Dynamical Modelling and Estimation in Wastewater Treatment Processes

Denis Dochain and Peter A. Vanrolleghem

IWA
Publishing

Dynamical Modelling and Estimation in Wastewater Treatment Processes

Dynamical Modelling and Estimation in Wastewater Treatment Processes

---

Denis Dochain
CESAME
Université Catholique de Louvain
Louvain-La-Neuve, Belgium

and

Peter Vanrolleghem
BIOMATH
Universiteit Gent
Gent, Belgium

IWA Publishing

Published by IWA Publishing, Alliance House, 12 Caxton Street, London SW1H 0QS, UK
Telephone: +44 (0) 20 7654 5500; Fax: +44 (0) 20 7654 5555; Email: publications@iwap.co.uk
Web: www.iwapublishing.com

First published 2001

Printed by TJ International (Ltd), Padstow, Cornwall, UK

British Library Cataloguing in Publication Data
A CIP catalogue record for this book is available from the British Library

Library of Congress Cataloging- in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 1 900222 50 7

Contents

Preface xi

1 Dynamical Modelling 1
- 1.1 Introduction 1
- 1.2 Classification of Mathematical Models 4
  - 1.2.1 Model Constituents 4
  - 1.2.2 Model Attributes 7
- 1.3 The Model Building Exercise 10
- 1.4 Optimal Experiment Design: Basic Ideas 15
- 1.5 Conclusion 17

2 Dynamical Mass Balance Model Building and Analysis 18
- 2.1 Introduction 18
- 2.2 The Notion of Mass Balances 19
  - 2.2.1 Stirred Tank Reactors 19
  - 2.2.2 A Simple Biological Reactor Model 22
  - 2.2.3 Biomass Death and Substrate Maintenance 24
  - 2.2.4 (Liquid and Gaseous) Product Dynamics 25
  - 2.2.5 Oxygen Dynamics 25
  - 2.2.6 Models of Reaction Rates 27
- 2.3 Examples of Biological Wastewater Treatment Process Models 34
  - 2.3.1 Anaerobic Digestion: The 4 Population Model 34
  - 2.3.2 Activated Sludge Process: The Basic Model 37
  - 2.3.3 Activated Sludge Process: The IWA Activated Sludge Model No. 1 38
  - 2.3.4 Two Step Nitrification 41

- 2.3.5 Two Step Denitrification 42
- 2.4 General Dynamical Model of Stirred Tank Reactors 42
  - 2.4.1 The General Dynamical Model 43
  - 2.4.2 Examples (Continued) 44
- 2.5 Multi Tank and Non Completely Mixed Reactors 47
  - 2.5.1 Sequential Reactors 48
  - 2.5.2 Fixed Bed Reactor: The Basic Mass Balance Model 50
  - 2.5.3 General Dynamical Model of Fixed and Fluidised Bed Reactors 52
  - 2.5.4 Settlers 58
- 2.6 Linear vs Nonlinear Models 69
- 2.7 Equilibrium Points, Linearisation and Stability Analysis 71
  - 2.7.1 First Order Kinetics 72
  - 2.7.2 Monod and Haldane Kinetics 72
  - 2.7.3 Linearised Tangent Model of Nonlinear Models 74
  - 2.7.4 Stability of Equilibrium Points 75
- 2.8 A Key State Transformation 76
  - 2.8.1 Definition of the State Transformation 76
  - 2.8.2 Example 1: Two Step Denitrification 78
  - 2.8.3 Example 2: Activated Sludge Process: The Basic Model 79
  - 2.8.4 Example 3: Activated Sludge Process: The IWA Activated Sludge Model No. 1 79
  - 2.8.5 Example 4: Fixed Bed Reactor Model With a Growth Reaction and a Death/Detachment Reaction 81
- 2.9 Model Order Reduction 81
  - 2.9.1 Singular Perturbation Technique for Low Solubility Products 82
  - 2.9.2 Singular Perturbation Technique for Substrates of Fast Reactions 83
  - 2.9.3 A General Rule for Order Reduction 85
  - 2.9.4 Example: the Anaerobic Digestion 85
  - 2.9.5 Specific Approach for Model Reduction of PDE Models 87
- 2.10 Connection Between Plug Flow Reactors and CSTRs: A Laplace Transform Approach 89
- 2.11 Conclusions 91

3 Structure Characterisation (SC) 94
3.1 Introduction 94
3.2 A Model Case Study 95
3.3 Structure Characterisation Methods 97
3.3.1 A Priori SC 97
3.3.2 A Posteriori SC 100
3.4 Optimal Experiment Design for Structure Characterisation 106
3.4.1 Theoretical Background of OED/SC 106
3.4.2 Application: Real-time OED/SC in a Respirometer 109
3.5 Conclusions 114

4 Structural Identifiability 115
4.1 Introduction 115
4.2 Theoretical Framework 117
4.3 Notion of Structural Identifiability of Linear Systems 117
4.3.1 A Simple Example 117
4.3.2 The Laplace Method 119
4.3.3 Some Generalisations and Definitions 121
4.3.4 A Second-Order Example: The Two Interconnected CSTRs Model 121
4.4 Methods for Testing Structural Identifiability of Nonlinear Systems 124
4.4.1 Taylor Series Expansion 125
4.4.2 Generating Series 126
4.4.3 Local State Isomorphism 127
4.4.4 Transformation of Nonlinear Models 127
4.4.5 A Simple Example 128
4.5 The Lyapunov-Based Method: An Historical Perspective with the Monod Model 132
4.6 Example #2: Respirometer-based Models 135
4.6.1 Identifiability of the First Order Kinetics Model (Laplace Transform) 136
4.6.2 Identifiability of the Single Monod Model (Taylor Series Expansion) 137
4.6.3 Identifiability of the Double Monod Model (Nonlinear Transformation) 140
4.6.4 Identifiability of a Modified ASM1 Model (Nonlinear Transformation and Generating Series) 142
4.6.5 Summary of the Results and Discussion 146

4.7 General Structural Identifiability Results for the ASM-type Models 148
4.8 Overparametrisation: An Illustrative Example 148
4.9 Conclusions 151

5 Practical Identifiability and Optimal Experiment Design for Parameter Estimation (OED/PE) 155
5.1 Introduction 155
5.2 Practical Identifiability 157
5.2.1 Theoretical Framework 157
5.2.2 Confidence Region of the Parameter Estimates 159
5.2.3 Sensitivity Functions 159
5.3 Optimal Experiment Design for Parameter Estimation (OED/PE) 163
5.3.1 Theoretical Background of OED/PE 164
5.4 Examples of OED/PE 166
5.5 Application: Real-time OED/PE in a Respirometer 168
5.5.1 Degrees of Freedom and Constraints for OED/PE 168
5.5.2 OED/PE for the Single Monod Model 169
5.5.3 Experimental Validation of OED/PE for the Single Monod Model 180
5.6 Optimal Experiment Design for the Dual Problem of Structure Characterisation and Parameter Estimation 185
5.7 Conclusions 187

6 Estimation of Model Parameters 188
6.1 Introduction 188
6.2 Introductory Examples 189
6.2.1 Example 1: Estimating the Mean of a Data Set 189
6.2.2 Example 2: A Simple Nonlinear Parameter Estimation Problem 193
6.3 Preliminary Steps in Parameter Estimation 194
6.3.1 Selecting Data Subsets for Estimation and Validation 194
6.3.2 Selection of Parameters to be Estimated 195
6.3.3 Differentiating Linear and Nonlinear Parameters 197
6.3.4 Use and Misuse of Linearised Forms of Nonlinear Models 198
6.3.5 Reparameterisation of Models with Nonlinear Parameters 200

6.3.6 Initial Estimates of the Parameters 201
6.3.7 Inequality Constraints on the Parameters 203
6.4 Error Characteristics 206
6.4.1 Measurement Errors and Residuals 206
6.4.2 Autocorrelated Residuals 206
6.4.3 Estimation of the Measurement Error Covariance Matrix 208
6.4.4 Errors-in-Variables Problems 210
6.5 Objectives in Parameter Estimation: Estimators 215
6.5.1 Maximum Likelihood Estimation 215
6.5.2 Weighted Least Squares (WLS) or χ^2 Estimation 216
6.5.3 Ordinary Least Squares (OLS) Estimation 217
6.5.4 Bayesian Estimation 218
6.5.5 Robust Estimation 218
6.5.6 Alternative Objective Functions 219
6.5.7 Selecting a Set of Feasible Parameters 220
6.5.8 Multi-Objective Functions 222
6.5.9 Multivariate Estimation 224
6.5.10 Multiresponse Estimation 224
6.6 Minimisation Approach 225
6.6.1 Linear Parameter Estimation 225
6.6.2 Nonlinear Parameter Estimation 226
6.6.3 Local Minimisation Algorithms Using Derivative Information 229
6.6.4 Derivative-Free Local Minimisation Algorithms 232
6.6.5 Global Minimisation Algorithms 236
6.6.6 Comparison of Different Minimisation Algorithms 238
6.7 Evaluation of Parameter Estimation Quality 239
6.7.1 Residuals Analysis 240
6.7.2 Parameter Estimation Error Covariance Matrix Determination 241
6.7.3 Confidence Regions 244
6.7.4 Significance of a Single Parameter 247
6.7.5 Significance of the Parameter Set 248
6.7.6 Correlation Among Parameters 248
6.8 Conclusions 249

7 Recursive State and Parameter Estimation 250
7.1 Introduction 250

7.2 State Observability 252
  7.2.1 Example #1: Two Step Nitrification Process 254
  7.2.2 Simple Local Observability Tests 255
  7.2.3 Example #1: Two Step Nitrification Process (Continued) 256
  7.2.4 Example #2: Simple Microbial Growth Process 256
7.3 Classical Observers 258
  7.3.1 The Basic Structure of a State Observer 258
  7.3.2 Extended Luenberger Observer 259
  7.3.3 Extended Kalman Observer 260
  7.3.4 Application to the General Dynamical Model 260
  7.3.5 Performance of Classical Observers in Presence of Model Uncertainties 266
7.4 Asymptotic Observers 268
  7.4.1 Asymptotic Observers for Single Tank Bioprocesses 269
  7.4.2 Asymptotic Observers for Multi-Tank Reactors 278
  7.4.3 Asymptotic Observers for Tubular Bioreactors 281
  7.4.4 Asymptotic Observers as a Tool for Model Selection 287
7.5 Observers for Processes with Badly Known Kinetics 291
  7.5.1 State Observer with the Unknown Parameters as Design Parameters 291
  7.5.2 Adaptive State Observer 293
  7.5.3 Generalisation 298
7.6 On-line Parameter Estimation 299
  7.6.1 The Observer-Based Estimator 299
  7.6.2 The Recursive Least Squares Estimator 306
7.7 Conclusions 310

Appendix A: Glossary 312
A.1 Model Constituents 312
A.2 Model Attributes 313
A.3 Terms of Model Building 315

Appendix B: Nomenclature 317
B.1 Greek Letters 319
B.2 Indices 320
B.3 Abbreviations 320

References 322

Index 341

Preface

The present book has a rather long history that dates back to a meeting of the European COST action 682 in San Sebastian, Espana in September 1994. At that time, the action was entitled "Optimizing the Design and Operation of Biological Wastewater Treatment Plants Through the Use of Computer Programs Based on a Dynamic Modelling of the Process". Later on, it has been renamed with a more comprehensive title: "Integrated Wastewater Management". The first title was given by our hosts in San Sebastian, i.e. Jaime Garcia-Heras and his group at the CEIT who were at the origin of this COST action. This particular meeting is certainly memorable for those who were lucky enough to participate. But for us, it was also the occasion to launch the idea of writing a book on identification in wastewater treatment processes. Indeed we had the feeling that this matter was a key topic in wastewater treatment for many reasons: the models of WWTP are complex; their dynamics are often badly grasped, a characteristic shared with the other biological processes but with the specific aspect that both the kinetics and the hydrodynamics may be badly defined; the data necessary for model identification suffer from a lot of drawbacks, they are basically too rare in terms of the number of measured variables and in terms of time distribution, and they are often not informative enough to perform a reliable model identification. And we had realised that there was no monograph that could fit to the needs of the practitioners and to the students in WWTP engineering. Available textbooks were too mathematically involved and/or not dedicated to the specific problems encountered in the identification of WWTP, or they were too much practice oriented and gave insufficient insight in the quantitative methods used. What was left was a large body of detailed papers and a few good review papers. We all know how difficult it is to get a comprehensive understanding of a topic as wide as WWTP modelling and estimation.

To start the writing we had basically already some good material coming from our individual and mutual scientific experience. We already had the chance to test the basic ideas of the book in a course that we had been asked to offer at a general meeting of the same COST action a few months before in Kollekolle, Denmark.

After six years of intensive writing, we are now able to deliver a manuscript that aims at gathering the main aspects of dynamical modelling, model building and selection, experiment design, parameter and state estimation in a comprehensive and pedagogical presentation. This is at least our intention. We have tried to select at each step (W)WTP examples that are drawn from our experience in these fields.

Obviously the scientific experience that we have gained in these fields is due to many people with whom we have interacted. These include scientific collaborators and students. From the interaction with each of them we have been able to gather the scientific expertise compiled in the present manuscript. We hope that we have not forgotten anyone in the list below: Spyros Agathos, Jens Alex, Kokou Avowlanou, Eduardo Ayesa, Jean-Pierre Babary, Georges Bastin, Danielle Baetens, Willy Bauwens, Bruce Beck, Ahmed Benhammou, Olivier Bernard, Herwig Bogaert, Geert Boeije, Jean-Francois Béteau, Benjamin Bonnet, Cédric Bonvillain, Bouchaib Bouaziz, Sylvie Bourrel, Juan Canals, Bernard Capdeville, Bengt Carlsson, Ronald Carrette, Jacob Carstensen, Gilda Carvalho, Chantal Cenens, Filip Claeys, Filip Coen, John Copp, Diedert Debusscher, Bob De Clercq, Bart De heyder, Carl Demuynck, Dirk De Pauw, Martijn Devisscher, Jeremy Dudley, Maxime Estaben, Amaya “Cindirella” Franco, Chantal Fronteau, Jaime Garcia-Heras Antoine Genovesi, Patrick Gérain, Krist Gernaey, Sylvie Gillot, Jean-Luc Gouzé, Sandra Grégoire, Koen Grijspeerdt, Serge Guiot, Willi Gujer, Zakaria Hadj-Sadok, Jérôme Harmand, Poul Harremoës, Sef Heijnen, Chris Hellinga, Mogens Henze, Lisa Hopkins, Steven Isaacs, Ulf Jeppsson, Sabine Julien, Asma Karama, Karel Keesman, Bas Kops, Peter Krebs, Patrick Labat, Luis Larrea, Juan Lema, Paul Lessard, Carl-Fredrik Lindberg, Stefano Marsili-Libelli, Jurgen Meirlaen, Henri Naveau, Fatiha Nejjari, Bart Nicolai, Ingmar Nopens, Jacques Nyns, Gustaf Olsson, Etienne Paul, André Pauss, Marco Pengov, Davide Perez Alvarino, Michel Perrier, Britta Petersen, Martin Pléau, Monique Polit, Marie-Noëlle Pons, Ana Punal, Isabelle Queinnec, Wolfgang Rauch, Peter Reichert, Enrique Roca, Alberto Rozzi, Manfred Schuetze, Ivan Simeonov, Laszlo Somlyody, Henri Spanjers, Jan Spriet, Dirk Stevens, Jean-Philippe Steyer, Hans Stigter, Imre Takacs, Nadia Tali-Maamar, Olivier Thas, Vincent Van Breusegem, Marc Van Daele, Bart Vanderhaegen, Alexis Vanderhasselt, Marijke Van de Steene, Hans Vangheluwe, Henk Vanhooren, Jan Van Impe, Mark van Loosdrecht, Lieven Van Vooren, Frederik Verdonck, Lode

Vermeersch, Willy Verstraete, Karina Versyck, Stefan Weijers, Aris Witteborg, Zhiguo Yuan and all the ones we have forgotten to mention here.

Several people have carefully read early versions of (parts of) the book: in particular we would like to thank the sometimes tedious efforts of Jean-Philippe Steyer and Michel Perrier who read unachieved bits of text.

Another person has played an important role in the writing of the present monograph, in particular via numerous discussions on the topic and his sensible comments: Paul Lessard. The first time we met him was at the already mentioned COST action meeting in San Sebastian. By that time, Paul was on sabbatical leave at the INSA of Toulouse in France. Since then, both of us have developed intense, friendly relations with Paul. We have both been impressed by his scientific rigour and his comments have always been highly welcomed.

Denis Dochain and Peter Vanrolleghem
Montréal and Gent, 1 May 2001

To our parents,
to Annick, Chantal, Kevin, Simon, Arnaud and Nicolas.

1

Dynamical Modelling

1.1 Introduction

As for any industrial process, models in biological wastewater treatment have always played an important role in the process design and in the determination of optimal operating conditions. Modelling of biological wastewater treatment processes is certainly a very active research area today. The models, originally developed for characterising the carbon removal in Continuous Stirred Tank Reactors (CSTRs), are constantly reviewed and revised in order to reflect the improvements achieved in the knowledge about the mechanistic behaviour of the involved processes. However, the lack of understanding of the underlying biochemical processes has always been a major limitation for the optimisation and effluent quality control in biological wastewater treatment processes. Hydrodynamics also play an essential role in the dynamics of wastewater treatment processes. In simple situations like CSTRs, these are well known and characterised. Yet in many instances, the behaviour of wastewater treatment processes deviates from the ideal simple CSTR. For instance, the complete mixing assumption may fail to be fulfilled in practical operation: the stirring system may not be sufficiently efficient to have a homogeneous medium in the tank. Also, the large dimensions of the plants imply that the matter may have a long way to go between the plant input and output: then transport delays cannot be neglected anymore. The hydrodynamics become an even more complex matter in other reactors than the CSTR (fixed or fluidised bed reactors, airlift loop reactors, settlers,...). This advanced reactor technology is increasingly operated for wastewater treatment plants.

The above means that better understanding of the complex biological processes (like nitrification, denitrification or phosphorus removal in activated sludge processes or the role of hydrogen in anaerobic digestion) as well as of the hydrodynamics of the reactors in which the processes are operated, appear to be essential to improve the operation and control of wastewater treatment plants. The objective of this book, and of Chapter 2 in particular, is to introduce some modelling concepts and, when possible (for some classes of processes), systematic model formalisms which will be useful in the context of parameter identification and monitoring of biological wastewater treatment processes.

Mankind is using models in nearly every aspect of human life as the principal vehicle to describe reality. In wastewater treatment, they are currently used in manifold applications, e.g.

1. to increase the understanding of the underlying bioprocess mechanisms;
2. in the design of (full-scale) processes, and of control and operating strategies;
3. for the training of operators and process engineers.

In these examples the mathematical model can only be successfully applied if it is a proper description, in terms of both model structure and model parameters, of the underlying process. Theoretical (deductive) modelling is most often not enough. There is also a need for identification of the model from experimental data (induction).

The goal of model identification is to find and calibrate a model for the system under investigation that is adequate for its intended purpose.

Although mathematical models can be classified according to many different criteria, an interesting classification is to distinguish between *models for understanding* and *models for prediction/forecasting* [53].

Models applied for understanding aim to increase knowledge of system behaviour. The objective is to develop a simple, though universal model of the system under consideration that gives an adequate description of reality as it was observed [213]. The use of models for the purpose of understanding is most frequent in research and education. These models may not reach the goal their name implies: full understanding is often too ambitious. These models should therefore be considered no more than a (dense) description of hypotheses or conjectures that are acceptable explanations of process behaviour. Such models will hold as long as there are no significant deviations between model predictions and measurements as the model is put into jeopardy under different conditions [41]. In other words, a model for understanding can never be validated, it can only be disproved or confirmed [121]. The confidence in the model assumptions (and the mechanisms they encapsulate), however, increases as the model passes more and more severe tests. The bottom line is that accepting a hypothesis never means that the hypothesis is proved, but only that it was not possible to contradict the hypothesis with the aid of the available data.

The prediction of either future or hypothetical system behaviour is one of the basic tasks in practice. Models applied for prediction aim at providing an accurate and fast image of a real system's behaviour under different conditions than the ones prevailing during model building. The model used can either aim at forecasting future states of the system (simulation with new inputs) or at predicting system behaviour under hypothetical scenarios (simulation with new parameter values). The latter mentioned application is most frequent in design. It is important to stress that prediction, e.g. in control applications, is possible with calibrated so-called *black box* models without understanding the basic mechanisms of the system. This approach is sometimes even preferable as the development time of such models is substantially lower.

Before going any further, let us draw the attention of the reader to the following key points:

1. The models that we shall derive are *dynamical* models, i.e. models aimed at describing the time evolution of the processes. The *static* or *steady state* models will then be deduced as a particular case (when there is no time variation, or more precisely the time derivatives are equal to zero). This approach is essential for us, especially in the context of system analysis, and its use for process monitoring and control: static models are a particular case of dynamical models, and not the opposite. A typical example of this "traditional" attitude is the dynamical modelling of settlers (see Chapter 2): from the solid flux theory a basically static model has been derived without any a priori reference to dynamics. It is only in a second step that the results have been incorporated in dynamical models for settlers. Unfortunately, because of the above argument, this can possibly make the applicability of such models for representing the time evolution of the process questionable.
2. Because the use of physical laws ("first principles") is a priori the best reference for deriving reliable models, we shall start the modelling by introducing the notion of mass balances. As we shall see, a fundamental aspect of any reactor dynamical model is that its dynamics is composed basically of two terms, transport dynamics (including hydrodynamics and mass transfer), and conversion (i.e. the specific influence of the biological reactions).

Before going into the mathematical derivation of the dynamical models, let us first introduce a classification of mathematical models that will (can) be used for parameter and state estimation (Section 1.2), and consider the question of model building (Section 1.3). The basic ideas of optimal experiment design, i.e.

how to design experiments so as to obtain the best data possible to solve properly the model building exercise, will be introduced in Section 1.4. Finally, because people active within the field of water quality management have very different backgrounds and jargon [51], a glossary with terminology typically used in this discipline has been put in the Glossary.

1.2 Classification of Mathematical Models

The abstract representation of a real system by the ideas on its constituents and functional relationships is called a **conceptual model**. Mathematical formulation of these ideas leads to a mathematical model that can be used to give quantitative answers to questions on its behaviour under given external conditions. Such mathematical models are referred to simply as “models” in this book. Because environmental systems are much too complicated to be described in detail, models must be drastically simplified descriptions of reality. Since the aspect of a system that is relevant depends on the question to be answered, a unique model for an environmental system does not exist, but different models must be used for different purposes, and even in a given context several adequate descriptions are possible.

1.2.1 *Model Constituents*

Ultimately, a model is a ‘machine’ that transforms inputs (u) to outputs (y) by defined relations [52], [60], where u and y are, when discretised, sequences of either scalars or vectors. (In continuous time formulation inputs u may be also a vector of **forcing functions** from outside forces.) The features of the input-output relations determine the basic structure type of the model, which is either an input/output or a state-space description [52]. The outputs are these variables the model user is interested in. The inputs of a model consist of disturbances and manipulated variables that affect the outputs.

An **input-output model** is, strictly speaking, only the set of **transfer functions** (g) that transform the inputs u directly to outputs y. An example of such a model (in discrete time formulation) where the output y at time t_k depends on past and present inputs $u(t_i)$, is:

$$y(t_k) = \frac{B(q)}{A(q)} u(t_k) \tag{1.1}$$

where $A(q)$ and $B(q)$ are polynomials in the backward shift operator q, i.e.

$$q^{-j}\,(y(t_i)) = y(t_{i-j}) \tag{1.2}$$

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_n q^{-n} \tag{1.3}$$

$$B(q) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \ldots + b_m q^{-m} \tag{1.4}$$

The order of the polynomials n and m and the model parameters a_i and b_i are to be estimated from a set of input-output data to make the model complete.

Alternative black box models are artificial neural network models that have been shown to be capable to describe any nonlinear input-output mapping. This feature has attracted a lot of attention and intense research is going on for their use in biotechnological applications. More details are given in te Braake *et al.* [246].

The most important feature of a **state-space model** is the introduction of **state variables** (vector x) which act as mediators between the inputs and the outputs. These state variables are additional model constituents and the system description is consequently addressed as internal. It is characterised by the fact that x is obtained from present and past x and u by means of the **state-transition equation** and y is generated from x by means of the **observation equation**.

The **state** of the system (as described by the model) is defined as the values of the state variables at any instant of time. Note that by definition it is neither required that state variables are measurable nor that they are meaningful in terms of natural science (although they frequently are the latter). It should also be stressed that the equations of a model can be formulated as either algebraic or differential equations.

In many cases, the dynamics of the state variables x considered important for the adequate description of the process can be described by the following state-space model:

$$\frac{dx}{dt} = Ax + Bu, \quad x(t = 0) = x_0 \tag{1.5}$$

and the output observations y are given by

$$y = Cx \tag{1.6}$$

In this model A, B and C are matrices containing the characteristic parameters of the system.

Nonlinearities of the bioprocesses, however, often ask for a different representation than the linear one given above. A more general model is:

$$\frac{dx}{dt} = f(x, u, t, \theta), \quad x(t = 0) = x_0 \tag{1.7}$$

$$y = h(x, u, t, \theta) \tag{1.8}$$

One can observe the nonlinear relations f and h between the state variables, inputs and outputs and the model parameters θ.

Whereas the dynamics of stirred tank reactors (STRs) are characterised by ordinary differential equations (ODEs)(like (1.7)), the dynamics of non completely mixed reactors (fixed bed reactors, settlers...) are generally speaking characterised by a presence of spatial gradients, i.e. a dependence of the key variables (typically the process component concentrations) not only on time but also on the spatial position in the tank: their dynamics are described by partial differential equations

(PDEs)(see Section 2.5). The systems described by ODEs are called “lumped parameter systems”, while those described by PDEs are called “distributed parameter systems”. We shall discuss PDEs in detail later, but let us anticipate with a simple illustrative example: the mass balance equation in a plug flow reactor of the concentration of a reactant x involved in one reaction with first order kinetics. The dynamics are then described by the following PDE:

$$\frac{\partial x}{\partial t} = -v\frac{\partial x}{\partial z} - k_0 x \tag{1.9}$$

$$x(t=0,z) = x_0(z) \tag{1.10}$$

$$x(t,z=0) = x_{in}(t) \tag{1.11}$$

where z is the space variable ($0 \leq z \leq$ L, with L the reactor length), v is the fluid superficial velocity, k_0 is the kinetic constant, and x_{in} is the inlet reactant concentration. Note that:

- x does not only depend on time t but also on space z (otherwise the partial derivatives would be meaningless);
- the model needs boundary conditions to be complete (here only one is necessary since the derivative with respect to z is of order 1).

The picture of mathematical models is only complete if one also considers *grey box* models. These are mainly considered as models in which part of the model structure is based on a priori knowledge of the process while another part consists of black box descriptions such as empirical rules. Such hybrid models combine the advantages of both approaches: identifiability and extrapolative power [50] [248].

Fuzzy models may also be considered grey box models though in a different sense. This is because these models incorporate insights in the internal working of the processes under study in a qualitative way. Therefore these models are particularly useful for description of systems where the process mechanisms are not (yet) completely understood. The mathematics of fuzzy set theory and some applications are introduced in te Braake *et al.* [246].

Irrespective of the **model structure**, the mathematical equations that relate inputs to outputs contain three types of constituents, that of **variables**, **constants** and **parameters**. Inputs, outputs and states are seen as variables in the equations. The difference between constants and parameters is less evident and gave rise to some confusion in modelling terminology. In the following we denote all model constituents that never change their value throughout all possible applications of the model as constants (e.g π, e, g,...).

A parameter, on the other hand, is a model constituent whose value may vary with the circumstance of its application. Hence, the value needs to be determined for each particular application of the model.

In some cases this strict definition is not followed for ease-of-use and communication reasons. Indeed, the value of some of those parameters may need to be modified during a specific application, e.g. adaptive control. Also, some 'time-varying' models exist in which a parameter is introduced in the model which is a priori known to be time-varying due to a dependency, e.g. on temperature. However, for the sake of clarity, this dependency is not explicitly included in the equations, but it is hidden elsewhere, leading to an apparent time-varying parameter. For some applications a parameter can be replaced by a function describing its time and space dependency but in fact that should be seen as a model extension. Henze *et al*. [120], [121] incorrectly use the term constant for parameter in the widely used IWA Activated Sludge Model No. 1 and 2 as they are application dependent.

1.2.2 *Model Attributes*

The model constituents describe the fundamental elements of a model, while the characteristics of a model can be addressed by a number of descriptive modelling terms. These terms are called model attributes. Some attributes have a clear and stringent definition (e.g. linear versus nonlinear), while others have a weak or relative definition (e.g. phenomenological versus mechanistic). In the following the model attributes are referred to as being strong or weak depending on the stringency of their definition. It is obvious that the stronger model attributes are applied, the better a model is characterised.

Strong model attributes

The model attributes **linear** and **nonlinear** relate to the structure of the model equations. The model may be linear with respect to the variables or to the parameters. Thus, a model can be nonlinear in the parameters but linear in the variables and vice versa. Linearity is a basic characteristic of a model that has quite some impact on the properties of solution, e.g. linear models are frequently used, because the analytical solution can be found. For nonlinear models numerical solutions are predominant.

The evaluation of the linearity of a model is conveniently performed by differentiating the function with respect to the variable or parameter of interest and evaluating whether the derivative is still function of the same variable or parameter. If this is the case, the model is nonlinear in the variable or parameter.

In wastewater treatment processes models are often characterised as **dynamic** in the sense that the variables evolve over time. A model which is not **dynamic** is called **static** or **steady state**. Thus, dynamic relates to a time dependency in the model which can be formulated as dynamic input variables and/or state variables. The output of a dynamic model is often called time series (e.g. [159]).

If the model parameters are constant in time the model is characterised as being **time-invariant**. Strictly speaking any model therefore is time-invariant. However, as discussed above (section 1.2.1), sometimes parameters are conveniently taken as time varying to increase clarity in the model description, e.g. the temperature dependency of the growth rates is well-known and induces a time-varying growth rate but this is not reflected in the specification of the growth rate equation. A model can have a space-dependency (e.g. a clarifier model). Such models are referred to as **distributed parameter** models. Currently, in wastewater treatment modelling only time- and space-derivatives are concerned, and dynamic distributed models are normally formulated as partial differential equations. However, in other biotechnological applications (and expected to become applied in wastewater treatment modelling in the near future), another type of partial differential equation formulations is used. In these models time is one independent variable, but the age of the biomass population is also considered and seen as an independent variable [127]. This leads to another type of distributed parameter model as the state variables have a time- and age-dependency. In pure culture fermentation biotechnology, these models have been termed **segregated models** because biomass is segregated into different age classes. Similarly particle sizing may be used to segregate the biomass and lead to PDEs.

Models can be termed discrete or continuous, and in most cases these attributes relate to the model formulation of difference/differential equations with respect to time, i.e. the correct terminology should be **discrete-time** or **continuous-time** model. However, **discrete-space** or **continuous-space** are two other model attributes describing a space relationship for up to three dimensions in the model formulation. A continuous-time differential equation is either solved analytically or discretised into a discrete-time difference equation (according to an Euler, Runge-Kutta, etc. scheme) which is solved numerically. It should be stressed that most computer programs apply a discretisation to the continuous-time and continuous-space differential equations, and the discretised equations are solved as algebraic equations.

If a model contains elements of randomness, it is called **stochastic** otherwise **deterministic**. The uncertainty encapsulated in any model is due to a combination ([189], [21], [23]) of

1. uncertainty in input variables,
2. uncertainty in parameter values, and
3. uncertainty in model structure.

Measurement uncertainty is embedded in these three different forms of uncertainty. In case all uncertainty aspects are neglected, the model is deterministic and the output is determined uniquely by input and initial conditions. The output of a stochastic model can be described as a probability density function. The terms stochastic and statistical are occasionally confused, and the use of statistical as model attribute should be avoided, since the term statistical is referring to methods of analysing data. Likewise, the term deterministic is occasionally confused with mechanistic, physical or white-box implying that deterministic models are always based on physical, chemical and biological laws. This is not true, however, because a black-box model may be deterministic as well, e.g. a neural network or a spline.

Stochastic input to a model is denoted as innovations, realisations, disturbances or perturbations.

The strong model attributes are highly recommended for characterising models as the meaning of the terms is well defined. However, the use of additional adjectives such as purely, totally or completely (e.g. ‘purely deterministic nor purely stochastic model’ in Harremoës and Cartensen [112] has no meaning when the stringent definitions of the strong model attributes are referred to – in fact, these combinations should be avoided as they are confusing. The majority of models for wastewater treatment so far are formulated as nonlinear dynamic continuous-time deterministic models [219], [137].

Weak model attributes

These terms have less clarity in their interpretation and may in the lack of strong model attributes potentially lead to confusion. However, provided that the terms are used correctly these attributes also signify the background of the model. To a large extent many of the weak model attributes describe almost the same model property, i.e. the degree of conceptualism, basis in physical, chemical and biological laws, simplification level, etc.

Words like **mechanistic**, **physical** and **white-box** are used to describe that the model’s structure is based on physical, chemical and biological laws. A priori knowledge is the only information source during the creation of such a model, i.e. a deductive modelling strategy is adopted. The attribute **transparent** has the same interpretation as **white-box**, which means that every detail in the model has a mechanistic explanation. The extreme is **reductionist** models (should not be confused with reduced order models, see below) that are based on the attempt to include as many details as possible. The term causal is also used with the same meaning as mechanistic, but in some scientific fields a causal model is strictly defined as a model which only depends on past observations and inputs. Thus, the use of causal as model attribute should be avoided. **Phenomenological, black-box**, **empirical** and **heuristic** (by rules of thumb) are used as model attributes to describe that the model is based on empiricism rather than laws. It is data-driven, i.e. inductive modelling is adopted. A black-box model has not necessarily a structure compatible with the underlying physical, chemical or biological reality [252]. The essential feature of black-box models is that they assume no knowledge of physical or internal relationships between the system inputs and outputs other than that the inputs should produce observable responses in the output. Hence, the system is considered ‘black box’ and no use is made of the available a priori knowledge (Casti [52] denotes it as an external system description), i.e. it omits the consideration of the mechanism by which inputs are related to outputs.

A combination of the mechanistic and phenomenological approach is frequently called **grey-box** modelling. Holst *et al.* [128] refer to grey-box models as reflecting a priori knowledge as well as black-box parts, while in Carstensen *et al.* [49], grey-box models are given two virtues – the properties of parameter interpretability and parameter identifiability. Ljung [159] refers to this approach as **semi-physical** modelling.

Another issue for characterising models is the degree of simplicity or complexity in the model. A **simple** model is characterised by few equations and parameters while a **complex** model has many equations and parameters, but it remains unclear when a given model should be termed simple or complex. As a rule of thumb, mechanistic and phenomenological models are normally formulated with a high and low degree of complexity, respectively. However, an artificial neural network is considered to be a black-box model but may at the same time have a high level of complexity [122]. The terms simple or complex for characterising a model indicates that the model is derived from a basis model or compared to another model. The confusion occurs when this reference model is not given or just assumed to be well-known. The level of model complexity/simplicity can also be addressed with the attributes **aggregated/segregated**. Jeppsson and Olsson [134] use a **reduced order** model to describe the model's derivation as being **lumped** or aggregated. A model is lumped or aggregated compared to an original or base model when model variables or equations are united in a simplified description. At this stage it is good to note that lumping of variables inevitably results in modelling errors. These errors end up in the simplified model, typically in the parameters thereof. This aspect will be dealt with to some detail in Chapter 2.

Even though it is obvious to some people that a model is complex based on the state-of-the-art in modelling today, this fact is very likely unclear to researchers unfamiliar with the specific topic or researchers within the field 10 years from now.

1.3 The Model Building Exercise

Using the 'story' of a model building exercise, a number of terms involved in this activity will be introduced within their appropriate context. At the same time a short review is given on the current state-of-the-art in modelling methodology. The diagram in Figure 1.1 summarises the aspects of model building which are described in detail below. Once the steps in the figure have been fulfilled successfully, the model can be applied for its intended purpose. These applications typically involve **simulation** that may be regarded as virtual experimentation with the virtual reality of the model.

Problem formulation

An often forgotten task in a model building exercise is the clear formulation of the goal of the model that is to be constructed. While in most cases this task is rather intuitively performed by the modeller in case he is also the problem supplier, problem formulation or goal incorporation is much more difficult in case these people are not the same. In this case, an important effort must be spent to answer such questions as desired accuracy of the results, degree of uncertainty in the provided

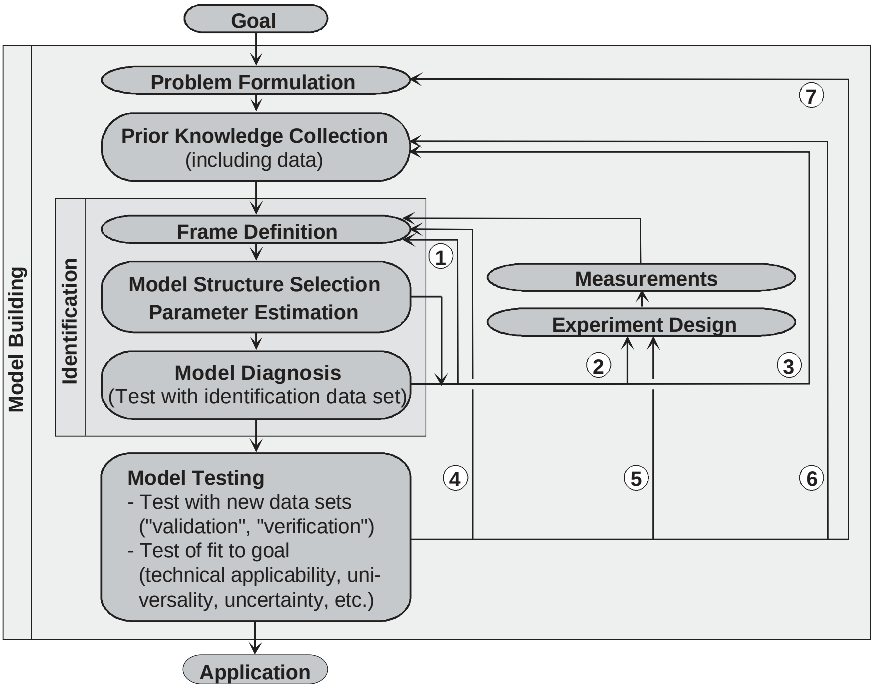

FIG. 1.1. The model building exercise [51].

answers, time scale of the solution, system boundaries, important variables, environmental conditions under which questions will be asked for which the model must give an answer, etc.

Prior knowledge collection

The next task is to collect the available, relevant a priori knowledge from literature and experts or from a model building environment that supports re-use of model-encapsulated knowledge [256]. At this (early) stage of the exercise, some experiments may be conducted or some data collected during previous experiments may be retrieved and stored in the experimental database.

Frame definition

As soon as these two tasks have been performed once, a first iteration of the model building can start. The **frame definition** phase aims to delineate the conditions under which a model will be used (e.g. temperature), to choose the class of models that seems fit for the task (time series, state-space, distributed parameter, stochastic...), to specify the variables that seem important to find a solution to the formulated problem (inputs, outputs, states), the range of time constants that need to be covered by the model, etc.

Candidate model set creation

Once the frame is defined, the purposes set and the a priori knowledge collected, the model builder is asked to create one or more possible candidate models for the system at hand. Two primary types of reasoning are used to create a candidate model [20]. The first involves the assembling of all prior hypotheses made on the mechanisms and phenomena that are believed to govern the system's behaviour. This is the synthesis part of the exercise. The analysis part consists of refuting or confirming these hypotheses on the basis of a set of field data.

Typically during the initial phases of model building, linguistic models are built using qualitative reasoning. There are however no methods or advices available to guide the modeller in this exercise. The translation of these conceptual models into mathematical models is better supported. Textbooks provide plenty of basic kinetic expressions for the biological, physical and chemical interactions among the variables of interest. It is of course up to the analyst to extract the proper ones.

While the above task is the creative one of the model builder, the model building exercise is only half way at this stage. It is equally important to demonstrate whether a particular candidate model (among many proposals that are being made) is a good or a poor approximation of reality. This second stage of evaluating the model against experimental data is best described as system **identification** (the term used in control theory). Simply stated, system identification may be considered as model **calibration**. However, as stated in Beck [20],

> *the word calibration suggests an instrument already well-designed and in need only of small, fine-tuning adjustments. Such a view weakens the role of experimental observation, relegating it to a minor test on what is the more or less inevitable path to applying the model with which the model-building procedure was first begun. It encourages the attitude that there is little to be learned from the field data that was not known before. Our preferred view of system identification is quite the contrary: it is an integral part of the process of developing scientific theories about the behaviour of a system; it, too, is about knowledge acquisition and hypothesis generation.*

This statement brings us back to the fact that modelling is an iterative cycle in which experiments play the role of indicating areas of model deficiency that are subsequently tackled in a new hypothesis generating step. The statement clearly originated from an environmental systems analyst. Indeed, for most physical and chemical applications, the a priori knowledge is of such high quality that the system framework and most of the model structure can be deduced from it. The modelling methodology developed for these systems is adequate to estimate the parameters and solve the minor uncertainties in the model structure by using final validation experiments and eventually iterating a small number of times through the procedure.

In contrast with this, the inherent characteristics of bioprocesses, i.e. their nonlinearity and nonstationarity, coupled with the lack of adequate measuring techniques, is such that this mathematical modelling methodology cannot be applied without modification [267]: more emphasis must be given to inductive reasoning to infer a larger part of the model structure from the scarce (or harder to obtain) experimental data. Consequently, structure characterisation (or model selection) methods become a more important tool, because the chance of obtaining an invalid model is much larger and, hence, the number of modelling iterations may increase substantially.

Model structure selection

The goal of **classical model structure selection** (**model structure characterisation**) is to select a unique model structure according to the principles of a quality of fit and parcimony [238], [113]. However, it is also possible to select a set of models that are attributed different weights reflecting their probability of appropriateness [79], [214]. Most model structure selection criteria require the parameter values to be estimated, e.g information criteria such as AIC [6] and BIC [228]. However, structural selection criteria that only require basic data analysis also exist for particular applications [238], [261]. A detailed discussion of this model building task will be given in Chapter 3.

Parameter estimation

Parameter estimation is typically based on the maximisation or minimisation of a goodness-of-fit criterion such as Least Squares, Weighted Least Squares, Maximum Likelihood, etc. and aims to provide values for the parameters in the model and, in case a state space representation is adopted, also values for the initial (and boundary) conditions of the state variables. Although several powerful estimation (nonlinear **regression**) algorithms are available (both for **recursive** and **batch estimation** depending on the number of data points used for a parameter update, for details, see Chapters 6 and 7), their success is highly dependent on the experimental dataset available. The identifiability analysis performed prior to the parameter estimation can provide answers to the key question whether, given a set of measured variables, unique parameter values can be obtained. Two types of answers can be given depending on the applied method. In case **structural** (also termed **theoretical** or **a priori**) **identifiability** [100] [185] is evaluated the answer is either yes or no, respectively meaning that the parameters can be given unique values or not at all [74]. However, it is not ensured that the data always contain sufficient information to provide reliable and unique estimates. Methods for the **practical** (also termed as **numerical** or **posteriori**) **identifiability** study allow to evaluate the information content of the dataset intended for parameter estimation [258]. The basis of these methods is also underlying methods that can provide a solution to a practical identifiability problem, i.e. **optimal experiment design.** This design procedure uses the model for which unique parameters are to be found to calculate experimental conditions such that sufficient information is contained in the data. One should note that a structural identifiability problem encountered cannot be solved without altering the candidate model or frame definition (e.g. include other variables in the system description). **Model reduction** can lead to models that become less ‘data-hungry’ and hence their identifiability properties may improve [134].

Model diagnosis

Once the parameters are estimated it remains to investigate whether the identified model violates the assumptions made in the frame definition. For instance, statistical tests of systematic deviations between model results and measurements (residuals) and their distribution are frequently used [38], [232]. Also, evaluation whether non-sense parameter values (e.g. negative affinity constants) or initial or boundary values are obtained can allow diagnosis of potential violation of the experimental frame.

Model testing

Fitness of a model can be evaluated by comparing its performance with data obtained under different conditions than the ones prevailing at the time of the data collection performed for model identification. This process of putting the model in jeopardy [41] or, in other words, straining the model to its limits, may reveal model inadequacies that may be sufficient to conclude that the model is no longer ‘fit’ for the purpose it was intended for. Hence, the whole model building process may have to start all over. Sometimes this may even lead to a reformulation of the problem as the modelling exercise may have provided considerable insight in the system under study and its behaviour. This process of putting the model into jeopardy by confronting it with new data is most often called **model validation**, but serious arguments are put forward against this term. As a model only describes part of the reality (the one defined in the frame) in a simplified manner, it is obvious that a model never can describe reality completely. Therefore, there will always exist experimental conditions for which the model is not valid. Hence, validation of a model is utopian! A completely other approach is to term this process of jeopardising the model a model **falsification** step [53] [212], which if answered negatively, provides more confidence in the selected model. However, the term falsification appears too negative and one has therefore looked for other terminology that is less pronounced (quantitative) as validation but still gives a qualitative insight in the level of confidence one has in a selected identified model. The terms put forth for this are **corroboration** and **confirmation** [200]. Finally, the term **verification** is frequently interchanged with validation, but in the quite related field of simulation methodology, verification has a different, but specific meaning. In this field, verification is the evaluation whether the coding of a model has been performed correctly, i.e. whether the implemented model equals the created model [156]. From this discussion it is advocated to use the term validation as it is the most widespread used and has no dubious meaning.

1.4 Optimal Experiment Design: Basic Ideas

This book devotes a lot of attention to the main tasks of a model building exercise, model selection and parameter estimation. However, these tasks can only be performed successfully if the sources of information, i.e. goal, a priori knowledge and experimental data, are of sufficiently high quality. Probably the definition of the objective is not the most difficult task albeit sometimes overlooked and not stated in a quantitative way. The analysis of the a priori knowledge leading to the construction of the set of candidate models may take some more time as knowledge must be acquired from the literature, experts and so forth. The main efforts however are, certainly for bioprocesses, devoted to the collection of data. The design of the experiments that will provide the data is an increasingly important task for the model builder for the following reasons:

- the limited resources available, e.g. experimentation time and expenses;
- the increased degrees of freedom that have to be exploited in an optimal way;
- the trade-off that has to be made between the information content of the data and the pursued goal;
- the requirement of good information content for the data for obtaining reliable models.

Goals pursued during an experiment design procedure can be categorised in the following areas:

- experiment design for a reliable selection of an adequate mathematical model structure of the process;
- the design of experiments for precise estimation of model parameters;
- the dual problem of structure characterisation and parameter estimation.

These will be considered separately in the rest of the book.

Each of these goals is associated with one or another quantitative function which will determine the focus of the experimental efforts. Some authors also suggest to separately consider [159]:

- the optimal sampling frequency;
- the best location of sensors and sampling points in a reactor system.

However, these items should be regarded as a part of the degrees of freedom available and the constraints to which one has to adhere during the experimental design process. An experiment design typically consists of three steps:

1. definition of an objective function that is the mathematical translation of the pursued goal of data usage;
2. enumeration of the available degrees of freedom and experimental constraints;
3. extremisation of the objective function by appropriate variations of the degrees of freedom without violation of the applicable constraints.

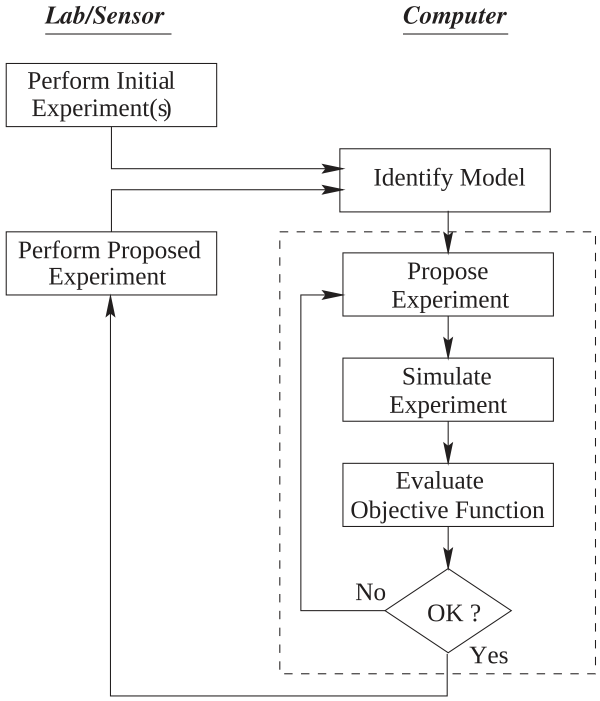

FIG. 1.2. General procedure for experiment design.

Figure 1.2 schematises the procedure used for (automated) optimisation of the experimental conditions with relation to (w.r.t.) the defined objective function. An essential feature of the procedure is that a preliminary model is available that is identified on the basis of previously acquired data. Essentially, the algorithm performing the design performs “simulated experiments” to quantify the potential effect of proposed experimental conditions on the objective function. Using a (nonlinear) optimisation algorithm that is able to take into account the experimental constraints, an “optimal” experiment is proposed that is subsequently applied in real-life. Real-life may be a specific laboratory experiment or a full-scale process on which some degrees of freedom are available that can be exploited to gather informative data.

Since the experiment design procedure depends to a large extent on a model of the process under study, it is evident that a prerequisite to good design is the availability of a good process model.

Questions that may be of guidance when initiating an experiment design study are the following ones.

What to measure? This question stresses the importance of a good system definition (input, output and state variables...).

Where to measure? This addresses the problem of sensor location.

When to measure? Here the sampling strategy is paid attention to.

Which manipulations? In many cases this is the most important question and requires creative work related to the design of excitation signals that will act on the process in such a way that highly qualitative information can be gathered. Examples are sine-waves, pulses, PRBS (pseudo random binary sequences [163]), etc.

What data treatment? Though most of the time not considered part of the experiment design, pretreatment of the data is often applied before model identification starts. It is therefore part of the collection of data for model identification purposes. Typical data pretreatment includes noise rejection, outlier detection/removal and elimination of “uninteresting” dynamics.

1.5 Conclusion

The objective of this chapter was to introduce the basic concepts of dynamical modelling. An important issue is the notion of dynamics, necessary to present the time evolution of processes. In Section 1.2 we have introduced a classification of mathematical models in terms of model constitutents (variables, constants and parameters) and of model attributes (in particular, linear vs nonlinear, dynamic vs static, lumped parameter vs distributed parameter, discrete-time vs continuous-time, deterministic vs stochastic, white-box, grey-box, or black-box). Subsequently, we have introduced the different elements of the model building exercise, from problem formulation to model testing. Finally the basic ideas of optimal experiment design, i.e. how to design experiments so as to obtain the best data possible to solve the model building exercise properly, have been introduced in Section 1.4.

2

Dynamical Mass Balance Model Building and Analysis

2.1 Introduction

The preceding chapter has introduced the basic concepts of dynamical modelling. In this chapter we shall go deeper into the subject of dynamical modelling. Because the use of physical laws (“first principles”) is a priori the best reference for deriving reliable models, we shall concentrate in this chapter on the notion of mass balances and related derivation and dynamical analysis of dynamical models. It is also worth noting that the mass balance models will play a key role in the rest of the book.

Section 2.2 will introduce the basic concepts of mass balances (“first principle”), and apply them to some simple bioprocess dynamics situations. The end of Section 2.2 will be dedicated to kinetics with a particular emphasis on mathematical models of the specific growth rate. In Section 2.3, we shall introduce several process examples, and their dynamical mass balance models in stirred tank reactors (STRs). In Section 2.4, we shall extend and generalise the mass balance dynamical model developed in Section 2.2 and applied in Section 2.3 by introducing a General Dynamical Model framework for completely mixed reactors (STRs). This will be further extended to non completely mixed reactors (plug flow reactors, fixed bed reactors, settlers) in Section 2.5. In Section 2.6, we will discuss the respective notions of linear and nonlinear systems a little further. In Section 2.7, we shall concentrate on the notion of equilibrium points, on the linearisation of nonlinear models around equilibrium points and on their stability properties. The last sections of the chapter will concentrate on the properties and the analysis of a model: a state transformation (Section 2.8), a model reduction method based on singular perturbation theory (Section 2.9), and the use of Laplace transformation in models described by PDEs (i.e. non completely mixed reactors).

2.2 The Notion of Mass Balances

2.2.1 *Stirred Tank Reactors*

Basically the reactor in which the wastewater treatment process is operated is a tank in which one or several biological reactions occur simultaneously. Let us first consider a stirred tank reactor, as schematised in Figure 2.1. The tank is characterised by a liquid medium volume V, and inflow and outflow rate, F_{in} and F_{out}, respectively. Let us consider that one (bio)chemical reaction is taking place in the tank. The mass balance of a component involved in the reaction, and characterised by a concentration C in the reactor can be formalised as follows:

$$\begin{pmatrix} Time\,variation \\ of\,the\,mass\,of \\ the\,component \\ in\,the\,tank \end{pmatrix} = \begin{pmatrix} Mass\,of\,the \\ component \\ that\,goes\,in \\ the\,tank \end{pmatrix} - \begin{pmatrix} Mass\,of\,the \\ component \\ that\,goes\,out \\ of\,the\,tank \end{pmatrix} \pm \begin{pmatrix} Mass\,of\,the \\ component \\ produced\,or \\ consumed\,by \\ the\,reaction \end{pmatrix} \tag{2.1}$$

The assumption of perfect mixing in the tank implies that the concentration C is homogeneous in the tank, and in particular its value at the reactor output is equal to the one anywhere in the tank. We now have all the elements to write the mass balance equation for the component:

$$\frac{d(VC)}{dt} = F_{in}C_{in} - F_{out}C \pm V\phi \tag{2.2}$$

where C_{in} is the component concentration in the influent, and ϕ the rate (per unit volume!) at which the component is produced ($+\phi$) or consumed ($-\phi$), i.e. the conversion rate of C. As we shall see later, the conversion rate ϕ is the product of a (positive) yield coefficient (ideally, a stoichiometric coefficient) with the reaction rate ρ, which generally speaking is a positive function of different process variables, e.g. the reactant concentrations, the product and biomass concentrations, or

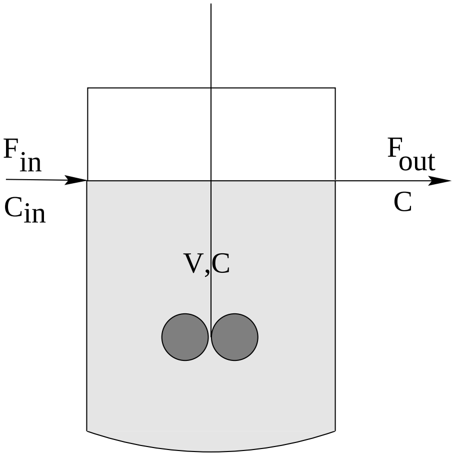

FIG. 2.1. Schematic view of a stirred tank reactor.

the physico-chemical conditions like the temperature or the pH. Let us rewrite the mass balance equation (2.2) by expanding the left hand side:

$$V\frac{dC}{dt} + C\frac{dV}{dt} = F_{in}C_{in} - F_{out}C \pm V\phi \tag{2.3}$$

and by dividing both sides by V and moving the second term of the left hand side to the right hand side, we obtain the differential equation describing the concentration dynamics:

$$\frac{dC}{dt} = \frac{F_{in}}{V}C_{in} - \frac{F_{out}}{V}C - \frac{C}{V}\frac{dV}{dt} \pm \phi \tag{2.4}$$

An important comment at this point is that the dynamics in a reactor is composed of two types of terms:

- transport dynamics ($\frac{F_{in}}{V}C_{in} - \frac{F_{out}}{V}C + \frac{C}{V}\frac{dV}{dt}$);
- conversion ($\pm\phi$).

By defining the dilution rate D:

$$D = \frac{F_{in}}{V} \tag{2.5}$$

a useful alternative formulation of equation (2.4) can be written as follows:

$$\frac{dC}{dt} = DC_{in} - DC \pm \phi \tag{2.6}$$

which covers the following important type of operations of STRs: batch, fedbatch (which may be encountered e.g. in Sequential Batch Reactors (SBRs)), and continuous (without and with volume variation). These are shortly described below.

Batch reactors. In batch reactors, there is no inflow nor outflow:

$$F_{in} = F_{out} = 0 \tag{2.7}$$

and equation (2.4) reduces to:

$$\frac{dC}{dt} = \pm\phi \tag{2.8}$$

which can also be obtained from (2.6) by setting D to zero ($D = 0$).

Fedbatch reactors. In fedbatch reactors, the tank is initially filled with some amount of reactants and catalysts, and is progressively filled with reactants (this is what happens in SBRs when the filling occurs, or when additional substrate is introduced during a denitrification phase): this means that there is an inflow but no outflow, and the time variation of the liquid volume V is equal to the inflow rate F_{in} (total mass balance):

$$F_{out} = 0, \ \frac{dV}{dt} = F_{in} \ (= DV) \tag{2.9}$$

If we add the above volume equation in (2.4), the mass balance equation (2.6) fits also to fedbatch operation.

Continuous reactors. In continuous tanks, the reactor is “continuously” fed with the reactants. Since the tank is filled, the inflow and outflow rates are equal, and the volume is constant:

$$F_{in} = F_{out}, \ \frac{dV}{dt} = 0 \tag{2.10}$$

and once again the mass balance in continuous stirred tanks fits into equation (2.6).

Continuous reactors with volume variation. The above argument corresponds to the classical continuous stirred tank (CSTR). But it may happen in some process configurations (like in Figure 2.2 which shows three types of overflow weir designs (see [188])) that the reactor is basically a continuous reactor (with inflow and outflow), but for which during transients (e.g. variations of the influent flow rate F_{in}), the volume is varying, and the outflow rate F_{out} is different from the inflow rate F_{in}. Then the above equations (2.10) have to be replaced by the following ones:

$$\frac{dV}{dt} = F_{in} - F_{out} \tag{2.11}$$

More precisely, we can have the following relationship between the (out)flow rate and the height h:

$$F_{out} = c + Nah^{b} \tag{2.12}$$

where N is the number of weirs, a and c are functions of weir type or width (c is only different from zero for the Sutro weir), and b is a function of the weir design

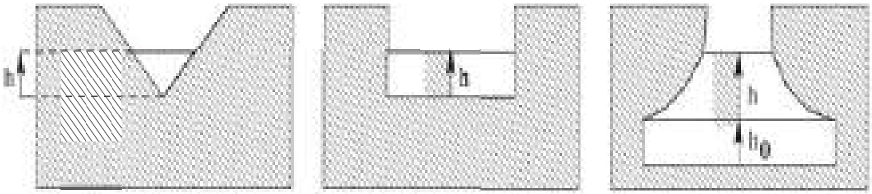

FIG. 2.2. V-notch, rectangular and Sutro weirs.

(b = 2.5, 1.5 and 1 for the V-notch, the rectangular and the Sutro weir, respectively, (see [231], [187]). Besides, note that:

$$V = Az = A(z_0 + h) \tag{2.13}$$

where A is the reactor section, z is the total depth, and z_0 the (constant) distance between the bottom of the tank and the lower edge of the weir. Then if we consider a tank with a constant section A, the mass balance equation (2.11) becomes:

$$\frac{dh}{dt} = -\frac{Na}{A}h^b - \frac{c}{A} + \frac{F_{in}}{A} \tag{2.14}$$

As was pointed out in [188](and as suggested by the above equation (2.14)), the weir designs may have an important influence on the propagation of hydraulics variations and disturbances in the plant.

The models mentioned above and their parameters a, b, c can be estimated from relatively simple experiments in which the pump flow rates are changed in a step-wise manner. In Figure 2.3, the simulation results are given of a variable volume model that is fitted to experimental data that were collected by measuring the effluent flow rate of an activated sludge system subjected to a sequence of step changes to influent and recycle flow rate [64].

2.2.2 *A Simple Biological Reactor Model*

Let us now consider an example more directly linked to biological reactors. Assume that the following microbial growth reaction is taking place in the reactor:

$$S \longrightarrow X \tag{2.15}$$

where S is the limiting substrate and X is the biomass. Note that the biomass X plays the role of an autocatalyst (it is both a product and catalyst) in the above reaction (sometimes a feedback arrow is added on top of the reaction arrow to emphasise this, see [14]). It is usually assumed (but this is not always correct, as mentioned by Henze [118], important amounts of biomass can be present in the influent) that only substrate is present in the reactor incoming flow: let us

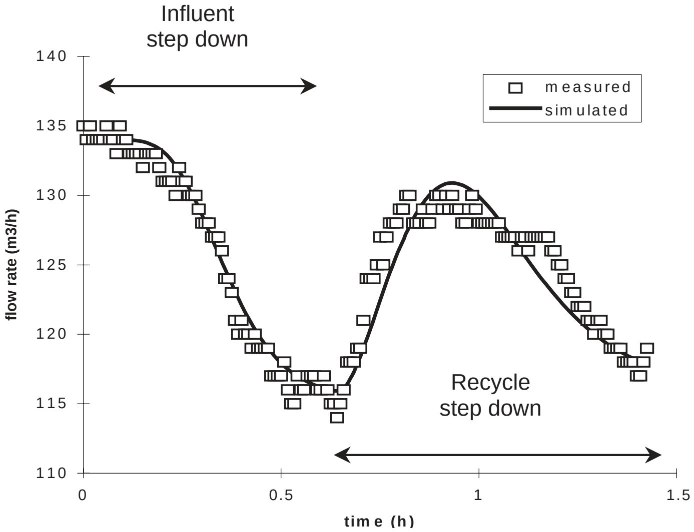

FIG. 2.3. Hydraulic propagation experiment [64].

denote by S_{in} the influent substrate concentration. In line with a commonly accepted assumption since Monod introduced it in 1942, we consider that the reaction rate ρ of the above microbial growth is the product of the autocatalyst X and a proportionality coefficient called specific (i.e. per unit biomass) growth rate μ ($\rho = \mu X$).

Then by considering (2.1), the mass balance equations for S and X are written as follows:

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\mu X \tag{2.16}$$

$$\frac{dX}{dt} = -DX + \mu X \tag{2.17}$$

Let us now take a closer look at this model and highlight some important features.

The conversion rate of the biomass, ϕ_X, is simply the reaction rate (i.e. the growth rate)($\phi_X = \rho = \mu X$). Indeed this corresponds to a standardisation of the yield coefficients with respect to one component (here, the biomass). This standardisation is obviously a priori arbitrary (it could as well have been performed with respect to the substrate, in the above example), although the use in biochemical engineering consists of standardising the yield coefficients of growth reactions with respect to the related biomass.

It is also usual in biochemical engineering to define yield coefficients in growth reactions as ratios of biomass production over substrate consumption (as Y above), or of product synthesis over biomass production (see section 2.2.4 below): this implies that in the conversion terms of the (dynamical) mass balance equations, the yield coefficient divides the reaction rate when the component is a substrate of the reaction, and it multiplies the reaction rate when it is a reaction product. Hence, in equation (2.16), the conversion term ϕ_S is the product of the yield coefficient $\frac{1}{Y}$ and of the reaction rate μX ($\phi_S = \frac{1}{Y}\rho = \frac{1}{Y}\mu X$).

Strictly speaking, if we refer to the reaction network formalism considered in (2.15), the yield coefficient Y should be a stoichiometric coefficient, i.e. here Y is the quantity of biomass X (in appropriate dimensions, e.g. in grammes) which is produced when one (mass) unit of substrate S is consumed *by the reaction (2.15)*. However, this would only be true if the conversion process is completely defined by the reaction (2.15), i.e. if

1. this is the only reaction (i.e. in particular there are no side reactions);
2. the only components are S and X;
3. and if the elemental content (in carbon, nitrogen, oxygen, hydrogen) of S and X are exactly known.

Most often, at least one of the above conditions are not fulfilled in biochemical processes. The underlying biological process is most often much more complex than the process (2.15), and it is also quite difficult in most instances to precisely characterise the elemental content of the biomass. One important consequence is that, instead of being constant (like they should if they were stoichiometric coefficients), the yield coefficients in biochemical processes may be in practice changing with time (within some bounds) e.g. when the wastewater composition or operating conditions are (significantly) changing, e.g. because side reactions which were previously neglected are no longer negligible.

2.2.3 *Biomass Death and Substrate Maintenance*

Biomass death and/or substrate maintenance terms are sometimes added in the simple microbial growth models. These reactions can be formalised as follows (see also [14]):

$$\text{biomass death:} \qquad X \longrightarrow X_d \tag{2.18}$$

$$\text{maintenance:} \qquad S + X \longrightarrow X \tag{2.19}$$

Note that in the above reactions, X no longer plays the role of an autocatalyst: it is a simple reactant (transformed in dead biomass X_d) in reaction (2.18), and a catalyst in reaction (2.19): then it is neither consumed or produced, that's why we have put X on both sides of the reaction scheme (2.19). The dynamical equations (2.16)(2.17) are then modified as follows:

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\mu X - m_S X \tag{2.20}$$

$$\frac{dX}{dt} = -DX + \mu X - bX \tag{2.21}$$

where m_S and b are maintenance and death coefficients, respectively.

2.2.4 *(Liquid and Gaseous) Product Dynamics*

The growth of microorganisms in bioreactors is often accompanied by the formation of products which are either soluble in the culture or which are given off in gaseous form. The mass balance relative to the product in the bioreactor is given by (assuming no product in the inlet):

$$\frac{dP}{dt} = -DP - Q + \nu X \tag{2.22}$$

with P being the product concentration (in the liquid phase), Q the rate of mass outflow of the product from the reactor in gaseous form, and ν the specific production rate.

The term νX in (2.22) represents the rate of product formation: it expresses the fact that the production is, in some way, "catalysed" by the biomass X.

In some instances (e.g. methane CH_4), the liquid concentration can be assumed to be negligible ($P \approx 0$). The gaseous outflow rate is then usually considered as being equal to the production rate and is written as follows:

$$Q = \nu X \tag{2.23}$$

We shall later (Section 2.9) introduce mathematical tools that give a sound basis to "reduce" the mass balance equation (2.22) to quation (2.23).

An important special case arises when the product formation is *growth-associated*, i.e. the specific production rate is assumed to be proportional to the specific growth rate:

$$\nu = Y_P \mu \tag{2.24}$$

with Y_P a yield coefficient. However, the specific production rate may also be completely or partially independent of the specific growth rate. A classical example is the lactic fermentation for which Luedeking and Piret (1959) [164] have proposed the following expression:

$$\nu = Y_P \mu + \eta \tag{2.25}$$

where η is the non-growth associated specific production rate.

2.2.5 *Oxygen Dynamics*

Aerobic bioprocesses are processes in which the microorganisms need oxygen in order to develop properly. A typical example is the activated sludge process. In such cases dissolved oxygen in the mixed liquor can be considered as an additional substrate.

$$S + S_O \longrightarrow X \tag{2.26}$$

The dissolved oxygen (DO) mass balance in the bioreactor is described as follows:

$$\frac{dS_O}{dt} = DS_{O,in} - DS_O + OTR - OUR \tag{2.27}$$

where S_O is the DO concentration in the reactor, OTR is the oxygen transfer rate and OUR is the oxygen uptake rate. Note the presence of the liquid inlet term $DS_{O,in}$ in the mass balance equation of oxygen. This term (which is often neglected) is introduced here to emphasise the possible presence of dissolved oxygen in the influent e.g. as a consequence of the oxygen transfer to the wastewater in the sewer system before it is fed to the tank.

The oxygen uptake rate OUR obviously depends on the growth of the biomass. This is often expressed as follows:

$$OUR = \frac{1}{Y_{O_2}}\mu X \tag{2.28}$$

with Y_{O_2} being a yield coefficient. Note that the usual formalism in activated sludge processes consists of considering that if Y is the biomass growth yield, Y biomass units (in COD (Chemical Oxygen Demand) units) are produced from one unit (in COD units) of the substrate, i.e. S, and the rest of the available COD, i.e. $1 - Y$ (in COD units) is oxidised to produce carbon dioxide CO_2. This implies by considering the above line of reasoning that the yield coefficient Y_{O_2} is equal to:

$$Y_{O_2} = \frac{Y}{1 - Y} \tag{2.29}$$

A term proportional to the biomass concentration ($b_{O_2}X$) is often included in equation (2.28) to account for the oxygen consumption for endogenous metabolism (to maintain viability of the cell). Hence, the total oxygen uptake rate OUR is composed of two terms, a growth related “exogenous” respiration rate OUR_{ex} and a cell maintenance related “endogenous” respiration rate OUR_{end}:

$$OUR = OUR_{ex} + OUR_{end} = \frac{1 - Y}{Y}\mu X + b_{O_2}X \tag{2.30}$$

By using a line of reasoning based on Henry’s law to model the liquid-gas transfer dynamics, the oxygen transfer rate, OTR, is expressed as follows:

$$OTR = k_La(S_O^* - S_O) \tag{2.31}$$

where k_La is the mass transfer coefficient and S_O^* is the oxygen saturation concentration.

However expression (2.31) may have to be used with precaution because the coefficients S_O^* and $k_L a$ may be unknown and may vary greatly with time: it is well known that the oxygen saturation concentration S_O^* depends on variables such as the oxygen partial pressure in the surrounding atmosphere, temperature, salinity and concentration of chemicals in the liquid, and that factors such as the type and geometry of the aerator, the air flow rate, the presence of surfactants or the biomass concentration determine the value of $k_L a$. In some applications, however, the input and output gaseous oxygen mass flow rates are measured on-line using off-gas analysis equipment [116]. Hence, provided the liquid-gas transfer dynamics are negligible, the OTR can simply be expressed from the gaseous oxygen balance:

$$OTR = Q_{O_2,in} - Q_{O_2,out} \tag{2.32}$$

where $Q_{O_2,in}$ and $Q_{O_2,out}$ are respectively the input and output oxygen flow rates (per reactor volume unit).

2.2.6 *Models of Reaction Rates*

It can clearly be seen from equations (2.16), (2.17), (2.24), (2.25) and (2.28) that the reaction rate ρ, and more particularly the specific growth rate μ is a key parameter for the description of biomass growth, substrate consumption and product formation.

Let us briefly introduce some basic notions of kinetics, before introducing some typical models of the specific growth rate (see also e.g. [88], [14], [91] for surveys on the subject; note also that more than 60 kinetic models for bioprocesses are listed in Bastin and Dochain [14]).

Basic kinetic law. The basic kinetic model for (bio)chemical reactions takes the following form:

$$\rho = k_0 \prod_i C_i^{\alpha_i} \tag{2.33}$$

where C_i are the concentrations of the reactants involved in the reaction, k_0 is the kinetic constant, and α_i are defined as the orders of the reaction with respect to the i^{th} reactant. The total *reaction order* is equal to the sum of the α_i ($\sum_i \alpha_i$). When the reaction orders α_i are equal to the stoichiometric coefficients, the reaction is referred to as an *elementary* reaction.

Note that, for instance, for the biomass death reaction (2.18), if the death coefficient b is constant, then (2.18) is a first order reaction.

One of the essential consequences of the above kinetic law (2.33) is that the reaction rate ρ will be equal to zero if the concentration of one of the reactants involved in the reaction is equal to zero:

$$\rho = 0 \quad \text{if} \quad C_i = 0 \tag{2.34}$$

It is obvious that this is a minimal requirement for any reaction rate model, i.e. also for specific growth rate models.

Models of the specific growth rate. Biochemical experiments carried out over more than half a century on pure cultures as well as on open cultures (with non-sterile substrates) have clearly indicated that the parameter μ varies with time and is influenced by many physico-chemical and biological environmental factors among which the most important ones are: substrate concentration, biomass concentration, product concentration, pH, temperature, dissolved oxygen concentration, light intensity and various inhibitors of microbial growth.

The specific growth rate is then commonly (but not systematically, see e.g. the Contois model here below) expressed by the multiplication of individual terms, each of them referring to one of the influencing factors:

$$\mu(t) = \mu(S)\mu(X)\mu(P)\mu(S_O)\mu(pH)\mu(T)... \tag{2.35}$$

where X, S, P, S_O have been defined above while T refers to temperature.

We shall present, in the following paragraphs, some of the most commonly used kinetic models for the different terms of equation (2.35).

It is important to draw the attention that any specific growth rate must follow some basic modelling rules in order to be a correct representation of the phenomenon that it is supposed to describe. First of all, recall that the specific growth rate is part of the reaction rate ($\rho = \mu X$) of a reaction which transforms substrate(s) into biomass (maybe some additional products). This means that the specific growth rate has to be positive (otherwise it would represent the reverse reaction of transformation of biomass into substrate(s)!):

$$\mu \geq 0, \qquad \text{for all time} \tag{2.36}$$

Secondly it must fulfill the basic kinetic law condition (2.34), i.e. the specific growth rate must equal to zero when the concentration of one of the substrates S_i (i = 1, 2,...) of the reaction is equal to zero.

$$\mu(S_i) = 0 \qquad \text{if} \qquad S_i = 0 \tag{2.37}$$

Influence of the substrate concentration S. The most widespread analytical specific growth rate model is certainly the “Michaelis-Menten law”, also often called the “Monod law”, which expresses the dependence of μ on the substrate concentration S as follows (Figure 2.4):

$$\mu = \frac{\mu_{max} S}{K_S + S} \tag{2.38}$$

where μ_{max} is the maximum growth rate and K_S is the “Michaelis-Menten” or saturation constant. The subscript S in the parameter K_S refers to the involved substrate. Note that the value of K_S corresponds to the substrate concentration for which the specific growth rate μ is half its maximum value μ_{max} (see Figure 2.4):

$$\mu(S = K_S) = \frac{\mu_{max}}{2} \tag{2.39}$$

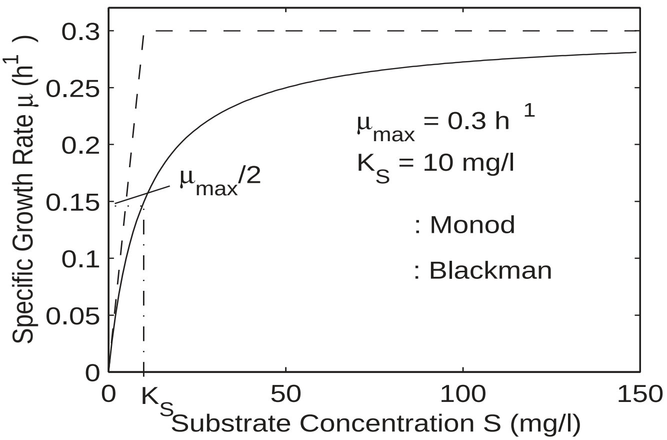

FIG. 2.4. The Monod law and the Blackman law.

In fact, this expression was initially proposed by Michaelis and Menten in 1913 [173] and physically justified by Briggs and Haldane [43] later, in 1925, to express the reaction rate of enzyme-catalysed reactions with a single substrate. It was extended by Monod in 1942 [176] to the case of microorganism growth, but without any specific physical justification.

Expression (2.38) was adopted by Monod because it fits experimental data well. But, as Monod himself recognised, “toute courbe d’allure sigmoïde pourrait être ajustée aux données expérimentales” (any sigmoidal curve could be fitted to the experimental data)[176]. Besides, another expression was suggested by Tessier [247] in the same year 1942:

$$\mu = \mu_{max}(1 - e^{\frac{-S}{K_S}}) \tag{2.40}$$

This equation could fit the Monod data equally well. Many different (and more or less esoteric) formulas have been proposed since then (see [14]).

Another simple and often used model for the specific growth rate dependence with respect to the substrate concentration is the Blackman law (Figure 2.4):

$$\mu = \frac{\mu_{max}}{K_S} S \qquad \text{if } S \leq K_S \tag{2.41}$$

$$= \mu_{max} \qquad \text{if } S > K_S \tag{2.42}$$

which is a combination of first order kinetics with respect to S for $S \leq K_S$, and zero order kinetics for $S > K_S$.

A drawback of the Monod, Tessier or Blackman laws is that they do not allow any account to be taken of possible substrate inhibitory effects at high concentrations (overloading). Typical examples in biological wastewater treatment are

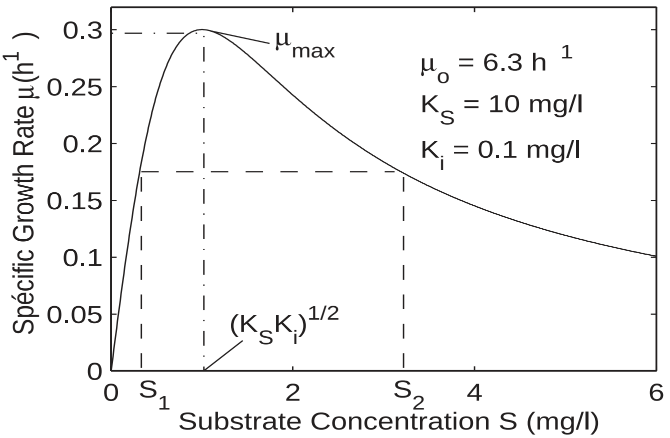

FIG. 2.5. The Haldane law.

NH_4-inhibition of nitrification or self-inhibition of phenol degradation. Andrews [8] suggested that substrate inhibition be treated by the “Haldane law” which was initially derived by Haldane [107] to describe inhibition in enzyme-substrate reactions (Figure 2.5):

$$\mu = \frac{\mu_0 S}{K_S + S + \frac{S^2}{K_i}} \tag{2.43}$$

where K_i is the “inhibition parameter” and:

$$\mu_0 = \mu_{max}(1 + 2\sqrt{\frac{K_S}{K_i}}) \tag{2.44}$$

If the substrate inhibition is neglected ($K_i \rightarrow \infty$), the Haldane law reduces to the Monod law.

One important feature of the Haldane model is shown in Figure 2.5: for each value of the specific growth rate, there exist two possible substrate concentrations (denoted by S_1 and S_2 in Figure 2.5), which are distributed on both sides of the maximum specific growth rate (which corresponds to the value of S equal to $\sqrt{K_S K_i}$). This feature has the following consequences in CSTRs with a simple microbial growth reaction (2.15) (i.e. described by equations (2.16)(2.17)). First, it allows the emphasis of the possible existence of three equilibrium points: beside the wash-out ($X = 0$, $S = S_{in}$), there are two possible steady states, each one corresponding to one of the two values S_1 and S_2. Secondly, the Haldane relation allows the emphasis of the possible existence of instability in bioprocesses: indeed, it will be shown in Section 2.7 that in a CSTR (2.16)(2.17), the steady state corresponding to S_1 is a stable equilibrium point, and the one corresponding to S_2

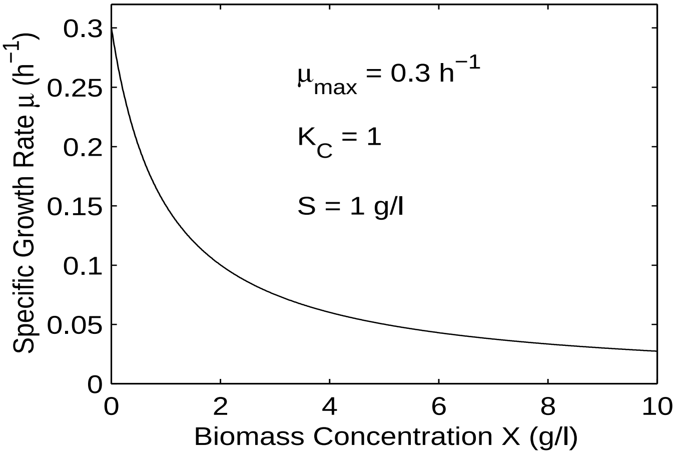

FIG. 2.6. The Contois law.

is an unstable one. The latter unstable state corresponds to the experimentally observed situation when there is accumulation of inhibiting substrate which results in a wash-out of the biomass. A typical example is the accumulation of volatile fatty acids in anaerobic digestion processes (e.g. [29], [92]).

Influence of the biomass concentration X. The biomass growth is often observed to be slowed down at high biomass concentration (and this has been experimentally observed in particular instances). A simple model that accomodates for this situation assumes that the specific growth rate decreases linearly with increasing biomass concentration:

$$\mu(X) = \mu_{max}(1 - aX) \tag{2.45}$$

where μ_{max} is the maximum growth rate and a (= $\frac{1}{X_{max}}$) the inhibition constant. It is often called the “logistic model” and was proposed by Verhulst in 1838 [271]. Another model which is a function of both S and X is the following:

$$\mu = \frac{\mu_{max} S}{K_c X + S} \tag{2.46}$$

with K_c constant. This model was proposed by Contois in 1959 [61], and is illustrated in Figure 2.6, which shows the inhibition dependence of μ with respect to X (for constant S).

Influence of the product concentration P. It is a well known fact that, in particular bioreactions, the product of the reaction can also inhibit the biomass growth. A typical example is anaerobic digestion when acetate/propionate and/or hydrogen

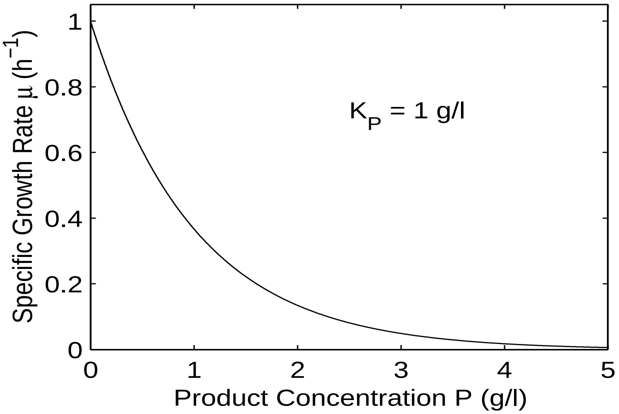

FIG. 2.7. Product inhibition model $\mu(P) = e^{-K_P P}$.

are inhibitory to acidogenesis reactions. Typical models for product inhibition are e.g. (Figure 2.7) [4]:

$$\mu(P) = \frac{K_p}{K_p + P} \tag{2.47}$$

$$\mu(P) = e^{-K_p P} \tag{2.48}$$

with K_p constant.

These inhibition models are also used to emphasise for instance the inhibition of denitrification by oxygen and are often termed “switching functions”.

Influence of pH. As we have indicated here above, the biomass growth can actually takes place only if pH and temperature lie within (usually small) ranges of admissible values.

The pH often inhibits the biological activity due to non dissociated acids or bases (ammonium, fatty acids, nitrite) in the mixed liquor. Up to now, few good models for this pH-dependency are presented. One possibility is a bell-shaped function:

$$K_{pH} = K_{\max} \cdot \frac{1}{1 + 10^{pK_1 - pH} + 10^{pH - pK_2}}$$

where pK_1 and pK_2 are the low and the high pH with half of the maximum activity (Figure 2.8).

In anaerobic digestion, for instance, the process is known to operate correctly only for nearly neutral pH (= 7). For this process, Rozzi [222] proposes treating the influence of pH by a parabolic law derived from experimental data:

$$\mu(pH) = a\, pH^2 + b\, pH + c \qquad \text{if} \quad pH_{min} \leq pH \leq pH_{max} \tag{2.49}$$

$$= 0 \qquad \text{otherwise} \tag{2.50}$$

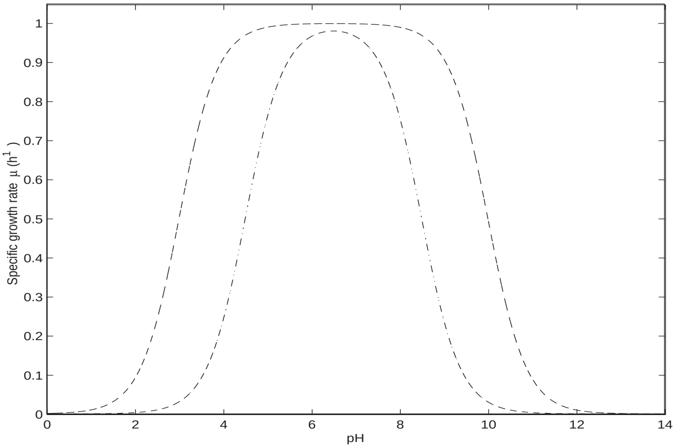

FIG. 2.8. Effect of pH on biological activity (.-: pK$_1$ = 4.5, pK$_2$ = 8.5; - -: pK$_1$ = 3, pK$_2$ = 10).

with a, b, c constants, and pH_{min} and pH_{max} are the pH bounds within which growth is possible.

Influence of temperature. On the other hand, the influence of temperature is most often modelled by an Arrhenius-type law, as has been done, for instance, by Topiwala and Sinclair [250]:

$$\mu(T) = a_1 e^{-\frac{E_1}{RT}} - a_2 e^{-\frac{E_2}{RT}} - b \tag{2.51}$$

where E_1, E_2 are activation energies, R is the gas constant (8.314 J.g.mol/K), and a_1, a_2, b are constants. This expression shows that the specific growth rate increases continuously with temperature up to a maximum value T_{max} (at which the cells die) (see Figure 2.9).

This is often simplified to:

$$\mu(T) = K_{T1} \cdot 10^{-\alpha(T_1 - T_2)} \text{ or } K_{T1} \cdot \theta^{-(T_1 - T_2)} \tag{2.52}$$

where θ is around 1.03 to 1.05 for most processes (corresponding to a doubling of the reaction rate with a temperature increase of 10°). On the other hand, nitrification is strongly dependent on temperature (θ = 1.15). More detailed models for description of the effects of temperature can be found in the overview by Zwietering *et al.* [291]. See also [115], [147], [210].

2.3 Examples of Biological Wastewater Treatment Process Models

In this section, we introduce some typical examples of biological wastewater processes and their mass balance dynamical models:

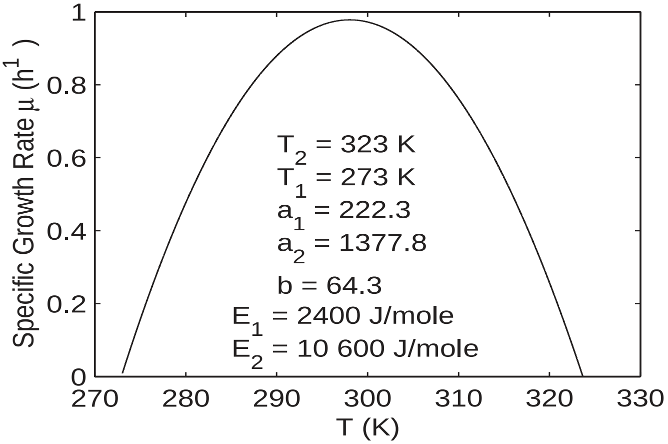

FIG. 2.9. Temperature model $\mu(T)$.

- Anaerobic digestion: The 4 population model
- Activated sludge process: The basic model
- Activated sludge process: The IWA Activated Sludge Model No. 1
- Two step nitrification
- Two step denitrification

The section will be used to extend the ideas developed in Section 2.2 for simple examples to more complex ones. As it has been suggested in the preceding section, the model building basically considers the following two steps:

1. building of an appropriate reaction network
2. derivation of the dynamical mass balances

We shall use the first example (anaerobic digestion) to illustrate the model building procedure with mass balances for processes which involve a large number of components (10) and reactions (4).

2.3.1 *Anaerobic Digestion: The 4 Population Model*

Anaerobic digestion is a biological wastewater treatment process which produces methane. Four metabolic paths [178] can be identified in this process: two for acidogenesis and two for methanisation (Figure 2.10).

In the first acidogenic path (Path 1), glucose (or another complex substrate) is decomposed into volatile fatty acids (VFAs)(acetate, propionate), hydrogen and inorganic carbon by acidogenic bacteria. In the second acidogenic path (Path 2), Obligate Hydrogen Producing Acetogens (OHPA) decompose propionate into acetate, hydrogen and inorganic carbon. In a first methanisation path (Path 3), acetate is transformed into methane and inorganic carbon by acetoclastic methanogenic

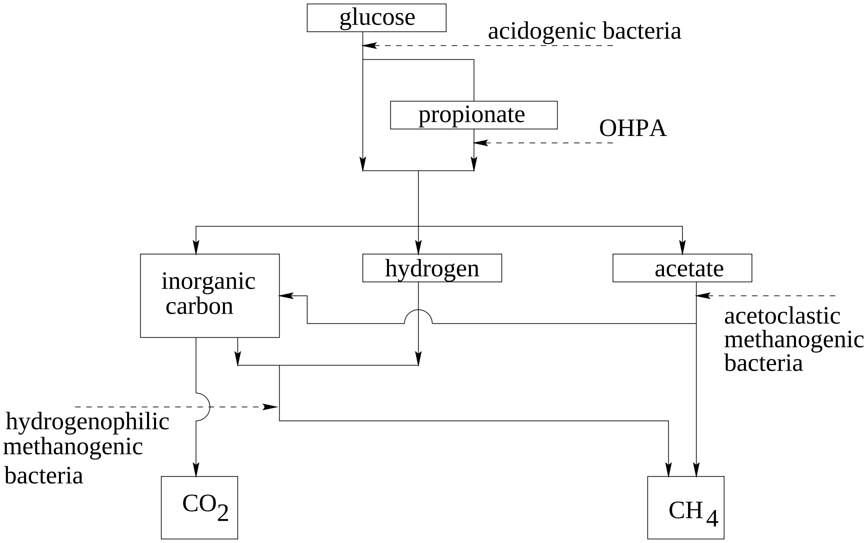

FIG. 2.10. Scheme of the anaerobic digestion.

bacteria, while in the second methanisation path (Path 4), hydrogen combines with inorganic carbon to produce methane under the action of hydrogenophilic methanogenic bacteria. The process can then be described by the following reaction network:

$$S_1 \longrightarrow X_1 + S_2 + S_3 + S_4 + S_5 \tag{2.53}$$

$$S_2 \longrightarrow X_2 + S_3 + S_4 + S_5 \tag{2.54}$$

$$S_3 \longrightarrow X_3 + S_5 + P_1 \tag{2.55}$$

$$S_4 + S_5 \longrightarrow X_4 + P_1 \tag{2.56}$$

where S_1, S_2, S_3, S_4, S_5, X_1, X_2, X_3, X_4 and P_1 are respectively glucose, propionate, acetate, hydrogen, inorganic carbon, acidogenic bacteria, OHPA, acetoclastic methanogenic bacteria, hydrogenophilic methanogenic bacteria and methane.

The conversion terms for the above reactions (2.53) to (2.56) are characterised by:

1. a reaction rate, denoted ρ_i, the index i = 1, 2, 3, 4 being related to reaction (2.53), (2.54), (2.55), (2.56), respectively. Each reaction rate is a growth rate, and each growth reaction involves a different microorganism population: therefore in line with the usual biochemical engineering notations, each reaction rate may be written as the product of a specific growth rate by the concentration of the biomass involved in the reaction:

$$\rho_i = \mu_i X_i, \; i = 1 \; to \; 4 \tag{2.57}$$

2. stoichiometry (yield coefficients) associated to each component of each reaction. Here we have considered to use the symbols Y_{ji} for the yield coefficients, where the index j corresponds to the component (ranging in the following order: X_1, S_1, X_2, S_2, X_3, S_3, X_4, S_4, S_5 and P_1), and the index i to the reaction (ranged as above). This choice looks a priori arbitrary, but, as we shall see in Section 2.4, corresponds to the position of elements in a matrix.

Beside the conversion terms, the mass balance model also contains transport terms: let us consider that the anaerobic digestion process is operated in a CSTR, and that the only inlet substrate is the organic matter S_1.

We now have all the necessary information to derive the dynamical mass balance model of the process. The dynamics of the anaerobic digestion process are then described by the following equations:

$$\frac{dX_1}{dt} = \mu_1 X_1 - DX_1 \tag{2.58}$$

$$\frac{dS_1}{dt} = -\frac{1}{Y_{21}}\mu_1 X_1 + DS_{in} - DS_1 \tag{2.59}$$

$$\frac{dX_2}{dt} = \mu_2 X_2 - DX_2 \tag{2.60}$$

$$\frac{dS_2}{dt} = Y_{41}\mu_1 X_1 - \frac{1}{Y_{42}}\mu_2 X_2 - DS_2 \tag{2.61}$$

$$\frac{dX_3}{dt} = \mu_3 X_3 - DX_3 \tag{2.62}$$

$$\frac{dS_3}{dt} = Y_{61}\mu_1 X_1 + Y_{62}\mu_2 X_2 - \frac{1}{Y_{63}}\mu_3 X_3 - DS_3 \tag{2.63}$$

$$\frac{dX_4}{dt} = \mu_4 X_4 - DX_4 \tag{2.64}$$

$$\frac{dS_4}{dt} = Y_{81}\mu_1 X_1 + Y_{82}\mu_2 X_2 - \frac{1}{Y_{84}}\mu_4 X_4 - DS_4 - Q_1 \tag{2.65}$$

$$\frac{dS_5}{dt} = Y_{91}\mu_1 X_1 + Y_{92}\mu_2 X_2 + Y_{93}\mu_3 X_3 - \frac{1}{Y_{94}}\mu_4 X_4 - DS_4 - Q_2 \tag{2.66}$$

$$\frac{dP_1}{dt} = Y_{03}\mu_3 X_3 + Y_{04}\mu_4 X_4 - DP_1 - Q_3 \tag{2.67}$$

where μ_1, μ_2, μ_3, μ_4 are the specific growth rates (h^{-1}) of reactions (2.53), (2.54), (2.55), (2.56), respectively, and S_{in}, Q_1, Q_2 and Q_3 represent respectively the influent glucose concentration (g/L) and the gaseous outflow rates of H_2, CO_2 and CH_4 (g/L.h).

Note that if we intend to use the model for (dynamic) simulation, we still need expressions for the specific growth rates. This can be done by choosing e.g. one

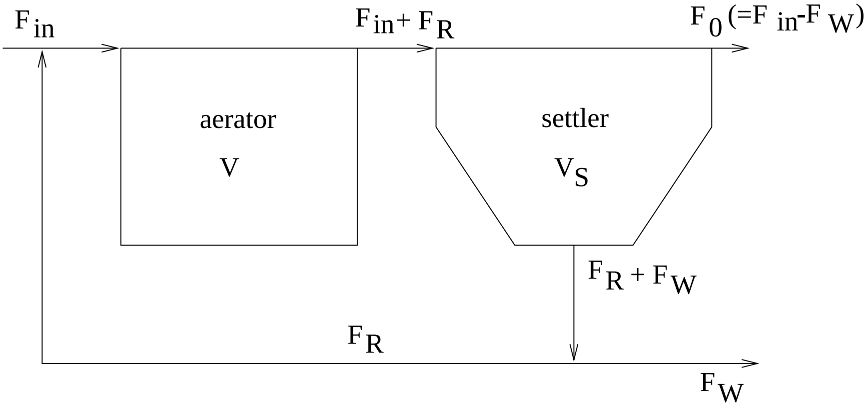

FIG. 2.11. Schematic view of an activated sludge process.

(or a combination) of the expressions presented in Section 2.2. The problem of choosing appropriate expressions for specific growth rates and calibrating parameter values will be discussed in Chapters 3, 4 and 5.

2.3.2 *Activated Sludge Process: The Basic Model*

The activated sludge process is one other classical biological (but aerobic) wastewater treatment process. It is usually operated in (at least) two sequential tanks (see Figure 2.11): an aerator (in which the degradation of the pollutants S takes place) and a settler (which is used to separate biomass X from the treated wastewater and recycle it to the aerator).

The reaction in the aerator may be described by a simple microbial growth reaction (see e.g. [108], [169]).

$$S + S_O \longrightarrow X \tag{2.68}$$

It is often assumed that the settler works perfectly, i.e. there is no biomass in the overflow of the settler (characterised by the flow rate F_o (= $F_{in} - F_W$)). The dynamics of the plant (aerator + settler) are then described by the following mass balance equations:

$$\frac{dS}{dt} = \frac{-1}{Y_S}\mu X - \frac{F_R + F_{in}}{V}S + \frac{F_{in}}{V}S_{in} \tag{2.69}$$

$$\frac{dS_O}{dt} = \frac{-1}{Y_0}\mu X + \frac{F_{in}}{V}S_{O,in} - \frac{F_R + F_{in}}{V}S_O + k_L a(S_O^* - S_O) \tag{2.70}$$

$$\frac{dX}{dt} = \mu X + \frac{F_R}{V}X_R - \frac{F_R + F_{in}}{V}X \tag{2.71}$$

$$\frac{dX_R}{dt} = \frac{F_{in} + F_R}{V_S}X - \frac{F_R + F_W}{V_S}X_R \tag{2.72}$$

where X_R is the concentration of the recycled biomass (g/L), F_{in}, F_R and F_W are the influent, recycle and waste flow rates (g/L.h), respectively, and V and V_S the aerator and settler volumes (L), respectively.

2.3.3 *Activated Sludge Process: The IWA Activated Sludge Model No. 1*

Generally speaking, the simple aerobic microbial growth reaction (2.68) is not the only one which takes place in an activated sludge process. The simple reaction results in a rather simple dynamical model which may be convenient in many instances, e.g. for control design. However, in other instances, e.g. for an exhaustive process description, the dynamical model should include the influences of the different reactions that are taking place in activated sludge processes. One of the objectives of the Task Group that eventually produced the IWA Activated Sludge Model No. 1 [120] was to gather these reactions (or at least the most important ones) into one model framework. There are 13 important components and 8 important reactions, which are carried out by two types of microorganisms: heterotrophs (their carbon comes from an organic compound) and autotrophs (they require only CO_2 to supply their carbon needs). Next to their growth (for the heterotrophs both under aerobic and anoxic conditions) and decay reactions (reactions 1-5), 3 other reactions take place: ammonification of soluble organic nitrogen (6), hydrolysis of entrapped organics (7) and of entrapped organic nitrogen (8). These can be formalised into the following reaction network:

1. Aerobic growth of heterotrophs: $S_S + S_O + S_{NH} \longrightarrow X_{B,H}$
2. Anoxic growth of heterotrophs: $S_S + S_{NO} + S_{NH} \longrightarrow X_{B,H}$
3. Aerobic growth of autotrophs: $S_O + S_{NH} \longrightarrow X_{B,A} + S_{NO}$
4. Decay of heterotrophs: $X_{B,H} \longrightarrow X_P + X_S + X_{ND}$
5. Decay of autotrophs: $X_{B,A} \longrightarrow X_P + X_S + X_{ND}$
6. Ammonification of soluble organic nitrogen: $S_{ND} \longrightarrow S_{NH}$
7. Hydrolysis of entrapped organics: $X_S \longrightarrow S_S$
8. Hydrolysis of entrapped organic nitrogen: $X_{ND} \longrightarrow S_{ND}$

where S_S, S_O, S_{NH}, S_{NO}, S_{ND}, X_{ND}, $X_{B,H}$, $X_{B,A}$, X_P, X_S are the readily biodegradable substrate, the dissolved oxygen concentration, the soluble ammonia nitrogen, the nitrate nitrogen, the soluble organic nitrogen, the particulate biodegradable organic nitrogen, the heterotrophic bacteria, the autotrophic bacteria, the particulate products arising from biomass decay, and the slowly biodegradable substrate, respectively.

The dynamics in a stirred tank reactor are given by the following mass balance equations (with the yield coefficient nomenclature of the IWA Activated Sludge Model No. 1):

$$\frac{dS_S}{dt} = -\frac{1}{Y_H}\mu_1 X_{B,H} - \frac{1}{Y_H}\mu_2 X_{BH} + \rho_7 + DS_{S,in} - DS_S \tag{2.73}$$

$$\frac{dX_S}{dt} = (1 - f_P)b_H X_{B,H} + (1 - f_P)b_A X_{B,A} - \rho_7 + DX_{S,in} - DX_S \tag{2.74}$$

$$\frac{dX_{B,H}}{dt} = \mu_1 X_{B,H} + \mu_2 X_{B,H} - b_H X_{B,H} + DX_{B,H,in} - DX_{B,H} \tag{2.75}$$

$$\frac{dX_{B,A}}{dt} = \mu_3 X_{B,A} - b_A X_{B,A} + DX_{B,A,in} - DX_{B,A} \tag{2.76}$$

$$\frac{dX_P}{dt} = f_P b_H X_{B,H} + f_P b_A X_{B,A} + DX_{P,in} - DX_P \tag{2.77}$$

$$\begin{aligned}\frac{dS_O}{dt} &= -\frac{1 - Y_H}{Y_H}\mu_1 X_{B,H} - \frac{4.57 - Y_A}{Y_A}\mu_3 X_{B,A} \\ &\quad + k_L a(S_O^* - S_O) + DS_{O,in} - DS_O \end{aligned}\tag{2.78}$$

$$\frac{dS_{NO}}{dt} = -\frac{1 - Y_H}{2.86Y_H}\mu_2 X_{B,H} + \frac{1}{Y_A}\mu_3 X_{B,A} + DS_{NO,in} - DS_{NO} \tag{2.79}$$

$$\begin{aligned}\frac{dS_{NH}}{dt} &= -i_{XB}\mu_1 X_{B,H} - i_{XB}\mu_2 X_{B,H} - (i_{XB} + \frac{1}{Y_A})\mu_3 X_{B,A} + DS_{NH,in} \\ &\quad - DS_{NH}\end{aligned}\tag{2.80}$$

$$\frac{dS_{ND}}{dt} = \rho_8 - k_a S_{ND} X_{B,H} + DS_{ND,in} - DS_{ND} \tag{2.81}$$

$$\begin{aligned}\frac{dX_{ND}}{dt} &= (i_{XB} - f_P i_{XP})b_H X_{B,H} + (i_{XB} - f_P i_{XP})b_A X_{B,A} - \rho_8 \\ &\quad + DX_{ND,in} - DX_{ND}\end{aligned}\tag{2.82}$$

$$\rho_7 = k_h \frac{X_S/X_{B,H}}{K_X + X_S/X_{B,H}} X_{B,H} \tag{2.83}$$

$$\rho_8 = \rho_7 \frac{X_{ND}}{X_S} \tag{2.84}$$

In the above equations, ρ_7 and ρ_8 hold for the reaction rates of reaction 7(hydrolysis of entrapped organics) and 8 (hydrolysis of entrapped organic nitrogen). In the IWA Activated Sludge Model No. 1, the model equations are summarised in a table (see Table 2.1[1]) which is very convenient in the sense that it clearly emphasises

---

[1] "S_I and X_I are not involved in any conversion process. Nevertheless they are included because they are important to the performance of the process"; "Inclusion of alkalinity (S_{ALK}) is not essential, but its inclusion is desirable because it provides information whereby undue changes in pH can be predicted" [120].

Table 2.1. Tabular format of the stoichiometry of the IWA Activated Sludge Model No. 1

	Components → i	1	2	3	4	5	6	7	8	9	10
j	Process ↓	S_S	X_S	$X_{B,H}$	$X_{B,A}$	X_P	S_0	S_{N0}	S_{NH}	S_{ND}	X_{ND}
1	Aerobic growth of heterotrophs	$-\frac{1}{Y_H}$		1			$-\frac{1-Y_H}{Y_H}$		$-i_{XB}$		
2	Anoxic growth of heterotrophs	$-\frac{1}{Y_H}$		1				$-\frac{1-Y_H}{2.86Y_H}$	$-i_{XB}$		
3	Aerobic growth of autotrophs				1		$-\frac{4.57-Y_A}{Y_A}$	$-\frac{1}{Y_A}$	$-i_{XB}-\frac{1}{Y_A}$		
4	‘Decay’ of heterotrophs		$1-f_P$	-1		f_P					$i_{XB}-f_P i_{XP}$
5	‘Decay’ of autotrophs		$1-f_P$		-1	f_P					$i_{XB}-f_P i_{XP}$
6	Ammonification of soluble organic nitrogen								1	-1	
7	‘Hydrolysis’ of entrapped organics	1	-1								
8	‘Hydrolysis’ of entrapped organic nitrogen									1	-1

the interaction and interconnections of the different process components within the different process reactions. The adoption of a Petersen matrix [196] format in which components, reaction kinetics and stoichiometry are clearly separated has shown to be very fruitful. In the General Dynamical Model section (Section 2.4), we shall introduce another, yet very similar, compact form of mass balance models (i.e. in a mathematical matrix form) which will be very convenient for mathematical manipulations (e.g. for model reduction and monitoring design).

2.3.4 *Two Step Nitrification*

Nitrification is the biological oxidation of ammonium-nitrogen to nitrate-nitrogen. Yet nitrification is known to be a two-step reaction (e.g. [5], [11], [286]):

$$S_{NH} \longrightarrow X_1 + S_{NO_2} \tag{2.85}$$

$$S_{NO_2} \longrightarrow X_2 + S_{NO_3} \tag{2.86}$$

where S_{NH}, S_{NO_2}, S_{NO_3}, X_1 and X_2 hold respectively for ammonium-nitrogen ($NH_4{-}N$), nitrite-nitrogen ($NO_2{-}N$), nitrate-nitrogen ($NO_3{-}N$), *Nitrosomonas* bacteria and *Nitrobacter* bacteria, respectively.

In the IWA Activated Sludge Model No. 1, nitrification is represented by only one reaction: reaction 3 (aerobic growth of autotrophs, see previous section). Indeed, the two sequential reactions (2.85)(2.86) of nitrification can be reduced to the single reaction 3 if the reaction (2.85) is assumed to be slower than reaction (2.86). In the model order reduction section (Section 2.9) we shall introduce a systematic approach for reducing reaction systems with slow and fast reactions.

The dynamics of the two-step nitrification in a stirred tank reactor fed with only ammonium containing wastewater are given by the following mass balance equations:

$$\frac{dS_{NH}}{dt} = -\frac{1}{Y_1}\mu_1 X_1 + DS_{NH,in} - DS_{NH} \tag{2.87}$$

$$\frac{dS_{NO_2}}{dt} = -\frac{1}{Y_2}\mu_2 X_2 + Y_3\mu_1 X_1 - DS_{NO_2} \tag{2.88}$$

$$\frac{dS_{NO_3}}{dt} = Y_4\mu_2 X_2 - DS_{NO_3} \tag{2.89}$$

$$\frac{dX_1}{dt} = \mu_1 X_1 - DX_1 \tag{2.90}$$

$$\frac{dX_2}{dt} = \mu_2 X_2 - DX_2 \tag{2.91}$$

where μ_1 and μ_2 are the specific growth rates of reaction (2.85) and (2.86), respectively, Y_1 to Y_4 are yield coefficients, and $S_{NH,in}$ the influent concentration of S_{NH}.

2.3.5 *Two Step Denitrification*

Denitrification plays an important role not only in wastewater treatment (see reaction 2 in the IWA Activated Sludge Model No. 1) but also in the production of drinking water. As this is a multi-step process and some toxic intermediates are formed, nitrite in particular, a process description involving nitrite is needed as one desires its complete absence from drinking water. In wastewater treatment, important cost savings can be obtained if one adheres to what has been termed "Nitrit Verfahren" for nitrogen removal, i.e. partial nitrification to nitrite and denitrification from nitrite to nitrogen gas [1],[5], [117].

For such processes the denitrification reaction scheme is also composed of two sequential reactions:

$$S_{NO_3} + S \longrightarrow X + S_{NO_2} \tag{2.92}$$

$$S_{NO_2} + S \longrightarrow X + N_2 \tag{2.93}$$

where S_{NO_3}, S_{NO_2}, S, N_2 and X are nitrate (NO_3^-), nitrite ($N0_2^-$), organic carbon, nitrogen, and denitrifying bacteria, respectively. In the above reaction scheme, the denitrifying bacteria are of the same type and are growing in a similar environment: that's why only one biomass X is considered in both reactions.

The dynamical mass balances in a stirred tank reactor are then equal to:

$$\frac{dS_{NO_3}}{dt} = -\frac{1}{Y_1}\mu_1 X + DS_{NO_3,in} - DS_{NO_3} \tag{2.94}$$

$$\frac{dS_{NO_2}}{dt} = -\frac{1}{Y_2}\mu_2 X + Y_3\mu_1 X - DS_{NO_2} \tag{2.95}$$

$$\frac{dS}{dt} = -\frac{1}{Y_4}\mu_1 X - \frac{1}{Y_5}\mu_2 X + DS_{in} - DS \tag{2.96}$$

$$\frac{dX}{dt} = \mu_1 X + \mu_2 X - DX \tag{2.97}$$

$$\frac{dN_2}{dt} = Y_6\mu_2 X - DN_2 - Q_{N_2} \tag{2.98}$$

where μ_1 and μ_2 are the specific growth rates of reaction (2.92) and (2.93), respectively, Y_1 to Y_6 are yield coefficients, $S_{NO_3,in}$ and S_{in} the influent concentration of nitrate S_{NO_3} and organic carbon S, and Q the gaseous nitrogen outflow rate.

2.4 General Dynamical Model of Stirred Tank Reactors

So far, we have formalised each biological wastewater treatment process example by considering reaction networks. The notion of reaction scheme (as it is suggested in the above examples) is a useful tool to derive the dynamical mass balance model of the process: indeed, the dynamical model can be derived on the basis of the reaction network and put in a matrix form that is now introduced in this section. But before, let us draw the attention on the following points.

- As suggested in the above examples, a reaction system (and in particular a biological wastewater treatment process) can be viewed in the context of mass balance modelling as a set of M (bio)chemical reactions involving N components. The reactions most often encountered in bioprocesses are microbial growth (in which the biomass plays the role of an autocatalyst) and enzyme catalysed reactions (in which the biomass can be viewed as a simple catalyst); but many other reactions can also take place, like microorganism death, maintenance,... The main components are microorganism populations, enzymes, substrates and products.
- The reaction schemes as introduced here do not represent a stoichiometric relation between the process components, in contrast to the common practice in chemical kinetics. They are sometimes qualitative (in line with the comments on yield coefficients of Section 2.2.2). Yet this approach is not exclusive, and the user has the flexibility to explicitly include yield or stoichiometric coefficients in the reaction scheme, if he feels this more convenient.
- The reaction network is basically a tool for process description and model derivation, and therefore does not need to be an exhaustive (i.e. with *all* the reactions and *all* the components) description of the process. For instance, side reactions and by-products may not appear in the reaction network if it appears that their role is negligible in the process description, or even more precisely, in the context of the considered study.

2.4.1 *The General Dynamical Model*

Let us now go back, for instance, to the denitrification example (2.94)-(2.98). The model can be rewritten in the following matrix form:

$$\frac{d}{dt}\begin{pmatrix} S_{NO_3} \\ S_{NO_2} \\ S \\ X \\ N_2 \end{pmatrix} = -D\begin{pmatrix} S_{NO_3} \\ S_{NO_2} \\ S \\ X \\ N_2 \end{pmatrix} + \begin{pmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ -\frac{1}{Y_4} & -\frac{1}{Y_5} \\ 1 & 1 \\ 0 & Y_6 \end{pmatrix}\begin{pmatrix} \mu_1 X \\ \mu_2 X \end{pmatrix} + \begin{pmatrix} DS_{NO_3,in} \\ 0 \\ DS_{in} \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Q_{N_2} \end{pmatrix} \tag{2.99}$$

This suggests to rewrite the dynamical mass balance model in the following compact matrix form:

$$\frac{d\xi}{dt} = -D\xi + Y\rho(\xi) + F - Q(\xi) \tag{2.100}$$

where ξ is the vector of the bioprocess component concentrations (dim(ξ) = N), Y is the yield coefficient matrix (dim(Y) = N x M), $\rho(\xi)$ is the vector of reaction rates (that may be function of the component concentrations ξ (dim (ρ) = M), F is the feed rate vector and Q the vector of gaseous outflow rates that also may be influenced by the component concentrations ξ (dim(F) = dim(Q) = N).

The model (2.100) has been called the *General Dynamical Model* for stirred tank bioreactors (see [14]). The derivation of the dynamical mass balance model from a reaction network is straightforward by noting that each component Y_{ij} of the yield coefficient matrix

$$Y = [Y_{ij}] \qquad i = 1 \; to \; N, \;\; j = 1 \; to \; M \tag{2.101}$$

is representative of the i^{th} component: it is negative if the component is a reactant, it is positive if it is a product and it is equal to zero if the component does not intervene in the reaction.

Note that the mass balance general dynamical model is composed of the two terms that we have mentioned in the introduction and in section 2.2.1:

- transport dynamics: $-D\xi + F - Q(\xi)$,
- conversion: $Y\rho(\xi)$

As it has been pointed out in section 2.2.1, note finally that the General Dynamical Model remains valid for processes with variable volumes via the addition of a dynamical equation describing the volume dynamics.

2.4.2 *Examples (Continued)*

Let us see how the different process examples that we have previously presented fit into the General Dynamical Model (2.100) framework.

Anaerobic digestion: The 4 population model. Equations (2.58) to (2.67) can be rewritten in the matrix form (2.100) by considering the following vectors and matrices (N = 4, M = 10):

$$\xi = \begin{pmatrix} X_1 \\ S_1 \\ X_2 \\ S_2 \\ X_3 \\ S_3 \\ X_4 \\ S_4 \\ S_5 \\ P_1 \end{pmatrix}, \; Y = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{Y_{21}} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ Y_{41} & - - \frac{1}{Y_{42}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ Y_{61} & Y_{62} & 0 & -\frac{1}{Y_{63}} \\ 0 & 0 & 0 & 1 \\ Y_{81} & Y_{82} & 0 & -\frac{1}{Y_{84}} \\ Y_{91} & Y_{92} & Y_{93} & -\frac{1}{Y_{94}} \\ 0 & 0 & Y_{03} & Y_{04} \end{pmatrix} \tag{2.102}$$

$$F = \begin{pmatrix} 0 \\ DS_{in} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ Q = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ Q_1 \\ Q_2 \\ Q_3 \end{pmatrix}, \ \rho = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{pmatrix} = \begin{pmatrix} \mu_1 X_1 \\ \mu_2 X_2 \\ \mu_3 X_3 \\ \mu_4 X_4 \end{pmatrix} \tag{2.103}$$

Activated sludge process: The basic model. An interesting feature of the basic model of activated sludge process is that due to its multi-tank characteristics (aerator + settler), the dilution rate becomes a matrix. In the General Dynamical Model (2.100) framework, the mass balance equations (2.69) to (2.72) become (N = 4, M = 1):

$$\xi = \begin{pmatrix} S \\ S_O \\ X \\ X_R \end{pmatrix}, \ Y = \begin{pmatrix} -\frac{1}{Y_S} \\ -\frac{1}{Y_O} \\ 1 \\ 0 \end{pmatrix}, \ F = \begin{pmatrix} D_{in} S_{in} \\ D_{in} S_{O,in} + k_L a (S_O^* - S_O) \\ 0 \\ 0 \end{pmatrix} \tag{2.104}$$

$$\rho = \mu X, \ Q = 0, \ D = \begin{pmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_1 & 0 & 0 \\ 0 & 0 & D_1 & -D_2 \\ 0 & 0 & -D_3 & D_4 \end{pmatrix} \tag{2.105}$$

with the following definitions for D_{in}, D_1, D_2, D_3, and D_4:

$$D_{in} = \frac{F_{in}}{V}, \ D_2 = \frac{F_R}{V}, \ D_1 = D_{in} + D_2, \ D_3 = \frac{F_{in} + F_R}{V_S}, \ D_4 = \frac{F_R + F_W}{V_S} \tag{2.106}$$

Activated sludge process: The IWA Activated Sludge Model No. 1. For the IWA Activated Sludge Model No. 1, the model equations are written in the General Dynamical Model (2.100) framework with the following definitions (N = 10, M = 8):

$$\xi = \begin{pmatrix} S_S \\ X_S \\ X_{B,H} \\ X_{B,A} \\ X_P \\ S_O \\ S_{NO} \\ S_{NH} \\ S_{ND} \\ X_{ND} \end{pmatrix}, \quad F = \begin{pmatrix} DS_{S,in} \\ 0 \\ 0 \\ 0 \\ 0 \\ DS_{O,in} + k_L a(S_O^* - S_O) \\ 0 \\ DS_{NH,in} \\ 0 \\ 0 \end{pmatrix}, \tag{2.107}$$

$$Y = \begin{pmatrix} -\frac{1}{Y_H} & -\frac{1}{Y_H} & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1-f_P & 1-f_P & 0 & -1 & 0 \\ 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_P & f_P & 0 & 0 & 0 \\ -\frac{1-Y_H}{Y_H} & 0 & -\frac{4.57-Y_A}{Y_A} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1-Y_H}{2.86Y_H} & \frac{1}{Y_A} & 0 & 0 & 0 & 0 & 0 \\ -i_{XB} & -i_{XB} & -i_{XA} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & Y_B & Y_B & 0 & 0 & -1 \end{pmatrix} \tag{2.108}$$

$$\rho = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \end{pmatrix} = \begin{pmatrix} \mu_1 X_{B,H} \\ \mu_2 X_{B,H} \\ \mu_3 X_{B,A} \\ b_H X_{B,H} \\ b_A X_{B,A} \\ k_a S_{ND} X_{B,H} \\ \rho_{X7} X_{B,H} \\ \rho_{X8} X_{B,H} \end{pmatrix}, \quad Q = 0 \tag{2.109}$$

with ρ_{X7} and ρ_{X8} the specific (i.e. per $X_{B,H}$ units) hydrolysis rates of reactions (6) and (7), $Y_B = i_{XB} - f_P i_{XP}$, and $i_{XA} = i_{XB} + \frac{1}{Y_A}$.

Note the great similarity between the Petersen matrix of the IWA Activated Sludge Model No. 1 and the matrix Y. The matrix is simply the transpose[2] of the Petersen matrix where columns 1, 3 and 13 have been deleted (see footnote of Section 2.3.3).

Two step nitrification. In the nitrification example, the different terms of (2.100) are defined as follows (N = 5, M = 2):

---

[2] Indeed the rows correspond to the reactions in the IWA Activated Sludge Model No. 1, and to the process components in the matrix Y of the General Dynamical Model (2.100). Conversely, the columns correspond to the process components in the matrix Y of the General Dynamical Model (2.100), and to the reactions in the IWA Activated Sludge Model No. 1.

$$\xi = \begin{pmatrix} S_{NH} \\ S_{NO_2} \\ S_{NO_3} \\ X_1 \\ X_2 \end{pmatrix}, \; Y = \begin{pmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ 0 & Y_4 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \; F = \begin{pmatrix} DS_{NH,in} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.110}$$

$$\rho = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \mu_1 X_1 \\ \mu_2 X_2 \end{pmatrix}, \; Q = 0 \tag{2.111}$$

Two step denitrification. In the denitrification example, the different terms of (2.100) are defined as follows:

$$\xi = \begin{pmatrix} S_{NO_3} \\ S_{NO_2} \\ S \\ X \\ N_2 \end{pmatrix}, \; Y = \begin{pmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ -\frac{1}{Y_4} & -\frac{1}{Y_5} \\ 1 & 1 \\ 0 & Y_6 \end{pmatrix}, \; F = \begin{pmatrix} DS_{NO_3,in} \\ 0 \\ DS_{in} \\ 0 \\ 0 \end{pmatrix}, \; Q = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Q_{N_2} \end{pmatrix}$$

$$\rho = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \mu_1 X \\ \mu_2 X \end{pmatrix} \tag{2.112}$$

2.5 Multi Tank and Non Completely Mixed Reactors

In the preceding sections, we have considered the mass balances in tanks where the medium is homogeneous. However in many practical situations, this is not the case due to a lack of sufficient mixing: this is typically the situation of large-scale reactors, for which it is almost impossible to provide proper stirring to homogenise the reactor medium. However, the non complete mixing may also be deliberate; there is even a trend for increased use of processes which intrinsically are characterised by non homogeneous conditions. Typical examples are fixed bed and fluidised bed reactors, in which the (auto)catalysts are either fixed on some support, or maintained in suspension in the tank. Another obvious example is the settler which is used to separate biomass from the treated wastewater. Finally, plug flow reactors are also characterised by non-perfect mixing conditions resulting in (deliberate) concentration profiles along the length of the reactor.

Well-known from engineering text books is the simplest way to model non homogeneous tanks by considering them as a sequence of reactors. More fundamentally, the dynamics of fixed bed or fluidised bed reactors and settlers can be derived by considering mass balances in small (“infinitesimal”) volumes: this will result in partial differential equation (PDE) models, i.e. models in which the state variables (the process components) do not only depend on time but also on the position in the tank; these models also most often include partial derivatives of the state variables with respect to time and space. The finite difference of fixed bed reactor or settler PDE models will result in sequential reactor models: this clearly emphasises the relation between both types of models. In the following, we shall first consider

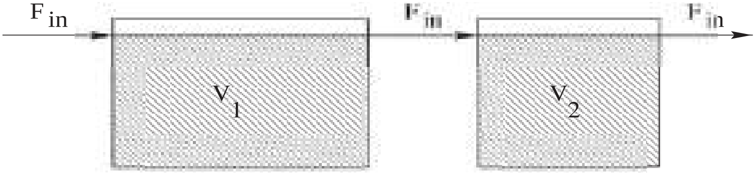

FIG. 2.12. Two sequential tanks.

the sequential reactor model, then the PDE models for fixed bed reactors (with and without axial and radial dispersion), fluidised bed reactors, and settlers.

2.5.1 *Sequential Reactors*

Let us first consider two sequential stirred tank reactors (Figure 2.12) of volumes V_1 and V_2, respectively, and in which a simple microbial growth reaction $S \longrightarrow X$ takes place. Let us denote each tank by an index i: i = 1 for the first tank, and 2 for the second. The mass balance for the substrate concentration S in both tanks is written as follows:

$$\frac{dS_1}{dt} = \frac{F_{in}}{V_1}S_{in} - \frac{F_{in}}{V_1}S_1 - \frac{1}{Y}\mu_1 X_1 \tag{2.113}$$

$$\frac{dS_2}{dt} = \frac{F_{in}}{V_2}S_1 - \frac{F_{in}}{V_2}S_2 - \frac{1}{Y}\mu_2 X_2 \tag{2.114}$$

Similarly for the biomass, the mass balance in both tanks gives rise to the following set of differential equations:

$$\frac{dX_1}{dt} = -\frac{F_{in}}{V_1}X_1 + \mu_1 X_1 \tag{2.115}$$

$$\frac{dX_2}{dt} = \frac{F_{in}}{V_2}X_1 - \frac{F_{in}}{V_2}X_2 + \mu_2 X_2 \tag{2.116}$$

in which μ_i is a function of the variables in tank i (i = 1 or 2), e.g. of S_1 and X_1 for μ_1, and of S_2 and X_2 for μ_2. An important feature of the substrate concentration equations is that the influent substrate concentration in tank 2 is the effluent substrate concentration from tank 1. The same remark holds for the biomass.

The generalisation to N (> 2) sequential reactors is then straigthforward. For instance, the mass balance of the substrate concentration in N STRs is written as follows:

$$\frac{dS_1}{dt} = \frac{F_{in}}{V_1}S_{in} - \frac{F_{in}}{V_1}S_1 - \frac{1}{Y}\mu_1 X_1 \tag{2.117}$$

$$\frac{dS_2}{dt} = \frac{F_{in}}{V_2}S_1 - \frac{F_{in}}{V_2}S_2 - \frac{1}{Y}\mu_2 X_2 \tag{2.118}$$

$$\vdots$$

$$\frac{dS_N}{dt} = \frac{F_{in}}{V_N} S_{N-1} - \frac{F_{in}}{V_N} S_N - \frac{1}{Y} \mu_N X_N \tag{2.119}$$

For the biomass, it is equal to:

$$\begin{aligned} \frac{dX_1}{dt} &= -\frac{F_{in}}{V_1} X_1 + \mu_1 X_1 \\ \frac{dX_2}{dt} &= \frac{F_{in}}{V_2} X_1 - \frac{F_{in}}{V_2} X_2 + \mu_2 X_2 \end{aligned} \tag{2.120}$$

$$\vdots \tag{2.121}$$

$$\frac{dX_N}{dt} = \frac{F_{in}}{V_N} X_{N-1} - \frac{F_{in}}{V_N} X_N + \mu_N X_N \tag{2.122}$$

The General Dynamical Model (2.100) formalism is still valid. Indeed, if for instance we choose to group the components by tank, i.e. if we define the component concentration vector ξ as follows:

$$\xi^T = \begin{pmatrix} S_1 & X_1 & S_2 & X_2 & \cdots & S_N & X_N \end{pmatrix} \tag{2.123}$$

the above mass balance equations can be written in the General Dynamical Model (2.100) framework by considering the following definitions:

$$Y = \begin{pmatrix} -\frac{1}{Y} & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & -\frac{1}{Y} & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -\frac{1}{Y} \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{2.124}$$

$$D = \begin{pmatrix} \frac{F_{in}}{V_1} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{F_{in}}{V_1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{F_{in}}{V_2} & 0 & \frac{F_{in}}{V_2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\frac{F_{in}}{V_2} & 0 & \frac{F_{in}}{V_2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{F_{in}}{V_N} & 0 & \frac{F_{in}}{V_N} & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\frac{F_{in}}{V_N} & 0 & \frac{F_{in}}{V_N} \end{pmatrix} \tag{2.125}$$

$$\rho = \begin{pmatrix} \mu_1 X_1 \\ \mu_2 X_2 \\ \vdots \\ \mu_N X_N \end{pmatrix}, \quad F = \begin{pmatrix} DS_{in} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Q = 0 \tag{2.126}$$

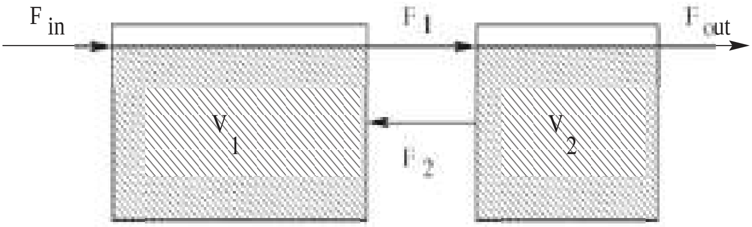

FIG. 2.13. Two interconnected tanks.

A first difference between the single tank model and the multi-tank one is, as it was already mentioned in the basic activated sludge model, that now the dilution rate is a matrix. A second difference is the increase of the dimension of the reaction rate ρ, and in consequence, of the yield coefficient matrix Y. Note finally the above matrix presentation is arbitrary: we could also have chosen to write the vector ξ by grouping the terms by component, i.e. the first rows for S and the remaining ones for X (yet the General Dynamical Model (2.100) formalism works!)(see also [56] for further details on multi-tank bioreactor models).

Let us end this section by the model of two interconnected CSTRs (Figure 2.13). Note that if the volumes V_1 and V_2 are constant, then $F_{out} = F_{in}$ and $F_2 = F_1 - F_{in}$. If we consider as before a simple microbial growth reaction in both tanks, the mass balance of the substrate concentration is then equal to:

$$\frac{dS_1}{dt} = \frac{F_{in}}{V_1}S_{in} + \frac{F_1 - F_{in}}{V_1}S_2 - \frac{F_1}{V_1}S_1 - \frac{1}{Y}\mu_1 X_1 \tag{2.127}$$

$$\frac{dS_2}{dt} = \frac{F_1}{V_2}S_1 - \frac{F_1}{V_2}S_2 - \frac{1}{Y}\mu_2 X_2 \tag{2.128}$$

The use of interconnected CSTR equations like (2.127)(2.128) is often considered to simulate the lack of medium homogeneity, e.g. due to insufficient stirring in reactors.

2.5.2 *Fixed Bed Reactor: The Basic Mass Balance Model*

Let us consider an example of a non completely mixed reactor: the fixed bed reactor, i.e. a tank in which the (auto)catalysts are *fixed* on some (solid) support and in which the reactants (in the liquid phase) are flowing through and are transformed when in contact with the (fixed) (auto)catalysts. Recently, a number of such reactor systems have been developed as they are capable of maintaining high autocatalyst concentrations and, hence, can sustain a high volumetric loading rate. Examples are submerged trickling filters for denitrification.

There is no specific stirring mechanism, and generally speaking, the reactor is not in completely mixed conditions anymore. Therefore in terms of (dynamic) modelling, the approach has to be modified: the mass balance has to be computed

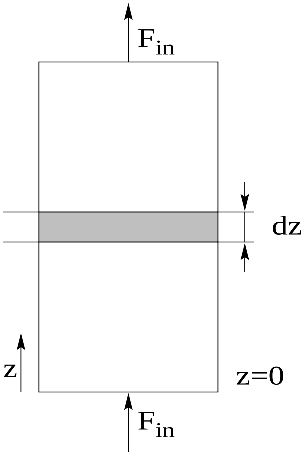

FIG. 2.14. Schematic view of a fixed bed reactor.

on a thin section dz of the reactor (and not anymore on the whole volume of the reactor) (see Figure 2.14).

Let us first concentrate on the dynamical model derivation for fixed bed reactors with axial dispersion, and for which radial dispersion is negligible. Assume that the dispersion phenomenon obeys Fick's diffusion law which expresses that the perpendicular flux of particles j through a unit surface is proportional to the gradient of concentration at the surface according to the following equation:

$$j = -D\frac{\partial C}{\partial z} \tag{2.129}$$

where D is a diffusion coefficient and z is the direction of the flux j. Indeed the "microscopic" phenomenon of diffusion has been by analogy extended to the "macroscopic" reactor situation, and the terminology "dispersion" is more often used than "diffusion" in fixed bed or fluidised bed reactor modelling. Let us note D_{ma} $(m^2.s^{-1})$ the axial mass dispersion coefficient.

For simplicity, we also consider that the reactor is tubular; this implies in particular that the cross-section of the reactor is constant (equal to A (m^2)). Furthermore this section is the sum of the section occupied by the (auto)catalysts (i.e. a "solid" section, A_S (m^2)) and of the section left free for the flow of (liquid) reactants and products, A_L (m^2) $(A = A_S + A_L)$.

The reactions are assumed to be autocatalytic and to take place on the *solid* catalytic support. Let us further assume that the mass transfer dynamics between the solid and liquid phases are negligible (This assumption may be formally justified by considering a *quasi steady state approximation*, e.g. [85], see also Section 2.9)

In order to explicit the model derivation, let us begin with the mass balance on the thin section dz of a component of concentration C ($kg.m^{-3}$) which intervenes in one reaction characterised by a reaction rate ρ ($kg.m^{-3}.s^{-1}$) (since the reaction takes place on the *solid* support, the dimension of the reaction is mass per time unit per *solid* volume (i.e. $A_S dz$):

$$\underset{\substack{\text{time variation}\\ \text{of the mass of C}}}{\frac{\partial}{\partial t}[A_L dz C]} = \underset{\substack{\text{inflow}\\ \text{at position z}}}{F_{in}C} \quad \underset{\substack{\text{outflow}\\ \text{at position z + dz}}}{-F_{in}(C+\frac{\partial C}{\partial z}dz)}$$

$$\underset{\substack{\text{diffusion}\\ \text{at position z}}}{-D_{ma}A_L\frac{\partial C}{\partial z}} \quad \underset{\substack{\text{diffusion}\\ \text{at position z + dz}}}{-(-D_{ma}A_L\frac{\partial}{\partial z}[C+\frac{\partial C}{\partial z}dz])}$$

$$\underset{\substack{\text{reaction rate}\\ \text{between z and z + dz}}}{+A_S dz\ \rho(z,t)}$$

where F_{in} is the influent flow rate ($m^3.s^{-1}$). If one divides both sides by $A_L dz$ and defines the void fraction ϵ and the fluid superficial velocity u ($m.s^{-1}$):

$$\epsilon = \frac{A_L}{A}, \ u = \frac{F_{in}}{A_L} \tag{2.130}$$

the above equation can be reduced to the following expression:

$$\frac{\partial C}{\partial t} = -u\frac{\partial C}{\partial z} + D_{ma}\frac{\partial^2 C}{\partial z^2} + \frac{1-\epsilon}{\epsilon}r(z,t) \tag{2.131}$$

Note that the above development assumes a constant cross-sectional area A and constant individual terms A_L and A_S.

2.5.3 *General Dynamical Model of Fixed and Fluidised Bed Reactors*

Fixed Bed Reactors with Axial Dispersion. Let us now generalise the model for fixed bed reactors with N components involved in M reactions [73]. Let us first assume that among the N process components, N_{fi} are microorganisms entrapped or fixed on some support and, hence, remain within the reactor, and N_{fl} other reactants (essentially substrates and products) flow through the reactor. In addition, let us define two associated component vectors, x_{fi} and x_{fl}, respectively.

From mass balance considerations on a section dz, we can deduce the following *General Dynamical Model for fixed bed reactors with axial dispersion*:

$$\frac{\partial x_{fi}}{\partial t} = Y_{fi}\ \rho(x_{fi}, x_{fl}) \tag{2.132}$$

$$\frac{\partial x_{fl}}{\partial t} = -u\frac{\partial x_{fl}}{\partial z} + D_{ma}\frac{\partial^2 x_{fl}}{\partial z^2} + \frac{1-\epsilon}{\epsilon} Y_{fl}\ \rho(x_{fi}, x_{fl}) \tag{2.133}$$

where $\rho(x_{fi}, x_{fl})$ is the reaction rate vector ($kg.m^{-3}.s^{-1}$):

$$\rho^T(x_{fi}, x_{fl}) = [\rho_1(x_{fi}, x_{fl}), \rho_2(x_{fi}, x_{fl}), ..., \rho_M(x_{fi}, x_{fl})] \tag{2.134}$$

and Y_{fi} and Y_{fl} are the yield coefficient matrices.

Remark: Note that, as in the completely mixed reactors, the above matrix formulation emphasises the presence of two terms in the process dynamics:

- *conversion*: $Y_{fi}\ \rho(x_{fi}, x_{fl})$ and $\frac{1-\epsilon}{\epsilon} Y_{fl}\ \rho(x_{fi}, x_{fl})$;
- *transport dynamics*: $-u\frac{\partial x_{fl}}{\partial z} + D_{ma}\frac{\partial^2 x_{fl}}{\partial z^2}$.

Boundary conditions have still to be added to complete the model. Because of the presence of the second order derivative with respect to z, two boundary conditions are required: typically, one at the reactor inlet ($z = 0$) and one at the outlet ($z = L$). Although they are the object of criticism in some instances (e.g. [85]), the most largely used boundary conditions are those of Danckwerts [63]:

$$D_{ma}\frac{\partial x_{fl}}{\partial z} = -u(x_{in} - x_{fl}) \text{ at } z = 0 \tag{2.135}$$

$$\frac{\partial x_{fl}}{\partial z} = 0 \quad \text{at } z = L \tag{2.136}$$

with x_{in} the influent value of x_{fl}.

Example: biochemical reactor with a growth reaction and a death/detachment reaction

Let us now consider a fixed bed reactor with a growth reaction and a death/detachment reaction:

$$\text{growth}: S \longrightarrow X + P \tag{2.137}$$

$$\text{death}: X \longrightarrow X_d \tag{2.138}$$

If we assume that the non-active microorganisms detach and leave the bioreactor, the dynamics of the bioprocess (2.137)(2.138) will be described by the following equations:

$$\frac{\partial X}{\partial t} = \mu X - bX \tag{2.139}$$

$$\frac{\partial S}{\partial t} = -u\frac{\partial S}{\partial z} + D_{ma}\frac{\partial^2 S}{\partial z^2} - \frac{1-\epsilon}{\epsilon}\frac{1}{Y_S}\mu X \tag{2.140}$$

$$\frac{\partial P}{\partial t} = -u\frac{\partial P}{\partial z} + D_{ma}\frac{\partial^2 P}{\partial z^2} + \frac{1-\epsilon}{\epsilon}Y_P \mu X \tag{2.141}$$

$$\frac{\partial X_d}{\partial t} = -u\frac{\partial X_d}{\partial z} + D_{ma}\frac{\partial^2 X_d}{\partial z^2} + \frac{1-\epsilon}{\epsilon}bX \tag{2.142}$$

$$\begin{aligned} z = 0 \ : \ & D_{ma}\frac{\partial S(0,t)}{\partial z} = -u(S_{in}(t) - S(0,t) \\ & D_{ma}\frac{\partial P(0,t)}{\partial z} = uP(0,t) \\ & D_{ma}\frac{\partial X_d(0,t)}{\partial z} = uX_d(0,t) \end{aligned} \tag{2.143}$$

$$z = L \ : \ \frac{\partial S(L,t)}{\partial z} = \frac{\partial P(L,t)}{\partial z} = \frac{\partial X_d(L,t)}{\partial z} = 0 \tag{2.144}$$

The above equations can be rewritten in the formalism of equations (2.132)(2.133) (2.135)(2.136) by considering the following definitions:

$$x_{fi} = X, \ x_{fl} = \begin{pmatrix} S \\ P \\ X_d \end{pmatrix} \tag{2.145}$$

$$x_{in} = \begin{pmatrix} S_{in} \\ 0 \\ 0 \end{pmatrix}, \ \rho = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \mu X \\ bX \end{pmatrix} \tag{2.146}$$

$$Y_{fi} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \ Y_{fl} = \begin{pmatrix} -\frac{1}{Y_S} & 0 \\ Y_P & 0 \\ 0 & 1 \end{pmatrix} \tag{2.147}$$

Extension 1: Plug Flow Reactor. The dynamical model of the fixed bed reactor in absence of dispersion, i.e. of the plug flow reactor, is readily obtained from the equations (2.132)(2.133)(2.135) by simply setting the dispersion coefficient D_{ma} to zero ($D_{ma} = 0$). Since there is then only a first order derivative of the state variable x_{fl} with respect to the space variable z, only the boundary condition at the reactor input ($z = 0$) (2.135) is kept, and the boundary condition at the reactor output (2.136) is dropped. The model equations are then:

$$\frac{\partial x_{fi}}{\partial t} = Y_{fi} \ \rho(x_{fi}, x_{fl}) \tag{2.148}$$

$$\frac{\partial x_{fl}}{\partial t} = -u\frac{\partial x_{fl}}{\partial z} + \frac{1-\epsilon}{\epsilon}Y_{fl} \ \rho(x_{fi}, x_{fl}) \tag{2.149}$$

$$x_{fl} = x_{in} \qquad at \ z = 0 \tag{2.150}$$

The frequent description of an activated sludge plant as a plug flow system is just a special case of the above in which no fixed biomass is present and hence equation(2.148) vanishes.

Extension 2: Gas phase. Assume that one product gives off in the gaseous phase, as in high rate fixed film anaerobic digesters (CO_2, CH_4) [67] or denitrification reactors (N_2) [102], [131]; or that the fixed bed reactor is aerated for aerobic carbon removal [172] or nitrification [283]. Note that the very widespread trickling filter also has to be described as a three phase fixed bed reactor [224]. The equations for fixed bed systems (2.132)-(2.136) will be modified by introducing the gaseous flow rate vector Q ($kg.s^{-1}$). By introducing similar arguments as in equation (2.131), a new term, the derivative of Q with respect to z, is introduced in the model formulation:

$$\frac{\partial x_{fi}}{\partial t} = Y_{fi}\rho(x_{fi}, x_{fl}) \tag{2.151}$$

$$\frac{\partial x_{fl}}{\partial t} = -u\frac{\partial x_{fl}}{\partial z} + D_{ma}\frac{\partial^2 x_{fl}}{\partial z^2} + \frac{1-\epsilon}{\epsilon}Y_{fl}\ \rho(x_{fi}, x_{fl}) - \frac{1}{A_L}\frac{\partial Q}{\partial z} \tag{2.152}$$

with the same boundary conditions (2.135)(2.136) as above. Note that the section A is now the sum of three sections, the solid, liquid and gaseous sections: $A = A_S + A_L + A_G$. In the above model formulation, it has been implicitly assumed that the section A_G is negligible with respect to A_S and A_L. If this assumption is not correct, as in trickling filters, then the term $(1-\epsilon)/\epsilon$ should be replaced by A_S/A_L.

Note that the above model extension to gas production is only valid if the dynamics between the liquid and gas phases are fast enough to be considered as being negligible, i.e. the gas and liquid phases are assumed to be in equilibrium.

As a matter of illustration, assume that in the bioreactor example (2.137) (2.138), the product P gives off in the gaseous phase. Then the model is completed with the following gaseous flow rate vector:

$$Q = \begin{pmatrix} 0 \\ Q_P \\ 0 \end{pmatrix} \tag{2.153}$$

where Q_P is the gaseous flow rate of the product P.

Extension 3: Radial Dispersion. Here again the extension (e.g. [85], [145]) is straightforward if one introduces a new spatial variable, the radial coordinate r in case of a tubular reactor. The model is readily obtained by using arguments similar to those used for the derivation of equations (2.132)(2.133)(2.135)(2.136): it results in an addition of a radial dispersion term and of two boundary conditions (at $r = 0$ and $r = R$, with R the radius of the reactor, see Figure 2.15). In the presence of radial dispersion, the two-phase model (2.132)(2.133)(2.135)(2.136) (i.e. without the gas phase) becomes:

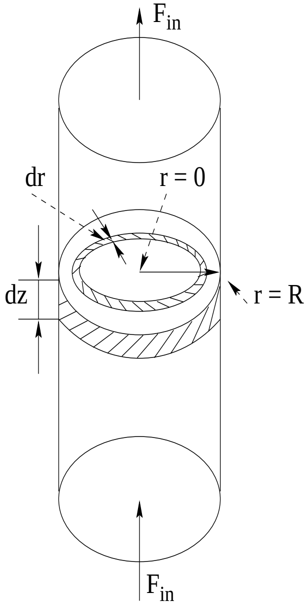

FIG. 2.15. Schematic view of a fixed bed reactor with radial dispersion.

$$\frac{\partial x_{fi}}{\partial t} = Y_{fi}\ \rho(x_{fi}, x_{fl}) \tag{2.154}$$

$$\begin{aligned}\frac{\partial x_{fl}}{\partial t} &= -u\frac{\partial x_{fl}}{\partial z} + D_{ma}\frac{\partial^2 x_{fl}}{\partial z^2} + D_{mr}\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial x_{fl}}{\partial r}) \\ &\quad +\frac{1-\epsilon}{\epsilon}Y_{fl}\ \rho(x_{fi}, x_{fl})\end{aligned} \tag{2.155}$$

$$D_{ma}\frac{\partial x_{fl}}{\partial z} = -u(x_{in} - x_{fl}) \qquad at\ z = 0 \tag{2.156}$$

$$\frac{\partial x_{fl}}{\partial z} = 0 \qquad at\ z = L \tag{2.157}$$

$$\frac{\partial x_{fl}}{\partial r} = 0 \qquad at\ r = 0 \tag{2.158}$$

$$\frac{\partial x_{fl}}{\partial r} = 0 \qquad at\ r = R \tag{2.159}$$

where D_{mr} $(m^2.s^{-1})$ is the radial mass dispersion coefficient.

Extension 4: Fluidised Bed Reactors. Fluidised bed technology as many others stems from chemical engineering technology, and has its roots in solid bed reactors

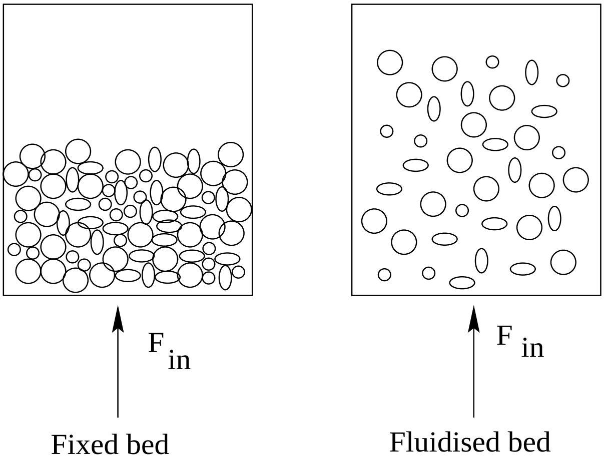

FIG. 2.16. Fixed bed versus fluidised bed.

[177]. In fluidised bed reactors, the solid (catalytic) particles are “fluidised” by the stream of liquid (most common) or gas [54] [249] from below the reactor as the upflow velocity increases. Figure 2.16 illustrates the difference between a fixed (or “packed”) bed reactor and a fluidised bed. If the fluid velocity is low, the particles are packed at the bottom of the reactor: the reactor behaves like a fixed bed reactor since the solid particles do not move upwards. For larger values of the fluid or gas velocity, the solid particles will be “pushed” upwards and may be maintained in suspension for appropriate hydrodynamical conditions, i.e. an expanded bed is formed.

The oldest application of fluidisation [227] in biotechnology is in the area of nitrification and denitrification of wastewater and drinking water treatment [243] [84]. Another important application area of biofluidisation is in the field of anaerobic digestion with the very widespread use of the UASB (Upflow Anaerobic Sludge Bed) reactors [67] [82].

The dynamical modelling of fluidised bed reactors requires even more attention and care. In terms of dynamical modelling, the basic difference between fixed bed and fluidised bed reactors is that in the latter, the solid phase (basically the catalyst on its support) is not fixed but is composed of particles in suspension in the reactor. A major implication is that now, generally speaking, the void fraction ϵ cannot be considered as being constant, especially during expansion and contraction of the bed.

At this point there are two possible options.

1. Either to consider that the transient phase during bed expansion and contraction is very fast and to neglect its dynamics. Then the void fraction ϵ in steady state can be deduced by considering the (heuristic but largely used)

Richardson and Zaki's law [216]:

$$\epsilon = (\frac{u_0}{U_T})^{\frac{1}{n}} \tag{2.160}$$

where u_0 is the fluid superficial velocity in the absence of solid particles, U_T is the terminal settling velocity of the particles and n is the expansion index (which typically depends on the type of particles and medium).

2. Or to consider the dynamics of the bed expansion and contraction. Then extra equations are required. Typically the mass balances equations (also sometimes called the *continuity equations*)(for instance, (2.132)(2.133) or (2.148)(2.149)) are completed with momentum equations [89]. In the plug flow case, these are written as follows for the liquid and particle phases, respectively:

$$\rho_L \frac{\partial}{\partial t}[\epsilon u] = -\frac{\partial}{\partial z}[\epsilon(p_L + \frac{1}{2}\rho_L u^2)] - \epsilon\ \rho_L\ g - F_i$$

$$\rho_S \frac{\partial}{\partial t}[(1-\epsilon)u_S] = -\frac{\partial}{\partial z}[(1-\epsilon)(p_S + \frac{1}{2}\rho_L u^2)] - (1-\epsilon)\ \rho_S\ g + F_i$$

where ρ_L, ρ_S, p_L, p_S, g, F_i and u_S are the liquid phase and particle phase densities ($kg.m^{-3}$), the liquid and particle pressures ($N.m^{-2}$), the gravity constant ($m.s^{-2}$), the liquid-particle interaction force per unit volume ($N.m^{-3}$) and the particles' velocity in the reactor ($m.s^{-1}$), respectively. Moreover the Richardson and Zaki's equation has to be modified to account for the transient of the solid particles (e.g. [199]:

$$u - u_S\epsilon = U_T\epsilon^{n-1} \tag{2.161}$$

The choice of appropriate expressions for F_i, p_L and p_S is still a matter of discussion in the scientific literature. As a matter of illustration, the following expressions are recommended e.g. in Foscolo and Gibilaro [89], and Jean and Fan [132] for F_i, p_L and p_S:

$$F_i = (1-\epsilon)\rho_L g + (1-\epsilon)(\rho_S - \rho_L)g(\frac{u_0 - u_S}{U_T})^{\frac{4.8}{n}}\epsilon^{-4.8} \tag{2.162}$$

$$p_S = 0,\ \epsilon p_L = -(\epsilon\rho_L + (1-\epsilon)\rho_S)gz \tag{2.163}$$

Another important equation is the volume balance at any position in the reactor:

$$\epsilon u + (1-\epsilon)u_S = u_0 \tag{2.164}$$

Because of the much lower complexity of the model, the first option is very often considered in many applications.

An example of dynamical model of an anaerobic digestion process in a fluidised bed reactor is given in Bonnet *et al.* [30].

2.5.4 *Settlers*

Solid flux theory basic model. In Section 2.5.2, we have considered that the dynamics of the settler (Figure 2.17) are often described by those of a CSTR. Largely because the model neglects sedimentation effects, the CSTR assumption is more and more an object of criticism in the scientific community, and alternative models are presently recommended. Let us first recall that the main functions of secondary settlers in wastewater treatment plants are double:

1. **clarification**, i.e. separation of the biomass from the treated wastewater in order to produce a solid-free effluent;
2. **thickening** of the biomass at the bottom of the settler to be recycled back into the aerator.

In some cases the settler is also considered for more advanced use, i.e.

1. **sludge storage**, i.e. in the bottom part of the settler sludge is stored for subsequent use under high waste load conditions;
2. **the settler as a reactor** where additional aerobic conversions can occur or where denitrification may take place [230] (with the danger of formation of nitrogen gas bubbles and rising sludge problems [119]).

Hence, to be able to evaluate the functioning of the settler under dynamic conditions, models must be able to describe these different aspects in more or less detail [83].

The alternative models are usually based on (or at least refer to) mass balances (here again, the terminology “continuity equation” is largely used) and on the Solid Flux Theory [69]. The mass balance is computed not anymore on the whole tank but on a small section dz (see Figure 2.17). If we recall the mass balance considerations of Section 2.5.2, the dynamical model in its simplest form (e.g. [170]) is written as follows:

$$\frac{\partial X}{\partial t} = -\frac{\partial F_l}{\partial z} \tag{2.165}$$

$$X(z = H, t) = X_R(t) \tag{2.166}$$

where F_l is the biomass solid flux ($kg.m^{-2}.h^{-1}$]) at depth z of the settler ($z \in [0,\ H]$, where 0 and H correspond to the top and bottom of the settler, respectively). The flux is basically due to two phenomena: sedimentation (*settling*) due to gravity, and “bulk” flow due to sludge withdrawal at the bottom. Note that, in the above equation (2.165), only convection ($\partial F_l/\partial z$) is considered.

The total flux $F_l(t, z)$ can be expressed as follows:

$$F_l(z, t) = \nu X + uX \tag{2.167}$$

where ν is the settling velocity, and u is the bulk velocity, i.e. the velocity of the recycled biomass mixed liquor due to underflow pumping. With u simply the ratio of the flow rate $F_R + F_w$ over the settler section A_{se}:

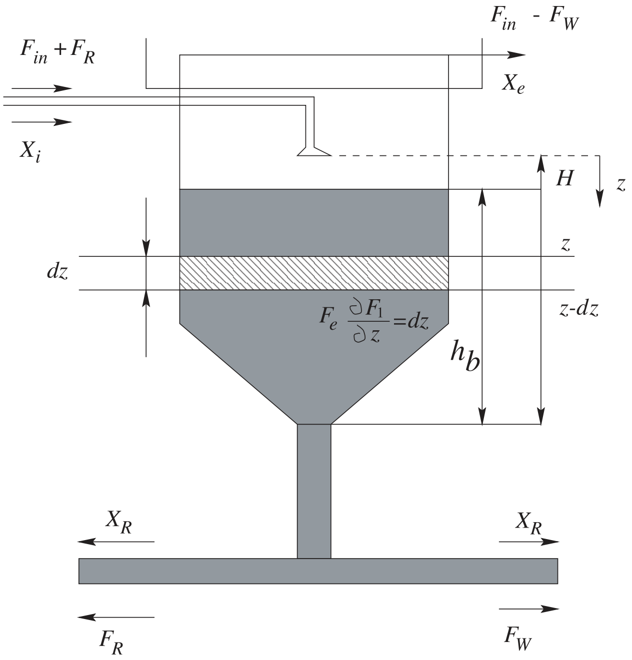

FIG. 2.17. Schematic view of a secondary settler.

$$u = \frac{F_R + F_w}{A_{se}} \tag{2.168}$$

and if we denote:

$$\nu X = f(X) \tag{2.169}$$

then the mass balance equations (2.165)(2.166) are rewritten as follows:

$$\frac{\partial X}{\partial t} = -(\frac{F_R + F_w}{A_{se}} + \frac{\partial f(X)}{\partial X})\frac{\partial X}{\partial z} + X\frac{F_R + F_w}{A_{se}^2}\frac{dA_{se}}{dz} \tag{2.170}$$

$$X(H, t) = X_R(t) \tag{2.171}$$

There exists in the literature a number of different models for the settling velocity ν, e.g.:

- Vesilind exponential model:

$$\nu = \nu_0 e^{-aX}, \qquad \nu_0, a > 0 \tag{2.172}$$

- Power law:

$$\nu = nX^{-a}, \qquad n, a > 0 \tag{2.173}$$

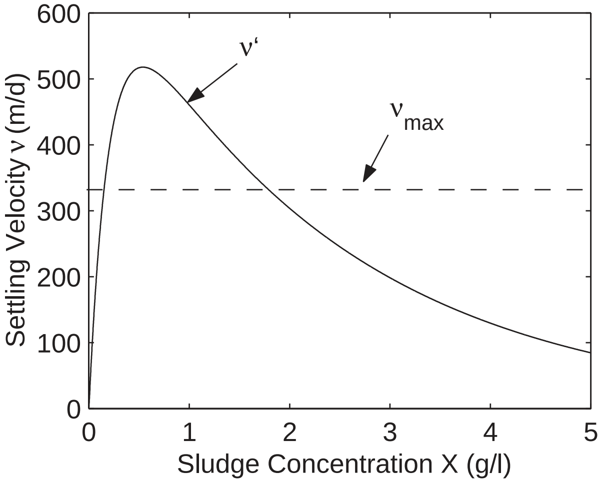

FIG. 2.18. Settling velocity: Takacs model.

- Inverse power law:

$$\nu = \frac{a}{b + X^n}, \qquad n, a, b > 0 \tag{2.174}$$

- Takacs model ("generalised Vesilind model") [245]:

$$\nu = \nu' \qquad if\, \nu' \leq \nu_{max} \tag{2.175}$$

$$= \nu_{max} \qquad if\, \nu' > \nu_{max} \tag{2.176}$$

$$\text{with} \quad \nu' = \nu_0 (e^{-r_h(X - X_{min})} - e^{-r_p(X - X_{min})}), \quad \nu_0, X_{min} > 0,\ r_p > r_h > 0 \tag{2.177}$$

with ν_{max} the maximum attainable settling velocity, X_{min} the minimum attainable suspended solids concentration, and r_h and r_p the settling characteristics of the hindered settling zone and of the low solids concentration, respectively. The model is graphically represented in Figure 2.18. Note that with that model, the maximum attainable settling velocity ν_{max} is generally speaking different from (lower than) the maximum of ν' which is reached for a value of X equal to:

$$X = X_{min} + \frac{1}{r_p - r_h} \ln\left(\frac{r_p}{r_h}\right) \tag{2.178}$$

- Cho model [59]

$$\nu = \nu_0 \frac{e^{-aX}}{X}, \qquad \nu_0, a > 0 \tag{2.179}$$

Remark: the model (2.170)(2.171) only considers one spatial coordinate: this means that we have implicitly assumed that the settler dynamics are uniform at each section dz. More complicated models can also be derived when this assumption is not fulfilled in practice, with 2 or 3 (for non-circular section settlers) space coordinates (see for instance [244]): this type of model also includes momentum conservation equations.

Steady-state curves of the settler model. Before going further, let us see what is the relative importance of the settling term and of the "bulk" flow term, at least in steady state. Figure 2.19 exhibits some curves, in steady state, of solid fluxes F_l as a function of the biomass X for the Vesilind model (2.172) and with the bulk velocity u as an operational parameter. It shows that for a given value of u, the recycled biomass concentration X_R is uniquely determined by the solid flux established in the settler. Indeed the minimum solid flux F^* represents the maximum rate of solids transferred through the liquid/solid interface for a given settler geometry and sludge settling behaviour. Excess solids which cannot be transferred to the bottom accumulate above the interface. On the other hand, in steady state, the underflow solids concentration is not a dynamic variable but depends on the limiting flux and available mass above the interface. If the settler is critically loaded, the total limiting flux through the thickening zone, i.e. $A_{se}F^*$ is equal to the underflow withdrawal, i.e. $(F_R + F_W)X_R$:

$$A_{se}F^* = (F_R + F_W)X_R \tag{2.180}$$

This gives the value of the recycled biomass:

$$X_R = \frac{A_{se}F^*}{F_R + F_W} \tag{2.181}$$

For the Vesilind model, ν tends to zero if X tends to infinity:

$$\lim_{X \to \infty} \nu = 0 \tag{2.182}$$

This implies that the solid flux F_l tends to uX for large values of X. Therefore the value of X_R can be graphically (Figure 2.19) determined from the intersection of F^* and uX (which is the asymptote of F_l for X tending to infinity).

Limitations of the solid flux theory basic model. The major drawback of the Solid Flux Theory basic model (2.165)(2.166) is its inability to emphasise concentration gradients at least in steady state. Indeed in steady state ($\frac{\partial X}{\partial t} = 0$) for settlers with constant section, we have:

$$\frac{dF}{dz} = 0 \tag{2.183}$$

i.e., since all the proposed models for the settling velocity ν are only functions of the sludge concentration X:

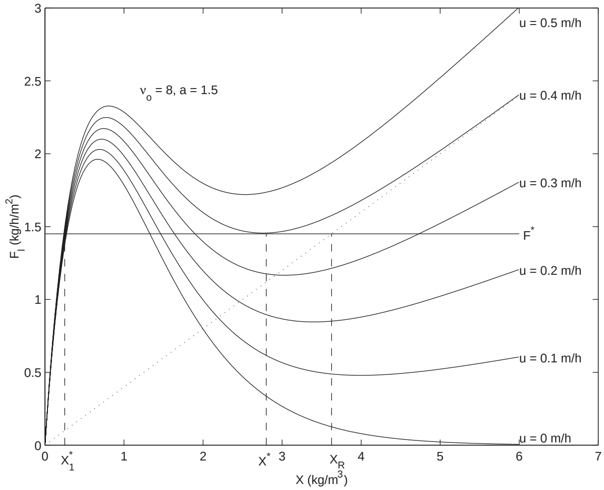

FIG. 2.19. Solid flux vs biomass concentration.

$$(u + \nu + X\frac{\partial \nu}{\partial X})\frac{dX}{dz} = 0 \tag{2.184}$$

There are therefore a priori (mathematically) two possible steady state solutions of (2.183) [75], [205]:

$$1)\ u + \nu + X\tfrac{\partial \nu}{\partial X} = 0 \tag{2.185}$$

$$2)\qquad \tfrac{dX}{dz} = 0 \tag{2.186}$$

Generally speaking, the first steady state has to be rejected, because it implies that:

$$\nu = -u + K\frac{1}{X}, \qquad K \text{ constant} \tag{2.187}$$

e.g. that ν is a function of u in steady state. This also imposes one model structure for the dependence of the settling velocity ν with respect to the sludge concentration X and the superficial fluid velocity u. ν will even be negative for sufficiently large values of X ($> u/K$).

Therefore we end up with one possible steady state,

$$\frac{dX}{dz} = 0 \tag{2.188}$$

which means that there is no spatial profile in steady state, in major contradiction with the experimental evidence (as illustrated in Figure 2.21 here below).

To circumvent the difficulty, different modifications have been proposed to the mass balance model. The most popular and most successful presently is the model proposed by e.g. [278],[245]. The model is based on two important assumptions: the first one is to consider a layered settler model (Figure 2.20), which can be viewed as a finite difference approximation of the space derivative in the mass balance model; the second one is a limitation of flux from one layer to another ("The mass flux into a layer cannot exceed the mass flux the volume is capable of passing, nor can it exceed the mass flux which the volume immediately below it is capable of passing").

The gap between the layered model and its assumed original model, the Solid Flux Theory based mass balance equations (2.165)(2.166)(2.167), with only a convection term ($\frac{\partial F_l}{\partial z}$), is indeed quite large. It is well known (e.g. [93], [73] and also Section 2.9.5) that a large number of "layers" (typically larger than 100) are necessary to obtain a satisfactory approximation of hyperbolic systems like (2.165)(2.166)(2.167) or plug flow reactor models (2.149)(2.150) that basically behave like delay systems (see also section 2.10), while a lower number of "layers" are necessary for approximating correctly parabolic systems with second order derivatives like the modified settler model (2.190) here below or the axial dispersion model (2.133)(2.135)(2.136). In practice, a limited number of layers are usually considered in the application of the layered model. This, combined with the flux limitation, contributes to make the layered model closer to a model with the second order derivatives (as proposed in the modified version of the settler model). Strangely, the bad level of approximation of the original model (i.e. a selection of not enough layers for a good approximation) results to a simulation model that tends to mimic and therefore behaves closer to a model that is able to emphasize the spatial gradient in settlers. This probably explains at least partially the popularity of the layered models in numerical simulation applications. We see therefore that the original model has been approximated in order to obtain a model that behaves more closely to the physical reality, but the connection of the resulting model with the original model is difficult to see. This renders even more difficult the interpretation of the resulting model with respect to the physical phenomenon that it is supposed to describe. It further motivates to go back to the basic physical phenomena to build physically sound models.

As we have just said, although the model proves quite successful in applications, the physical basis of the model is questionable. As clearly shown by Jeppsson ([133], there is no rigourous connection between the layered settler model and the "convection-type" Kynch theory derived model. The main problem is that the flux limitation should be inherently connected to the mass balance, and not introduced in the lumped model after the approximation procedure. In other words, the question could be formulated as follows: is it possible to derive the mass balance model such that a limitation of flux is implicitly included (i.e. a consequence of the structure of the physical model)?

A *possible* solution to the above question and to that of the presence of spatial gradient in settlers is indeed to introduce a second order derivative term with re-

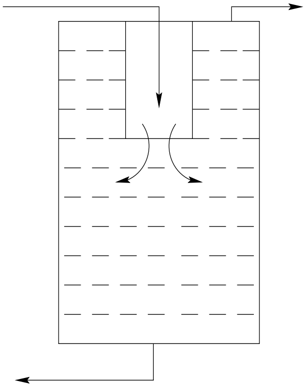

FIG. 2.20. Layered settler scheme.

spect to z in the mass balance model. A possible interpretation of such a term is that it represents diffusion following Fick's law. This type of term is largely used e.g. in chemical engineering to emphasise non perfect plug flow hydrodynamics and the presence of "back-mixing" effects in reactors like fixed-bed reactors. In such a case, the flux F is written as follows:

$$F = uX + \nu X - D_a \frac{\partial X}{\partial z} \tag{2.189}$$

and the mass balance model becomes:

$$\frac{\partial X}{\partial t} = -u \frac{\partial X}{\partial z} - \frac{\partial(\nu X)}{\partial z} + D_a \frac{\partial^2 X}{\partial z^2} \tag{2.190}$$

Let us now explain why this is a possible solution to introduce a spatial gradient in the settler model. In order to keep the line of reasoning as simple as possible, let us assume that the settling velocity ν is constant (this assumption is only needed for this purpose, and this does not mean that we impose ν to be constant in the settler model).

Let us then calculate the steady state solution (i.e. when $\frac{\partial X}{\partial t} = 0$) of the PDE equation with the second order derivative term. The above assumption (constant ν) allows us to have an analytical solution of the steady state (it is indeed its main merit). In steady state, the mass balance equation then becomes:

$$D_a \frac{d^2 X}{dz^2} - (u + \nu) \frac{dX}{dz} = 0 \tag{2.191}$$

Let us consider here the following boundary conditions [3]:

---

[3] These have been chosen in accordance with the physical reality, but will not be very useful in practice for simulating the model, because it requires the values of the sludge concentrations at the settler's input and output.

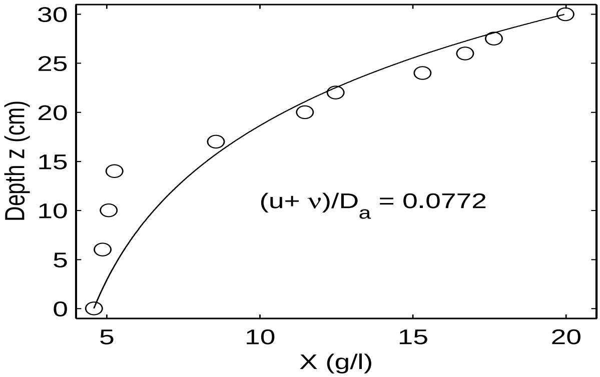

FIG. 2.21. Identification of model (2.193) with experimental data.

$$X(z = H) = X_R, \ X(z = 0) = X_i \tag{2.192}$$

where X_R and X_i are the sludge concentrations at the output and at the input of the settler, respectively. Then the analytical solution has the following form:

$$X(z) = \frac{1}{1 - e^{\frac{u+v}{D_a}H}}[(X_i(e^{\frac{u+v}{D_a}z} - e^{\frac{u+v}{D_a}H}) - X_R(e^{\frac{u+v}{D_a}z} - 1)] \tag{2.193}$$

Note that the above steady state expression of X depends explicitly on the spatial z ($e^{\frac{u+v}{D_a}z}$) in accordance with the experimental evidence (accumulation of sludge from the top to the bottom of the settler). This is illustrated on one typical steady state data set (from a cylindrical laboratory settler from the CEIT (San Sebastian, Spain) [253]): Figure 2.21 presents the results of the identification of the model equation (2.193) via a Levenberg-Marquardt identification routine compared to the data. The identification procedure gives the following value to the parameter:

$$\frac{u+v}{D_a} = 0.0772 \tag{2.194}$$

It is worth noting that the above suggested approach is in line with a number of recent works in the area. A similar model with a dispersion term is also considered in [109]. The report of Vanrolleghem *et al.* [260] is dealing with the parameter identification of such a model. And another approach which results in a model which has some similarities is considered by Cacossa and Vaccari [46]. In their paper, the authors consider the following settling velocity expression:

$$v = v_m(1 - \frac{1}{K}\frac{\partial X}{\partial z}) \tag{2.195}$$

where v_m and K are the maximum settling velocity and the compressibility function, respectively. Then the dynamical model of the settler is written as follows:

$$\frac{\partial X}{\partial t} = -u\frac{\partial X}{\partial z} - \nu_m\frac{\partial X}{\partial z} + \frac{\nu_m}{K}(\frac{\partial X}{\partial z})^2 + \frac{\nu_m}{K}X\frac{\partial^2 X}{\partial z^2} \tag{2.196}$$

Here again a second order derivative term with respect to z has been introduced in the model.

The settler model equation (2.190) has only been considered to draw attention to the fact that with a second order derivative term with respect to z, a gradient in the sludge concentration can be emphasised. This is its prime advantage, but we do not suggest modelling settler dynamics using this format for, because it has at least two drawbacks to be feasible in more general practical situations.

1. The settling dynamics with respect to X are linear (ν is constant): all the models proposed in the literature are nonlinear.
2. The boundary conditions are not very convenient, at least for numerical simulation (indeed the values of the concentrations at both inputs and outputs of the settler are needed to solve the equations, while the objective should be to be able to predict one (and also the distribution inside the settler) from the other). That's why we suggest to use, in line with those used for non completely mixed reactors (see Section 1.5.2), the following boundary conditions:

$$X(z=0) = X_i, \ \frac{\partial X}{\partial z}(z=H) = 0 \tag{2.197}$$

A clarifier/sedimentation model. By using the arguments considered above, we can derive a model that combines the clarification and sedimentation aspects in the settler. The model derivation is based on Figure 2.22. It is also a one-dimensional model with respect to space. An important question is that of the boundary conditions at the inlet. Because the model considers only one dimension (vertical), the interface corresponding to the inlet can be considered as homogeneous. If we denote by X_1 and X_2 the sludge concentration in the settler and in the clarifier, respectively, the mass balance equations for clarifiers/settlers with constant cross sections are written as follows:

- Settler

$$\frac{\partial X_1}{\partial t} = -\frac{F_R + F_W}{A_{se}}\frac{\partial X_1}{\partial z} - \frac{\partial(\nu_1 X_1)}{\partial z} + D_a\frac{\partial^2 X_1}{\partial z^2} \tag{2.198}$$

$$\frac{\partial X_1}{\partial z} = 0 \qquad \text{for } z = H_1 \tag{2.199}$$

$$D_a\frac{\partial X_1}{\partial z} = \frac{F_R + F_W}{A_{se}}(X_1 - X_i) - \nu_i X_i + \nu_1 X_1 \qquad \text{for } z = 0 \tag{2.200}$$

- Clarifier

$$\frac{\partial X_2}{\partial t} = -\frac{F_{in} - F_W}{A_{se}}\frac{\partial X_2}{\partial z} - \frac{\partial(\nu_2 X_2)}{\partial z} - D_a'\frac{\partial^2 X_2}{\partial z^2} \tag{2.201}$$

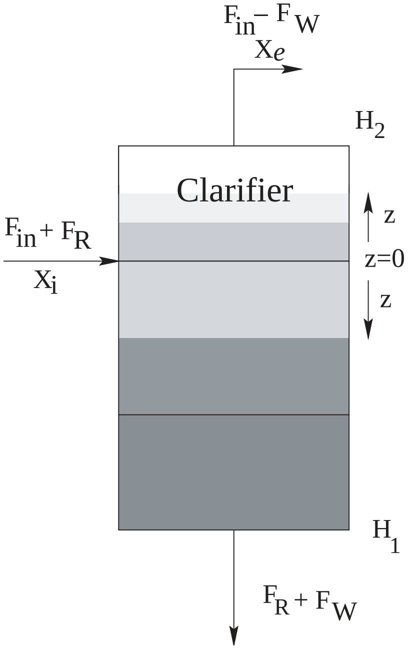

FIG. 2.22. Clarifier/settler scheme.

$$\frac{\partial X_2}{\partial z} = 0 \qquad \text{for } z = H_2 \quad (2.202)$$

$$D_a' \frac{\partial X_2}{\partial z} = \frac{F_{in} - F_W}{A_{se}}(X_2 - X_i) - \nu_i X_i + \nu_2 X_2 \qquad \text{for } z = 0 \quad (2.203)$$

In line with other works [154], [281], we have introduced the flexibility of the potential existence of different dispersion coefficients for the clarifier and for the settler.

Determination of the sludge blanket height. There is still an important variable: the sludge blanket height h_b. The easiest way is to calculate it as the location where the concentration X becomes larger than a certain threshold value when starting from the top of the settler. One possible alternative to determine it has been proposed by Stehfest [241]. It is based on the following line of reasoning (the technical mathematical details are extensively discussed in [241]). Let us consider the mass balance equation (2.165) around the interface liquid/solid, and integrate it with respect to the spatial coordinate z (with z_a and z_b (unknown) on each side of h_b):

$$\frac{d}{dt}\int_{z_a}^{z_b} X dz = F(X(z_b)) - F(X(z_a)) \quad (2.204)$$

If the concentration X_- and X_+ are the values of X for $z \to h_b$ from below and above, then the above equation becomes:

$$\frac{dh_b}{dt} = \frac{F(X_+) - F(X_-)}{X_- - X_+} \tag{2.205}$$

The computation of h_b also requires the knowledge of X_- and X_+. These are given by mass balance equations for the quantities of biomass above and below h_b. In order to have a simple operational model, the biomass concentrations are assumed to be homogeneous in both zones. This means that the quantities of biomass in the zones above and below h_b are equal to $(H - h_b)X_a$ and $h_b X_b$. After some mathematical manipulations, the dynamical model for determining h_b proposed by [241] has the following form:

$$\frac{dX_a}{dt} = \frac{1}{H - h_b}(F(X_i) - F(X_a)) \tag{2.206}$$

$$\frac{dX_b}{dt} = \frac{1}{h_b}(F(X_a) - F(X_0) + (X_b - X_a)\frac{F(X_a) - F(X_-)}{X_a - X_-} \tag{2.207}$$

$$\frac{dh_b}{dt} = \frac{F(X_a) - F(X_-)}{X_- - X_a} \tag{2.208}$$

with:

$$X_- = min(X_b, X_1^*), \;\; X_0 = max(X_b, X^*) \tag{2.209}$$

where X_1^* and X^* are the steady-state values defined in Figure 2.19 and Section 2.5.4 (subsection “steady-state curves of the settler model”).

2.6 Linear vs Nonlinear Models

Up to now we have introduced a number of models, and we have formalised all the models based on mass balance in one general dynamical model framework, for stirred tank reactors on one hand, and for non completely mixed reactors on the other hand. In Section 1.2, we have introduced a brief classification of models. In particular, we have made a distinction between linear and nonlinear models. Strictly speaking, what we had considered was the notion of systems linear (or nonlinear) *in the state x and the input u*, and also linear (or nonlinear) *in the parameters*. Let us illustrate further these notions by an example: let us consider the simple microbial growth model (2.16) (2.17) repeated below as (2.210) (2.211):

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\mu X \tag{2.210}$$

$$\frac{dX}{dt} = -DX + \mu X \tag{2.211}$$

With Monod kinetics, the above equations are written as follows:

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\frac{\mu_{max} S}{K_S + S}X \tag{2.212}$$

$$\frac{dX}{dt} = -DX + \frac{\mu_{max} S}{K_S + S} X \tag{2.213}$$

These equations are clearly nonlinear in the state (via the conversion term which includes the multiplication of two “states” S and X, nd the division by a function of S, i.e. $K_S + S$). It is also nonlinear in the parameters Y, μ_{max} and K_S. And if the dilution rate D is considered as an input, then there is a bilinear term appearing in both equations: DS and DX, respectively.

Let us consider another viewpoint. For instance, let us consider zero order kinetics with respect to S for the growth rate ρ ($= \mu X$ in the above equations (2.210) (2.211)) (because for instance in the considered application, the substrate concentration S is large compared to the substrate affinity constant K_S). Hence, the specific growth rate μ can be assumed to be almost constant and equal to the maximum specific growth rate μ_{max}. This means that the growth rate can be written as follows:

$$\rho = \mu_{max} \tag{2.214}$$

and the mass balance equations take the following form:

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y} \mu_{max} X \tag{2.215}$$

$$\frac{dX}{dt} = -DX + \mu_{max} X \tag{2.216}$$

Now the model is linear in the states S and X. However, it is nonlinear in the parameters Y and μ_{max} (division by Y and “multiplication” $\frac{1}{Y}\mu_{max}$. Yet, by considering the following (one-to-one) transformation:

$$\theta_1 = \frac{1}{Y} \mu_{max}, \ \theta_2 = \mu_{max} \tag{2.217}$$

the model can be rewritten as follows:

$$\frac{dS}{dt} = DS_{in} - DS - \theta_1 X \tag{2.218}$$

$$\frac{dX}{dt} = -DX + \theta_2 X \tag{2.219}$$

which is linear in the parameters θ_1 and θ_2. (Note that we could have used the same argument with the Monod kinetics mass balance model if the problem had been to estimate the parameters $\frac{1}{Y}$ and μ_{max} under the assumption that K_S is known, albeit that this model remains nonlinear in the states.)

Finally, assume that the operating conditions are such that the dilution rate D is fixed and that the influent substrate concentration S_{in} can be manipulated, then S_{in} is the system input and D can be considered as a parameter: this implies that the model is also linear in the input. Then the model (2.218)(2.219) can be written in the linear model framework:

$$\frac{d}{dt}\begin{pmatrix} S \\ X \end{pmatrix} = \begin{pmatrix} -D-\theta_1 & 0 \\ \theta_2 & -D \end{pmatrix}\begin{pmatrix} S \\ X \end{pmatrix} + \begin{pmatrix} D \\ 0 \end{pmatrix} S_{in} \tag{2.220}$$

The objective of the above discussion is motivated by the following arguments:

- on one hand, as suggested above, the models used in biological wastewater treatment are, generally speaking, nonlinear;
- on the other hand, the tools of system analysis are, generally speaking, much simpler, easier to apply and/or of larger use for linear systems than for nonlinear ones.

That's why we shall prefer the use of linear system methods whenever possible, either because the model is linear in the context of the study or because by a proper manipulation we have been able to transform the nonlinear model into a linear one. We have already considered a transformation for the parameters. In other instances, we might have to use another transformation, the most largely used one consists of linearising the nonlinear model around some equilibrium point (or steady state).

Finally, it is important to point out that all the analysis results which can be obtained from a "linearised" model are usually not global (i.e. not valid for all the values of the variables and parameters) but local: for instance, analysis results based on a model linearised around some equilibrium points will only be valid close to these points. Note that even the above proposed parameter transformation exhibits a singular point: the combination ($\theta_1 = 0$, $\theta_2 \neq 0$) gives $Y = \infty$! And therefore strictly speaking the transformation is one-to-one except at that singular point.

2.7 Equilibrium Points, Linearisation and Stability Analysis

An equilibrium point (or steady state) of a dynamical model is, by definition, a constant state, i.e. its time derivative is equal to zero. If we consider the general dynamical model (2.100), the equilibrium points (that we shall denote $\bar{\xi}$) are such that:

$$\frac{d\bar{\xi}}{dt} = 0 \tag{2.221}$$

This implies that equilibrium points are solutions of the algebraic equation:

$$-\bar{D}\bar{\xi} + Y\rho(\bar{\xi}) + \bar{F} - Q(\bar{\xi}) = 0 \tag{2.222}$$

for given constant values of $\bar{D}$ and $\bar{F}$ of the dilution rate D and of the feed rates F. The problem of calculating the equilibrium points $\bar{\xi}$ of a biological wastewater treatment process whose dynamics are described by equations (2.100) is that of solving equation (2.222). The latter has no general analytical solution and can only be solved in specific applications.

Furthermore an important feature of (bio)chemical processes is that they can exhibit multiple steady states, i.e. several solutions to (2.222) exist. Let us illustrate this via the simple microbial growth model with different kinetic models (first order, Monod, and Haldane).

2.7.1 *First Order Kinetics*

Let us consider the mass balance model (2.215) (2.216) of the preceding section:

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\mu_{max}X \tag{2.223}$$

$$\frac{dX}{dt} = -DX + \mu_{max}X \tag{2.224}$$

which, as we already said, is linear. As a consequence, there will be no multiple equilibrium points, i.e. to one set of values of $(\bar{D}, \bar{S}_{in})$ corresponds one and only one set of values of $\bar{S}$ and $\bar{X}$. Indeed the equilibrium points are here the solution of the equations:

$$0 = \bar{D}\bar{S}_{in} - \bar{D}\bar{S} - \frac{1}{Y}\mu_{max}\bar{X} \tag{2.225}$$

$$0 = -\bar{D}\bar{X} + \mu_{max}\bar{X} \tag{2.226}$$

The equilibrium point is then given by the following expressions:

$$\bar{S} = \frac{\bar{D}}{\bar{D} + \frac{1}{Y}\mu_{max}}\bar{S}_{in} \tag{2.227}$$

$$\bar{X} = \frac{k_0}{\bar{D} + \frac{1}{Y}\mu_{max}}\bar{S}_{in} \tag{2.228}$$

2.7.2 *Monod and Haldane Kinetics*

Let us recall the dynamical mass balance equations of the simple microbial growth process (2.13) (2.14):

$$\frac{dS}{dt} = DS_{in} - DS - \frac{1}{Y}\mu X \tag{2.229}$$

$$\frac{dX}{dt} = -DX + \mu X \tag{2.230}$$

The equilibrium points are therefore the solutions of the following set of algebraic equations:

$$0 = \bar{D}\bar{S}_{in} - \bar{D}\bar{S} - \frac{1}{Y}\bar{\mu}\bar{X} \tag{2.231}$$

$$0 = -\bar{D}\bar{X} + \bar{\mu}\bar{X} \tag{2.232}$$

From the second equation, we note that there are two possible solutions:

1. $\bar{X} = 0$
2. $\bar{\mu} = \bar{D}$

The first equilibrium point corresponds to the *wash-out* steady state:

$$\bar{X} = 0, \ \bar{S} = \bar{S}_{in} \tag{2.233}$$

It corresponds to a wash-out of the biomass from the reactor. It can occur for any value of $\bar{D}$ and $\bar{S}_{in}$. It is also the only possible equilibrium point when $\bar{D} > \mu_{max}$. It is obvious that wash-out is undesirable and should be avoided as far as possible in practice.

The second equilibrium point corresponds to operational steady states. Expressions of the specific growth rate are necessary to obtain explicit expressions of the equilibrium points. Let us consider two typical specific growth rate expressions.

Monod kinetics. The operational steady state equation $\bar{\mu} = \bar{D}$ with Monod kinetics is written as follows:

$$\frac{\mu_{max}\bar{S}}{K_S + \bar{S}} = \bar{D} \tag{2.234}$$

We can deduce the following equilibrium point from the above equation and the steady state equation (2.231) of $\bar{S}$:

$$\bar{S} = \frac{\bar{D}K_S}{\mu_{max} - \bar{D}} \tag{2.235}$$

$$\bar{X} = Y(S_{in} - S) = Y(S_{in} - \frac{\bar{D}K_S}{\mu_{max} - \bar{D}}) \tag{2.236}$$

Note in particular that the steady state of the substrate concentration $\bar{S}$ is independent of the influent substrate concentration $\bar{S}_{in}$. This property is indeed generic for any model of the specific growth rate which *only* depends on the substrate concentration. (Exercise: calculate the steady states of the simple microbial growth process with Tessier and with Contois specific growth rate models, and compare the results.)

Haldane kinetics. For the Monod kinetics we have obtained two possible equilibrium points: one wash-out steady state, and one operational steady state. Let us now consider the same microbial growth process but with Haldane kinetics (2.43). There is still obviously a wash-out equilibrium point. As suggested in Figure 2.5, there are now two operational equilibrium points (instead of one with the Monod model). Indeed the equation $\bar{\mu} = \bar{D}$ with the Haldane model is written as follows:

$$\frac{\mu_{max}\bar{S}}{K_S + \bar{S} + \frac{\bar{S}^2}{K_i}} = \bar{X} \tag{2.237}$$

Therefore the steady state value $\bar{S}$ is the solution of the following equation:

$$\frac{\bar{D}}{K_i}\bar{S}^2 + (\bar{D} - \mu_{max})\bar{S} + \bar{K}_S = 0 \tag{2.238}$$

and the two solutions S_1 and S_2 in Figure 2.5 are equal in steady state to:

$$\bar{S}_1 = \frac{(\mu_{max} - \bar{D})K_i}{2\bar{D}} - \frac{K_i}{2\bar{D}}\sqrt{(\mu_{max} - \bar{D})^2 - 4\bar{D}^2\frac{K_S}{K_i}} \tag{2.239}$$

$$\bar{S}_2 = \frac{(\mu_{max} - \bar{D})K_i}{2\bar{D}} + \frac{K_i}{2\bar{D}}\sqrt{(\mu_{max} - \bar{D})^2 - 4\bar{D}^2\frac{K_S}{K_i}} \tag{2.240}$$

The above two operational equilibrium points have fundamentally different dynamical properties in terms of stability. This is what we shall briefly analyse in Section 2.7.4. However since the analysis is based on a linearised model of the process, we shall first introduce the linearised tangent model of nonlinear models.

2.7.3 *Linearised Tangent Model of Nonlinear Models*

Let us consider a nonlinear dynamical model:

$$\frac{dx}{dt} = f(x, u) \tag{2.241}$$

where x is the state vector and u the input vector. The equilibrium point(s) are the values of $\bar{x}$, the solution of:

$$0 = f(\bar{x}, \bar{u}) \tag{2.242}$$

for constant values of $\bar{u}$. The derivation of the linearised tangent model around the steady state $\bar{x}$ is based on the Taylor series' expansion of the function f around this steady state for x and u.

Let us define the deviation variables $\tilde{x}$ and $\tilde{u}$:

$$\tilde{x} = x - \bar{x}, \ \tilde{u} = u - \bar{u} \tag{2.243}$$

The Taylor series expansion of f around $x = \bar{x}$ and $u = \bar{u}$ is equal to:

$$f(x, u) = f(\bar{x}, \bar{u}) + \frac{\partial f(\bar{x}, \bar{u})}{\partial x}\tilde{x} + \frac{\partial f(\bar{x}, \bar{u})}{\partial u}\tilde{u} + \frac{1}{2!}\frac{\partial^2 f(\bar{x}, \bar{u})}{\partial x^2}\tilde{x}^2 + \frac{1}{2!}\frac{\partial^2 f(\bar{x}, \bar{u})}{\partial u^2}\tilde{u}^2 + ... \tag{2.244}$$

If we stop the expansion at the first order derivative and we consider the deviation state variable $\tilde{x}$, the nonlinear dynamical model (2.241) can be rewritten under the following linear form:

$$\frac{d\tilde{x}}{dt} = A\tilde{x} + B\tilde{u} \tag{2.245}$$

with:

$$A = \frac{\partial f(\bar{x}, \bar{u})}{\partial x}, \quad B = \frac{\partial f(\bar{x}, \bar{u})}{\partial u} \tag{2.246}$$

The above model (2.245)(2.246) is the linearised tangent model of the nonlinear model (2.241). (It is indeed tangent of the nonlinear model at the equilibrium point $\bar{x}$).

As a matter of illustration, let us apply the above linearisation procedure to the simple microbial growth process equations (2.16) (2.17). Assume that the specific growth rate is, generally speaking, a function of S and X ($\mu(S, X)$), and D and S_{in} are both inputs. Then the linearised tangent model around operational equilibrium points $S = \bar{S}$ and $X = \bar{X}$ ($\bar{\mu} = \bar{D}$!) is equal to:

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix} \tilde{s} \\ \tilde{x} \end{pmatrix} &= \begin{pmatrix} -\bar{D} - \frac{1}{Y}\bar{\mu}_S\bar{X} & -\frac{1}{Y}(\bar{D} + \bar{\mu}_X\bar{X}) \\ \bar{\mu}_S\bar{X} & 0 \end{pmatrix}\begin{pmatrix} \tilde{s} \\ \tilde{x} \end{pmatrix} \\ &+ \begin{pmatrix} \bar{D} & \bar{S}_{in} - \bar{S} \\ 0 & -\bar{X} \end{pmatrix}\begin{pmatrix} \tilde{s}_{in} \\ \tilde{d} \end{pmatrix} \end{aligned} \tag{2.247}$$

with:

$$\tilde{s} = S - \bar{S}, \; \tilde{x} = X - \bar{X}, \; \tilde{s}_{in} = S_{in} - \bar{S}_{in}, \; \tilde{d} = D - \bar{D} \tag{2.248}$$

$$\bar{\mu}_S = \frac{\partial \mu(\bar{S}, \bar{X})}{\partial S}, \; \bar{\mu}_X = \frac{\partial \mu(\bar{S}, \bar{X})}{\partial X} \tag{2.249}$$

For the Haldane model for instance:

$$\bar{\mu}_X = 0, \; \bar{\mu}_S = \frac{\mu_{max}(K_S - \frac{\bar{S}^2}{K_i})}{(K_S + \bar{S} + \frac{\bar{S}^2}{K_i})^2} \tag{2.250}$$

and the state matrix A is then equal to:

$$A = \begin{pmatrix} -\bar{D} - \frac{1}{Y}\bar{\mu}_S\bar{X} & -\frac{1}{Y}\bar{D} \\ \bar{\mu}_S\bar{X} & 0 \end{pmatrix} \tag{2.251}$$

2.7.4 *Stability of Equilibrium Points*

Stability theory of dynamical systems is a very old subject. Yet, significant advances in the stability analysis are due to the Russian mathematician Lyapunov (1892), whose results were largely confined to Eastern European countries until about 1960. Lyapunov's first method [284] utilises the eigenvalues of the state matrix A of the linear model (2.245) to check the stability of the equilibrium state. Recall that the eigenvalues λ of a square matrix A are the roots of the characteristic polynomial:

$$det(\lambda I - A) = 0, \qquad I : \text{identity matrix} \tag{2.252}$$

If the real parts of all the eigenvalues are negative, the equilibrium point is stable. If any of the real parts of the eigenvalues are positive, the equilibrium is unstable. No conclusion can be drawn in case of eigenvalues having zero real parts.

Let us apply the analysis to the simple microbial growth model with Haldane kinetics. Let us calculate the characteristic polynomial of the linearised tangent model state matrix (2.251):

$$det(\lambda I - A) = \lambda^2 + (\bar{D} + \frac{1}{Y}\mu_S \bar{X})\lambda + \frac{1}{Y}\mu_S \bar{X}\bar{D} \qquad (2.253)$$

A necessary and sufficient condition for a second order polynomial to have roots with all negative real parts is that all the coefficients of the polynomial have the same sign. This means that here in order to have stability, $\bar{D} + \frac{1}{Y}\bar{\mu}_S \bar{X}$ and $\frac{1}{Y}\bar{\mu}_S \bar{X}\bar{D}$ must be positive. If we recall the value of $\bar{\mu}_S$ (2.250), this condition will be fulfilled if and only if $\bar{S} < \sqrt{K_S K_i}$, i.e. for S_1 in Figure 2.5. Conversely, the equilibrium point $\bar{S}_2$ will be unstable since $\bar{S} > \sqrt{K_S K_i}$, and (at least) one coefficient in the polynomial will be negative.

This is a quite interesting feature of the Haldane model that it is able to emphasise the possible presence of unstable steady states. These states might be of practical interest in industrial applications but are not reachable without any appropriate external control. It is clearly a major issue in automatic control to regulate and stabilise processes around equilibrium points which are open-loop (i.e. without external action) unstable (see e.g. [14]).

2.8 A Key State Transformation

The key result of this section is the introduction of a state transformation by which part of the dynamical model (2.100) becomes independent of the reaction kinetics ρ (see [14], [56]). This transformation will play a very important role in the design of asymptotic observers (Chapter 7). The proposed transformation readily derives from the notion of invariants in reaction systems (see e.g. [95], [86]).

2.8.1 *Definition of the State Transformation*

The transformation is defined as follows. Let us denote rank(Y) = R and consider a state partition (in which the state vector is decomposed in two sub-vectors ξ_a and ξ_b):

$$\xi = \begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix} \qquad (2.254)$$

where ξ_a contains R (arbitrarily chosen) process variables and ξ_b the others, but such that the corresponding submatrix Y_a is full rank (rank(Y_a) = R). Let us define the state transformation into the auxiliary variable ζ (dim(ζ) = N-R):

$$\zeta = C_a \xi_a + C_b \xi_b \qquad (2.255)$$

where C_a and C_b are solutions of the matrix equation:

$$C_a Y_a + C_b Y_b = 0 \qquad (2.256)$$

In the particular (but quite general) situation of M independent irreversible reactions, then R = M and C_b may be chosen to be a full rank square matrix (the simplest choice being obviously the identity matrix I). Then C_a is found from

$$C_a = -C_b Y_b {Y_a}^{-1} \tag{2.257}$$

Let us illustrate the dynamics of ζ from the General Dynamical Model (2.100) or (2.132)(2.133) and the definition (2.255). Note that the vectors F, Q, D are partitioned according to the partitioning of the state vector ξ.

1) Single reactor: D = scalar

$$\frac{d\zeta}{dt} = -D\zeta + C_a(F_a - Q_a) + C_b(F_b - Q_b) \tag{2.258}$$

2) Multi-reactor: D = matrix

$$\begin{aligned}\frac{d\zeta}{dt} = &-(C_b D_{bb} + C_a D_{ab})C_b^{-1}\zeta + C_a(F_a - Q_a) + C_b(F_b - Q_b) \\ &+[(C_b D_{bb} + C_a D_{ab})C_b^{-1} - C_b D_{ba} - C_a D_{aa}]\xi_a\end{aligned} \tag{2.259}$$

with:

$$D = \begin{pmatrix} D_{aa} & D_{ab} \\ D_{ba} & D_{bb} \end{pmatrix} \tag{2.260}$$

3) Fixed bed reactor

For simplicity reasons, let us consider here $C_b = I$ and let us put the vector ξ_{fi} of the fixed components in ξ_b:

$$\xi_b = \begin{pmatrix} \xi_{bf} \\ \xi_{fi} \end{pmatrix} \tag{2.261}$$

Then we can rewrite the auxiliary variable ζ as follows:

$$\zeta = \begin{pmatrix} \zeta_{fl} \\ \zeta_{fi} \end{pmatrix} = \begin{pmatrix} \xi_{bf} \\ \xi_{fi} \end{pmatrix} + \begin{pmatrix} C_{af} \\ C_{ae} \end{pmatrix} \xi_a \tag{2.262}$$

The dynamics of ζ can then be written as follows:

$$\frac{\partial \zeta_{fl}}{\partial t} = -\frac{F_{in}}{A}\frac{\partial \zeta_{fl}}{\partial z} + D_{am}\frac{\partial^2 \zeta_{fl}}{\partial z^2} \tag{2.263}$$

$$\frac{\partial \zeta_{fi}}{\partial t} = -\frac{F_{in}}{A}C_{ae}\frac{\partial \xi_a}{\partial z} + D_{am}C_{ae}\frac{\partial^2 \xi_a}{\partial z^2} \tag{2.264}$$

Note that the dynamical equations of ζ (2.258), (2.259) and (2.263)(2.264) are independent of the reaction kinetics $\rho(\xi)$.

2.8.2 *Example 1: Two Step Denitrification*

Let us consider as a first example the denitrification process described by equations (2.99). One possible choice for the state partition is the following:

$$\xi_a = \begin{pmatrix} X \\ N_2 \end{pmatrix}, \ \xi_b = \begin{pmatrix} S_{NO_3} \\ S_{NO_2} \\ S \end{pmatrix} \tag{2.265}$$

Indeed, the submatrices Y_a and Y_b are then equal to:

$$Y_a = \begin{pmatrix} 1 & 1 \\ 0 & Y_6 \end{pmatrix}, Y_b = \begin{pmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ -\frac{1}{Y_4} & -\frac{1}{Y_5} \end{pmatrix} \tag{2.266}$$

and Y_a is full rank (since Y_6 is by definition strictly positive). If we consider the simplest possible choice for C_b (= I), then C_a is equal to:

$$C_a = -Y_b Y_a^{-1} = \begin{pmatrix} \frac{1}{Y_1} & -\frac{1}{Y_1 Y_6} \\ -Y_3 & \frac{Y_2 Y_3 + 1}{Y_2 Y_6} \\ \frac{1}{Y_4} & \frac{Y_5 - Y_4}{Y_4 Y_5 Y_6} \end{pmatrix} \tag{2.267}$$

Therefore the auxiliary variable ζ is equal to:

$$\zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} = \begin{pmatrix} S_{NO_3} + \frac{1}{Y_1} X - \frac{1}{Y_1 Y_6} N_2 \\ S_{NO_2} - Y_3 X + \frac{Y_2 Y_3 + 1}{Y_2 Y_6} N_2 \\ S + \frac{1}{Y_4} X + \frac{Y_5 - Y_4}{Y_4 Y_5 Y_6} N_2 \end{pmatrix} \tag{2.268}$$

and its dynamics are written as follows:

$$\frac{d}{dt} \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} = -D \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} + \begin{pmatrix} D S_{NO_3,in} + \frac{1}{Y_1 Y_6} Q_{N_2} \\ -\frac{Y_2 Y_3 + 1}{Y_2 Y_6} Q_{N_2} \\ D S_{in} - \frac{Y_5 - Y_4}{Y_4 Y_5 Y_6} Q_{N_2} \end{pmatrix} \tag{2.269}$$

Remark: all the above calculations are based on an arbitrary choice of the state partition ξ_a, ξ_b. Many other choices would have been appropriate. For instance we could have considered the first two entries of the state vector, i.e. S_{NO_3} and S_{NO_2} for ξ_a and the other for ξ_b, and this would have resulted in the following state transformation:

$$\zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} = \begin{pmatrix} S - \frac{Y_1(Y_5 + Y_2 Y_3 Y_4)}{Y_4 Y_5} S_{NO_3} - \frac{Y_2}{Y_5} S_{NO_2} \\ X + Y_1(1 + Y_2 Y_3) S_{NO_3} + Y_2 S_{NO_2} \\ N_2 + Y_1 Y_2 Y_3 Y_6 S_{NO_2} + Y_2 Y_6 S_{NO_3} \end{pmatrix} \tag{2.270}$$

It is important to remember that the time evolution of these auxiliary variables can be calculated without knowledge of the reaction kinetics, i.e. without any requirement concerning the knowledge about its model structure and/or its parameters.

2.8.3 *Example 2: Activated Sludge Process: The Basic Model*

As already mentioned, the basic activated sludge process model is a two-tank reactor model, with a dilution rate matrix D.

There is one reaction: ξ_a will be a scalar. Let us for instance consider X for ξ_a (note that we could also have taken S or S_O but not X_R). Then we have the following state partition:

$$\xi_a = X, \ \xi_b = \begin{pmatrix} S \\ S_O \\ X_R \end{pmatrix} \tag{2.271}$$

This means that the different entries D_{aa}, D_{ab}, D_{ba} and D_{bb} of the matrix D are equal to:

$$D_{aa} = D_1, \ D_{ab} = \begin{pmatrix} 0 \ 0 \ -D_2 \end{pmatrix} \tag{2.272}$$

$$D_{ba} = \begin{pmatrix} 0 \\ 0 \\ -D_3 \end{pmatrix}, \ D_{bb} = \begin{pmatrix} D_1 & 0 & 0 \\ 0 & D_1 & 0 \\ 0 & 0 & D_4 \end{pmatrix} \tag{2.273}$$

The matrices Y_a, Y_b and C_a are equal to:

$$Y_a = 1, \ Y_b = \begin{pmatrix} -\frac{1}{Y_S} \\ -\frac{1}{Y_O} \\ 0 \end{pmatrix}, \ C_a = -Y_b Y_a^{-1} = \begin{pmatrix} \frac{1}{Y_S} \\ \frac{1}{Y_O} \\ 0 \end{pmatrix} \ (C_B = I) \tag{2.274}$$

i.e. ζ is equal to:

$$\zeta = \begin{pmatrix} S + \frac{1}{Y_S} X \\ S_O + \frac{1}{Y_O} X \\ X_R \end{pmatrix} \tag{2.275}$$

2.8.4 *Example 3: Activated Sludge Process: The IWA Activated Sludge Model No. 1*

A careful look at the reaction network of the IWA Activated Sludge Model No. 1 reveals that there is a loop with reactions 1, 4 and 7. Indeed the readily biodegradable substrate S_S is transformed in reaction 1 in heterotrophic bacteria X_{BH}, which in turn is transformed in the decay reaction 4 in slowly biodegradable substrate X_S. And in reaction 7 X_S is hydrolysed in S_S.

This reaction loop has an important consequence on the above state transformation. Indeed the process components and reactions are not completely independent from each other. This means that not any choice of state partition will give a full rank Y_a. For instance, if one takes the first 8 components for ξ_a, the related Y_a will be singular, since row 7 will be a linear combination of rows 1 to 4 and row 6, or in other words, in terms of reaction kinetics, S_{NO} is a linear combination of S_S, X_S, X_{BH}, X_{BA} and S_O.

A possible choice for the state partition is to take the first 6 components in ξ (2.107) plus S_{NH} and S_{ND} for the subvector ξ_a:

$$\xi_a = \begin{pmatrix} S_S \\ X_S \\ X_{B,H} \\ X_{B,A} \\ X_P \\ S_O \\ S_{NH} \\ S_{ND} \end{pmatrix}, \ \xi_b = \begin{pmatrix} S_{NO} \\ X_{ND} \end{pmatrix} \tag{2.276}$$

$$Y_a = \begin{pmatrix} -\frac{1}{Y_H} & -\frac{1}{Y_H} & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1-f_P & 1-f_P & 0 & -1 & 0 \\ 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_P & f_P & 0 & 0 & 0 \\ -\frac{1-Y_H}{Y_H} & 0 & -\frac{4.57-Y_A}{Y_A} & 0 & 0 & 0 & 0 & 0 \\ -i_{XB} & -i_{XB} & -i_{XB}-\frac{1}{Y_A} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix} \tag{2.277}$$

$$Y_b = \begin{pmatrix} 0 & -\frac{1-Y_H}{2.86Y_H} & \frac{1}{Y_A} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i_{XB}-f_P i_{XP} & i_{XB}-f_P i_{XP} & 0 & 0 & -1 \end{pmatrix} \tag{2.278}$$

Then the matrix C_a is equal to:

$$C_a = \begin{pmatrix} c_1 & c_1 & c_2 & c_2 & c_3 & \frac{1}{2.86} & 0 & 0 \\ c_4 & c_4 & c_5 & c_5 & c_6 & 0 & 1 & 1 \end{pmatrix} \tag{2.279}$$

with

$$c_1 = \frac{Y_H}{2.86}\left(-\frac{1-Y_H}{Y_H} + \frac{1.71-Y_A}{Y_A}\right) \tag{2.280}$$

$$c_2 = \frac{1.71-Y_A}{2.86Y_A} \tag{2.281}$$

$$c_3 = \frac{(1-f_P)(Y_A - 1.71Y_H) + 1.71Y_A}{2.86 f_P Y_A} \tag{2.282}$$

$$c_4 = \frac{Y_H}{Y_A} \tag{2.283}$$

$$c_5 = i_{XB} + \frac{1}{Y_A} \tag{2.284}$$

$$c_6 = \frac{1 - Y_H + f_P(Y_H - Y_A i_{XP})}{f_P Y_A} \tag{2.285}$$

and the auxiliary variables ζ are written as follows:

$$\zeta = \begin{pmatrix} S_{NO} + c_1(S_S + X_S) + c_2(X_{B,H} + X_{B,A}) + c_3 X_P + \frac{1}{2.86} S_O \\ X_{ND} + c_4 S_S + c_4 X_S + c_5 X_{B,H} + c_5 X_{B,A} + c_6 X_P + S_{NH} + S_{ND} \end{pmatrix} \tag{2.286}$$

2.8.5 *Example 4: Fixed Bed Reactor Model With a Growth Reaction and a Death/Detachment Reaction*

Let us consider for instance the following state partition:

$$x_a = \begin{pmatrix} S \\ X_d \end{pmatrix}, \; x_b = \begin{pmatrix} P \\ X \end{pmatrix} \tag{2.287}$$

The matrices Y_a and Y_b are equal to:

$$Y_a = \begin{pmatrix} -\frac{1}{Y_S} & 0 \\ 0 & 1 \end{pmatrix}, \; Y_b = \begin{pmatrix} Y_P & 0 \\ \frac{\epsilon}{1-\epsilon} & -\frac{\epsilon}{1-\epsilon} \end{pmatrix} \tag{2.288}$$

Therefore the matrix C_a is equal to:

$$C_a = -Y_b Y_a^{-1} = \begin{pmatrix} Y_S Y_P & 0 \\ \frac{Y_S \epsilon}{(1-\epsilon)} & \frac{\epsilon}{1-\epsilon} \end{pmatrix} \tag{2.289}$$

ζ is then equal to:

$$\zeta = \begin{pmatrix} P + Y_P Y_S S \\ X + \frac{Y_S \epsilon}{(1-\epsilon)} S + \frac{\epsilon}{1-\epsilon} X_d \end{pmatrix} \tag{2.290}$$

2.9 Model Order Reduction

The examples of bioprocesses presented in the preceding sections have shown that a bioreactor dynamical model may be fairly complex in some instances and involve a large number of differential equations. But there are many practical applications where a simplified reduced order model is sufficient from an engineering viewpoint. One possible systematic approach to achieve model simplification is to use the singular perturbation method, which is a technique that allows to transform a set of $n + m$ differential equations into a set of n differential equations and a set of m algebraic equations. It is based on the partition of the state equations into two sets of dynamical equations characterised by the variables x_1 and x_2, of dimension n-m and m, respectively, i.e.:

$$\frac{dx_1}{dt} = f_1(x_1, x_2, u) \tag{2.291}$$

$$\delta \frac{dx_2}{dt} = f_2(x_1, x_2, u, \delta) \tag{2.292}$$

i.e. the time derivative of some of the state variables are multiplied by a small parameter δ (also sometimes called the perturbation parameter). For sufficiently small parameter δ, the parameter δ can considered as being negligible, and the singular perturbation consists of setting δ to zero, and to replace the $n + m$ differential equations (2.292) by algebraic equations:

$$f_2(x_1, x_2, u, \delta) = 0 \tag{2.293}$$

More precisely, if $\bar{x}_2$ is the solution of the above algebraic equation, then the model (2.291)(2.292) is replaced by the following set of equations:

$$\frac{dx_1}{dt} = f_1(x_1, \bar{x}_2, u) \tag{2.294}$$

$$\bar{x}_2 = g(x_1, u) \tag{2.295}$$

The model (2.294) is sometimes called a *quasi-steady state* model, which is often considered for instance in (bio)chemical engineering, e.g. [37], [114], [255].

This technique is suitable when neglecting the dynamics of products with low solubility in the liquid phase or of substrates in fast reactions. The method will be illustrated with two specific examples (low solubility product, and substrates in a fast reaction) before stating the general rule for order reduction.

2.9.1 *Singular Perturbation Technique for Low Solubility Products*

Let us consider a reaction described by the following reaction scheme:

$$S \longrightarrow P \tag{2.296}$$

where P is a volatile product which can be given off in gaseous form and has low solubility in the liquid phase (e.g. H_2 in anaerobic digestion). The dynamical model is as follows:

$$\frac{dS}{dt} = -\rho - DS + DS_{in} \tag{2.297}$$

$$\frac{dP}{dt} = Y\rho - DP - Q \tag{2.298}$$

For the consistency of this model, the product concentration P can be expressed with respect to the saturation concentration representative of the product solubility as follows:

$$P = \Pi P_{sat}, \qquad 0 \leq \Pi(t) \tag{2.299}$$

where P_{sat} is the saturation concentration which is constant in a stable physico-chemical environment. The model (2.297)(2.298) is rewritten in the standard singular perturbation form, with $\delta = P_{sat}$:

$$\frac{dS}{dt} = -\rho - DS + DS_{in} \tag{2.300}$$

$$\delta\frac{d\Pi}{dt} = Y\rho - \delta D\Pi - Q \tag{2.301}$$

If the solubility is very low, we obtain a reduced order model by setting $\delta = 0$ and replacing the differential equation (2.301) by the algebraic one:

$$Q = Y\rho \tag{2.302}$$

2.9.2 *Singular Perturbation Technique for Substrates of Fast Reactions*

Singular perturbation can also be applied to reduce the order of the model dynamics for reactors with fast and slow reactions. For simplicity we shall concentrate on an example, i.e. the nitrification process which is characterised by the following two sequential reactions (2.85) (2.86) repeated below as (2.303)(2.304):

$$S_{NH} \longrightarrow X_1 + S_{NO_2} \tag{2.303}$$

$$S_{NO_2} \longrightarrow X_2 + S_{NO_3} \tag{2.304}$$

The dynamics in a CSTR are given by the following equations:

$$\frac{dS_{NH}}{dt} = -\frac{1}{Y_1}\rho_1 + DS_{NH,in} - DS_{NH} \tag{2.305}$$

$$\frac{dS_{NO_2}}{dt} = -\frac{1}{Y_2}\rho_2 + Y_3\rho_1 - DS_{NO_2} \tag{2.306}$$

$$\frac{dS_{NO_3}}{dt} = Y_4\rho_2 - DS_{NO_3} \tag{2.307}$$

$$\frac{dX_1}{dt} = \rho_1 - DX_1 \tag{2.308}$$

$$\frac{dX_2}{dt} = \rho_2 - DX_2 \tag{2.309}$$

In line with the basic kinetics rules (Section 2.2.6), the reaction rates ρ_1 and ρ_2 can be written as follows:

$$\rho_1 = k_1 S_{NH}^{\alpha_1}\phi_1(S_{NH}, X_1, S_{NO_2}) \tag{2.310}$$

$$\rho_2 = k_2 S_{NO_2}^{\alpha_2}\phi_2(S_{NO_2}, X_2, S_{NO_3}) \tag{2.311}$$

where k_i (i = 1, 2) are the maximum reaction rates, α_i (i = 1, 2) are coefficients (which might be the reaction orders), and ϕ_i (i = 1, 2) positive functions of the state ($\phi_i > 0$). Assume now that the first reaction is slow and the second one is fast. This can be formalised by considering that the maximum reaction rate of the second reaction k_2 is much larger than the one of the first reaction k_1:

$$k_2 \gg k_1 \tag{2.312}$$

Let us now consider the following state transformation (the general formulation is given in [255]):

$$\zeta_1 = Y_4 Y_2 S_{NO_2} + S_{NO_3} \tag{2.313}$$

$$\zeta_2 = Y_2 S_{NO_2} + X_2 \tag{2.314}$$

Then we can rewrite the process dynamics as follows:

$$\frac{dS_{NH}}{dt} = -\frac{1}{Y_1}\rho_1 + DS_{NH,in} - DS_{NH} \tag{2.315}$$

$$\frac{dS_{NO_2}}{dt} = -\frac{1}{Y_2}\rho_2 + Y_3\rho_1 - DS_{NO_2} \tag{2.316}$$

$$\frac{d\zeta_1}{dt} = Y_2 Y_3 Y_4 \rho_1 - D\zeta_1 \tag{2.317}$$

$$\frac{dX_1}{dt} = \rho_1 - DX_1 \tag{2.318}$$

$$\frac{d\zeta_2}{dt} = Y_2 Y_3 \rho_1 - D\zeta_2 \tag{2.319}$$

The choice of the above transformation can be briefly intuitively motivated as follows. First, the second equation (2.316) is the only one that still contains the kinetics of both reactions, while the other four equations only contain the reaction rate of the slow reaction. Secondly the auxiliary variables ζ_1 and ζ_2 are the algebraic sum of S_{NO_2} (i.e. the “intermediate” component: product of the first reaction, and substrate of the second), and of the products of the second (fast) reaction, S_{NO_3} and X_2 respectively.

Let us define the singular perturbation δ as the inverse of the maximum reaction rate k_2:

$$\delta = \frac{1}{k_2} \tag{2.320}$$

By considering the expression (2.311) and the singular perturbation parameter δ, the dynamical equation of S_{NO_2} can be rewritten as follows:

$$\delta\frac{dS_{NO_2}}{dt} = -\frac{1}{Y_2}S_{NO_2}^{\alpha_2}\phi_2(S_{NO_2}, X_2, S_{NO_3}) + \delta Y_3\rho_1 - \delta DS_{NO_2} \tag{2.321}$$

If the second reaction is sufficiently fast, then we can apply the singular perturbation and set δ to zero. This implies that the above equation reduces to:

$$\frac{1}{Y_2}S_{NO_2}^{\alpha_2}\phi_2(S_{NO_2}, X_2, S_{NO_3}) = 0 \tag{2.322}$$

Since Y_2 and ϕ_2 are strictly positive, this implies: $S_{NO_2} = 0$. Then by recalling the definition of ζ, the equations (2.315)(2.317)(2.318)(2.319) of the dynamical model can be rewritten as follows:

$$\frac{dS_{NH}}{dt} = -\frac{1}{Y_1}\rho_1 + DS_{NH,in} - DS_{NH} \tag{2.323}$$

$$\frac{dS_{NO_3}}{dt} = Y_2Y_3Y_4\rho_1 - DS_{NO_3} \tag{2.324}$$

$$\frac{dX_1}{dt} = \rho_1 - DX_1 \tag{2.325}$$

$$\frac{dX_2}{dt} = Y_2Y_3\rho_1 - DX_2 \tag{2.326}$$

i.e. it is as if the two sequential reactions (2.303)(2.304) have been reduced to one reaction:

$$S_{NH} \longrightarrow X_1 + X_2 + S_{NO_3} \tag{2.327}$$

2.9.3 *A General Rule for Order Reduction*

The above examples show that the rule for model simplification is actually very simple and that an explicit singular perturbation analysis is not really needed. Consider that, for some i, the dynamics of the component ξ_i are to be neglected. The dynamics of ξ_i are described by equation (2.100):

$$\frac{d\xi_i}{dt} = -D\xi_i + Y_i\rho + F_i - Q_i \tag{2.328}$$

where Y_i is the row of Y corresponding to the component ξ_i. The simplification is then achieved by setting ξ_i and $d\xi_i/dt$ to zero i.e. by replacing the differential equation (2.328) by the following algebraic equation:

$$Y_i\rho = -F_i + Q_i \tag{2.329}$$

It has been shown that the above model order reduction rule is not only valid for low solubility products but also for bioprocesses with fast and slow reactions. Then the above order reduction rule (2.329) applies to substrates of fast reactions (as long as they intervene only in fast reactions) (see [255] for further details).

Note the close connection between the singular perturbation reduction and the quasi steady state (QSS) approximation, which is largely used in (bio)chemical engineering. This suggests the following comment: singular perturbation can be viewed as an efficient mathematical tool to rigourously justify QSS approximations on a systematic basis via an appropriate analysis (including the choice of an appropriate *small* perturbation parameter).

The above considerations also apply to models with other hydrodynamics, like the fixed bed reactor models with or without dispersion, mutatis mutandis, i.e. by setting the different derivative ($\frac{\partial x_i}{\partial t}$, $\frac{\partial x_i}{\partial z}$, $\frac{\partial^2 x_i}{\partial z^2}$) of the low solubility product or of the “fast substrate” to zero in its mass balance equation.

2.9.4 *Example: the Anaerobic Digestion*

Let us see how to apply the above model order reduction rule (2.329) to a specific example, the anaerobic digestion. First of all, it is well-known that methane is a low solubility product. Therefore the above procedure applies. Furthermore, assume that the second methanisation path (hydrogen consumption) is limiting, i.e. that the first three reactions (2.53)(2.54)(2.55) are fast and the fourth one (2.56) is slow. We can then apply the model order reduction rule (2.329) to the the glucose concentration S_1, the propionate concentration S_2, the acetate concentration S_3 and the dissolved methane concentration P_1. By setting their values and their time derivatives to zero:

$$S_1 = S_2 = S_3 = P_1, \; \frac{dS_1}{dt} = \frac{dS_2}{dt} = \frac{dS_3}{dt} = \frac{dP_1}{dt} = 0 \tag{2.330}$$

we reduce their differential equations to the following set of algebraic equations:

$$\begin{pmatrix} -\frac{1}{Y_{21}} & 0 & 0 & 0 \\ Y_{41} & -\frac{1}{Y_{42}} & 0 & 0 \\ Y_{61} & Y_{62} & -\frac{1}{Y_{63}} & 0 \\ 0 & 0 & Y_{03} & Y_{04} \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{pmatrix} = \begin{pmatrix} -DS_{in} \\ 0 \\ 0 \\ Q_3 \end{pmatrix} \tag{2.331}$$

By inverting the submatrix of the yield coefficients of the left hand side of (2.331), we can express the reaction rates ρ_1, ρ_2, ρ_3 and ρ_4 as functions of the feedrate DS_{in} and of the gaseous methane outflow rate Q_3:

$$\rho_1 = Y_{21} DS_{in} \tag{2.332}$$

$$\rho_2 = Y_{21} Y_{41} Y_{42} DS_{in} \tag{2.333}$$

$$\rho_3 = (Y_{41} Y_{42} Y_{62} + Y_{61}) Y_{21} Y_{63} DS_{in} \tag{2.334}$$

$$\rho_4 = \frac{1}{Y_{04}} Q_3 - \frac{Y_{03}}{Y_{04}} (Y_{41} Y_{42} Y_{62} + Y_{61}) Y_{21} Y_{63} DS_{in} \tag{2.335}$$

Let us replace the reaction rates ρ_1, ρ_2 and ρ_4 by their above expressions (2.332), (2.333), (2.335) in the dynamical equation of the hydrogen concentration S_4, which is then rewritten as follows:

$$\frac{dS_4}{dt} = -DS_4 - Q_1 - Y_1 Q_3 + Y_2 DS_{in} \tag{2.336}$$

where Y_1 and Y_2 are defined as follows:

$$Y_1 = \frac{1}{Y_{84} Y_{04}} \tag{2.337}$$

$$Y_2 = Y_{81} Y_{21} + Y_{82} Y_{21} Y_{41} Y_{42} + \frac{Y_{03}}{Y_{04} Y_{84}} (Y_{41} Y_{42} Y_{62} + Y_{61}) Y_{21} Y_{63} \tag{2.338}$$

Note that the coefficients Y_1 and Y_2 are nonlinear combinations of the yield coefficients Y_{ij}.

2.9.5 *Specific Approach for Model Reduction of PDE Models*

The model reduction of PDEs is also a key question in dynamical modelling, estimation and control. In Section 2.5, we have already suggested that the fixed bed reactor is often approximated by a sequence of STRs. Here we shall present two approaches (singular perturbation, and Laplace transform) which also allow the connection to be made between the PDE model of fixed bed reactors to STR models. The two approaches are used to reduce the *infinite-dimensional* Distributed Parameter Systems (DPS) to *finite-dimensional* lumped parameter models.

Application of the Singular Perturbations to the Fixed Bed Reactor Model. Here the perturbation method is used in a somewhat different perspective as before (see also [289], [72], [32]). A small parameter δ is also introduced, but here the application of the perturbation method is based on an asymptotic expansion of the state variables in powers of δ, see [148]:

$$x(z,t) = \sum_{i=0}^{\infty} \delta^i x_i(z,t) \tag{2.339}$$

where $x_i(z,t)$ represent the approximation of the variable $x(z,t)$ at order i (as we shall see here below, the order 0 in the approximation of the axial dispersion model, for instance, corresponds to the STR model). The approximation methods are often referred to (but the terminology is not uniform in the literature!) as *perturbation methods* when the series expansion is convergent (or is expected to converge), and as *asymptotic methods* when the series is divergent but asymptotic (so that the first few terms yield a good approximation for very low values of δ). It is a generally admitted viewpoint that proving that the series is either convergent or asymptotic is not essential. Also, it is a general property of asymptotic series that finding additional terms not necessarily improves the approximation since the series is generally divergent. As the perturbation parameter δ decreases, the approximation provided by the series gets better. Thus a key assumption in the following is that δ is very small (see e.g. [290], [25]).

As explained e.g. in [270], the axial dispersion model (presented in a general form (2.132) (2.133)) is an intermediate model between two “extreme” models: the STR model and the plug flow reactor model. For highly diffusive reactors, the behaviour of the axial dispersion reactor tends to the one of a STR, while for low dispersion coefficients, the axial dispersion model tends to a plug flow model.

In the following, we shall consider fixed bed reactors in which the mass dispersion may be assumed as being very important: D_{ma} is very large (with respect to uL). Let us define the parameter δ as the (dimensionless) mass Peclet number Pe_m:

$$\delta = Pe_m = \frac{uL}{D_{ma}} \tag{2.340}$$

Before going any further, note that the singular perturbation will only apply to the variables x_{fl}: therefore in order to keep the notations as simple as possible, the index fl will be dropped in the following. Moreover, we shall compact the writing of the conversion term by defining $Y = \frac{1-\epsilon}{\epsilon} Y_{fl}$.

Then the equations (2.133)(2.135)(2.136) become:

$$\delta \frac{\partial x}{\partial t} = uL \frac{\partial^2 x}{\partial z^2} + \delta(-u \frac{\partial x}{\partial z} + Y\rho) \tag{2.341}$$

$$\frac{\partial x}{\partial z} = -\frac{\delta}{L}(x_{in} - x) \qquad \text{for } z = 0 \tag{2.342}$$

$$\frac{\partial x}{\partial z} = 0 \qquad \text{for } z = L \tag{2.343}$$

By using singular perturbation techniques, away from the boundary layer, the solution of (2.341) is given by:

$$x(z,t) = \sum_{n=0}^{\infty} \delta^n x_n(z,t) \tag{2.344}$$

$$\rho(x) = \sum_{n=0}^{\infty} \delta^n \rho_n(x_0, x_1, ..., x_n) \tag{2.345}$$

with:

$$\rho_0(x_0) = \rho(x_0), \qquad \rho_1(x_0, x_1) = (\frac{\partial \rho}{\partial x})_{(x_0)} x_1 \tag{2.346}$$

The differential equations for the different terms x_0, x_1 of the expansion (2.344) (2.345) are thus obtained by substituting these expansions in the equations (2.341) (2.342) (2.343) and equating the terms of the same power in δ, so we have:

- **zero order approximation**:

$$\frac{\partial^2 x_0}{\partial z^2} = 0 \qquad \text{for } 0 \leq z \leq L \tag{2.347}$$

$$\frac{\partial x_0}{\partial z} = 0 \qquad \text{for } z = 0 \text{ and } z = L \tag{2.348}$$

- **first order approximation**:

$$uL \frac{\partial^2 x_1}{\partial z^2} = u \frac{\partial x_0}{\partial z} + \frac{\partial x_0}{\partial t} - Y\rho_0 \qquad \text{for } 0 \leq z \leq L \tag{2.349}$$

$$\frac{\partial x_1^o}{\partial z} = \frac{1}{L}(x_0 - x_{in}) \qquad \text{for } z = 0 \tag{2.350}$$

$$\frac{\partial x_1}{\partial z} = 0 \qquad \text{for } z = L \tag{2.351}$$

- **approximation of order n $\geq$ 2**:

$$uL\frac{\partial^2 x_n}{\partial z^2} = u\frac{\partial x_{n-1}}{\partial z} + \frac{\partial x_{n-1}}{\partial t} - Y\rho_{n-1} \quad \text{for } 0 \leq z \leq L \quad (2.352)$$

$$\frac{\partial x_n}{\partial z} = \frac{1}{L}x_{n-1} \quad \text{for } z = 0 \quad (2.353)$$

$$\frac{\partial x_n}{\partial z} = 0 \quad \text{for } z = L \quad (2.354)$$

By integrating the equation (2.347)-(2.348) with respect to the spatial coordinate z, we obtain:

$$x_0(z,t) = x_0(t) \quad (2.355)$$

$$\frac{dx_0}{dt} = \frac{u}{L}(x_{in} - x_0) + Y\rho(x_0) \quad (2.356)$$

The equations (2.355)-(2.356) show that the first term in the approximation, say x_0, is independent on the spatial coordinate z, and consequently, for the zero-order approximation, the process behaves like a stirred tank reactor. So singular perturbation when applied to PDEs also allows the reduction of the PDE fixed bed reactor model to a STR model and emphasises the connection between both types of models on a systematic basis.

2.10 Connection Between Plug Flow Reactors and CSTRs: A Laplace Transform Approach

Laplace transform can be a very interesting tool for transforming a linear distributed parameter model into a lumped parameter one. One of its advantages is that it is an exact lumping method and does not involve any approximation. However, it may look difficult to generalise the approach, especially for nonlinear systems, since the Laplace transform is basically applicable to linear systems only. Let us just illustrate the use of Laplace transform in a simple example: the dynamics of the substrate's concentration in a plug flow reactor where a first order reaction takes place. As we shall see, the Laplace transform will be helpful in systematising the physically admitted idea that if the input is the influent substrate concentration and the output the effluent substrate, then the process basically behaves like the combination of a stirred tank and a time delay (due to the transport of the substrate from the reactor input to the reactor output) (see also [105]).

The dynamics of the reactant concentration C is described by the following equation:

$$\frac{\partial C}{\partial t} = -u\frac{\partial C}{\partial z} - k_0 C \quad (2.357)$$

(where k_0 is the kinetic constant), with the following boundary condition:

$$C(z = 0, t) = C_{in}(t) \tag{2.358}$$

Let us consider the Laplace transform of C(z,t) with respect to time t:

$$\mathcal{L}[C(z, t)] = \mathcal{C}(z, s) \tag{2.359}$$

Let us apply the Laplace transform to the above equation in which we neglect the initial conditions $C(z, t = 0) = 0$:

$$s\mathcal{C}(z, s) = -u\frac{\partial \mathcal{C}}{\partial z} - k_0\mathcal{C} \tag{2.360}$$

which can be rewritten as follows:

$$\frac{\partial \mathcal{C}}{\partial z} = -\frac{s + k_0}{u}\mathcal{C} \tag{2.361}$$

The solution of the above differential equation (in the independent variable z) is equal to:

$$\mathcal{C}(z, s) = \mathcal{C}(0, s)e^{-(s+k_0)z/u} \tag{2.362}$$

which can be rewritten as follows:

$$\mathcal{C}(z, s) = e^{-sz/u}[\mathcal{F}(z, s)] \tag{2.363}$$

$$\mathcal{F}(z, s) = \mathcal{C}(0, s)e^{-k_0 z/u} \tag{2.364}$$

If one recalls the shift property of the Laplace transform, we can notice that the above equation is characterised by a time delay z/u. Indeed:

$$\mathcal{L}[f(z, t - \frac{z}{u})] = e^{-sz/u}[\mathcal{F}(z, s)] \tag{2.365}$$

Moreover, if we consider the boundary condition (2.358), the function $\mathcal{F}(z, s)$ is rewritten as follows:

$$\mathcal{F}(z, s) = \mathcal{C}_{in}(s)e^{-k_0 z/u} \tag{2.366}$$

The above equations (2.363)(2.366) can be used e.g.

1. *to calculate time responses*
   For instance, if the initial state is assumed to be zero ($C(t = 0, z) = 0$), the response to a step $C_{in}1_+(t)$ is equal to:

$$C(z, t) = 0 \qquad \text{for } t < \frac{z}{u} \tag{2.367}$$

$$= C_{in}e^{-k_0 z/u} \qquad \text{for } t \geq \frac{z}{u} \tag{2.368}$$

2. *to compare the plug flow reactor model with the CSTR one*
   Then it is first worth noting that the equation (2.366) is also the expression of the equilibrium state of $C(z,t)$ for C_{in} constant, i.e. the solution of:

$$0 = -u\frac{\partial C}{\partial z} - k_0 C \tag{2.369}$$

   Let us compare this expression with the steady state value of $C(t)$ in a CSTR, which is the solution of the following equation:

$$0 = \frac{F}{V}(C_{in} - C) - k_0 C \tag{2.370}$$

   i.e.

$$\bar{C} = \frac{C_{in}}{1 + \frac{k_0 V}{F}} \tag{2.371}$$

   It is worth noting that by definition:

$$\frac{F}{V} = \frac{u}{L} \tag{2.372}$$

   We note then that there is a great similarity between both expressions $e^{-k_0 L/u}$ and $\frac{1}{1+\frac{k_0 V}{F}}$ at the reactor output $z = L$. Note in particular that both expressions have the same first order term in the Taylor series expansion:

$$1 - \frac{k_0 L}{u} \tag{2.373}$$

   This similarity is illustrated in Figure 2.23, where the expressions $e^{-k_0 L/u}$ and $\frac{1}{1+\frac{k_0 V}{F}}$ have been drawn. This illustrates the analogy, in that specific case (when the input is the influent concentration), between the plug flow reactor and the CSTR + a time delay.

The above simple example illustrates the advantage of the Laplace transform to study the properties of distributed parameter models and to reduce them exactly to lumped parameter models. It also emphasises the possible limitations of the method. For instance how can we extend the above results to more complex dynamics, e.g. for nonlinear kinetics?

2.11 Conclusions

This chapter has been dedicated to mass balance modelling. Starting from simple examples, we have emphasised the two basic items in mass balance models in WWTP: transport dynamics, and conversion. We have also used the notion of reaction schemes as a basis for modelling, and illustrated the concept via a number of

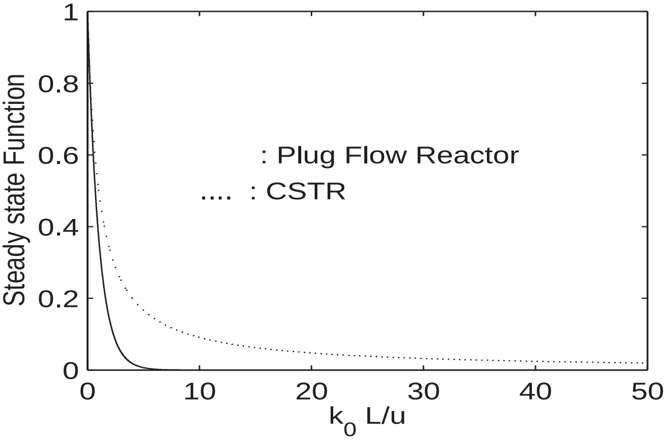

FIG. 2.23. Steady state functions of the CSTR and the plug flow reactor.

WWTP examples. We introduced the General Dynamical Model; this formalism is rather similar to the tabular format (Petersen matrix) used for the IWA model. The General Dynamical Model format is convenient for algebraic manipulations, as has already been illustrated in the present chapter via the state tranformation of Section 2.8 (which will be used later to design software sensors for process monitoring) and the model order reduction approach of Section 2.9. We have illustrated dynamical modelling by using several examples of wastewater treatment processes: anaerobic digestion, activated sludge process (basic model and IWA model), two step nitrification, two step denitrification.

Other examples like waste stabilisation ponds (lagoons) and soil decontamination could also have been considered. A model of detoxification is given in Chapter 7 (Section 7.4.4) in the context of model building of the reaction network independently of the process kinetics, via the use of the asymptotic observers introduced in that chapter. A full study, from model building to parameter identification, of a lagoon is presented in Bonvillain [31] and Grégoire *et al.* [106]. The model is derived from mass balance considerations by considering three types of microorganisms (micro-algae, aerobic bacteria, anaerobic bacteria) working at different depths in the pond. The resulting model contains 36 parameters, but only 3 variables (temperature, dissolved oxygen and light intensity) were available for measurement. Grégoire *et al.* [106] summarizes the identification results for the model parameters, which is clearly a challenging issue and has therefore to be very carefully performed. With that respect, the results presented in Grégoire *et al.* [106] are illustrative and exemplary of a situation largely encountered in parameter identification of WWTP.

We would also like to mention the possibility of including alkalinity in anaerobic digestion models via the introduction of electro-chemical equilibrium equations [223]. A complete study from model building to parameter identification has been performed in the context of an EEC project (AMOCO) for a pilot anaerobic digester [26] [27] [78]. One interesting point of the study is the design of experiments for parameter identification: in order to account for the specific constraints about the operation of the anaerobic digestion process (the process is rather slow and sensitive to disturbances that may easily drive the process to instability, i.e. wash-out), the experiment design for parameter identification has resulted in a sequence of step changes of the influent flow rate and of the influent substrate concentration (or equivalently, of the dilution rate and of the organic load with sufficiently long steady states in between) that basically covers all the operating regions that the process is expected to face. This experiment design can be viewed as an alternative strategy to the one presented in Chapter 5.

So far we had considered single stirred tank reactors. In Section 2.5, we have extended the dynamical models to multi-tank processes, and to non-completely mixed ones, like fixed bed reactors, fluidised bed reactors, and settlers. In the latter case, the dynamical equations are described by PDEs (partial differential equations) instead of ODEs (ordinary differential equations).

Sections 2.6 to 2.10 are dedicated to the dynamical properties of the mass balance models. In Section 2.6, we have drawn the attention to the difference between linear and nonlinear models, and to the fact that the same model can even be either linear or nonlinear depending on the considered input(s) and/or the kinetic models, for instance. In Section 2.7, we have briefly introduced the concept of stability, and the multiple steady states in WWTP: for one set of input values, the process can have different steady states; each of these steady states can even correspond to different stability characteristics, i.e. they can be either stable or unstable. Section 2.8 has introduced a key state transformation that will be very useful to analyse the process model and to design software sensors. Section 2.9 was concerned with a systematic approach for model order reduction (the singular perturbation approach) applied to ODE models as well as to PDE models. Finally, Section 2.10 has considered the use of the Laplace transform as a tool for comparing plug flow reactor models and CSTR models.

3

Structure Characterisation (SC)

3.1 Introduction

As mentioned in the first chapter, Structure Characterisation (SC) is a key step in the model building exercise (see also Figure 1.1 in Chapter 1). After a chapter on mass balance modelling, the present chapter will be concerned with structure characterisation. The objective is to infer the level of model complexity (dimension of the state vector...) and to determine relationships between variables; this can also be viewed, in other words, as a selection of the best model structure among different model structure candidates on the basis of experimental data.

Finding the "true model" $M_T(S_T, P_T)$ with model structure S_T and parameters P_T is utopian. Rather one must aim at finding – from a finite set of N noisy data points – the partial descriptions that are purposeful within the application [159]. Settling for the best possible model $M(S_N, P_N)$, however, induces an error that has two components:

$$\begin{aligned} M_T(S_T, P_T) - M(S_N, P_N) = {} & [M_T(S_T, P_T) - M(S_N, P^*)] \\ & +[M(S_N, P^*) - M(S_N, P_N)] \end{aligned} \qquad (3.1)$$

The first term is due to the error between the true model structure S_T and the model structure S_N chosen from the set of candidate models with restricted complexity. This so-called *bias error* reflects the unmodelled dynamics [104]. The second component, the *variance error*, is caused by the particular realisation of the noise in the limited number of data used in the system identification. Each data set will result in different parameter estimates P_N that will only tend to the real P^* (for this structure) when there is no noise or when the number of data points tends to infinity. This variance error also includes the effect of the overparametrisation: the more parameters included in the model, the more uncertain their values will be.

Typically the variance error decreases like $1/N$, but increases like p, with p the number of parameters in the model structure, a measure of its complexity. The bias error, on the other hand, will decrease as p increases, but is independent of N [157], [159]. Hence, as the aim is to obtain the model structure giving the lowest total error, the goal of model structure characterisation will be to find the compromise between bias error and variance error.

This chapter is organised as follows. Section 3.2 starts with the introduction of a model used in respirometry that we shall consider as a case study throughout Chapters 3, 4 and 5. Section 3.3 is dedicated to the presentation of the methods for structure characterisation, basically gathered in a priori and a posteriori methods. Finally, in Section 3.4, we shall deal with optimal experiment design for structure characterisation.

3.2 A Model Case Study

Let us consider the following candidate models that we shall build on the basis of on-line measurements of the oxygen uptake rate via a respirometer ([74]). The considered respirograms are indeed representative of the biodegradation kinetics in the WWTP, and the objective is to identify a kinetic model from the data of the respirometer. Recall that the model considered here expresses the dependence of the exogenous oxygen uptake rate OUR_{ex} on the biodegradation of k substrates S_i present in the mixed liquor:

$$OUR_{ex} = \sum_{i=1}^{k} (1 - Y_i) r_{S_i} \tag{3.2}$$

In the above expression, Y_i (the yield coefficient) is the fraction of pollutant S_i which is not oxidised but converted into new biocatalyst X, and r_{S_i} is the rate of consumption of S_i. The experiments that we shall consider are performed in batch conditions. In these conditions (and *only* in these conditions! Recall what has been said in Chapter 2, Section 2.2.1), we can write:

$$\frac{dS_i}{dt} = -r_{S_i} \tag{3.3}$$

The four types of wastewater/sludge interaction that have been included in the set of candidates of model structure are:

Type 1 (Exponential): One pollutant, first order kinetics (k=1)

$$\frac{dS_1}{dt} = -\frac{k_{max1}X}{Y_1}S_1 = -r_{S_1} \tag{3.4}$$

Type 2 (Single Monod): One pollutant, Monod kinetics (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}X}{Y_1}\frac{S_1}{K_{S1}+S_1} = -r_{S_1} \tag{3.5}$$

Type 3 (Double Monod): Two pollutants simultaneously degraded without mutual interaction, double Monod kinetics (k=2)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}X}{Y_1}\frac{S_1}{K_{S1}+S_1} = -r_{S_1} \tag{3.6}$$

$$\frac{dS_2}{dt} = -\frac{\mu_{max2}X}{Y_2}\frac{S_2}{K_{S2}+S_2} = -r_{S_2} \tag{3.7}$$

Type 4 (Modified IWA activated sludge model No. 1) [233]: 3 pollutants, 2 hydrolysed into the first substrate which is used for growth according to the Monod kinetics (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}X}{Y_1}\frac{S_1}{K_{S1}+S_1} + k_rX_r + k_sX_s = -r_{S_1} + k_rX_r + k_sX_s \tag{3.8}$$

$$\frac{dX_r}{dt} = -k_rX_r \tag{3.9}$$

$$\frac{dX_s}{dt} = -k_sX_s \tag{3.10}$$

Note that nitrification and its associated oxygen consumption have not been included in this case study. The k_{max1}, μ_{maxi} (i = 1, 2) and k_j (j = r, s) are rate constants, and the K_{Si} (i = 1, 2) are the affinity constants expressing the dependency of the degradation rate on the concentration of pollutant S_i (i = 1, 2). Recall also that the experiments on which the model identification is to be based, are performed in such a way that the change in biomass concentration can be assumed negligible (i.e. $\frac{dX}{dt}$ = 0) (which is a fair assumption provided S(t=0) $<<$ X(t=0)), that the oxygen uptake rate data are only due to exogenous (= substrate induced) respiration (OUR_{ex}), i.e. endogenous respiration is either assumed negligible or is eliminated from the data, and that the oxygen concentration is always maintained above 2 mg/L so that oxygen is never limiting.

The aim is to select between the above biodegradation models via measurements of the oxygen uptake rate OUR, as these characterise substrate degradation.

3.3 Structure Characterisation Methods

In this section a variety of objective decision tools will be introduced that enable to find the trade-off between bias and variance error. The picture of approaches, methods and results of structure characterisation (SC) is very diverse and scattered over several scientific disciplines. Although a number of studies have evaluated different SC techniques, none of them has resulted in a clear recommendation on a definitely superior method [6], [65], [146], [166], [239]. With the above case study, an analogous exercise was devoted to decide on a SC technique most adapted to the problem at hand. To illustrate the fact that this step of model building is not as mature as, for instance, parameter estimation, we use this case study to also show how new SC methods can be developed. A number of these are specific to the candidate models used in this case study, but others are potentially more generally applicable.

The SC techniques have been classified according to their impact on the total time needed for a model identification (i.e. model selection + parameter estimation).

1. Most existing methods for SC evaluate the quality of the different model structures after fitting each model to the data. Hence, these methods can be termed *a posteriori SC* [262].
2. Methods capable of selecting a model without the need of first estimating the parameters belong to the other class, the *a priori SC* techniques.

In view of the observation that parameter estimation of the nonlinear bioprocess models that are dealt with here takes considerable computing time (see Chapter 6), it is clearly advantageous to apply a priori methods, since they will only require the estimation of the parameters of the a priori selected model.

3.3.1 *A Priori SC*

The approach of a priori SC has also been termed model structure selection based on preliminary data analysis and has been reported to be an underdeveloped field [159].

Two groups of methods can be discerned: one type of method is generally applicable, while the other SC methods take advantage of specific features of the model structures present in the set of candidate models.

General Methods. The pattern recognition capabilities of neural networks [229] have incited a study on their applicability as SC technique. Details on the method developed for the case study can be found elsewhere [272], but the main principles are summarised here (Figure 3.1): A three-layer recurrent neural net architecture receiving preprocessed data was used. Different data preprocessing algorithms were tested, all aiming to perform data reduction (to decrease the number of input nodes to the neural net classifiers) without loss of structure specific information. The learning stage was performed with 750 training-patterns obtained from

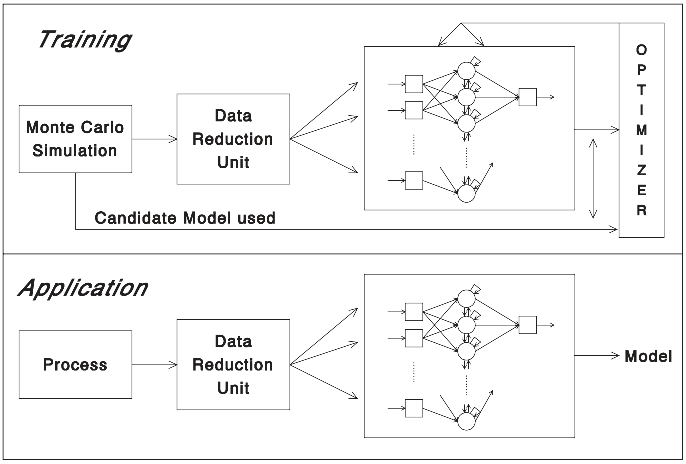

FIG. 3.1. Training and application stages of a neural net used for model selection.

Monte Carlo simulations of the different candidate models with different parameter values and initial conditions. Once the net was trained to recognise the most appropriate model, it could be used on new (preprocessed) data. The results for the case study are discussed below.

A next method that can be applied stems from recent developments in identification of linear state space models. Numerical algorithms for Subspace State Space System Identification (N4SID, read as *enforce it*) combine the estimation of the order of the state with the identification of the system matrices, but, in contrast to traditional identification schemes, the order is estimated first [257]. This aspect of this data-driven approach may be useful for non-linear models as well. Clearly, for the case study for instance, one cannot expect that N4SID will give a direct measure of the number of substrates to be included or the degradation mechanism involved. Rather it can be hoped that the order computed by the algorithm enables differentiation among the models. The order is determined on the basis of the number of non-zero singular values SV in a matrix H composed of the following Block Hankel matrices containing “past” (Y_p) and “future” (Y_f) data:

$$Y_p = \begin{pmatrix} OUR_1 \dots & OUR_j \\ OUR_2 \dots & OUR_{j+1} \\ \vdots & \vdots \\ OUR_i \dots & OUR_{j+i} \end{pmatrix} \tag{3.11}$$

$$Y_f = \begin{pmatrix} OUR_{i+1} & \dots & OUR_{j+i} \\ OUR_{i+2} & \dots & OUR_{j+i+1} \\ \vdots & & \vdots \\ OUR_{2i} & \dots & OUR_{j+2i} \end{pmatrix} \tag{3.12}$$

Different H-matrices can be evaluated:

$$H = Y_f Y_p^T \tag{3.13}$$

$$H = Y_f Y_p^T (Y_p Y_p^T)^{-1} Y_p \tag{3.14}$$

$$H = L_f^{-T} Y_f Y_p^T L_p^{-1} \tag{3.15}$$

with L_f and L_p Cholesky factorisations of $Y_f Y_f^T$ and $Y_p Y_p^T$.

The choices of i and j are important: j must be as large as possible (for the case study j was set $N - 2i$); i must be much smaller than j, but larger than the largest model order expected (i was set 10). For the case study the best results were obtained with the first form of H and singular values considered 0 when below 0.1: type 1 models were selected if $SV \ll 2$; type 2 if $SV = 2$ and type 3 if $SV \gg 2$.

Specific Methods. Structure characterisation on the basis of so-called parameter invariant model features has been advocated, but has found little application due to the difficulty in finding such features from the models under consideration. Indeed, from the data one must find a characteristic that is independent of the parameters of the corresponding model. For instance, we look for features which are scale independent. For the candidate models in the case study, such features can be found [259]: the number of inflection points is 0, 1 and 3 for models of type 1, 2 and 3 respectively (see also Figure 3.2). Though this is analytically correct, the determination of inflection points on noisy data is not straightforward. However, using a moving window regression method in which linear and quadratic regressions are compared with an F-test, the significance of the second derivative can be assessed, leading to a more reliable estimate of the number of inflection points.

Another model-specific approach applicable to the case study consists of fitting an exponential function, a hyperbolic tangent function and a Double hyperbolic tangent function to the data and comparing the resulting sum of square residuals (SSR). In other words, this a priori method for selecting among the biodegradation models is in fact an a posteriori method with respect to the Exp/Tanh/Double Tanh candidate model set. However, these models can be fitted much faster than the biokinetic models. SSR ratios larger than 10 were found necessary before a more complex model was accepted.

A third method developed for the models of the case study is based on the ratio between the area below the OUR_{ex} versus time data (and the square formed by the ($OUR_{ex,max}$, t_{max}) and (0, 0) corners). Some preprocessing of the OUR_{ex} data is performed prior to the determination of this criterion: first, the zero-tail (Figure 3.2) is cut from the data-set using a t-test on the mean value of a data-window and, second, the reduced OUR_{ex}, t-dataset is scaled to the unit square. Type 1 data have the lowest area ($\ll 0.3$) and type 2 the largest area ($\gg 0.5$) under the curve (Figure 3.2). These threshold values for the areas enable the selection among the candidate models type 1, 2 and 3 [259].

These examples of specific methods illustrate that one can devise effective methods for structure characterisation, but this requires special insight in the models to be selected.

3.3.2 *A Posteriori SC*

Most methods described in this section have found more widespread application. In some applications they seem less appropriate because they may become very time-consuming, especially as the number of candidate models increases. However, it is not inconceivable that a SC strategy is devised in which a priori methods are used to make up a first ranking of the different structures, after which a posteriori methods are called in order to make the final selection among the structures with the highest ranking. In that context, time constraints on the model building process and the selection reliability can be sufficed.

Criteria with Complexity Terms. The criteria on which most of the a posteriori SC methods are based take one of the following two forms [232]:

$$\frac{SSR}{N}\left[1+\beta(N,p)\right] \tag{3.16}$$

$$Nlog\left(\frac{SSR}{N}\right)+\gamma(N,p) \tag{3.17}$$

with SSR the sum of squared residuals, and N and p as defined before. For $N \gg p$ one can show that both representations are equivalent if [232]:

$$\gamma(N,p) = N\beta(N,p) \tag{3.18}$$

For both cases, the first term decreases with increasing p (increasing complexity) while the second term penalises too complex (overparametrised) models. The model structure with the smallest criterion value is selected. Different authors have proposed several functional forms for the model selection criteria depending on a theoretical starting point. The two best-known are the Final Prediction Error (FPE), with [159]:

$$\beta(N,p) = \frac{2p}{N-p} \tag{3.19}$$

and Akaike's Information Criterion (AIC), with

$$\gamma(N,p) = 2p \tag{3.20}$$

FPE and AIC have been proven not to be consistent (i.e. they do not guarantee that the probability of selecting the wrong model tends to zero as the number of data points tends to infinity) [140]. However, this disadvantage is compensated by the fact that AIC and FPE enjoy certain properties that allow the determination of good prediction models in case the true model does not belong to the set of candidate models [232].

An example of a consistent criterion is the Bayesian Information Criterion (BIC) in which [6], [228]:

$$\gamma(N, p) = p\, log(N) \tag{3.21}$$

If $p \geq 8$, it can easily be seen that BIC will tend to favour models of lower complexity than those chosen by AIC. Another consistent criterion is LILC where [110]:

$$\gamma(N, p) = p\, log\,(log(N)) \tag{3.22}$$

Criteria that Assess Undermodelling. Goodwin *et al.* [104] decomposed the total error between the true model and a candidate model into three components:

- the effect of the variance of the particular noise realisation
- the effect of the parameter errors due to the noise present in the identification data
- the effect of the undermodelling, corresponding with the bias term in (3.1).

Their General Information Criterion, GIC, is defined as:

$$GIC(N, p) = \hat{\sigma}_v^2 + \frac{p}{N}\hat{\sigma}_v^2 + undermodelling \tag{3.23}$$

In this equation it is essential to have an estimate of the residual error ${\sigma_v}^2$ that is independent of the undermodelling error. Goodwin *et al.* [104] obtained this by fitting a high-dimensional model (assuming it to be the true model) and using the residual variance as an estimate. In the case study of this chapter, such a model is not available, but, fortunately, the zero-tail end of the OUR_{ex}-dataset can be used for σ_v^2 estimation since this part of the data is characterised by the absence of biological dynamics [263]. As a result, the variance of this part of the data can be used for an accurate estimate of the residual error. Using the similarities shown by the authors between the expected value of FPE and GIC [104], it is even possible to write an explicit formula for the undermodelling term [259]:

$$undermodelling = \frac{SSR}{N} - \left(\hat{\sigma}_v^2 + \frac{p}{N}\hat{\sigma}_v^2\right)\frac{N-p}{N+p} \tag{3.24}$$

Structure characterisation is then performed by selecting the model with the lowest undermodelling value.

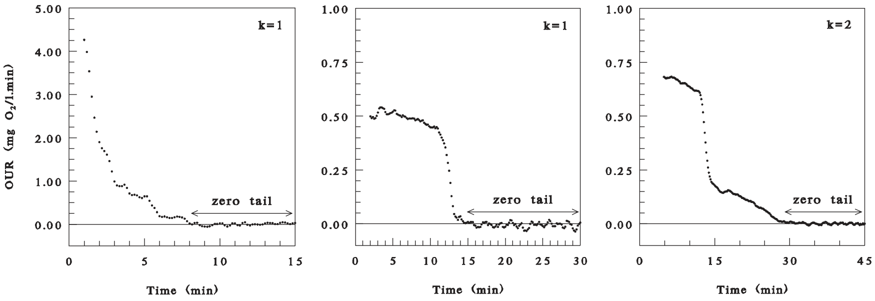

FIG. 3.2. Illustrative OUR_{ex}-data.

Statistical Hypothesis Tests. The F-test is probably the most frequently applied method to choose among model structures. The statistical test

$$\frac{\left(SSR_i - SSR_j\right) / \left(p_j - p_i\right)}{SSR_j / \left(N - p_j\right)} \tag{3.25}$$

is compared with the $F(p_j - p_i, N - p_j)$ distribution to decide whether the more complex model j is significantly (with a confidence level α) better than model i. The similarities that exist between F- and χ^2-tests and the equivalence between the AIC/FPE criteria and F-tests with a prespecified significance level have been shown [232].

Diagnostic Checking (Analysis of Residuals). When modelling, some assumptions are made concerning the properties of the noise. In most cases, the prediction errors $\epsilon(t)$ are assumed to be a realisation of independent random variables with zero mean and a defined distribution. Then, the quality of a model can be assessed by analysis of the properties of the calculated residuals. Two approaches can be used to check the independence of the residuals (white noise property): the autocorrelation and the run test. The autocorrelation test [232] is based on the fact that the correlation function for a white noise sequence $\epsilon(t)$:

$$\hat{r}_\epsilon(\tau) = \frac{\displaystyle\sum_{t=\tau}^{N-\tau} \frac{\epsilon(t-\tau)\epsilon(t)}{N-2\tau}}{\displaystyle\sum_{t=1}^{N} \frac{\epsilon^2(t)}{N}} \tag{3.26}$$

is zero except for $\tau = 0$. Structure characterisation with these tests is performed by selecting the model whose residuals are as white as possible. Two statistical tests are suggested to make objective decisions on the whiteness of the residuals.

- One test compares the correlation for each lag τ (3.26) with the limit value $\frac{N(0,1)}{\sqrt{N}}$, which for $\alpha = 0.05$ means that only 5 % of the autocorrelations may

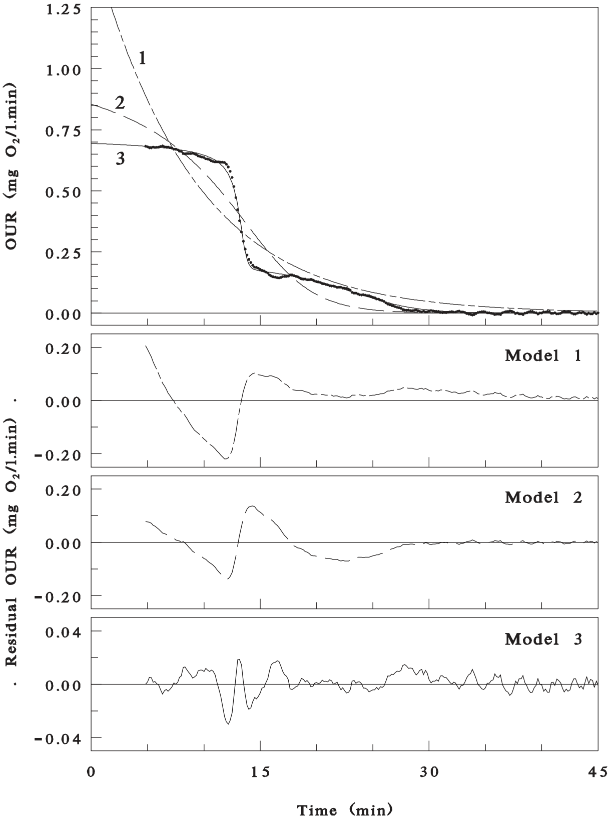

FIG. 3.3. Typical model fitting results for three models to a "Double Monod" style data set.

be larger than $\frac{1.96}{\sqrt{N}}$. In Figure 3.3, an example taken from the case study, the residuals for the three models fitted to a typical type 3 dataset are given. In Figure 3.4 the corresponding autocorrelation function for lags $\tau = 0$ to 20 is depicted. Clearly, the residuals for models 1 and 2 are highly dependent, while for model 3 only the first 4 correlations (20 %) are significantly higher than the prescribed level (indicated by the horizontal lines). Hence, the residuals do not originate from a white noise sequence, indicating some unmodelled dynamics. Looking into some more detail to the residuals of model 3 (Figure 3.3), one distinguishes an oscillatory pattern, probably causing the significant autocorrelation.

- The other autocorrelation test compares a combination of the first m covariances:

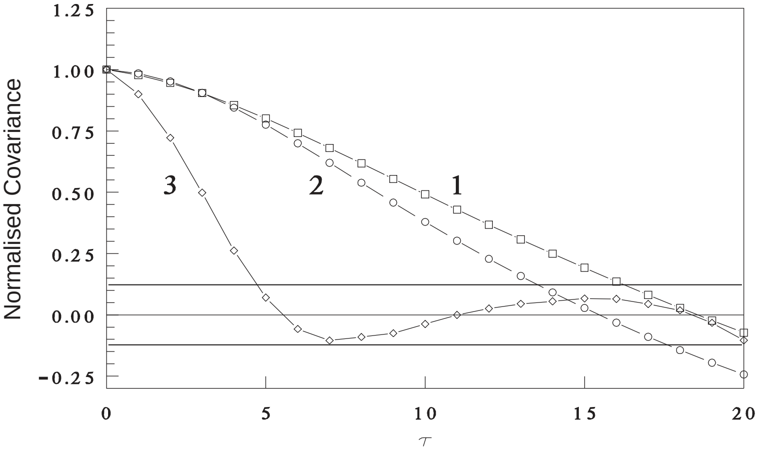

FIG. 3.4. Normalised covariance function of the residuals for the models in figure 3.3.

$$\frac{N}{\hat{r}_\epsilon^2(0)} \sum_{i=1}^{m} \hat{r}_\epsilon^2(i) \tag{3.27}$$

with the $\chi^2(m)$ distribution giving a significance level for the independence of the residuals [159].

Another residuals test is a so-called non-parametric test in which the number of runs R, calculated easily as the number of sign changes in the sequence of residuals, is evaluated against the expected number of runs, $N/2$ [232]. To assess the significance of a deviation from this number the statistical test

$$\frac{R - \frac{N}{2}}{\sqrt{N/2}} \tag{3.28}$$

can be compared with $N(0, 1)$. A posteriori SC with this method selects the model with the statistical test closest to zero or the simplest model with a non-significant statistical test.

Results for the Case Study. The different SC methods introduced above were evaluated on the basis of 8 typical real-life OUR_{ex}-datasets [259]. Only the type 1, 2 and 3 models were included in the candidate model set. In Table 3.1 the selected models are compared with the advice of a human expert. Dataset 4 is difficult to classify since it could be considered very close to both a type 1 and type 2 model. Hence, both were considered correct.

Except for the N4SID method, all a priori SC methods produced very good selection results. Among them, the Tanh and neural net approaches may be preferable considering the noisy data which may cause problems in estimating the number of

Table 3.1 Model selected on the basis of different a priori and a posteriori SC methods. The results are compared with the advice of a human expert (figures in bold indicate "right" choice, underfit and overfit refer to the complexity of the selected model compared to the choice of the expert)

Method of Structure Characterisation	*Dataset*								*Evaluation*		
	1	2	3	4	5	6	7	8	underfit	overfit	correct
N4SID	**1**	3	2	**1**	2	**2**	2	**2**	3	1	4
Neural Net	3	**2**	**3**	**1**	**3**	**2**	**3**	**2**	0	1	7
Inflection points	**1**	**2**	**3**	**2**	**3**	**2**	**3**	**2**	0	0	8
Tanh	**1**	**2**	**3**	**1**	**3**	**2**	**3**	**2**	0	0	8
OUR_{ex}-Area	**1**	**2**	**3**	3	**3**	**2**	**3**	**2**	0	1	7
AIC	3	3	**3**	3	**3**	3	**3**	3	0	5	3
FPE	3	3	**3**	3	**3**	3	**3**	3	0	5	3
BIC	3	3	**3**	3	**3**	3	**3**	3	0	5	3
LILC	3	3	**3**	3	**3**	3	**3**	3	0	5	3
Undermodelling-GIC	**1**	**2**	**3**	**1**	**3**	**2**	**3**	**2**	0	0	8
F-test	3	3	**3**	3	**3**	3	**3**	3	0	5	3
Autocorrelation	**1**	**2**	**3**	**2**	**3**	**2**	**3**	3	0	1	7
Run-test	**1**	**2**	**3**	3	**3**	**2**	**3**	3	0	2	6
Human Expert	**1**	**2**	**3**	**2(1)**	**3**	**2**	**3**	**2**	-	-	-

inflection points. Neural nets have an additional advantage as a potentially general tool for a priori SC.

For the a posteriori methods, an important finding was that all "information" criteria, i.e. AIC, FPE, BIC and LILC, with the notable exception of GIC, result in overfitting of the model compared to the "expert advice". This is probably due to the oscillations that can be observed in the OUR_{ex}-data (see Figure 3.2 and the residuals in Figure 3.3). Since the more complex models possess sufficient flexibility, some of these oscillations can be modelled. This reduces the residual error to such an extent that any penalty for model complexity is compensated for. As the observed oscillations cannot be explained by any biological process and are probably due to some hardware dependent process which is not of interest to the user, this "parasite process" should be eliminated from the data before model identification is initiated.

The F-test also suffers from the modelling of the parasite processes superimposed on the biological response. For the GIC-based method, these overfitting problems were nonexistent because the effect of the oscillations is included in the estimate of the variance (from the zero-tail). In this way only the undermodelling of the biological phenomena is retained leading to model selections congruent with the observation of the human expert.