describing the flow of biomass from the division classes to the corresponding birth classes, *i.e.*, the birth process.

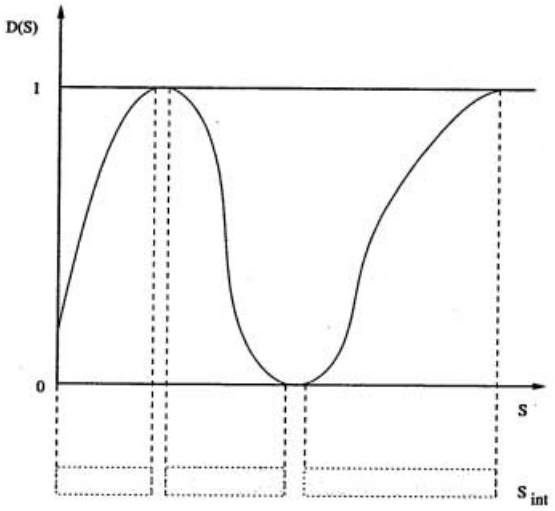

Fig. 1.3. A possible S_{int} set (for a given division class i)

Table 1.1. The transition matrix $A(S_t)$. The notations $P_t = f(S_t)(M-1)^{-1}$ and $D_i = D_i(S_t)$ have been used for clarity. The diagonal blocks are square, with sizes r_b, r_g and r_d.

$$
A(S_t) = \left[\begin{array}{ccc:ccc:ccc}
1-P_t & 0 & & & & & MP_tD_1 & & \\
MP_t & 1-P_t & & & & & & MP_tD_2 & \\
0 & \ddots & \ddots & & & & & \ddots & \\
& & 1-P_t & & & & & & MP_tD_{r_d-1} \\
& & & & & & & & MP_t \\
\hdashline
& & MP_t & 1-P_t & & & & & \\
& & & \ddots & \ddots & & & & \\
& & & & MP_t & 1-P_t & & & \\
\hdashline
& & & & & MP_t & 1-P_t & & \\
& & & & & & MP_t(1-D_1) & 1-P_t & \\
& & & & & & & \ddots & \ddots \\
& & & & & & & & 1-P_t \\
& & & & & & & & MP_t(1-D_{r_d-1}) \quad 1-P_t
\end{array}\right]
$$

In order to keep P_t in a reasonable interval (*i.e.*, $P_t \in [0, 1]$), some restrictions have to be made. The minimal doubling time D_{min} and the maximal growth rate μ_{max} (the continuous time equivalent of m) are two expressions of the same quantity (since $D_{min} = \ln 2/\mu_{max}$), and are species constants. So they cannot be tampered with.

Now, what happens in general cases? We are given a μ_{max} (or equivalently a D_{min}), we choose a number of classes, and want to study the behaviour of the system. Hence the determination of T is what comes last, and it is this quantity that we must constrain.

The first constraint to be verified is that $m(M-1)^{-1} \leq 1$. Therefore we obtain $T \leq \frac{2^{1/(r_b+r_g+1)}-1}{\mu_{max}}$. But the constraint $r_{double}T \leq D_{min}$ expressed in [13] also has to be satisfied, with r_{double} the number of classes that a cell has to span for its biomass to double. In [13], $r_{double} = r$, the total number of classes, while here $r_{double} = r_b + r_g + 1$. Therefore, written in terms of μ_{max} and adapted to our model, this condition reads $T \leq \frac{\ln 2}{(r_b+r_g+1)\mu_{max}}$.

Hence, supposing μ_{max} given and the number of classes chosen, we must require that

$$T \leq \min\left(\frac{2^{1/(r_b+r_g+1)} - 1}{\mu_{max}}, \frac{\ln 2}{(r_b + r_g + 1)\mu_{max}}\right) \tag{1.3}$$

Now another problem which could arise is that $U_t f(S_t) > S_t$ for some t. This would lead to negative values of S_{t+1}, which must of course be forbidden. In order to avoid this problem, U_t has to be constrained. We proceed as in [13]. We fix an $\eta \in (0, 1)$, and require that $f(S)U/S < \eta$. Defining $W > S^0$ an upper bound on $U + S$ that depends on the range of initial conditions that one wishes to accommodate, we require that

$$f'(0)W < \eta \tag{1.4}$$

Since we have imposed that $f'' < 0$, we have that $f'(s) \leq f'(0)$ for all $s \geq 0$, and this last assumption thus sets a bound on the maximal growth rate.

Observe that these conditions should not be viewed as conditions on m, but only on the iteration period T and the number of classes r.

2. Model behaviour

Let us begin by verifying that the *conservation principle* (or *mass conservation principle*) holds. We have

$$\mathbb{1}^T A(S) = [1 + (M-1)P_t]\mathbb{1}^T = (1 + f(S_t))\mathbb{1}^T$$

therefore

$$U_{t+1} = \mathbb{1}^T x_{t+1} = \mathbb{1}^T A(S_t)x_t = (1 - E)(1 + f(S_t))U_t$$

As a consequence,

$$U_{t+1} + S_{t+1} = (1 - E)(U_t + S_t) + ES^0 \tag{2.1}$$

which means that mass conservation is verified. In words all the mass that is present in the chemostat at time t is still present at time $t+1$, save for what enters (ES^0) and for the quantity $E(U_t + S_t)$ that is washed out.

Hence all the results in the single population case of [13] hold. The method that is used is the following.

- Since mass conservation is verified, it is quite easy to show that the total "mass" contained in the chemostat tends to a fixed quantity, namely S^0.
- Therefore, the dynamics of the system can be studied on the invariant set $U+S = S^0$, inside of which the system reduces to a one dimensional system.
- On this set, the global behaviour of the system can be studied. Smith showed that under certain conditions, there exists a globally stable equilibrium for this simplified system.
- Following this, the global dynamics of the two dimensional system in substrate and total biomass can be deduced. This system is shown to admit, under the same conditions as the previous one dimensional system, a globally stable non trivial equilibrium.
- Finally, using a result of Golubitsky *et al.* [9], the distribution of biomass in each one of the size classes can be deduced.

The proofs of the results will not be shown here, since they are identical to the ones of Smith. The interested reader should therefore refer to [13] for a complete derivation of the results.

First, we can solve (2.1) easily, and obtain

$$U_{t+1} + S_{t+1} = S^0 - (S^0 - U_0 - S_0)(1-E)^t, \quad t \geq 1 \tag{2.2}$$

Let us now define Γ, a positive bounded set, as

$$\Gamma = \{(x, S) \in \mathbb{R}_+^{r+1};\ \mathbb{1}^T x + S \leq W\}$$

where W is defined as in equation (1.4). We then have the following result concerning the total mass (both organic and inorganic) contained in the chemostat.

Proposition 2.1 ([13]). *If* $(x_0, S_0) \in \Gamma$, *then* $(x_t, S_t) \in \Gamma$ *for* $t \geq 1$, $S_t - U_t f(S_t) > 0$ *for* $t \geq 1$ *and*

$$S_t + U_t \to S^0, \quad t \to \infty \tag{2.3}$$

Let us now consider the system restricted to the positively invariant set $\{(U, S) \in \mathbb{R}_+^2;\ U+S = S^0\}$. On this set, we can use a technique standard to those systems that conserve the mass: we replace S by $S^0 - U$ (with $0 \leq U \leq S^0$ since S is positive). Thus, when restricted to this set, the system (1.2) becomes 1-dimensional:

$$U_{t+1} = (1-E)(1 + f(S^0 - U_t))U_t \tag{2.4}$$

and we have the following result.

Proposition 2.2 ([13]). *If* $(1-E)(1+f(S^0)) \leq 1$, *then* $\lim_{t\to\infty} U_t = 0$, *for all solutions of (2.4) with* $U_0 \in [0, S^0]$. *If* $(1-E)(1+f(S^0)) > 1$, *then* $\lim_{t\to\infty} U_t = \tilde{U}$, *for all solutions of (2.4) with* $U_0 \in [0, S^0]$.

Let $F(U) = (1 - E)(1 + f(S^0 - U))U$. In order to determine the value of $\tilde{U}$, one then has to compute the positive fixed point of F. Therefore, we define λ as the unique solution, when it exists, of $F(U) = U$, *i.e.*,

$$f(\lambda) = (1 - E)^{-1} - 1 \tag{2.5}$$

Hence $\lambda = f^{-1}((1 - E)^{-1} - 1)$. Recall that we noted $m = f(\infty)$ the maximal growth rate of the organisms, then $\lambda < \infty$ if $(1 - E)^{-1} - 1 < m$, and $\lambda = \infty$ otherwise. This is the classical chemostat behaviour: if the dilution rate E is larger than the maximal growth rate of the organisms, then the population cannot compensate the loss due to the outflow, and it becomes extinct. Finally, if $\lambda < S^0$, $\tilde{U}$ is given by

$$\tilde{U} = S^0 - \lambda \tag{2.6}$$

Now the dynamics of the 2 dimensional system

$$U_{t+1} = (1 - E)(1 + f(S_t))U_t \tag{2.7a}$$

$$S_{t+1} = (1 - E)(S_t - f(S_t)U_t) + ES^0 \tag{2.7b}$$

can be studied. Let $\Omega = \{(U, S) \in \mathbb{R}^2_+;\ U + S < W\}$.

Theorem 2.3 ([13]). *If $(1 - E)(1 + f(S^0)) < 1$, then for all solutions of (2.7) such that $(U_0, S_0) \in \Omega$,*

$$(U_t, S_t) \to (0, S^0), \quad t \to \infty$$

If $(1 - E)(1 + f(S^0)) > 1$, then there exists a non zero steady state, and for all solutions of (2.7) such that $(U_0, S_0) \in \Omega$,

$$(U_t, S_t) \to (S^0 - \lambda, \lambda), \quad t \to \infty$$

Now that we know the global behaviour of the 2 dimensional system, we can use the following result of Golubitsky *et al.* to derive the equilibrium distribution of the x_i's.

Theorem 2.4 ([9]). *Suppose that T_k is a sequence of nonnegative primitive matrices, and that $T_k \to T$ as $k \to \infty$, where T is also nonnegative and primitive. If e is the Perron-Frobenius eigenvector of T satisfying $\mathbb{1}^T e = 1$ and $\xi_{k+1} = T_k \xi_k$ is a sequence starting with $\xi_0 \geq 0$ and $\xi_0 \neq 0$, then*

$$\frac{\xi_k}{\mathbb{1}^T \xi_k} \to e, \quad k \to \infty$$

Therefore, we have the following result.

Theorem 2.5. *Let e be the Perron-Frobenius eigenvector of $(1 - E)A(\lambda)$ satisfying $\mathbb{1}^T e = 1$. If $(1 - E)(1 + f(S^0)) > 1$, $x_0 \neq 0$ and $\lambda \in S_{int}$, then the system (1.2a) (1.2b) admits one globally asymptotically stable non trivial equilibrium $(\tilde{x}, \tilde{S})$, where*

$$\frac{\tilde{x}}{\tilde{U}} = e \tag{2.8}$$

Proof. From Theorem 2.3 we know that $(U_t, S_t) \to (S^0 - \lambda, \lambda)$ as t tends to ∞. Therefore, there exists $\tau > 0$ and a neighbourhood $\mathcal{N}_1(\lambda)$ of λ such that, $\forall t \geq \tau$, $S_t \in \mathcal{N}_1(\lambda)$.

Now, under the hypotheses of the theorem, $\lambda \in S_{int}$. Since the functions $D_i(S)$ are continuous, there exists a neighbourhood $\mathcal{N}_2(\lambda)$ of λ such that for all $S \in \mathcal{N}_2(\lambda)$, $S \in S_{int}$. We can then choose τ sufficiently large so that $\mathcal{N}_1(\lambda) \subset \mathcal{N}_2(\lambda)$. Such a value of τ having been determined, we have for all $t \geq \tau$ and all $i < r_d$, $0 < D_i(S_t) < 1$.

Then, for all $t \geq \tau$, the matrices $A(S_t)$ and $A(\lambda)$ are primitive. Indeed, it is easy to check that any class can be reached from any other class in a finite number of steps, yielding the irreducibility of these matrices. Since the trace of $A(S_t)$ and $A(\lambda)$ are positive, we furthermore have primitivity [3, p. 34]. Finally, as the matrices $A(S_t)$ and $A(\lambda)$ are obviously nonnegative for all $t \geq 0$, the conditions of Theorem 2.4 are fulfilled for all $t \geq \tau$, *i.e.*, for $\xi_0 = x_\tau$, and the proof is done. □

Remark 2.6. Here the reasons that have led to the introduction of the set S_{int} clearly stand out. Convergence of S to λ is a consequence of the nature of the two dimensional system (a discrete time Monod model). Indeed, even with D_i's taking values of 0 or 1 for some i's, we always have $\mathbb{1}^T A(S_t) = (MP_t+1-P_t, \ldots, MP_t+1-P_t)$, so this convergence is independent of the values of the D_i's, because of mass conservation. But concerning the convergence of the distribution, we *must* have primitivity of the matrices $A(\lambda)$ and $A(S_t)$ (for S_t close to λ), and hence we must have $\lambda \in S_{int}$.

The Perron-Frobenius eigenvalue is $(1 - E)(1 - P + MP)$ $(= (1 - E)(1 + f(\tilde{S})))$, its associated eigenvector has the following form.

$$e = \frac{1}{r} \begin{pmatrix} D_1(\tilde{S}) \\ \vdots \\ 1 - \prod_{k=1}^{r_b - 1}(1 - D_k(\tilde{S})) \\ \hline 1 \\ \vdots \\ 1 \\ \hline 1 \\ 1 - D_1(\tilde{S}) \\ \vdots \\ \prod_{k=1}^{r_d - 1}(1 - D_k(\tilde{S})) \end{pmatrix} \tag{2.9}$$

where the blocks correspond to r_b, r_g and r_d rows. It is easily verified that $(1 - E)$ $A(\tilde{S})e = (1 - E)(1 - P + MP)e$.

Let us denote by x_i^b the i^{th} birth class $(i = 1, \ldots, r_b)$, by x_i^g the i^{th} growth class $(i = 1, \ldots, r_g)$, and by x_i^d the i^{th} division class $(i = 1, \ldots, r_d)$. Using (2.8) and (2.9), the stable distribution can be computed. This is expressed in the following result.

Proposition 2.7. *Suppose that the conditions leading to the existence of a non trivial equilibrium are satisfied. Then the equilibrium biomass distribution is given by the following formulas:*

– *for* $i = 1, \ldots, r_b - 1$ *(birth classes except the last one):*

$$\tilde{x}_i^b = \left(1 - \prod_{k=1}^{i}(1 - D_k(\tilde{S}))\right)\tilde{U} \tag{2.10}$$

– *for* $i = 1, \ldots, r_g$ *(growth classes), for the last birth class and the first division class:*

$$\tilde{x}_{r_b}^b = \tilde{x}_i^g = \tilde{x}_1^d = \tilde{U} \tag{2.11}$$

– *for* $i = 2, \ldots, r_d$ *(division classes except the first):*

$$\tilde{x}_i^g = \left(\prod_{k=1}^{i-1}(1 - D_k(\tilde{S}))\right)\tilde{U} \tag{2.12}$$

3. Numerical results

Before going any further, we will here give some precisions about the functions used in the following numerical experiments. First of all, the growth function f is taken to be a Michaelis-Menten function. This function is the most widely used in the chemostat literature. It is defined by

$$f(S) = m\frac{S}{k_S + S} \tag{3.1}$$

where m is the maximal growth rate per iteration period, and k_S is called the half-saturation constant.

Second and most important, let us give some more details about division proportions functions. For the simulations showed hereafter, we have used functions such as the ones shown in Figure 3.1. In order to cut down on the number of parameters in this system, we assume that the division proportions functions (with respect to size) follow a Gaussian distribution $\mathcal{N}(\mu(S), \sigma(S))$. Since the division proportions are substrate dependent, and in order to take into account the biological evidence discussed in Section 1.3, we have supposed that the mean and variance of these distributions are substrate dependent. For example in Figure 3.1 (where size is expressed in cell diameter), we have used a mean and standard deviation varying from $\mu = 13$ and $\sigma = 0.5$ when substrate concentration is zero, to $\mu = 9$ and $\sigma = 2.5$ when $S = S^0 = 260$, using affine variation of μ and σ between these two values of S.

One can see that using this type of division proportions, we will have bigger cells when the substrate concentration is low, and smaller cells as the substrate concentration rises.

Figure 3.2 shows a comparison of the stable distributions of numbers (computed using the approximation $n(i) = x_i/(M^{i-1}b_{min})$) as given by the homogeneous model of Gage *et al.* and by our model. In this example, we have assumed that the mean division (and consequently birth) biomass is located in the middle of the division classes.

Figure 3.3 shows the transient behaviour of the model. We can see that the biomass in each size class has damped oscillations prior to the equilibrium. The biological parameters used in this computation are the ones corresponding to the data set (and to the algae species) that will be shown in the next section.

Figure 3.4 shows the substrate concentration and total biomass, as well as the normalised number of cells, corresponding to the same simulation. The number of cells oscillates, and this behaviour lasts until the biomass distribution has reached its equilibrium

The transient behaviour of the system can be studied as in [8], by defining the wavelength of complex eigenvalue λ_i as

$$\omega_i = 2\pi T / \tan^{-1}\left(\frac{\Im(\lambda_i)}{\Re(\lambda_i)}\right)$$

with T the time step, $\Im(\cdot)$ and $\Re(\cdot)$ the imaginary and real parts respectively, of λ_i. In Figure 3.5, two numerical experiments are shown. On the left hand side, the wavelength of the complex eigenvalue of greatest magnitude is plotted, for an increasing number of division classes, while the total number of classes is constant. We can see that as the proportion of classes that are division classes rises, the wavelength of the first complex eigenvalue decreases.The right hand side of the figure shows that with a given proportion of the classes being division classes, the wavelength of the complex eigenvalue first decreases, then stabilises.

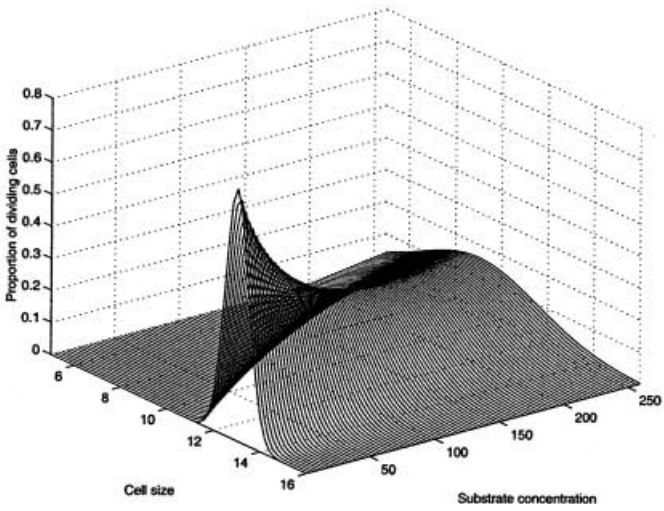

Fig. 3.1. An example of division proportions functions. Cell size is here expressed in diameter

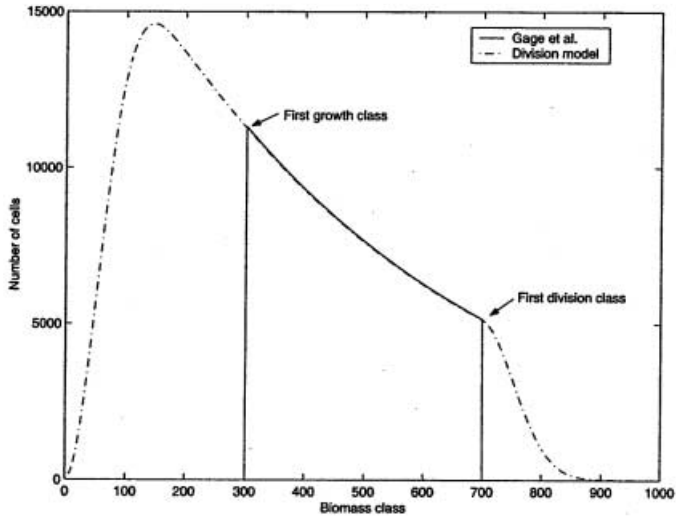

Fig. 3.2. Comparison of the equilibrium distributions (converted to numbers), as given by the two models. This example uses $r_b = r_d = 300$ and $r_g = 400$. The vertical lines show the passage from one type of class to another

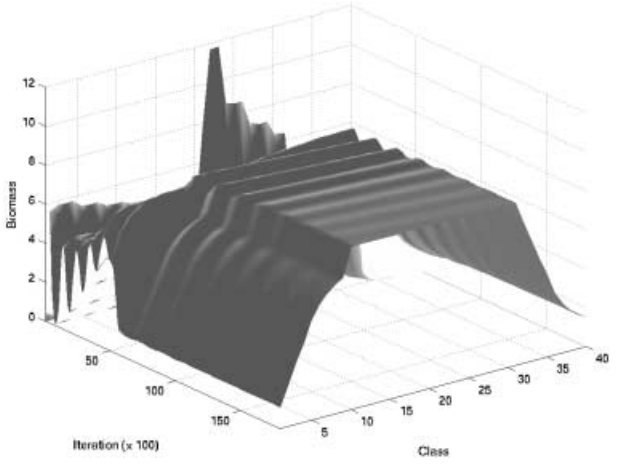

Fig. 3.3. Transient behaviour of the model following an initial condition of Dirac type in class 20 (of a total of 40). The parameters are $T = 0.001$, $r_b = r_d = 10$ and $r_g = 20$. The figure shows a sampling (every 100 iterations) out of a total 18000 iterations (corresponding to 18 days)

Hence it seems that the proportion of classes in which cells divide (and accordingly are born) is more determinant for the transient behaviour of the system than the total number of classes. Interpretation of this fact is rather straightforward: the more division classes (and accordingly birth classes), the faster the "dispersion" of cells.

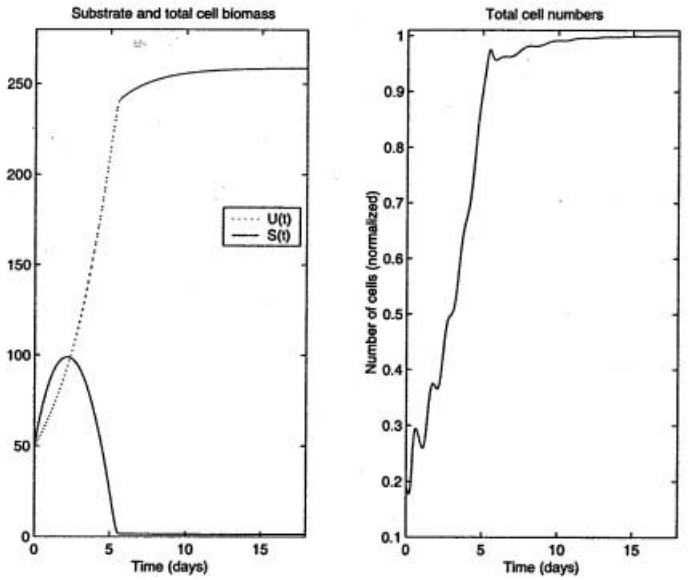

Fig. 3.4. Behaviour of the total biomass and of the substrate (left), and of the total number of cells (right), under the same conditions as in Figure 3.3

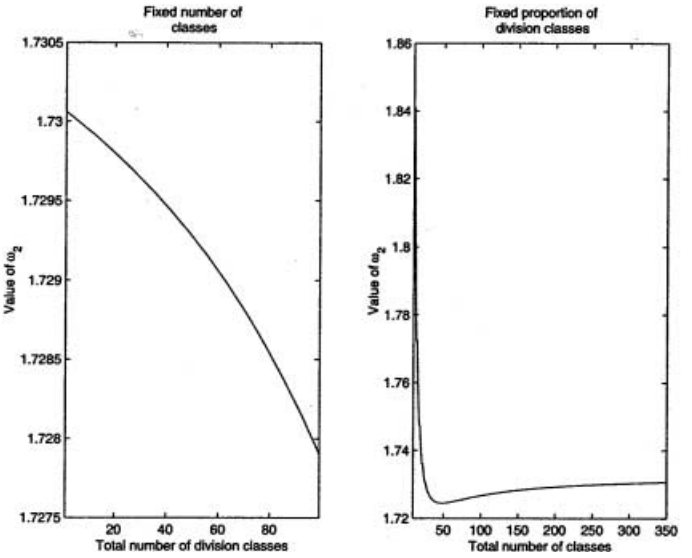

Fig. 3.5. Wavelength of the first complex eigenvalue: (left) for a fixed total number of classes ($r = 200$), as the number r_d of division classes progresses from 1 to 100; (right) for a fixed proportion (1/4) of division classes, as the total number r of classes progresses from 10 to 300

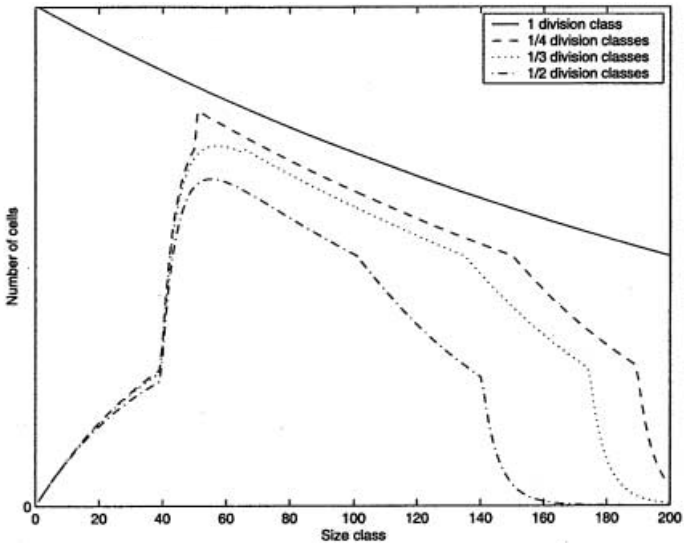

Fig. 3.6. Comparison of distributions obtained for 200 size classes, when the proportion of these classes that are division classes changes: one division class, $1/4 = 50$ division classes, $1/3 = 66$ division classes and $1/2 = 100$ division classes

The proportion of division classes also influences the equilibrium distribution, as can be seen in Figure 3.6. In this figure, we show, for a fixed total number of classes, the shape of the distribution (converted to numbers) as the proportion of division classes among the total number of classes rises. The plain curve corresponds to the prediction of the model of Gage *et al*. Then, as the proportion of division classes grows, the distribution becomes more and more narrow.

4. Comparison to experimental data

As was mentioned in the introduction, we have access to very long time series of structured data from a chemostat, such as the dataset shown in Figure 4.1. This data corresponds to the algae species *Cryptomonas*, grown in limiting conditions on nitrate. The maximal growth rate of this species is $\mu_{max} = 0.7$ day^{-1}, corresponding to a minimal doubling time of approximately one day, and the half saturation constant is 1 μMol. The dilution rate in this experiment was 0.4 day^{-1}, and the input concentration was 260 μMol.

Let us now briefly mention how this data is obtained. The chemostat is fully automated, both on the operating side and on the measurement side [5]. The dilution rate and the input nutrient concentration are computer controlled; this allows fluctuating inputs, as well as long time, nearly unattended functioning.

But the main feature of this device is that it allows nearly continuous monitoring of several variables: substrate concentration (by colorimetric methods, using a Technicon Auto-analyser), cell size and number distributions (by means of a particle counter (HIAC/ROYCO with laser sensor HRLD 400)), and even chlorophyll

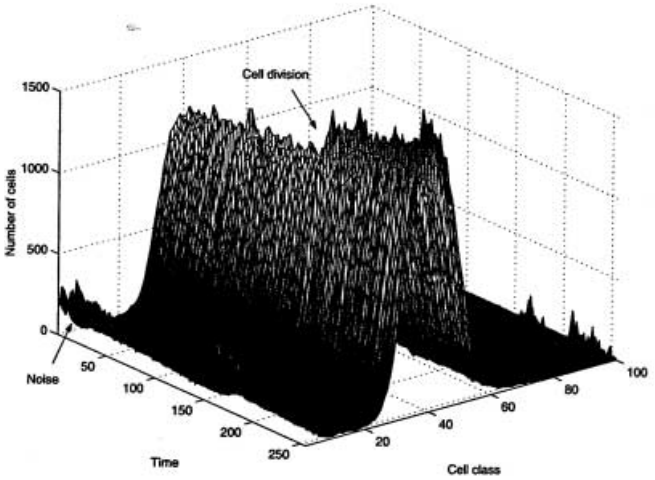

Fig. 4.1. Sample data set of *Cryptomonas sp*. This series represents three weeks of chemostat culture. There are 100 equally sized size classes, containing cells with diameters ranging from $1.5\mu m$ to $15\mu m$

concentration (using a spectrophotometer). All of these measures are computer controlled. Typical experiments last between one and four months, resulting in sets of one thousand to four thousand measures (each measure consisting of a size spectrum).

We now turn to the comparison of this data to the model, and to the identification of some of the model parameters.

As was mentioned before, parameters k_S and m of the Michaelis-Menten function (3.1) are known, since they are species constants. The dilution rate E is given for each experiment. These values are recalled in the following table:

Parameter	Unit	Value used
S^0	μM or μgat.l^{-1}	260
$D = E/T$	day^{-1}	0.4
$\mu_{max} = m/T$	day^{-1}	0.7
a	μMol	1

Since the equilibrium is stable, and that (2.10) (2.11) and (2.12) are given, we can at first restrict the identification problem to that of the identification of the steady state parameters.

Several parameters have a determinant influence on the steady state distribution. Here, we will be concerned with the determination of only a small number of them. Indeed, the most important part is the determination of the division proportion function at the steady state. However, without any prior hypotheses concerning this distribution, we would have to determine the value of a possibly large number of states (r_d). This is unrealistic when carrying out an identification process.

Hence, we will use the same assumption that was used in the previous section: we suppose that the division proportions (with respect to size) follow a Gaussian distribution. Since we are working at the steady state, S is very close to λ, and thus

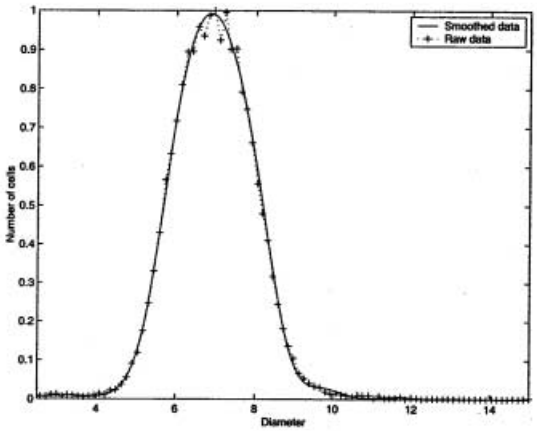

Fig. 4.2. Comparison between a smoothed data set and the original data set

we only have to identify the mean $\mu(\lambda)$ and the variance $\sigma^2(\lambda)$ of this distribution. The steady state biomass distribution is also characterised by the numbers r_d (or equivalently r_b) and r_g of division and growth classes, as well as the minimal birth biomass b_{min}.

In practice, the data is expressed in number of cells per diameter class. Hence we convert the predictions of the model into such units, and so identify the parameters μ_{diam}, σ_{diam} and d_{min}, diameter equivalents of the previous biomass quantities.

Suppose that $d_{i,m}$ is the value given by the model and $d_{i,d}$ the value of the data, for cells of size i. The set of possible sizes I is the reunion of the set $[d_{min,d}; d_{max,d}]$ of observed diameters in the data, and of the set $[d_{min,m}; d_{max,m}]$ of the diameters given by the model. The parameter identification problem can then be stated as follows.

(IP) Let Θ be the set of admissible parameter values. Find $\theta \in \Theta$ such that the functional

$$J(\theta) = \sum_{i \in I} \|d_{i,m} - d_{i,d}\|$$

is minimal.

There are constraints for the parameters. The minimal birth diameter d_{min} and the mean division diameter μ_{diam} must both belong to the set of observed diameters $[d_{min,d}, d_{max,d}]$. The minimisation procedure then is applied to smoothed data, using MATLAB's constrained nonlinear minimisation function `fmincon`. The reason for which smoothed data is used rather than the original, unsmoothed data, is that optimisation procedures require one to use a rather smooth "objectives" function. Smoothing of the data was carried out using B-spline functions (see [1]). One can see in Figure 4.2 that this smoothing preserves the main characteristics of the data distributions.

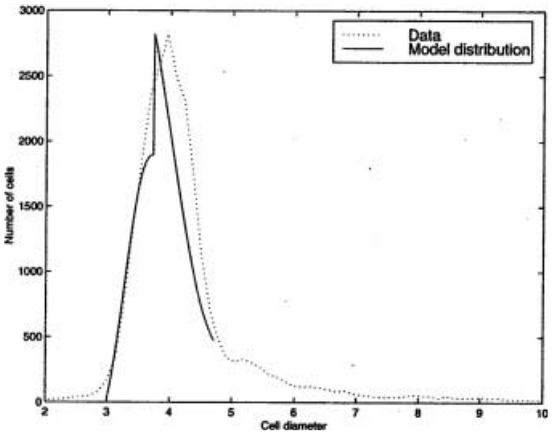

Fig. 4.3. Comparison of smoothed data and of the result of a parameter identification procedure (results being converted into number of cells per diameter class)

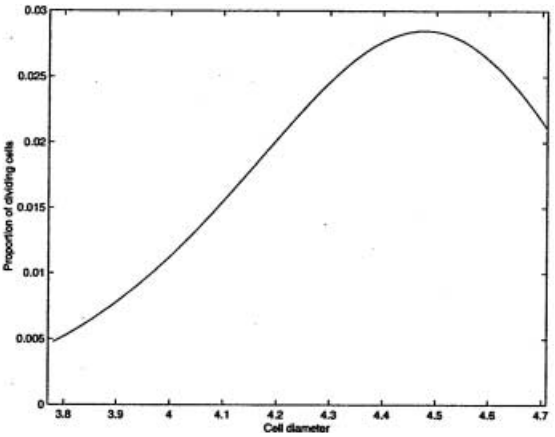

Fig. 4.4. Distribution of the division proportions obtained using parameters of the division function noted in the text

Figure 4.3 shows the result of an identification procedure on a data set, using the Euclidean norm. The obtained parameters are a total number of classes $r = 120$, with $r_d = 59$ classes (hence only one growth class). d_{min} was in this case found equal to 3, which was the lower bound that we had set on the birth diameter in the identification procedure. Parameters of the division proportion function were found to be $\mu_{diam} = 4.2616$ and $\sigma_{diam} = 14$, yielding the division function shown in Figure 4.4.

5. Discussion

We have shown that the introduction of size and substrate dependent cell division in the model of Gage *et al.* does not fundamentally modify its dynamical behaviour.

Since our model sums to a Monod type model, we are able following the analysis of [13], to prove the existence of a globally stable non trivial equilibrium. This is done by first studying the dimension one system consisting of the total mass (inorganic – substrate – and organic – total biomass –) in the chemostat, which is shown to have a globally stable equilibrium. From this knowledge, the global behaviour of the two dimensional system consisting of the total biomass and of the substrate is deduced, and the existence of a non trivial, globally stable equilibrium is also proved. Then using a weak ergodic theorem of Golubitsky *et al.*, we are able to derive the distribution of biomass in the size classes.

Then, we have presented several numerical simulations. These simulations show that the ratio of division classes to the total number of classes is determinant of the shape of the equilibrium distribution. The proportion of division classes also has an impact on the transient behaviour of the model: as it increases, the frequency of the oscillations in individual size classes reduces, as well as their duration. From a modelling point of view, the comparison of Figures 3.3 and 3.4 pleads for the use of structured models of the chemostat. Indeed, consider the total biomass curve in Figure 3.4. It is monotonically increasing, while in Figure 3.3 we observe damped oscillations: the summation hides the more complicated behaviour of the individual biomass classes.

Then in Section 4, we have briefly presented an automated chemostat, and have run a parameter identification procedure on data produced by this device. This procedure was, at present, limited to the identification of parameters of the system when the steady state distribution is reached. The results obtained are not very satisfactory, at least in a visual sense (as can be seen in Figure 4.3). But they raise an interesting question. Remember that we mentioned in Section 1.1 that the maximal biomass of a cell in the system is b_{max}. It is an easy constatation that it allows multiplication of a cell's biomass by a factor of at most 4. Indeed, a cell born in birth class 1 and dividing in division class r_d has a biomass going from b_{min} to $b_{max} = M^{r_b+r_g+r_d} b_{min}$ at the instant of its division. Since $M = 2^{1/(r_b+r_g+1)}$, this means that $b_{max} < 4b_{min}$ (b_{max} being maximal when $r_g = 0$). Now convert this maximal fourfold increase to diameter (considering that biovolume is a good approximation of cellular biomass). This gives a maximal multiplication of the cell diameter of $4^{1/3} \simeq 1.59$, while our data roughly suggests a diameter interval of 4 to 10 micrometers, which represents a multiplication by a factor of 2.5. While the size increase our model allows is better than that of the model of Gage *et al.* (which accounts for a mere $2^{1/3} \simeq 1.26$ multiplication of the cell diameter), it is still far from the experimentally observed range of cell diameters.

The obvious conclusion that one draws from this is that biological hypothesis (H3) – division of a mother cell into daughter cells of the same size – should be relaxed. Indeed, if cell division results in daughter cells of different sizes (by *asymmetric cellular division*), then the range of biomass (or equivalently of diameters) accounted for in the model will be much wider. While this behaviour has been often modelled in other contexts, such as cancerous cell populations [2], we know of only one work by Heijmans in [10] in the phytoplankton context. Introduction of asymmetric cell division in our model should therefore allow us to gain more insight on the population dynamics of microorganisms in a chemostat. This, however,

requires a complete and thorough rethinking of the model, and will be the object of further work. In our opinion, it is however a strength of the present model to show that inhomogeneous cell division size *is not sufficient* to account for experimentally observed cell size distributions.

To conclude this discussion, let us note that [13] introduced competition between two species in the chemostat. Part of this was raised to the n species case by Smith and Zhao [15]. We have not considered here the competitive case. But we do expect that if we were to do so, the results of the two previously mentioned papers should hold, thus leading to a competitive exclusion situation.

Acknowledgements. We thank the editor and two anonymous referees for their valuable remarks.

References

1. Arino, J.: *Modélisation structurée de la croissance du phytoplancton en chemostat*. PhD thesis, Université Grenoble **1**, (2001)
2. Arino, O., Kimmel, M.: Asymptotic analysis of a cell cycle model based on unequal division. *SIAM J. Appl. Math.*, **47**, 128–145 (1987)
3. Berman, A. Plemmons, R.: *Nonnegative matrices in the mathematical sciences*, **9** of *Classics in Applied Mathematics*. SIAM, 1994
4. Bernard, O., Gouzé, J.-L.: Non-linear qualitative signal processing for biological systems: application to the algal growth in bioreactors. *Mathematical Biosciences*, **157**, 357–372 (1999)
5. Bernard, O., Malara, G., Sciandra, A.: The effects of a controlled fluctuating nutrient environment on continuous cultures of phytoplankton monitored by computers. *Journal of Experimental Marine Biology and Ecology*, **197**, 263–278 (1996)
6. Cushing, J.M.: A competition model for size-structured species. *SIAM J. Appl. Math.*, **49**(3), 838–858 (1989)
7. Cushing, J.M.: *An introduction to structured population dynamics*, **71** of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1998
8. Gage, T.B., Williams, F., Horton, J.: Division synchrony and the dynamics of microbial populations: A size-specific model. *Theoretical Population Biology*, **26**, 296–314 (1984)
9. Golubitsky, M., Keeler, E.B., Rothschild, M.: Convergence of the age structure: applications of the projective metric. *Theoretical Population Biology*, **7**, 84–93 (1975)
10. Metz, J., Diekmann, O.: editors. *The dynamics of physiologically structured populations*, volume **68** of *Lecture Notes in Biomathematics*. Springer-Verlag, 1986
11. Pascual, M., Caswell, H.: From the cell cycle to population cycles in phytoplankton-nutrient interactions. *Ecology*, **78**(3), 897–912 (1997)
12. Sciandra, A., Lazzara, L., Claustre, H., Babin, M.: Responses of growth rate, pigment composition and optical properties of *Cryptomonas* sp. to light and nitrogen stresses. *Marine Ecology Progress Series*, **201**, 107–120 (2000)
13. Smith, H.L. A discrete, size-structured model of microbial growth and competition in the chemostat. *J. Math. Biol.*, **34**, 734–754 (1996)
14. Smith, H.L., Waltman, P.: *The theory of the chemostat. Dynamics of microbial competition.*, volume **13** of *Cambridge Studies in Mathematical Biology*. Cambridge University Press, 1995

15. Smith, H.L., Zhao, X.-Q.: Competitive exclusion in a discrete-time, size-structured chemostat model. *Discrete Contin. Dyn. Syst. Ser. B*, **1**(2), 183–191 (2001)
16. Turkia, J., Lepistö, L.: Size variations of planktonic *Aulacoseira* thwaites (diatomae) in water and in sediment from Finnish lakes of varying trophic state. *Journal of Plankton Research*, **21**(4), 757–770 (1999)
17. Williams, F.M.: Dynamics of microbial populations. In B. Patten, editor, *Systems Analysis and Simulation in Ecology*, 198–267. Academic Press, 1971