BIOMEDICAL ENGINEERING

Population Balances in Biomedical Engineering

Segregation Through the Distribution of Cell States

MARTIN A. HJORTSØ

Population Balances in Biomedical Engineering

Segregation through the Distribution of Cell States

Martin A. Hjortsø, Ph.D.

Department of Chemical Engineering
Louisiana State University
Baton Rouge, Louisiana

McGraw-Hill

New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

The McGraw-Hill Companies

CIP Data is on file with the Library of Congress

1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 1 0 9 8 7 6 5

ISBN 0-07-144768-7

The sponsoring editor for this book was Kenneth P. McCombs, the editing supervisor was Stephen M. Smith, and the production supervisor was Richard C. Ruzycka. It was set in Century Schoolbook by Digital Publishing Solutions. The art director for the cover was Anthony Landi.

Printed and bound by RR Donnelley.

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please write to the Director of Special Sales, McGraw-Hill Professional, Two Penn Plaza, New York, NY 10121-2298. Or contact your local bookstore.

This book is printed on recycled, acid-free paper containing a minimum of 50% recycled, de-inked fiber.

I dedicate this book to the memory of James Bailey, my dissertation advisor who first introduced me to population balance modeling.

Preface

Like many textbooks, this book evolved from a set of lecture notes that I wrote for a class. I first taught a class on population balances during a sabbatical at the Center for Process Biotechnology (now the Center for Microbial Biotechnology or CMB) at the Technical University of Denmark and have since taught the class at both Louisiana State University and CMB, again, during a second sabbatical. Over the periods that the class was taught I received many constructive ideas and suggestions that have helped to make this a better book. These contributions came from students and colleagues, too many to list by name, but all have my gratitude.

This book is intended for students and researchers from any discipline in which models of cellular systems are of relevance, from chemical engineering to medicine. However, when I taught this class in the past the audience was primarily chemical engineering students, and the book therefore contains many chemical engineering idioms and concepts. I hope that I have succeeded in explaining these in sufficient detail to make the material accessible to a wider range of students and researchers. The requirements of the reader are calculus up to and including ordinary differential equations and some exposure to simple growth models such as the celebrated Monod model.

Much of the work to finish the class notes and turn them into book form was done during a sabbatical at CMB, and I am grateful to the Otto Mønsted Foundation for financial support for this sabbatical stay. Finally, I want to extend my gratitude to Karen Jones, my partner in my private life, for careful reading of the manuscript and for the many good suggestions she made along the way.

Martin A. Hjortsø, Ph.D.

Nomenclature

Symbol	Variable	Units
a	Cell age	Time
C_P	Product concentration	Concentration
C_S	Substrate concentration	Concentration
C_{Sf}	Substrate feed concentration	Concentration
D	Dilution rate	Time^{-1}
$f(z, t)$	Normalized distribution of states with respect to cell state parameter z	Inverse units of z
$h(z)$	A priori distribution of division states	Inverse units of z
m	Cell mass	Mass
M_n	nth moment of $f(z, t)$	
$N(t)$	Cell number concentration	Inverse volume
$p(z, \bar{z})$	Distribution of birth states	Inverse units of z
r	Single-cell growth rate	(Unit of the state parameter)/time
t	Time	Time
$W(z, t)$	Cell number concentration distribution of states	$(\text{Volume} \cdot \text{units of } z)^{-1}$
Y	Yield	(Rate of formation of z)/(rate of substrate consumption)
z	Arbitrary cell state parameter	
$\Gamma(z)$	Division intensity	Time^{-1}
$\delta(z)$	Dirac delta function	
$\Theta(z)$	Death intensity	Time^{-1}
μ	Specific growth rate of a population of cells	Time^{-1}
ν	Specific growth rate of a single cell	Time^{-1}
$\Phi(z)$	Distribution of division states	Inverse units of z

Contents

Chapter

1

Introduction

This chapter aims to clarify the concept of *population balance model* or *population balance equation*, terms that are used almost interchangeably in this book. This is followed by a short narrative of the strengths and weaknesses of these models.

1.1 What Are Population Balance Models?

Population balance is not a well-defined concept in science and engineering, but means slightly different things to different people. During the fall of 2004, a Web search on the term "population balance model" gave more than 1 million hits, and a casual perusal of some of the Web pages obtained in this search makes clear this confusion of connotations. In this book, population balance models will connote the equations or sets of equations that model the dynamics of the distribution of states of a population of cells or particles.

Population balances are models describing how the number of individuals in a population and their properties change with time and with the conditions of growth. In engineering, population balances are used to model not just populations of living cells, but also populations of inanimate particles, such as the size and number of crystals in a crystalizer or the size, number, and composition of droplets in an aerosol.

Although an engineering concept, there is a population balance notion that is known to most people and that is the population pyramid. Age pyramids are histograms depicting the number of people in each of a set of age classes. Often, these histograms are split into two parts, one for males and one for females, and are placed with a common vertical

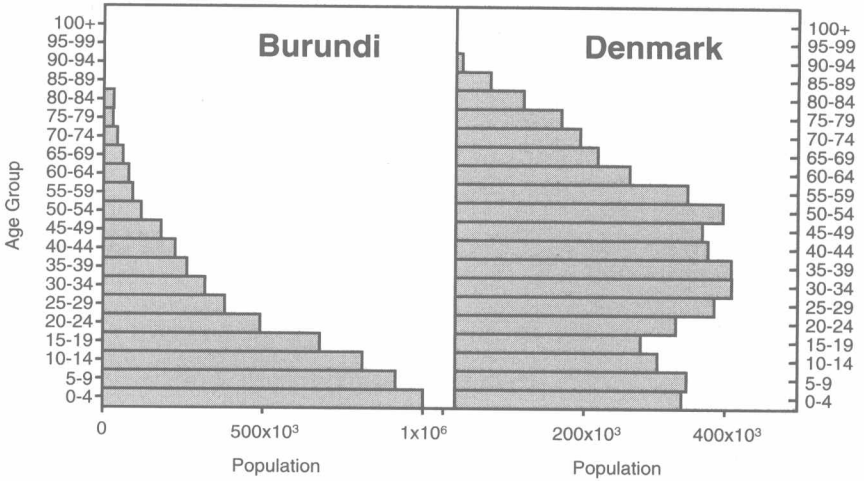

Figure 1.1 Population pyramids for Burundi and Denmark, 2000. (*Source: U.S. Census Bureau.*)

axis signifying age, and two horizontal axes, running in opposite direrctions for males and females, indicating number of individuals in each age class. This placement gives rise to a roughly triangular shape reminiscent of a pyramid, thus the name. The age pyramids for Burundi and Denmark for the year 2000 are shown in Fig. 1.1.

Without knowing anything about the mathematics of population balance models, most people will be able to look at these two pyramids and immediately conclude that

- The population of Burundi is increasing while the population of Denmark is not, or if so, only very slowly compared to the population of Burundi.
- Denmark experienced a baby boom after World War II while Burundi did not.
- The average life span in Denmark is longer than the average life span in Burundi.

The rate of population increase in Burundi can be inferred from the large number of people in the younger age groups as compared to the older groups, indicating a population with a large fraction of young individuals. This trend could conceivably be explained by a high rate of death for all of the age groups, but it is not a valid explanation in this case, since natural death in humans occurs predominantly at older ages. Instead, the large fraction of young people is a result of a high birth rate causing each generation to be larger than the previous and

thus the total population to increase with time. This trend turns out to hold for microbial populations as well: the higher the specific growth rate of the population, the larger the fraction of younger cells and vice versa. The population pyramid for Denmark, on the other hand, shows an approximately constant population size for age groups younger than 60. Only after this age does death cause a significant decrease in population size with age.

The Danish population pyramid is at its widest between ages 25 to 54; the age distribution has a local maximum in this interval of ages. This, of course, is a signature of the baby boom, the increase in birth rate that occurred in most of the western world after World War II, which was a period during which people postponed starting families. Although the Danish population pyramid indicates a population that is not changing rapidly in size, the baby boom hump shows that the age distribution in the population is not at a steady state. The baby boom subpopulation in the western world will, as time goes by, shift toward older ages, resulting in a population with a high fraction of senior citizens and giving rise to concerns about how society can cope with this increase in retirees. This connection between a temporary increase in birth rate and a local peak in the age distribution is also seen in the age distribution of microbial cultures. When such a peak is formed, the culture is said to be synchronized, or partially synchronized, and the sharper the peak in the age distribution, the higher the degree of synchrony is said to be.

The average age in Burundi and Denmark can be easily be calculated from the values of their respective population pyramids. The average age is simply the first moment of the age distribution, and the lower average age for Burundi as compared to Denmark reflects both a shorter life span and a more rapidly increasing population in Burundi.

Population balance models of the populations in Burundi and Denmark will allow for quantitative predictions about the future of the populations in the two countries rather than just the simple qualitative statements above. For instance, models would allow one to predict or estimate future population sizes in Burundi or the fraction of retirees in Denmark, both estimates that are valuable for reaching political decisions about how to manage future changes in the populations. However, the focus of this book is not on models of human populations but of models of cultures of cells, be they single-celled procaryotes, eucaryotes, or even the cells that make up tissues.

Most growth models of cell cultures can be classified as either structured or unstructured, and as distributed or segregated [94]. The term "structured model" refers to a model where more than one variable is used to specify the composition of the biophase. Typically, these

variables are the chemical compounds of the biophase. To keep the number of model variables manageable, models make frequent use of pseudocomponents, functionally similar compounds that have been lumped into groups such as proteins, various types of RNAs, and lipid content. Unstructured models, on the other hand, characterize the biophase by a single variable such as the amount of biomass. Distributed models are models that make the simplifying assumption that the cells in a culture form a single well-mixed biophase, while segregated models are more realistic and take into account the fact that the biological material is segregated into individual cells that are not necessarily identical in composition. In segregated models, the biophase is described by a *distribution of cell states*, a frequency function that indicates the probability that a cell, picked at random, is in a specified state. This specific state can be any measure of the cell state: cell size, cell mass, cell age, DNA content, protein content, etc. The state of a cell can even be specified by using multiple variables such as DNA and protein content, in which case the distribution of states becomes a multidimensional frequency function.

Distributed models can be either structured or unstructured. An unstructured, distributed model consists of a balance on the biomass coupled with mass balances on the media component, and these balances form a set of coupled, ordinary differential equations. A structured, distributed model also consists of coupled ordinary differential equations, balances on the components in the biophase and balances on components in the media—identical to the balances one would write on any two-phase reactor.

Segregated models can be either structured or unstructured, depending on how many parameters are used to describe the state of a cell. They are usually much more complex than distributed models, typically consisting of partial differential, integral equations for the distribution of cell states, coupled to mass balances on the substrate components. Segregated models are a type of population balance model, but the concept of population balances encompasses many more systems than just cell cultures.

The population balance models that are the topic of this book are segregated models of microbial populations. They are not only age distribution models, but also models of the size or mass distribution, or multidimensional models involving several cell state parameters. As alluded to earlier, these models share some of the features and issues of models of human populations. To model either type of population, one will want to know when reproduction or cell division occurs, at what rate cells or individuals in different states die, the state (e.g., size or mass) of newborn cells, and the growth rate of individual cells. Of

course, for the age distribution problem, the last two issues are trivial; newborn cells have age zero and the age growth rate is unity. When other state parameters such as cell mass are used, it is more difficult to say something about the rate of growth of individual cells or the distribution of states of newborn cells.

1.2 The Distribution of States

The models of microbial populations that we will consider here will not be of the discretized version that is exhibited by the human population histogram in Fig. 1.1, but will assume that the state parameter (age, mass, etc.) is a continuous variable, giving rise to distributions of states that are usually smooth functions instead of the discontinuous bins that the histogram represents. (Of course, a smooth distribution can always be represented by a histogram if so desired.) The distributions of states can be scaled several ways, either as a frequency function such that the zeroth moment equals unity, or as a cell number distribution such that the zeroth moment equals the cell number concentration. We will adopt the nomenclature that $f(\cdot)$ indicates the normalized distribution of states and $W(\cdot)$ the cell number concentration distribution of states. Thus, if the state of a cell is given by z, then

$$f(z, t)dz = \text{fraction of cells with state } z \in [z, z + dz]$$

at time t and similarly

$$W(z, t)dz = \text{cell number concentration of cells with state } z \in [z, z + dz]$$

The two distributions scale such that

$$\int_z f(z, t)dz = 1$$

where the z subscript in the integral indicates that the integration is over all possible cell states z. Similarly

$$\int_z W(z, t)dz = N(t)$$

where $N(t)$ is the cell number concentration at time t. Clearly,

$$W(z, t) = N(t)f(z, t)$$

and the equations that describe how these functions evolve with time and under different growth conditions are the population balance

models that we seek. The fact that these distributions indicate that the number of individuals in a given group can be a fractional number does not contradict the fact that in real populations the number of individuals within a given group is always an integer because the distributions should be thought of in a statistical sense. They represent the probability that a cell chosen at random is in a given group or interval of states. Also, in most practical applications, the number of cells in a population is so huge that the difference between the true discrete population and the continuum approximation represented by the distribution of states becomes negligible.

Often one may want to find several different distributions of states for the same population. For instance, one may want to know both the distribution of cell mass and the distribution of cell age. Instead of solving for each distribution separately, one can, since a single state parameter is used, solve for either one and find the other by a variable transformation. For instance, consider a case where the age distribution is known and where the mass distribution is desired. All we need to know to carry out the transformation is the cell mass as a function of cell age. Call this function $m(a)$ and the inverse function $a(m)$; then

$$\text{Number of cells between } a \text{ and } a + da = f(a)da$$

$$\text{Number of cells between } m(a) \text{ and } m(a + da) = f(m)dm$$

and thus

$$f(a)da = f(m)dm \Rightarrow$$

$$f(m) = f(a(m))\frac{da}{dm}, \quad f(a) = f(m(a))\frac{dm}{da}$$

The distribution of states can be partially characterized by various scalar quantities such as the zeroth moment mentioned above. In general, the nth moment of $f(z, t)$ is

$$M_n(t) = \int_z z^n f(z, t)dz = \frac{\int_z z^n W(z, t)\,dz}{\int_z W(z, t)\,dz}$$

The first moment has a simple biological interpretation; it is the mean or average z value of the cells in the population, e.g., the average cell mass or cell size. The moments defined this way are mathematically important because an approximate distribution can often be reconstructed from the moments. However, in terms of descriptive value, the centered moments are preferred. These are defined as

$$\mathcal{M}_n = \int_z (z - M_1)^n f(z, t)\, dz$$

and many of these have common names such as the second centered moment or the variance σ^2,

$$\sigma^2 = \int_z (z - M_1)^2 f(z, t)\, dz = M_2 - M_1^2$$

which describes how broad or uniform the distribution is. For a perfectly synchronized distribution in which all cells are in the same cell state, the variance equals zero. The asymmetry of the distribution is measured by the skewness defined as

$$\gamma_1 = \int_z (z - M_1)^3 f(z, t)\, dz/\sigma^3 = \frac{M_3 - 3M_1M_2 + 2M_3^3}{(M_2 - M_1^2)^{3/2}}$$

The reason for division by σ^3 is that it renders the skewness dimensionless. If a distribution is symmetric, it has zero skewness; if it has a tail at values greater than its maximum, it has positive skewness; if the tail is at values less than the maximum, it has negative skewness. Finally, the kurtosis is defined in terms of the fourth centered moment as

$$\gamma_2 = \int_z (z - M_1)^4 f(z, t)\, dz \,/\, \sigma^4 - 3 = \frac{M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4}{M_2^2 - 2M_1^2M_2 + M_1^4} - 3$$

The reason for the –3 term in the definition is that it results in the normal distribution having a kurtosis of 0. The kurtosis defined above is therefore sometimes called the *kurtosis excess*, as opposed to the *kurtosis proper*, which is defined without the –3 term. The kurtosis is a measure of the degree of peakedness of a distribution. If the distribution is more concentrated around the mean than the normal distribution, then the kurtosis is positive, otherwise it is negative.

1.3 The Age Population Balance

Derivation of the age population balance is particularly easy and will be done first to illustrate the general concept of a particle balance. We can obtain the equation by doing a cell number balance on a group of cells with ages between b and c, where we assume $0 < b < c$. The age bracket that defines the cells is an example of a so-called *control volume*, the "volume" in state space over which a number balance, or any

other kind of conservation balance for that matter, can be written. The number of cells in the control volume is

$$\int_b^c W(a, t)\, da$$

This number changes with time, and the rate of change in the number of cells inside the control volume is the time derivate of the integral:

$$\text{Rate of change in cell number} = \frac{\partial}{\partial t}\int_b^c W(a, t)\, da = \int_b^c \frac{\partial W}{\partial t}\, da$$

The number of cells in the control volume changes through three processes: Cells leave the group as they grow older than c, younger cells enter the group as they grow older than b, and cells leave the group because they divide. The rates at which cells enter and leave the group by growth are $W(b, t)$ and $W(c, t)$, respectively. The rate at which cells of age a divide is harder to account for, and we will need to define a function, $\Gamma(a, t)$, such that $\Gamma(a, t)\ W(a, t)$ equals this rate. Γ is called the division intensity, and we shall return to this function later and discuss it in more detail. Thus, the rate at which cells leave the control volume through division equals the rate for cells of age a integrated over all the control volume ages:

$$\text{Rate of cell leaving by division} = \int_b^c \Gamma(a, t) W(a, t) da$$

The rate of change of the number of cells in the group can now be related to the rates at which cells enter and leave the group by a number balance:

Rate of change in cell number =

rate of cells entering – rate of cells leaving

or, as an equation,

$$\int_b^c \frac{\partial W}{\partial t}\, da = W(b, t) - W(c, t) - \int_b^c \Gamma(a, t) da$$

The cell balance is not particularly useful in this form, so we will rewrite it by first writing the difference $W(b, t) - W(c, t)$ as an integral,

$$\int_b^c \frac{\partial W}{\partial t}\, da = -\int_b^c \frac{\partial W}{\partial a} da - \int_b^c \Gamma\,(a, t)\, W(a, t) da$$

then collecting all the terms under a single integral sign,

$$\int_b^c \left\{ \frac{\partial W}{\partial t} + \frac{\partial W}{\partial a} + \Gamma(a, t) W(a, t) \right\} da = 0$$

As the limits of the integral are arbitrary, the integrand itself must be identically zero, giving the desired result:

$$\frac{\partial W}{\partial t} + \frac{\partial W}{\partial a} = -\Gamma(a, t) W(a, t) \tag{1.1}$$

Since this equation was obtained from a number balance on cells inside a specified age bracket or control volume, this equation (as well as other equations obtained by number balances) will be referred to as a *population balance equation* (PBE). By themselves, population balance equations do not present sufficient information to solve for the distribution of states. They must first be supplied with side conditions or boundary conditions, initial conditions, and typically equations for the concentrations of growth-limiting nutrients in the medium, as well as equations that relate these concentrations to the division intensity and other kinetic functions in the population balance equation. We will refer to the combination of the population balance equation and all its side conditions and supporting equations as a *population balance model* (PBM). The alternative term *corpuscular*[1] *models* has been suggested [81], but the term has never caught on, while the term segregated model is used in many biochemical engineering books for PBMs of cell cultures [3, 10, 66].

1.4 Other PBMs

The term "population balance model" was firmly established as the preferred term when a United Engineering Foundation conference in Kona, Hawaii, in the year 2000 titled itself the Engineering Foundation Conference on Population Balance Modeling and Applications, and when, shortly after this conference, Professor Doraiswami Ramkrishna published the first general textbook on population balances simply entitled *Population Balances* [74]. It is immediately obvious in looking through this book or through the papers from the Kona conference [47] that population balance models are not limited to populations of microbial cells. In fact, in engineering the term refers to any number balance over a particulate system, and population balance models have been formulated for aerosols, crystallizers, emulsions, soot formation, polymerization kinetics, and granulation operations. Even networks

[1]Pertaining to, or composed of, corpuscles, or small particles.